

An integrative approach to profiling the central nervous system

Melissa Grant-Peters

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford



Keble College
University of Oxford
Michaelmas Term, 2021

An integrative approach to profiling the central nervous system
Melissa Grant-Peters, Keble College, University of Oxford
Thesis submitted for the degree of Doctor of Philosophy
Michaelmas term, 2021

Abstract

With the current increase in life expectancy comes a higher prevalence of age-associated neurological conditions, thus increasing social and economic pressure to improve maintenance of neurological health. Limited sample accessibility and availability place a technological demand on methods compatible with profiling central nervous system tissue which maximise sample utility. Here we present a label-free multi-modal molecular profiling approach which yields spatially resolved transcriptomic, proteomic and biochemical profiles of brain and spinal cord tissue sections, thus addressing this technological demand. This approach was used to profile samples from individuals without neurological condition (n=19) and individuals with secondary progressive multiple sclerosis (SPMS) (n=18), which is characterised by significant architectural alterations to central nervous system identifiable by histological characterisation. Using machine learning and cluster correlation, we have identified two tissue profiles associated with controls which corresponded to white and grey matter regions based on macromolecular, compositional, transcriptomic and proteomic features. By combining the unsupervised clustering across modalities, we identified and profiled a total of 6 degrees of pathology severity.

Apart from overall tissue classification, this approach has also generated valuable mechanistic biological insight: we have been able to not only recapitulate previously reported SPMS pathological biochemical features, but have also linked these processes to specific cell types and tissue-specific processes. This includes evidence of oxidation and alterations to protein processing/folding in oligodendrocytes and neurons. The protein alterations we report include increased β-sheet content and decreased α-helix content in SPMS, a feature normally associated with neurodegenerative conditions with protein folding alterations, and which had never been, to our knowledge, reported in SPMS and adds to its neurodegenerative-like pathological features. We have also outlined processes in SPMS analogous to injury-associated central nervous system scarring, and pathology-associated deviations from typical responses. In particular, we observed niches with high expression of S100β, a protein which has variable cellular function highly dependent on its concentration within tissue: the highly localised profile of spatial gene expression we observe provides insight into a previously unappreciated mechanism by which gene expression can have a deleterious impact attributable to its spatial expression concentration. We believe these findings support the utility of this technological approach not only for disease characterisation and hypothesis generation, but also generating mechanistic and clinical insight for diseases with limited tissue availability, thus enabling the development of treatments to promote healthy ageing of the CNS.

Acknowledgements

This body of work could not have been accomplished without the support of many generous individuals. Firstly, I would like to thank my main supervisor, Calli Dendrou who patiently guided and mentored me through this exciting project, through many set-backs (including a global pandemic) tirelessly, continually encouraging me and making me a better scientist. Many thanks to both my co-supervisors Prof. Jin-Chong Tan and Gianfelice Cinque for all the support relating to the μFTIR experiments. Many thanks to the rest the Dendrou group both in the work and social aspects of the last four years. Specifically, to Dr. Victor Yeung for training me and supporting in the experimental work of this project and to Dr. Charlotte Rich-Griffin for guiding me through the transcriptomic analysis and overall support with bioinformatics. Also, to our youngest honorary group member, Archie, whose cameos in online meetings made the COVID lockdown endurable.

With such a wide range of methods and techniques, other individuals and groups which supported this work include: the Genomics Core, in particular Hubert Slawinski and Angela Lee; Djordje Gveric with the Parkinson's disease and Multiple sclerosis brain bank; Heather Harrington for discussing data analysis approaches; Leedham group for sharing equipment; the John Radcliffe neuropathology department for the use of their cryostat until 2019 and Ben Edwards for allowing us to use their cryostat at a critical time point

during the pandemic; Rhodri Wilson for support with imaging Visium optimization; Roman Fisher and Simon Davis for the support with proteomics work.

Behind the scenes during the course of this project I also received immense support which helped me to press on. I must express immeasurable gratitude to my husband Jonathan for embarking on the adventure of doing DPhils simultaneously – in true Samwise Gamgee style, you could not carry the DPhil for me but you did carry me. My family – mum, Melanie, Jo, Frank and all the Peters siblings, grandparents and extended family – for endless video calls keeping us sane, prayer and support which helped us keep going through the rough times. My friends and community in St. Andrew's – in particular, Hannah Eggimann for dinners and skype calls marking weekly rhythms of friendship, and Jessie Simkiss for the companionship in doing life together and endless encouragement (I hope I am half as good a scientist as you believe me to be).

Lastly but most importantly, none of this could have been done without the generous sacrifice of patients and relatives who donated post-mortem brains to the brain bank. I hope this work does justice to your gift of something so intrinsically personal, and serve as a building block towards effective therapies for MS in the future.

S.D.G.

List of abbreviations

ACN – Acetonitrile
ABH – Antibody high
AD – Alzheimer disease
ALS – Amyotrophic lateral sclerosis
ANOVA – Analysis of variance
ASIC – Acid-sensing ion channel
ATP – Adenosine triphosphate
BBB - Blood-brain barrier
BDNF – Brain-derived neurotrophic factor
CAM – Cell adhesion molecule
CEV - Cumulative explained variance
CIS – Clinically isolated syndrome
CμFTIR – Conventional Fourier transform infrared microspectroscopy
CLR - Centered log-ratio
CNS – Central nervous system
DA – Differential abundance
DGE – Differential gene expression
DNA – Deoxyribonucleic acid
DTT – Dithithreitol
EAE – Experimental autoimmune encephalomyelitis
EBNA1 –Epstein-Barr virus nuclear antigen 1
EBV – Epstein-Barr virus
EV - Explained variance
FISH – Fluorescence *in situ* hybridisation
FPA – Focal-plane array
μFTIR – Fourier transform infrared microspectroscopy
FTIR - Fourier transform infrared
GO – Gene ontology
GSEA - Gene set enrichment analysis
GM – Grey matter
GWAS – Genome-wide association study
H&E – Haematoxylin and eosin staining
HLA – Human leukocyte antigen
IAA – Iodoacetamide
IHC – Immunohistochemistry
IR – Infrared
LCM – Laser capture microdissection
LFB – Luxol fast blue

LISI – Local inverse Simpson index
MRI - Magnetic resonance imaging
MS – Multiple sclerosis
MOGAD – Myelin oligodendrocyte glycoprotein antibody disease
NATMI – Network analysis toolkit for multicellular interactions
NAWM – Normal-appearing white matter
NO – Nitric oxide
OCT – Optimal cutting temperature
OG – Oligodendrocyte
OPC – Oligodendrocyte progenitor cell
ORO – Oil-red-O
Pb – Pseudobulk
PC – Principal component
PCA – Principal component analysis
PD – Parkinson disease
PFA – Paraformaldehyde
PML – Promyelocytic leukaemia
PMS – Progressive multiple sclerosis
PPMS – Primary progressive multiple sclerosis
QC – Quality control
RIN – RNA integrity number
RIPA – Radioimmunoprecipitation assay buffer
RNA – Ribonucleic acid
RNS – Reactive nitrogen species
ROS – Reactive oxygen species
RRMS – Relapsing-remitting multiple sclerosis
SC – Spinal cord
scFEA – single cell flux estimation analysis
snRNAseq – Single nuclear RNA sequencing
SμFTIR – Synchrotron Fourier transform infrared microspectroscopy
SPMS – Secondary progressive multiple sclerosis
UMAP – Uniform manifold approximation and projection
WGCNA – Weighted gene correlation network analysis
WM – White matter

“καὶ θεραπεύετε τοὺς ἐν αὐτῇ ἀσθενεῖς

καὶ λέγετε αὐτοῖς Ἤγγικεν ἐφ’ ὑμᾶς ἡ βασιλεία τοῦ Θεοῦ.”

Λουκᾶς 10.9

List of publications

Scientific publications directly associated with this project

- Grant-Peters, M., Rich-Griffin, C., Grant-Peters, J. E., Cinque, G., Dendrou, C. A. "**Photizo: an open-source library for cross-sample analysis of FTIR spectroscopy data**". Bioinformatics, 2022

Other publications

- Grant-Peters, M., Passos, G.R.D., Yeung, H.-Y., Jacob, A., Huda, S., Leite, M.I., Dendrou, C.A. and Palace, J., "**No strong HLA association with MOG antibody disease in the UK population**". Ann Clin Trans Neurol, 2021

Prospective manuscripts

- Grant-Peters, M., Rich-Griffin, C., Hing-Yuen Yeung, Tom Thomas, Simon Davis, Mohammad Azizian, Joseph Fisher, Roman Fischer, Gianfelice Cinque, Calliope Dendrou, "**Biochemical and metabolic niches define lesional diversity in progressive multiple sclerosis**". *[In preparation for submission]*
- Grant-Peters, M., Cinque, G., Dendrou, C. A. "**A review of FTIR spectroscopy in the study of neurodegenerative disorders**". *[In preparation for submission]*

Statement about impact of the COVID-19 pandemic

The main regard in which the 2020 COVID-19 pandemic impacted this work was in the delay for data generation, thus reducing available time for analysis of these datasets. The 10x Visium and proteomic profiling experiments had been optimised and were going to be run in March/2020 when lockdown was imposed. This caused further delays due to expiration of reagents during lockdown. Moreover, the equipment used for several of these experiments was not available in our institute (e.g.: cryosectioning system, imaging system for Visium) and access as a visitor to other departments was restricted under new COVID-19 protocols.

Important goals of this project (e.g.: a chapter of integrative multi-omic analysis) have had to be revised and reconsidered due to reduced time for data analysis. I was granted a 12-week extension to work on this project, however there have ultimately been several delays which impacted publication timelines.

Table of Contents

ABSTRACT ..... **II**

ACKNOWLEDGEMENTS ..... **IV**

LIST OF ABBREVIATIONS ..... **VI**

LIST OF PUBLICATIONS ..... **IX**

STATEMENT ABOUT IMPACT OF THE COVID-19 PANDEMIC ..... **X**

CHAPTER 1: INTRODUCTION **1**

1.1 PROFILING THE HUMAN CENTRAL NERVOUS SYSTEM 1

1.1.1 Key benefits of omics approaches to CNS profiling *3*

1.2 MULTIPLE SCLEROSIS 7

1.2.1 Clinical presentation and disease course of multiple sclerosis *7*

1.2.2 Genetic and environmental risk factors of multiple sclerosis *9*

1.2.3 Cause of multiple sclerosis: 'outside-in' and 'inside-out' models *13*

1.2.4 MS lesion pathophysiology *15*

1.3 THERAPIES FOR MS 31

1.3.1 Therapies for RRMS *31*

1.3.2 Therapies for progressive MS *32*

1.4 PROJECT AIMS AND OBJECTIVES 37

CHAPTER 2: MATERIALS AND METHODS **41**

2.1 HUMAN POST MORTEM TISSUE 41

2.1.1 Sample acquisition and preparation *41*

2.1.2 Histological characterisation *41*

2.2 FTIR MICROSPECTROSCOPY 44

2.2.1 Sample preparation *44*

2.2.2 Conventional µFTIR *44*

2.2.3 Synchrotron µFTIR *45*

2.2.4 Data Analysis *46*

2.3 SINGLE-NUCLEAR RNA SEQUENCING 51

2.3.1 Nuclear isolation, hashing and sequencing *51*

2.3.2 Data analysis *53*

2.4 SPATIAL TRANSCRIPTOMICS: 10X GENOMICS VISIUM 63

2.4.1 Sample preparation *63*

2.4.2 Optimisation of permeabilisation conditions *64*

2.4.3 Gene expression protocol *64*

2.4.4 Data analysis *65*

2.5 SPATIAL PROTEOMICS 67

2.5.1 Sample preparation *67*

2.5.2 Laser capture microdissection *67*

2.5.3 Sample preparation for proteomic analysis *68*

2.5.4 Data analysis *69*

CHAPTER 3: MACROMOLECULAR ANALYSIS OF SPMS **71**

3.1 INTRODUCTION 71

3.2 RESULTS 76

3.2.1 Synchrotron-based µFTIR profiling of biochemical changes within the SPMS-affected brain *76*

3.2.2 Conventional µFTIR profiling of biochemical changes reveals lesion heterogeneity in SPMS-affected brain tissue *82*

3.2.3 Synchrotron-based µFTIR-based molecular profiling of spinal cord in SPMS and controls *107*

3.2.4 CµFTIR of the human spinal cord *113*

3.3 DISCUSSION 131

3.3.1 Outlining the relationship between lipid and protein content in SPMS pathology *131*

3.3.2 High-resolution lipid biochemistry profiling in the SPMS CNS *132*

3.3.2 High-resolution protein biochemistry profiling in the SPMS CNS. *137*

3.3.3 Improved instrumentation and updated µFTIR analysis yielded valuable biological findings 141

CHAPTER 4: SINGLE-NUCLEAR TRANSCRIPTOMIC ANALYSIS OF SPMS 142

4.1 INTRODUCTION 142

4.2 RESULTS 145

4.2.1 Meta-analysis of snRNAseq MS and control data uncovers high-resolution differences in cell type abundance and gene expression 145

4.2.1.5 Endothelial cells 189

4.2.2 Cell-cell interaction analysis 204

In addition to DA, DGE, WGCNA, and metabolite analyses, cell-cell interaction analyses can provide potential functional insights regarding which cells in a given tissue type or area engage with each other the most and via ligand-receptor (LR) interactions. 204

4.3 DISCUSSION 223

4.3.2 Evidence of OG and neuronal dysfunction consistent with fundamental neurodegenerative processes 232

CHAPTER 5: SPATIAL CONTEXTUALISATION AND VALIDATION 239

5.1 INTRODUCTION 239

5.1.1 Spatial mapping of CNS omics data 239

5.2 RESULTS 243

5.2.1 Spatial transcriptomics: mapping of differentially abundant cell types 243

5.2.2 Spatial transcriptomics: cell type co-localisation network analysis 246

5.2.3 Proteomic characterisation of SPMS and control CNS tissue 273

5.2.4 Comparison between biochemical metrics, and spatial proteomics and transcriptomics 285

5.2.5 Integrated sample annotation across modalities 306

5.2.6 Exemplary targeted integrated analysis 310

5.3 DISCUSSION 316

5.3.1 Spatial transcriptomics reveals differential cellular co-localisation networks in the SPMS CNS relative to controls that correlates with high-resolution biochemical profiles. 316

5.3.3 Correlating biochemical and proteomic changes *320*

5.3.4 Integrated annotation and degree of pathological severity classification *325*

5.3.5 Evidence of non-conventional CNS scarring in SPMS *327*

5.3.6 S100β expression interfaces with 'severe with astrogliosis' classification *333*

CHAPTER 6: CONCLUSIONS **336**

6.1 INTEGRATED SPATIALLY RESOLVED MULTIOMIC PROFILING: A PROOF-OF-CONCEPT 337

6.1.1 Unsupervised machine learning for classification of pathology *337*

6.1.2 Anchoring integrated analysis on spatial elements *339*

6.1.3 Limitations and future directions of spatially resolved multiomic approaches *339*

6.1.4 Sample availability limitations and contingency approaches *342*

6.1.5 Practical considerations for future multi-omic approaches *343*

6.2 A COMPREHENSIVE VIEW OF SPMS MOLECULAR LANDSCAPE 345

6.2.1 Biological contextualisation of biochemical alterations in SPMS *346*

6.2.2 CNS scar formation in SPMS *348*

6.2.3 Highly localised S100β gene expression *350*

6.2.4 SPMS-specific sample classification and heterogeneity *351*

6.3 FINAL REMARKS 353

BIBLIOGRAPHIC REFERENCES **355**

Chapter 1: Introduction

1.1 Profiling the human central nervous system

Among the key predicted scientific challenges of the 21st century is the maintenance of health in the ageing human central nervous system (CNS)[1]. Being the seat of sentience, consciousness and the molecular gateway through which humans can interact with their surrounding environment, the increasing societal impact of CNS disorders is an ominous prospect strongly linked to the unprecedented increase in average life expectancy. A main feature that sets health maintenance of the brain and spinal cord apart from many other systems in the human body is that, while other systems have a remarkable capacity to regenerate, the resilience of the CNS is substantially more limited. With a much slower cell turnover than other systems, the neuronal network of an adult is largely made up of the same neurons formed during development. This makes the CNS more prone to the accumulation of damage, with injury and disease having more lasting impact and, therefore, making the brain more likely to have a decrease in performance with age when compared to other systems[2, 3].

Profiling the human CNS anatomy, connectivity and cellular and molecular composition has significantly contributed to our understanding of health, disease and ageing. CNS profiling with non-invasive non-destructive *in vivo* methods has flourished in recent decades, particularly with the continually increasing computational processing power and improvements in instrumentation. Technological availability has moved the

field towards multi-modal enquiries over the last two decades, integrating multi-level cognitive, electrophysiological and biochemical data, often by anchoring the data integration on neuroanatomical features[4, 5].

However, despite such advancements, molecular profiling of the human CNS poses a particularly difficult challenge. Limited tissue accessibility and understandable ethical concerns make CNS biopsy material rare. This leaves the option of post-mortem sample processing, which has highly variable molecular profiling potential due to sample quality variation associated with the process and speed of post-mortem tissue preservation. Human post-mortem CNS tissue collection generally occurs through brain banks responsible for providing samples across groups and studies. While this standardises tissue processing and preservation methods across studies, it also limits the number of samples available to any single study[6]. Thus, an adequate method for cellular and molecular profiling of the CNS requires maximising the utility of relatively small amounts of starting material. In light of these and other methodological needs, the advent of 'omics' technologies has been received with optimism and already has evidence of being transformative to the study of the CNS[7-9].

Omics technologies are high-throughput biochemical assays that simultaneously capture all, or almost all, molecules of the same type in a biological sample. Genomics refers to capture of DNA, transcriptomics captures RNA transcripts, proteomics captures proteins and metabolomics captures metabolites[10].

While these methods can be applied individually to samples, the combination of two or more methods – referred to as 'multi-omics' - can provide a more encompassing view of several functional biological aspects of a given sample, serving also, in certain cases, as validation. This holistic approach moves away from single-hypothesis enquiries and towards data-driven studies which can have a significant role in the generation of hypotheses.

Multi-omics holds new analytical and computational challenges largely unprecedented in biological sciences, but it has already provided valuable opportunities for the study of mechanistically complex diseases of the CNS which have been difficult to study in a human setting, such as Alzheimer's disease (AD)[7, 8] and amyotrophic lateral sclerosis (ALS)[9].

1.1.1 Key benefits of omics approaches to CNS profiling

Omic approaches pose an extraordinary opportunity to tackle the study of CNS disorders with complex underlying mechanisms which have yet to be fully characterised, such as neurodegenerative disorders[11]. Inherent characteristics of this type of method - high sensitivity, high resolution, high throughput, among others - are strong contributing factors to this.

Until recently, most molecular profiling methods, had limited sensitivity and required large amounts of starting material to yield adequate results. Omics technologies are sensitive and require small amounts of starting material, optimising sample utility – particularly relevant for tissues of limited accessibility and for the study of rare diseases.

RNA capture efficiency is continually increasing for single-cell transcriptomic methods[12] as well as spatial transcriptomic methods[13]. This increase in sensitivity is also increasing resolution of data up to the level of a single cell[14] and, in certain cases, even subcellular level[15] - revealing the previously unappreciated variety in cell subtypes and small subpopulations which would have been lost in bulk methods. Being characterised by a notably large number of highly specialised cell types, CNS profiling benefits from this type of analysis – the cytological complexity of cell type and function can now be elucidated at an unprecedented level.

Improved sequencing techniques and data quality are contributing to making large scale studies increasingly technically and economically viable - each single-cell transcriptomics study is currently sequencing between tens of thousands and millions of cells[14]. While sample availability is still a limiting factor, studies are becoming increasingly statistically robust, in part due to meta-analysis potential using previously published datasets with newer analytical approaches.

As omics fields mature, protocols are becoming standardised and more reproducible. This, added to the cultural shift in scientific research towards data availability[10] is generating momentum for data integration methods to be developed and improved[16-18]. In many instances batch and sample processing effects can be challenging to distinguish from meaningful biological variation, however the increasing use of strategies such as sample multiplexing[19] and computational algorithms for batch

correction are making integration of single-modality data from various studies/sources a viable research strategy[16-18].

Experimentally, this allows analysts to overcome study-specific biases (e.g.: specific contaminants or cell type biases due to technical differences in sample preparation). This also overcomes economic barriers to scientific development – improved data reusability democratizes scientific output, enabling research groups with limited income (e.g.: in developing countries) to also contribute to data analysis and interpretation meaningfully, a valuable asset to have when tackling challenging organ systems and diseases. Moreover, this challenges scientific reproducibility as this model is continually testing reported findings.

1.1.1.1 High multi-modal integration potential

The multi-modal nature of omics technologies brings unique opportunities – alongside analytical and computational challenges – for integrative analysis approaches across different types of data[20]. Given the aforementioned small starting sample requirements, this may even be applied by using as input for different modalities serial sections of the same tissue sample[8, 9] or integration of data from different tissue types, depending on the scope of the scientific enquiry[7, 21]. More recently, novel approaches have become available which enable the profiling of more than one modality on the same cells or tissue section[22].

Advances in barcoding, sample preparation and instrumentation are slowly increasing the repertoire of omics technologies which preserve spatial configuration and

tissue architecture of the sample. The interest of this in CNS profiling is two-fold: (1) the complex architectural organisation of the CNS is intrinsically linked to CNS function and, therefore, its preservation can have significant impact on the study of both health and disease; (2) the neuroscientific field is particularly well equipped with a wealth of established neuroimaging techniques and has already made significant advancements in data processing and integration for other spatially resolved methods at an anatomical scale. Hence, the integration potential for spatially resolved molecular methods with other neuroimaging methods is promising. Particularly with the increasing resolution of modern neuroimaging methods – such as post-mortem 7T magnetic resonance imaging (MRI) - it is conceivable that molecular-level data could be directly integrated with other neuroimaging methods. Spatially resolved multi-omics could be one of the first approach with potential to bridge the gap between anatomy and molecular biology which bypasses the requirement for labelled inquiries characteristic of immunofluorescence techniques.

Early work using MRI data to inform omics analysis with longitudinal data is in its infancy but becoming increasingly employed as a methodological strategy[23], with brain banks now employing preservation strategies which include post-mortem MRI imaging and brain partitioning for several distinct types of analyses making this approach accessible[24]. This strategy holds significant promise for personalised medicine, albeit does not come without a steep logistical cost and analytical challenges, given the large longitudinal data sets it requires[25].

1.2 Multiple sclerosis

Multiple sclerosis (MS) is a chronic immune-mediated neurodegenerative disease that affects approximately 2.8 million people worldwide, of which ~75% are women[26]. Unlike most neurodegenerative disorders that have an increased risk in patients over 60 years of age, MS has a disease onset between 20 and 35 years of age[27]. Due to this, it has significant personal and social impacts since this life stage is marked by several life milestones such as career and family planning, as well as increased responsibility for ageing parents. It poses a severe personal burden for patients and their primary caregivers, partly due to physical symptoms - it is one of the greatest non-traumatic causes of disability in young adults – as well as challenging symptoms affecting cognition and mental health[27, 28]. The heterogeneous presentation and trajectory of MS can also make the personal experience of patients with the disease particularly isolating and psychologically taxing.

1.2.1 Clinical presentation and disease course of multiple sclerosis

MS is characterised by the presence of lesions – sites of inflammation and demyelination - in the brain and spinal cord which can be observed through MRI. The location and number of lesions is variable across patients and relates to clinical presentation for each patient in early disease[29]. This pathophysiological variation across

patients is linked to the heterogeneity of MS clinical presentation and disease progression.

Disease progression can be classified based on certain patterns and key events. The most common disease trajectory affecting 85% of patients begins with relapsing-remitting MS (RRMS), where at onset patients will experience temporary neurological deficits known as a clinically isolated syndrome (CIS). This will usually present in the form of sensory or motor symptoms (43% and 30-40% of patients, respectively) and last between a few days and a few weeks, and will be followed by cycles of relapse and remission of symptoms[27]. Relapses will generally correspond to the appearance of demyelinating lesions in white matter areas of the CNS identified by MRI. At early stages patients experience lesion recovery, although as the disease progresses recovery will become increasingly limited. This results in the accumulation of CNS damage and, consequently, disability[30].

Of these patients with RRMS, 80% will go on to develop secondary progressive MS (SPMS) 10-20 years post diagnosis. SPMS is characterised by the slow and continuous expansion of existing lesions and accumulation of cortical lesions, as well as diffused injury of neurological tissue, with neuroaxonal loss and consequential loss of overall brain volume (**Fig 1.1**). Clinically this manifests as increasing acquisition of disability without periods of remission. In the absence of remission, in SPMS the increase in lesion volume is ~3.6 times more rapid than that in RRMS[30].

Between 10 and 15% of MS patients are diagnosed with primary progressive MS (PPMS), which has a progressive disease course from onset and is functionally similar to SPMS. Interestingly, the transition from RRMS to SPMS occurs within a well-defined age window (35-50 years of age) which coincides with PPMS onset (~40 years)[27, 31]. In spite of the differences in trajectory, familial cases of MS do not support that PPMS is a distinct disease from RRMS/SPMS[32].The remaining 5% of patients are diagnosed with rarer forms of MS, such as progressive-relapsing, paediatric and opticospinal MS[33].

1.2.2 Genetic and environmental risk factors of multiple sclerosis

While a precise mechanistic cause for MS has remained largely elusive, identified risk factors feature a common immunological effect. Current evidence indicates an encompassing theme among risk factors of complex interactions between genetics and environment which have cumulative effects.

Supporting evidence for the role of genetic risk factors includes the incidence of familial cases, which constitute ~13% of cases[34]. As demonstrated by twin studies, risk of recurrence increases with percentage of genetic sharing: age-adjusted risk in monozygotic twins is 18-31% and ~2.7-3.8% in siblings[35, 36]. Heritability is not consistent with a monogenic disease, but rather with a polygenic condition involving polymorphisms which cumulatively contribute to MS risk. Genome wide association studies (GWAS) have identified >200 genetic risk variants associated with MS which are estimated to account for ~30% of disease risk[37, 38]. Of these genetic regions, the majority of risk genes are thought to have an immunological role, including the HLA class

I and HLA class II variants, such as the strongest risk determinant of MS, *HLA-DRB1*15:01*. Many of the identified variants also have a role in other autoimmune diseases, further suggesting a role for immune genes in MS risk [37, 39].

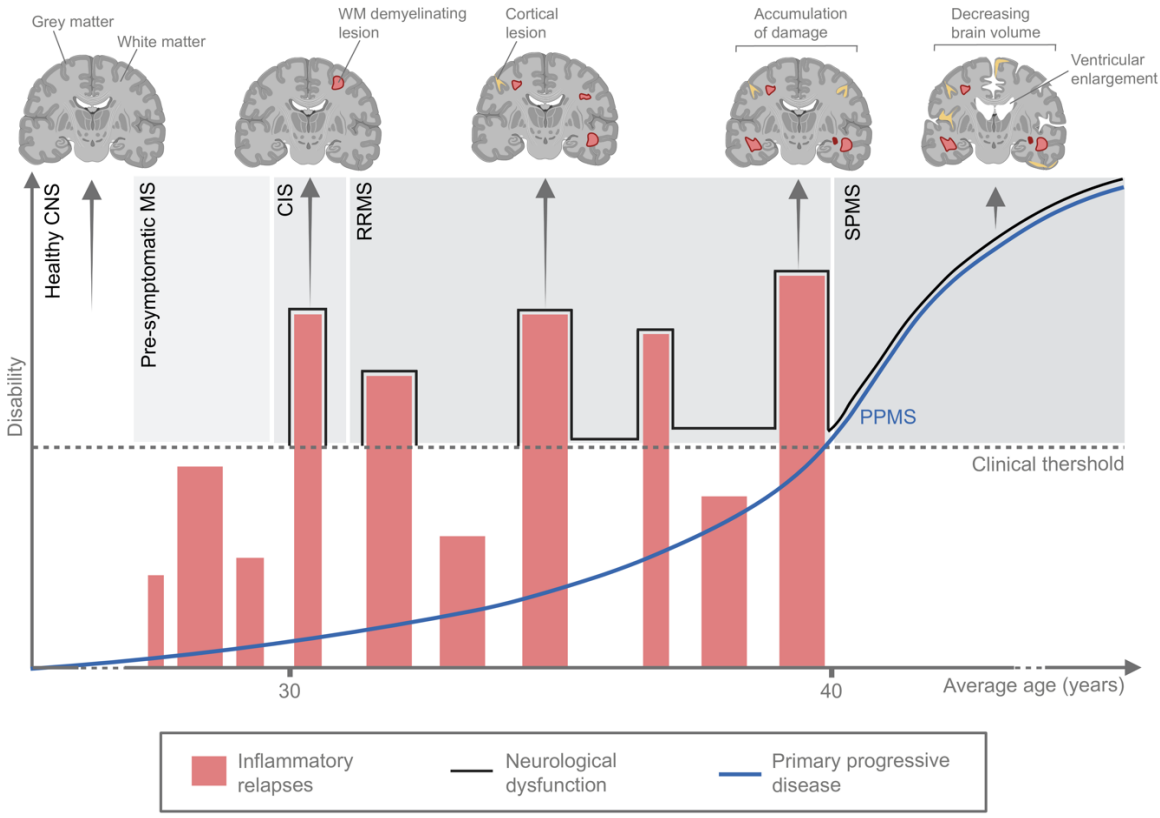


Figure 1.1: Overview of MS disease progression. The most prevalent MS trajectory is characterised by relapsing-remitting cycles in early disease, with corresponding lesions detectable by MRI which heal completely or partially in remission. Late disease is characterised by accumulation of disability and corresponding loss of brain volume. Adapted from Dendrou et al, 2015[40].

Differing ethnical incidence of MS has also been noted, reportedly being higher in patients of European descent, and rarely in Asian, Native American, Maori or Black populations, which could relate to genetic or environmental differences[41]. A study sampling individuals resident in one same geographical location found that there was a higher incidence in Black members of the community than white suggesting differences in incidence may not be limited to but may also be partly due to environmental factors[42].

The latitude gradient of MS incidence is one of the most prominent arguments for a role of environmental factors in MS disease risk[43]. Among Western countries, those with a higher latitude have an incidence 2.5 times higher than other with lower latitudes. It is likely this correlation has some genetic contribution due to the increased prevalence of MS-predisposing *HLA DRB1*15:01* allele enriched in Northern gradients migration studies, although migration data suggests it is unlikely to be the only factor. Migration studies indicate that individuals who migrated from low-risk to high-risk countries before adolescence have similar risk to those born and residing in the high-risk country[44-46]. Among candidate contributing factors, the involvement of ultraviolet radiation and vitamin D have been extensively studied.

The dependence on ultraviolet radiation of vitamin D converting to its active metabolite makes this relationship difficult to tease apart, although there is evidence

that separately these factors can have a protective role particularly if exposure occurs during adolescence[44]. Suggested mechanisms for this relate to the numerous roles of vitamin D in the immune system, since it contributes towards regulating the immune system and skewing immune cell proliferation and differentiation away from a pro-inflammatory phenotype[47].

Viral infections such as Epstein-Barr virus (EBV), and EBV-dependent infectious mononucleosis, have also been implicated in MS risk. A meta-analysis study suggests that infectious mononucleosis can increase risk of developing MS by >2-fold[48]. Elevated EBV nuclear antigen 1 (EBNA1) is also associated with MS[49]; a particularly striking observation in one study was the observation that all EBNA1-negative individuals had serologically converted to being anti-EBNA1 antibody positive before MS onset[50], supporting a functional involvement of anti-EBNA1 antibodies.

Interactions between viral infections and HLA MS risk variants have also been outlined: infectious mononucleosis and increased EBNA1 antibody titres interact with HLA MS risk variants to increase MS risk[51], specifically, infectious mononucleosis interacts with the *HLA-DRB1*15:01* variant[52]. While it is interesting to hypothesise that the genetic risk for MS might be mechanistically related to viral infection, this remains to be definitively proven.

Interestingly, cytomegalovirus infection has been suggested to have a protective role[53], although the small nature of these studies mean this has largely remained speculative and unclear. Among other environmental risk factors are early life obesity,

smoking, circadian disruption (e.g. shift-work), different gut microbiota, alcohol and salt intake, and a protective effect has been suggested for coffee[44]. Jointly, the current genetic and environmental body of evidence points towards an immunological trigger in MS disease pathology or at least significant immune involvement.

1.2.3 Cause of multiple sclerosis: 'outside-in' and 'inside-out' models

The likely protagonistic role of inflammation and autoimmune mechanisms in MS pathophysiology has led to the hypothesis that immune dysregulation is the main trigger for MS. Aforementioned genetic and environmental evidence, much of which can be traced to have an effect on the immune system, contributes compelling evidence supporting this theory.

As a consequence of this underlying assumption, MS research has been largely shaped by immune-related inquiries, and the 'outside-in' model of MS development is based on this and on the muring model of MS, experimental autoimmune encephalitis (EAE). The EAE model was first described in the 1930s, but is still widely used[54, 55]. In the EAE model, disease typically develops after rats or mice are immunised with CNS-derived protein or peptides with adjuvant and injected with pertussis toxin to promote permeabilisation of the blood-brain barrier (BBB); with this protocol the peripheral immune cells of genetically susceptible animals that are specific for the CNS antigen used for the immunization are directed to and attack the CNS, leading to MS-like disease characterised by tail and hindlimb paralysis. Target proteins include, but are not limited to, MBP[56, 57], MOG[58] and PLP[59], with differing presentation, disease course and

predominant adaptive immune cell types involved. This model thus suggests that possibility for an 'outside-in' triggering mechanism in humans if peripheral tolerance is broken promoting activation of CNS-specific immune cells (potentially during an infection due to mechanisms such as molecular mimicry), along with a concomitant breakdown of the BBB.

This model is being challenged by an alternative view that MS is primarily a neurodegenerative disorder involving cytodegeneration which is convoluted by secondary autoimmune response as a consequence of releasing highly antigenic constituents in a predisposed host[60]. Observational evidence supporting the 'inside-out' theory includes (1) findings indicating oligodendrocyte damage occurs in the most internal lamellae of myelinating processes, closest to axons[61] which would not be the first regions likely to be damaged by immune cell infiltration (2) the finding that myelin damage can occur outside of areas of inflammation, suggesting that alone inflammation may not be causal[61, 62] (3) analysis of early active lesions showing no T- or B cell infiltration, only macrophage infiltration and microglial activation which would be consistent with recruitment of phagocytosing cells for myelin debris clearance in a scenario where myelin damage is a primary, rather than secondary, event[63] (4) abnormal white matter in MS brains with progressive forms of MS in the absence of inflammation[64]. While these observations are compelling, post-mortem observation of early lesions is rare, making it difficult to validate the findings relating to this disease phase and ensure they are representative across MS cases. For instance, these findings

from early studies could also be explained by a misdiagnosis of CNS antibody-mediated disorders such as neuromyelitis optica spectrum disorder or myelin oligodendrocyte glycoprotein antibody disease (MOGAD) as MS[65].

Interestingly, the inside-out model, however, does not necessarily clash with current genetic evidence or with the observed efficacy of immunomodulatory drugs thought to act upon the peripheral immune system, although it requires an alternative interpretation. Such an interpretation could involve a CNS-directed infection or trauma causing damage within the CNS, promoting BBB permeability and leading to the release of CNS antigens into the periphery, that then facilitates the mounting of a CNS-directed immune response in genetically susceptible individuals.

As MS may be triggered a decade or more before the first clinical signs develop it may be impossible to definitely determine which model of triggering occurs, and it may indeed vary across patients. However, regardless of how MS triggers, the relative inability of immunodulatory drugs to substantially halt neurodegeneration and disease progression suggests that a better understanding of the pathophysiological processes that underpin these aspects of disease development is critical for improved therapeutic intervention.

1.2.4 MS lesion pathophysiology

MS pathology can be observed in the CNS at two main biological levels. The pathological and diagnostic hallmark of MS, demyelinating lesions, are heterogeneous both in their spatial presentation and their temporal trajectory. With varied

neuroanatomical distribution across patients, lesions can appear across the brain and spinal cord, initially in the white matter but also in the grey matter. While distinctly heterogeneous, the distinguishing hallmarks seen across lesion sites and stages are the presence of inflammation, focal plaque demyelination and microglial activation[66-68]. The inflammatory nature of each plaque varies according to the cellular presence in its vicinity. During phases of immune cell infiltration inflammatory processes are more acute, becoming more chronic during the progressive phase. Similarly, while demyelination is present in MS throughout all disease stages it is phase dependent. New appearing active lesions during RRMS phase may recover and remyelinate becoming a shadow plaque or may become reactivated at a future relapse, making the myelination trajectory of a lesion variable. Lesions in progressive MS have a more linear progression with gradient-like expansion of pre-existing lesions, with an expanding rim of demyelination and axonal loss. Lesion volume increases more rapidly during this phase[30]. This is associated with a diffused global tissue injury which has features shared across neurodegenerative conditions, characterised by axonal loss and inflammation, resulting in loss of CNS volume, brain atrophy and ventricular enlargement[69, 70].

The transition from one lesion type to another does not occur simultaneously across all lesions in a patient and it is common for active lesions, chronic expanding, inactive and remyelinated lesions to coexist in the same patient. The types of lesions present will depend partly on the MS stage as well as the competing rates of damage versus remyelination/recovery[31]. Due to this distinction, when presenting pathological

mechanisms of MS lesions we will refer to them as 'early disease', meaning active lesion formation most common during early RRMS and 'late disease', meaning the chronic neurodegenerative lesion stage which is most commonly – but not exclusively – present in SPMS and PPMS (Fig 1.2).

1.2.4.1 Early disease

Early disease lesion processes include new lesion formation and reactivation of previously developed lesions. Immune cell infiltration is a key feature of early disease lesions and may occur via through three routes: (1) direct crossing of the BBB; (2) crossing of subarachnoid spaces; or (3) crossing the choroid plexus across the blood-cerebrospinal fluid barrier. In the early pre-phagocytic phase, CD8+ T cells are the most abundant infiltrates and will - alongside activated and recruited microglia/macrophages - cause the first inflammatory wave. The BBB integrity is further compromised with the ongoing inflammatory process and peripheral innate and adaptive immune cells are recruited onto the lesion site. These contribute to the second wave of inflammation, both directly and by continuing to activate and recruit CNS-resident cells sensitive to homeostatic changes, such as microglia and astrocytes. Significantly, astrocytes have a crucial role in formation of the CNS barrier by forming the glia limitans, while also being capable of mounting inflammatory responses and recruiting immune cells[71]; thus their activation and recruitment directly affects BBB integrity.

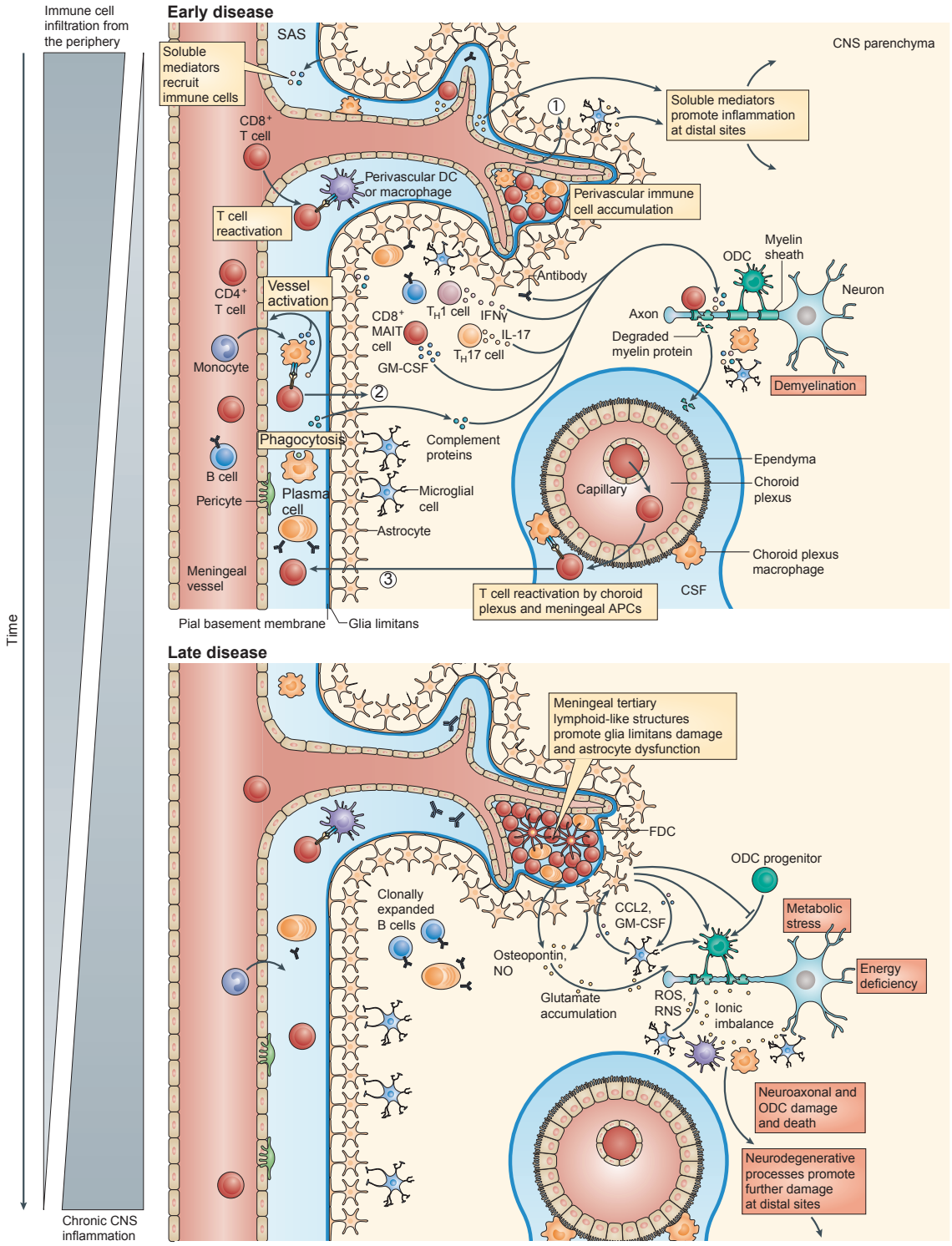


Figure 1.2: MS pathophysiology in early and late disease. Early disease is marked by immune cell infiltration, pro-inflammatory signalling and immune cell activation and recruitment. Less immunological in nature, late disease is characterised by a less acute phenotype with evidence of neurodegenerative-like features such as oxidation, metabolic stress, ionic imbalance. Adapted from Dendrou et al, 2015[40].

With prominent immune cell infiltration, there is an increase in secreted cytotoxic cytokines, excitotoxins and reactive oxygen species (ROS) or nitric oxide (NO) species released by activated macrophages and microglia that will cause a diffused chemical insult on the CNS. Concomitantly, activated cell types will also promote damage by cell-to-cell contact. Soluble mediators are continually recruiting immune cells, perpetuating damage in a continuous insult. The effects of this infiltration are distinct in the white matter and the grey matter, partly due to the differing cellular composition of each tissue type.

In the white matter, oligodendrocytes and neuronal axons are particularly affected by the process of demyelination. The myelin protective sheath generated by oligodendrocytes that provide structural and metabolic support to neurons is degraded and phagocytosed in this process, primarily damaging oligodendrocytes and having secondary effects on neurons. Apart from the loss of metabolic support from oligodendrocytes, neurons, damaged from the neurotoxic effect from the immune cell mediators, are also deprived of the neurotrophic factor secretion of oligodendrocytes which have a neuroprotective role in a setting of homeostatic stress. Jointly these events contribute to make this lesion stage the one with most pronounced axonal injury in the white matter. The extent of axonal damage correlates significantly with the number of infiltrating lymphocytes and activated microglia[67, 72]. Remyelination may occur, however this process is inhibited by effector T cells and IFNγ overexpression by astrocytes; IFNγ also reduces the overall number of oligodendrocytes and induces an

oligodendrocyte precursor cell (OPC) phenotype which may hamper remyelination[73]. Moreover, a recent animal study suggests that following damage, the myelination capability of surviving oligodendrocytes is impaired with a failure to target and remyelinate appropriately[74].

In the grey matter, axonal and synaptic loss are prevalent but due to the presence of neuronal soma in this region, neuronal damage and cell death also ensue and are particularly pronounced in early-stage cortical plaques. Grey matter lesions have been reported to have fewer lymphocytes and fewer macrophages/microglia[75]. This is particularly interesting in light of the fact that in a state of homeostasis a larger number of microglia reside in the grey matter.

1.2.4.2 Late disease

Late MS lesions still present with demyelination and inflammation, but through seemingly distinct mechanisms. In this phase, lesions are already established and identifiable based on areas of demyelination. Hypercellularity is no longer observable; BBB integrity has been mostly restored and only few immune cell infiltrates are observed, mostly in the vicinity of small veins and venules. Immune cell exhaustion associated with chronic antigenic exposure may play a role in the decreased infiltration[76]. Cortical lesions - particularly in the subpial cortical layers - are most abundant in this stage. Lesion accumulation especially in the sulci has led to the hypothesis of the presence of a soluble factor diffusing from the meninges into the

cortex and either inducing demyelination and damage directly or through activation of microglia[77, 78].

In SPMS, there is some evidence that this may be related to the formation of meningeal tertiary lymphoid-like structures where once infiltrating adaptive immune cells become CNS resident. These structures tucked in the meninges contain germinal T cells and B cells as well as follicular dendritic cells. Diffused soluble factors deriving from pro-inflammatory processes in these structures promote dysfunction and inflammation in resident microglia and astrocytes, causing glia limitans damage and astrocytic dysfunction[77]. However, it is unlikely this is the only pro-inflammatory mechanism in SPMS, since there is no evidence that such structures are present in PPMS in spite of the many similarities in disease trajectory, presentation and pathophysiology of the two MS subtypes[78].

Microglia activated by the infiltrating immune cells may in turn stimulate astrocytes, which then recruit and activate further microglia into the site. Activated microglia are linked to active demyelination and neurodegeneration. Healthy neurons express inhibitors which block the phagocytic capacity of microglia, but following the severe damage to the neuronal population in the early stage of disease, this function would be at least partially impaired, causing a cycle of activation and recruitment to go unchecked and unhindered[79]. Apart from the unchecked inflammation, recovery and remyelination are further hampered by astrocytes preventing remyelination at sites of neuroaxonal injury by inhibiting progenitor cells from developing into mature oligodendrocytes[80].

This disease stage is marked by common features with other neurodegenerative conditions including Alzheimer's disease and Parkinson's disease, such as presence of ROS, NO, increased levels of glutamate, metabolic stress, energy deficiency, ventricular enlargement and overall reduction in CNS volume[69].

1.2.4.3 Mechanisms of neurodegeneration

1.2.4.3.1 Mitochondrial injury, oxidative stress and iron accumulation

The high energetic demand of neurones requires an abundant and continuous ATP supply, normally provided by oxidative phosphorylation of glucose. The disruption of mitochondrial homeostasis observed in MS lesions causes oxidative stress, is linked to lysosomal dysfunction, and can even affect cellular response to proteotoxicity.

A mechanism by which mitochondria may become damaged is due to the presence of ROS and reactive nitrogen species (RNS), which can act as signalling molecules, but which at elevated concentrations can cause damage to biomolecules and mitochondrial dysfunction. The abundant presence of free radicals in MS lesions results from inflammation-driven oxidative stress, mitochondrial dysfunction and presence of metal ions. Activated macrophages have been shown to produce ROS and RNS in MS and EAE lesions[75], and their oxidation products correlate with inflammation [81, 82] (Fig 1.3). Reported effects of oxidation in MS include alterations of proteins of the mitochondrial respiratory chain[83, 84], DNA mitochondrial deletions in neurons [85], oxidised lipid and DNA oxidation [86-90].

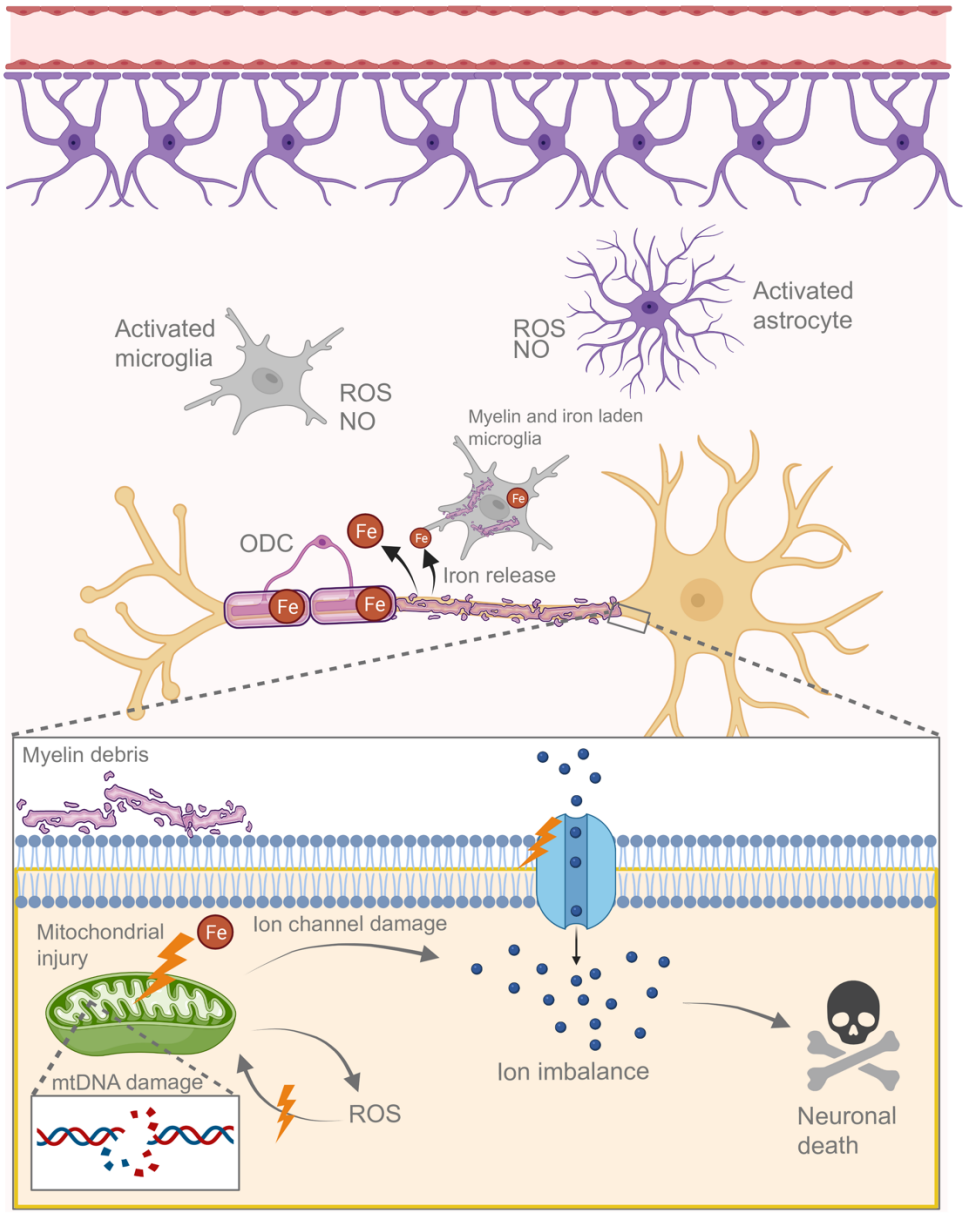


Figure 1.3: Mechanisms associated with oxidation in MS. In the context of neuroinflammation, activated astrocytes and microglia release reactive oxygen species and nitric oxide which promote oxidation. The process of demyelination also increases the concentration of iron ions, since this causes the release of iron stored in the oligodendrocytes. On a cellular level, this causes ion imbalance, mitochondrial injury and damage of mitochondrial DNA, which causes further release or reactive oxygen species and increased oxidation and energetic stress. Ion imbalances contribute to neuronal death.

Apart from general cellular and biomolecular effects, specific cell types have selective vulnerability to oxidative stress – this has been observed both across neuronal subtypes as well as glial cells [91-93]. Among glial cells, oligodendrocytes are particularly susceptible, possibly due to reduced antioxidant defence capacity as well as their inherently higher iron content due to their role in iron accumulation[93]. This means oligodendrocyte cell death significantly contributes to the perpetuation of ROS release, mitochondrial damage and oxidative stress: oligodendrocyte apoptosis releases accumulated Fe^{2+} into the extracellular space, Fe^{2+} acts as a catalyst in converting H_2O_2 to OH^-, thus amplifying oxidative damage. Fe^{2+} is then taken up by macrophages and microglia[94], although these cells ultimately become dystrophic, undergo fragmentation and cellular degeneration leading to another wave of Fe^{2+}. Due to their high energetic demand and the neurotoxic effect of ROS and NO species, neuronal populations are also vulnerable to oxidative stress through various mechanisms, such as alterations to ion homeostasis.

1.2.4.3.2 Altered axonal ion homeostasis

Disturbed ion homeostasis is increasingly gaining prominence in our understanding of the process of axonal damage and neuronal death. In particular in MS, the process of demyelination disrupts arrangements of cell adhesion molecules, ion channels and cytoskeletal proteins established during myelination. Studies have demonstrated the involvement of NA^+ channels[95], glutamate receptors[96] and voltage-gated Ca^{2+} channels[97] in EAE pathology, while bulk RNA-sequencing methods and snRNAseq are

revealing the dimension of alterations to ion channels in MS lesions, particularly those of a more chronic nature[98].

Many alterations in ion balance have a close relationship with energetic alterations and low supply of ATP, causing for instance Na^+/K^+ ATPase pump failure and a deleterious accumulation of Na^+ flux[99]. Ion imbalances have been suggested to relate to the presence of NO and alterations to mitochondrial respiration, leading to an accumulation of Na^+ and Ca^{2+} intracellularly in neurons as well.

Targeting ion channels has in fact been suggested as a neuroprotective, such as inhibition of Na^+/Ca^{2+} channels[100]. Specifically, TRMP4 and acid-sensing ion channels such as ASIC2 - cation channels sensitive to homeostatic changes in in their environment - increase Na^+ and Ca^{2+} levels in response to alterations in chronic CNS inflammation including in EAE models, contributing to axonal degeneration[99, 101]. Asic1$^{-/-}$ EAE mice had reduced clinical deficit and reduced axonal loss when compared to wild type; further work with pharmacological intervention blocking ASICs was neuroprotective was protective in wild-type EAE models[101]. *In vitro* assays suggest inhibition of TRMP4 also had protective effects to excitotoxic stress and energy deficiency; *in vivo* work in EAE models identified that pharmacological TRPM4 inhibition reduced axonal and neuronal degeneration and attenuated clinical scores, although without altering EAE-relevant immune function[102].

1.2.4.4 MS lesion characterisation and classification

In 1868, equipped only with carmine histological staining and a mid-19th century microscope, J.M. Charcot was the first individual known to observe the histology of MS lesions. Remarkably, in his work *Histologie de la sclérose en plaques* Charcot accurately described and hypothesised many important MS mechanisms which have stood the test of time such as demyelination and axonal loss, which he correctly linked to disability. Charcot also correctly predicted the possibility of remyelination and neuronal survival mechanisms as a potential cause for the long disease progression characteristic of MS[103].

Since Charcot's first observations, traditional staining has continued to play a central role in the study of MS lesions. The addition of targeted staining such as immunohistochemical (IHC)/immunofluorescence staining to the histological toolkit was one of the main technological innovations enabling identification of specific cell types. Traditional staining methods are still used to profile myelination with lipid-specific dyes such as luxol fast blue (LFB) and oil-red-O (ORO) for myelination and axonal survival can be visualised with Bielschowsky stain.

Equipped with these methods, multiple studies have sought to classify MS lesions. The main criteria for classification are based on myelination (myelinated, demyelinated or remyelinated), inflammation and presence immune cell infiltrates (active versus chronic).

Five main systems have been proposed – (1) Bö/Trapp, (2) Lassman/Lucchinetti/Brück, (3) Lassmann/Brück, (4) de Groot/van der Valk

modification of Bö/Trapp and (5) Vienna consensus system. These systems are not all mutually exclusive – for instance, while the Bö/Trapp system aims to classify lesions based on myelination and cellularity, the Lassma/Lucchinetti/Brück system focuses only on classifying active lesions, and has resulted in some of these being revised and combined. However, the field has not successfully settled on a single canonical classification system.

1.2.4.4.1 Bö/Trapp system

Not requiring knowledge of cell types present for lesion classification, this simple and objective system is based only on cellularity and myelination, using exclusively haematoxylin and eosin (H&E), LFB and ORO staining. Following identification based on demyelination, lesions are classified as active, chronic active or inactive: hypercellular lesions are identified as active-demyelinating; hypocellular lesions are considered chronic-inactive; and lesions with hypocellular centres and hypercellular rims are considered chronic-active[104].

1.2.4.4.2 Lassman/Lucchinetti/Brück system

This comprehensive classification system focuses on active lesions and is based on staining with H&E, LFB, periodic acid-Schiff reaction, Bielschowsky's silver as well as IHC. Targeted staining was used to identify T cells, B cells, plasma cells, monocytes, macrophages, activated macrophages, myelin, astrocytes and oligodendrocytes[105].

All lesion patterns have a common feature of T cells and macrophages/microglia being present. Based on distribution of myelin protein loss, plaque geography and

extension, pattern of oligodendrocyte destruction and evidence of immunoglobulin and activated complement deposits, four distinct types of lesions were described (**Table 1.1**).

Table 1.1: Lassman/Lucchinetti/Brück lesion pattern classification system[105]

	Pattern I	**Pattern II**	**Pattern III**	**Pattern IV**
Inflammatory infiltrates	T lymphocytes and macrophages	T lymphocytes and macrophages	T lymphocytes, macrophages and activated microglia	T lymphocytes and macrophages
Deposition of Igs and complement C9neo	Absent	Ig and complement C9neo antigen at sites of active myelin destruction	Absent	Absent
Demyelination	Centred around small veins	Centred around small veins	Preservation of rim myelin around inflamed vessels.	Associated with oligodendrocytic death in rim of lesion.
Lesion border	Well defined	Well defined	Diffused	Well defined
Oligodendrocytes	Present in plaque centre, variably on edges	Present in plaque centre, variably on edges	Signs of apoptosis	No signs of apoptosis
Myelin proteins	All simultaneously	All simultaneously	Preferential loss of MAG	No differences
Remyelination plaques	Present when oligodendrocytes present in centre of plaques	Present when oligodendrocytes present in centre of plaques	Absent	Absent

Due to the complex temporal development of MS lesions, this classification system was later suggested to in fact demonstrate not different types of lesions but lesions at various stages[106]. One particular challenge in the classification of active lesions such as this study proposed is that post-mortem MS samples frequently have more chronic lesions due to the advanced disease stage of the patient at the time of death. Therefore,

this could be a further complicating factor for the generation of a reproducible, all-encompassing classification system.

1.2.4.4.3 Lassman/Brück system

In this system lesions are classified as late or early based on the state of myelination (measured by LFB, myelin proteins such as MOG, MBP and PLP and myelin breakdown products such as MRP14 and 27E10), and active or inactive based on the presence of immune cell infiltrates – mainly macrophages and lymphocytes. The resulting groups are: early inactive, late active, inactive with macrophages, early remyelinating and late remyelinating[107, 108].

1.2.4.4.4 Vienna Consensus

The Vienna consensus is the result of merging the Bö/Trapp system and the Lassman/Brück lesion in an attempt to standardise classification using cellularity, myelination and inflammation as criteria. This category includes 6 lesion classifications: (1) inflammatory and demyelinating; (2) inflammatory and not demyelinating; (3) inflammatory rim, hypocellular centre and demyelinating; (4) inflammatory rim and hypocellular centre but not demyelinating; (5) no inflammation but demyelinating; and (6) no inflammation and not demyelinating[67].

1.2.4.4.5 de Groot/van der Valk modification of Bö/Trapp

This system is a review of the Bö/Trapp, Lassmann/Brück and Vienna consensus systems and is focused on time-staging the MS lesions based on more recent data. In this approach, chronic active and chronic inactive classification are unchanged from the

Bö/Trapp system, while active lesions are divided into demyelinating and not demyelinating, and this is the only system to include a 'pre-active' category, where white matter has observable abnormalities such as clusters of microglial cells, HLA-DR, CD45+ and inflammatory cell presence preceding demyelination[68].

At present, there is no consensus regarding classification systems. Both the inter-patient and intra-patient heterogeneity of MS pose an extraordinary challenge for understanding lesion composition and timeline. The variability of lesion location across patients is a further confounding factor. Given the inherent architectural complexity and corresponding cellular diversity of the CNS, the resident cell types in regions affected by an MS lesion varies greatly within patients and across patients. Due to the likely existence of selective vulnerability of cell types resident in certain regions, the overall lesion profile could be linked to the region of the CNS that is affected as well. Moreover, the non-linear progression of most MS cases makes the disease progression timeline challenging to deconvolute and to thus distinguish lesion stage differences from lesion type differences. Herein lies one of the great opportunities of novel spatially resolved molecular profiling methods. These data-driven methods have the potential to generate objective criteria for lesion classification, benefitting, for instance, from machine-learning based clustering approaches as well as computational tools such as pseudotime to attempt to predict the temporal trajectory of cells based on their gene expression. Moreover, these methods are often compatible with traditional staining methods such as

H&E, and can benefit from the legacy of previous classification systems in the process of furthering our understanding of MS lesions.

1.3 Therapies for MS

1.3.1 Therapies for RRMS

Current treatments for MS have variable efficacy and, while some are effective in reducing the frequency of relapses and slowing disease progression, there are no known therapies which halt disease progression.

The most effective currently MS disease-modifying therapies available target early MS. In light of the strong immune component in early MS pathophysiology, these treatments generally target either immune cell infiltration or have an anti-inflammatory role.

In the first category, two of the most effective drugs, ocrelizumab[109] and ofatumumab[110], target B cells with anti-CD20 antibodies. Fingolimod, which has a moderate efficacy, acts by blocking infiltration of autoreactive lymphocytes into the CNS. Natalizumab also may have a role in immune cell migration and BBB transmigration, although its effects are broader - it is an integrin inhibitor, meaning it targets cellular adhesion in general. It has been shown to be highly effective in reducing relapses and slowing disease progression[111, 112].

Strategies aiming at anti-inflammatory mechanisms, such as dimethyl fumerate[113], teriflunomide[114], IFNβ[115] and glatiramer acetate[116] are also employed, albeit with

more variable efficacy[117]. Joint strategies can target both inflammation and immune cell infiltration; for instance, IFNβ in conjunction with natalizumab has been reported as an effective therapeutic approach[111].

1.3.2 Therapies for progressive MS

Although most MS patients will experience a progressive form of MS, whether as a primary event (PPMS) or secondary event following the relapsing remitting phase (SPMS), MS therapeutics targeting progressive/late-stage disease are severely lacking. This is a stark contrast when compared to the successes in therapy for RRMS. Part of the challenge in developing and testing appropriate therapeutics for progressive MS relates to a lack of understanding of underlying causality. While the role of inflammation and neurodegeneration in acquisition of disability are widely accepted, there is no consensus as to whether inflammation and neurodegeneration are entirely causally linked.

This lack of understanding contributes to the challenge of developing appropriate animal models for progressive MS that fully recapitulate the complexity and heterogeneity of the human disease. The non-obese diabetic mouse model of the Biozzi antibody high (ABH) mouse model of EAE, for example, has a progressive disease course, which is largely driven by inflammation and may not necessarily accurately reflect the progressive disease that unfolds in humans over many years. Non-autoimmune models displaying white matter degeneration such as the mouse model of cuprizone-induced

demyelination also exist, although again this model may only reflect one facet of potential MS pathology.

Recent developments have been moderately encouraging, but are largely based on targeting MS patients with progressive disease who still show immune infiltration. This approach has led to ocrelizumab being approved for as a therapeutic for a subset of PPMS patients[118] and siponimod being approved for SPMS[119]. An ongoing trial, ChariotMS, aiming at tackling immune activity in earlier stages of progressive MS using cladribine to target T cell and B cell populations. This drug is already used as a treatment for active RRMS with success[120] and its trial for progressive MS is expected to be concluded in 2024.

Another important factor to take into consideration in treatment of progressive MS is that, unlike RRMS, at this stage the blood-brain-barrier integrity has been mostly restored and increases with age, thus providing an extra drug delivery hurdle to be overcome - particularly for late disease stages. In SPMS a potentially strategic window of opportunity for preventative drug delivery exists during the RRMS phase, while PPMS has even less frequent evidence of BBB leakage than SPMS. Based on the pathophysiology of progressive MS, treatment strategies generally fall into one of four categories: (1) targeting T cells/B cells; (2) neuroprotection; (3) anti-inflammatory approaches; or (4) promotion of remyelination.

1.3.2.1 Targeting neuroinflammation

Most of the current approved drug targets for progressive MS are aimed at reducing inflammation by targeting pro-inflammatory cell types such as macrophages, microglia, lymphocytes and astrocytes. While their success is far from a cure, these are encouraging clinical developments.

While T cell and B cell infiltration is not as prominent a feature of progressive forms of MS when compared to RRMS, infiltrating cells can reach the CNS through small blood vessels. Moreover, the presence of tertiary-lymphoid-like structures in SPMS in particular could point towards these once infiltrating-turned-resident cells contributing to ongoing inflammation by diffuse pro-inflammatory signalling across the meninges. CD20-targetting antibody ocrelizumab reduced risk of disability progression by 24% in PPMS, suggesting there is some therapeutic value in this type of approach[118]. Astrocytic involvement in MS inflammation is becoming increasingly evident. The S1PR modulator siponimod which blocks astrocyte activity and reduces astroglyosis has been shown to reduce risk of disability progression of SPMS by 21%[119]. As a result, this is now an approved therapeutic strategy.

Microglia and macrophages, which populate the rim of active progressive plaques, contribute to the pro-inflammatory environment in progressive MS and have also been targeted in a clinical trial. This was done by targeting macrophage migration inhibitory factor (MIF) with ibudilast which reduced the loss of brain volume, although further studies are required for this pharmacological route[121].

1.3.2.2 Neuroprotection

The high energetic demand of neurons for survival makes them particularly susceptible to oxidative stress, especially in the absence of metabolic support from oligodendrocytes due to the selective vulnerability of oligodendrocytes in MS. Oxidative stress is generally accompanied by mitochondrial damage, presence of ROS, NO and accumulation of iron and these factors can contribute to cascading effects of damage, ROS release which worsens oxidative stress. Ultimately this can compromise neuronal ion channel function, contributing to axonal damage and neuronal cell death. Thus, neuroprotective strategies have often aimed to intervene at several levels of these processes and have promising effects in early animal studies, but often have disappointing results when reaching phase II or phase III clinical trials[122].

Glutamate levels are also high in progressive MS, and at high concentrations glutamate can have a neurotoxic effect by increasing intracellular calcium levels. A small study demonstrated that glutamate release inhibition and modulation of glutamate uptake with riluzole reduced the rate of cervical cord atrophy and the development of T1 hypointense lesions, but not the development of T2 lesions[123]. Riluzole was tested in a phase IIb trial alongside amiloride and fluoxetine for SPMS, but the results suggest this strategy had no effect on SPMS progression[124].

Compounds with inherent neuroprotective/neurorescue effects such as neurotrophic factors could also be valuable in promoting neuronal survival, whether as a drug or by promoting activities which increase production of these factors. A study

identified that patients with inherently higher brain derived neurotrophic factor (BDNF) levels were relapse free during the duration of the study[125]. A drug free avenue for increasing BDNF in healthy ageing individuals is through moderate intensity endurance training. Based on this premise, an ongoing phase II trial is testing this as a therapeutic strategy for progressive MS[126].

1.3.2.3 Promoting remyelination

Several strategies for remyelination are also being tested to treat progressive MS. In spite of promising results in EAE models, targeting leucine-rich repeat neuronal protein 1 – a negative regulator of oligodendrocyte differentiation – has not had promising results in any of the phase II trials to date. Induction of prolactin production has also had promising results in EAE and is being investigated for SPMS in a phase II clinical trial. High dose biotin was investigated up to phase II with SPMS and PPMS and had slightly more promising results initially[127], but long term investigations proved this to be an unfruitful line of inquiry as well[128].

Animal studies indicate that surviving oligodendrocytes could have compromised myelinating capacity, with increased aberrant myelination of the neuronal soma rather than axons, and fewer myelinating processes per oligodendrocyte[74]. This could partly contribute to the lack of efficacy in remyelination strategies, particularly in late disease since many surviving oligodendrocytes could exhibit similar damage.

1.4 Project aims and objectives

It is clear that a deeper understanding of progressive MS is indispensable for identification of suitable drug targets or treatments. Given the advent of multi-modal omics technologies for profiling tissue architecture and composition at the cellular and molecular level, our overall goal is to assess how multi-omic tissue profiling – that allows analysis of the transcriptome, the proteome and biochemical macromolecular composition of tissue – can be used to characterise post-mortem CNS samples from patients with progressive MS (**Fig 1.4**). Where possible, technologies that retain the spatial architecture of the tissue have been employed in order to use the tissue architecture as the anchoring point for cluster-correlation analysis and comparisons across modalities.

The motivation for this spatially resolved strategy is two-fold: (1) the chosen methods are compatible with traditional histological staining methods; this enable our analyses to be contrasted with previous histological profiling methods; and (2) given the expanding nature of lesions in progressive MS, profiling the molecular components at play in the interface between the lesion and the surrounding normal appearing white matter could give clarity as to the damage-promoting signals as well as the inherent protective signals in these areas. We hypothesise that this could allow us to identify initial potential targets, mechanisms or pathways for future validation that that may be

therapeutically manipulated to help tip the balance away from damage and towards the slowing down or perhaps even the halting of disease progression.

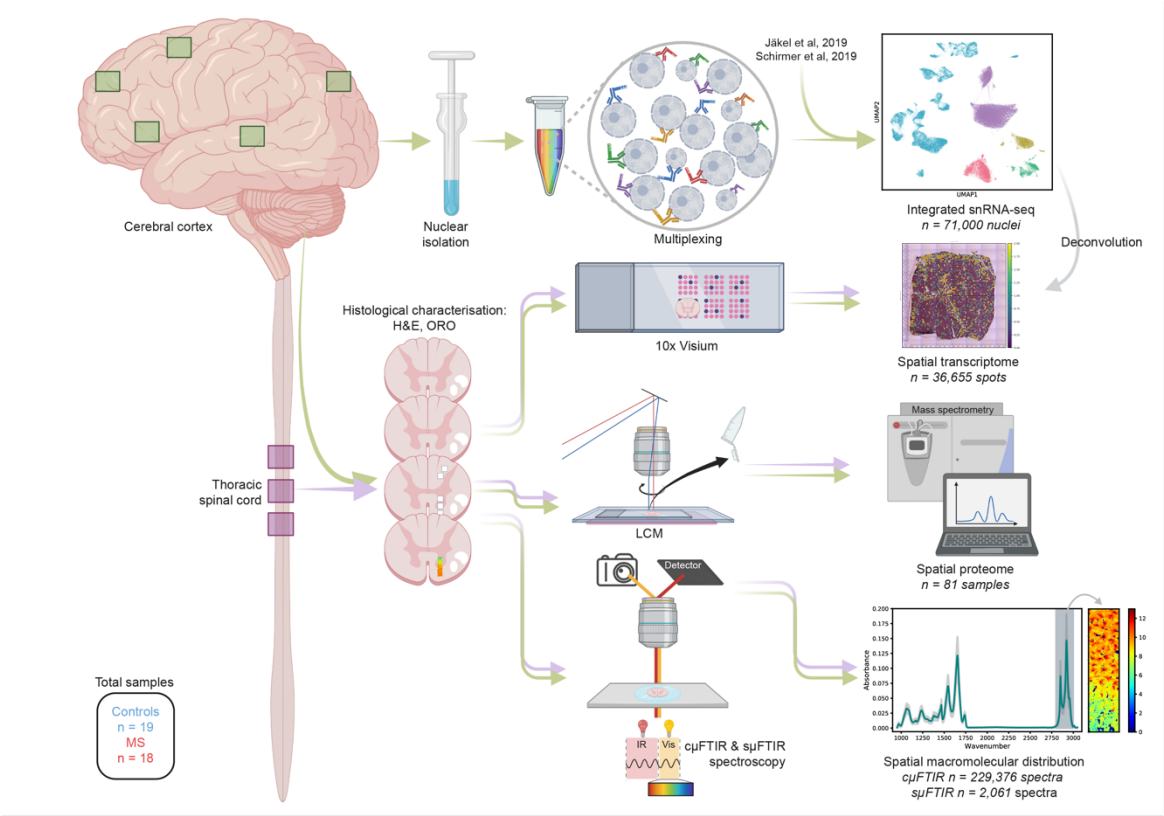


Figure 1.4: Experimental design for multi-omic molecular profiling of MS samples. In this multi-modal strategy we are profiling human post-mortem brain and spinal cord samples transcriptomically (snRNAseq and 10x genomics Visium), proteomically (mass spectrometry) and biochemically (μFTIR).

Firstly, in order to characterise the macromolecular composition of the tissue, we have used Fourier-transform infrared microspectroscopy (μFTIR). μFTIR is a high-resolution spatially resolved and label-free method which captures the macromolecular makeup of a sample. μFTIR does not technically fit the traditional definition of an 'omic' technique as it is based on the physical properties of the response of molecular bonds to an incident light source rather than the sequence of molecules, but it does enable high-resolution, unbiased profiling of the biochemical tissue content. Particularly of interest to this study, it captures the presence of molecules with vibrational mode within the infrared range, many of which are of biological relevance such as lipid bonds (CH_3, CH_2, C=O), as well as amides, and nucleic acids[129].

Secondly, having spatially characterised the biochemical composition of our CNS tissue samples, we then sought to characterise their cellular composition by employing a transcriptomic profiling strategy that combines the use of single-nuclear RNA sequencing (snRNAseq) and 10x Genomics Visium.

Single-nuclear RNA-seq is one of the more mature omics technologies, benefitting from analysis methods developed for scRNAseq but with applicability to fresh-frozen human brain samples. The availability of other MS snRNAseq datasets has allowed us to explore the integration and meta-analysis of data from a single modality[130, 131], and has also enabled an increase in power considering the number of collective cases and controls used for the meta-analysis. Integrating our data with additional datasets

allowed us to better investigate the inherent heterogeneity of MS, to uncover and overcome any study-specific biases, and to obtain a higher resolution cluster annotation compared to prior studies that has facilitated the identification of previously unappreciated cell subpopulations. By pairing snRNAseq to 10x Visium spatial transcriptomics, we have obtained transcriptomic information which combines single-nuclear resolution and the spatially resolved element, by utilising the matched snRNAseq data for spatial transcriptomics data deconvolution.

Finally, for proteomic validation, we have used spatial proteomics – a technique which pairs the use of laser capture microdissection and mass spectrometry, providing a spatially targeted, label-free proteomic analysis method[132].

With this multi-modal approach we are aiming to capture the multi-level functional range of the cell, from the transcriptomic, to proteomic and finally macromolecular biochemical changes in the CNS, in health and disease while preserving tissue architecture. Moreover, we aim to capture some of the diversity of progressive MS in relation to cellular composition, cell identity, cell-to-cell communication, cell activity, metabolism and resulting biochemical composition - all within the spatial context of the lesional and peri-lesional areas so as to begin to determine how the application of this multi-modal profiling approach may provide insights into pathophysiological mechanisms and therapeutic opportunities.

Chapter 2: Materials and Methods

All reported experiments and computational analyses were performed by Melissa Grant-Peters unless otherwise stated.

2.1 Human post mortem tissue

2.1.1 Sample acquisition and preparation

Post mortem human brain and spinal cord samples were obtained from the Multiple Sclerosis and Parkinson's Disease Brain Bank in Imperial College London (brain: n=23; spinal cord: n=11). This work was performed by the relevant Multicentre Research Etichs Committee (08/MRE09/31+5). Upon receipt, samples were embedded in optimal cutting temperature compound (OCT) for optimal preservation; they were then stored at -80°C and only removed for cryosectioning (chamber temperature ~22°C, specimen temperature ~17°C). Unless otherwise stated, cryosectioning was performed in a Shandon cryostat in the Oxford University Hospitals (John Radcliffe Hospital) Department of Neuropathology. Samples were sectioned to 10 µm thickness.

2.1.2 Histological characterisation

2.1.2.1 H&E staining

Haematoxylin and eosin (H&E) staining was performed as previously reported[133]. Briefly, tissue was fixed in 3.8% paraformaldehyde (PFA for 10 minutes and incubated for 1 minute in propan-2-ol. The tissue was then covered with hematoxylin (Dako,

#CS700) and incubated for 7 minutes, following washing, staining with bluing buffer (Dako, #CS702) and eosin (Sigma-Aldrich #HT110216) prepared in tris-acetic acid buffer (final concentration: 10% eosin). The samples were air dried, mounted with DPX (Sigma #6522) and imaged on an Aperio S2 slide scanner with 20X magnification.

2.1.2.2 Oil-Red-O staining

Oil-red-o staining was performed in accordance with the protocol recommended by the Multiple sclerosis and Parkinson Tissue Bank. In short, tissue was fixed in 10% formalin for 10 minutes, slides were washed in running tap water for 2 minutes with agitation every 10 seconds. Slides were dipped three times into 60% triethyl phosphate solution and subsequently immersed in Oil-Red-O solution (Sigma-Aldrich #O0625) for 15 minutes and agitated every 5 minutes. Slides were dipped three times in 60% triethyl phosphate and washed in running tap water for 2 minutes. Sections were counterstained in haematoxylin for 5 minutes. Lastly, tissue was washed in lukewarm running tap water for 5 minutes, air dried, mounted with 50% glycerol and imaged in an Aperio S2 slide scanner with 20X magnification.

2.1.2.3 RNA quality assessment: RIN values

In order to assess the RNA quality of the bulk tissue, sections were placed in a chilled Lysin D matrix tube; pre-chilled dissociation beads were then added to the tube followed by a suspension of Buffer RLT Plus and β-mercaptoethanol. This was then agitated with a FastPrep-24^{TM} homogeniser system for two seconds at setting 6. The flowthrough was collected and 100% ethanol was added to the suspension. RNA was

isolated using an RNEasy Plus Mini kit (Qiagen) following the manufacturer's instructions. The RNA quality was then assessed using an Agilent 2200 TapeStation with the High Sensitivity RNA reagents according to the manufacturer's instructions. Samples with RNA integrity number (RIN) values >5.5 were included in the study (Table 2.1).

Table 2.1: Summary of patients and samples included in this study

Case	Tissue Type	Block Position	Disease	Disease Type	Disease Duration (years)	Age at Death (years)	Sex	Cause of Death	TODF (hours)
C0005	Brain	P2D4	-	-	-	95	F	Bronchopneumonia	10
C014	Brain	P5C3	-	-	-	64	M	Cardiac failure	18
C028	Brain	P1C2	-	-	-	60	F	Ovarian cancer	13
C048	Brain	A3E5	-	-	-	68	M	Metastatic colon cancer	10
C064	Brain	A2D1	-	-	-	63	F	Multiple strokes, acute renal failure, pneumonia	21
C084	SC	T (SC2, SC3)	-	-	-	84	F	Old age	23
C085	SC	T (SCA2) L (SCA6)	-	-	-	81	F	Unknown	22
MS352	Brain	P1C5	MS	SP	18	43	M	Bronchopneumonia, MS	26
MS376	SC		MS	SP	20	48	F	MS	19
MS402	Brain	A6D4	MS	SP	20	46	M	MS, bronchopneumonia	12
MS425	Brain	A3A2	MS	SP	21	46	F	MS, pneumonia	25
MS432	SC		MS	SP	31	48	M	Bronchopneumonia	25
MS438	Brain	P2D4	MS	Unknown	18	53	F	MS	17
MS466	Brain	P2C5	MS	SP	36	65	F	MS, Pneumonia, Lung and womb cancer	25

A = Anterior; F = Female; L = Lumbar; M = Male; P = Posterior; T = Thoracic; SC = Spinal Cord; SP = Secondary Progressive; TODF = Time Of Death to Freezing

2.2 FTIR microspectroscopy

2.2.1 Sample preparation

Cryosectioned sections were placed on calcium fluoride infrared (IR) optical windows (Crystran #CAFP25-1). They were then fixed using 4% PFA at room temperature for 10 minutes and washed 3 times with water to ensure removal of remaining PFA and not affect FTIR results. Samples were then air-dried and stored at room temperature.

2.2.2 Conventional µFTIR

Data collection for Fourier transform infrared microspectroscopy (µFTIR) was performed at Diamond Light Source (Didcot, UK) in two separate sessions, one for brain samples and a separate session for spinal cord samples, each lasting 48 hours. This was performed under the supervision of Dr. Gianfelice Cinque in beamline B22 and with technical support from Dr. Calliope Dendrou and Dr. Victor Yeung.

Image data were collected at 20X on a Hyperion 3000 microscope. Spectroscopy data were collected using a Bruker Vertex 80V Fourier Transform IR Interferometer with a focal-plane array (FPA) detector (64x64, 40 µm pixel size, for near- and mid-IR) with 2x2 binning, the cµFTIR spectra capturing one data point per 1.928 cm^{-1} (spectral resolution).

Prior to data collection, the microscope was cooled with liquid nitrogen and left to stabilise in temperature for approximately 30 minutes. Following this, we performed

optical alignment and tested spectral acquisition and quality. Liquid nitrogen was added every 3.5-4 hours to maintain the low temperature and alignment was verified and adjusted as necessary every time samples were placed on the stage. Maximum absorbance was kept below 1.0.

A clean tissue-free area of the slide was selected for background measurements and scanned 64 times. The total area scanned for each brain sample was 398,881.3 μm^2, where 16,384 spectra were captured (spatial resolution = 24.34 μm^2 per spectra). Scanned SC areas varied between 0.5-2.5 times this area, depending on features of interest in the sample.

Due to minimal changes in atmospheric composition, alignment, light source and technical setup over each 48 hour block, joint analysis of all samples with data collection in the same session. Due to inherent tissue differences and the separate conditions for data collection, spinal cord and brain data were analysed separately.

2.2.3 Synchrotron µFTIR

Synchrotron based µFTIR (sµFTIR) data collection of brain data was performed at Diamond Light Source (Didcot, UK) under the supervision of Dr. Gianfelice Cinque over 5 days with Dr. Calliope Dendrou and Dr. Victor Yeung. Image data was collected at 36X magnification on a Hyperion 3000 microscope with 15 µm x 15 µm aperture, with 5 µm scanning area overlap with neighbouring collection areas for optimal spectral quality and in order to maximise resolution. Similarly to cµFTIR, the microscope was ensured to be

cooled with liquid nitrogen and stabilised in temperature and optical alignment was confirmed prior to data acquisition.

This method had higher spectral resolution, with one data point per 0.964 cm^{-1} in the spectrum. Each sample area was scanned 64 times. Similarly, a clean tissue-free area was selected and scanned 64 times as a background reference. In a separate three-day session data was collected for all the spinal cord samples under the same settings.

2.2.4 Data Analysis

2.2.4.1 Pre-processing

Following data collection, pre-processing was performed including atmospheric compensation and baseline correction (cμFTIR – concave rubberband correction, 3 iterations with 32 baseline points; sμFTIR - rubberband correction) on OPUS (v. 7.4). The data were then exported to python for analysis in a custom analysis pipeline built for this work (**Fig 2.1**).

The region between 2250 - 2400 cm^{-1} was excluded from analysis, since this spectral region is associated with atmospheric CO_2. Outlier spectra with lipid peak and protein peak simultaneously below 1/4 of maximum peak were excluded. Baseline variation is a feature associated with scattering, thus all spectra with a variation in baseline superior to 5 times the average baseline variation were excluded.

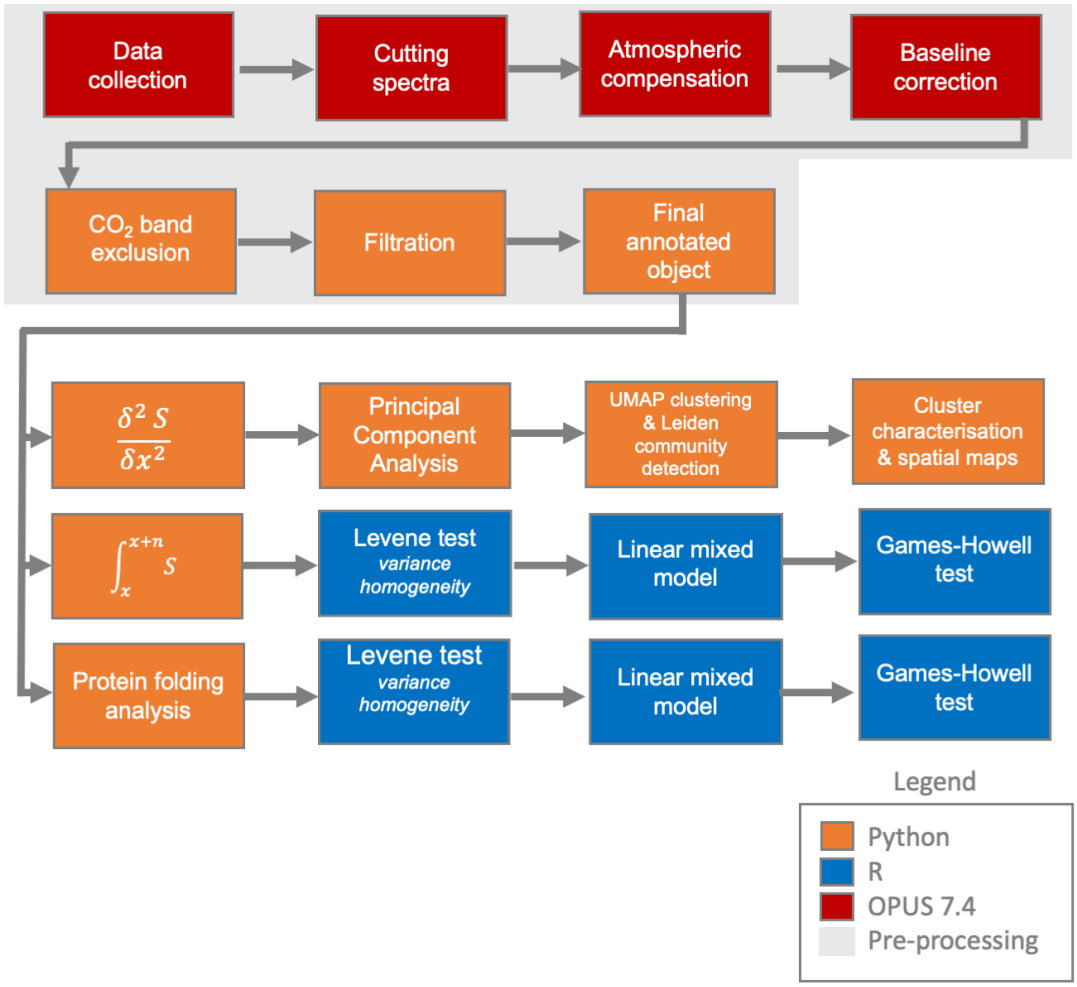


Figure 2.1: µFTIR analysis flowchart. All steps in python and R were performed in custom code for this project.

Prior to exclusion, visual inspection of microscopy images confirmed these regions corresponded to areas without tissue or regions where tissue was damaged by the fixation process. Following filtration, each spectrum was normalised by dividing it by the

norm of that sample, using the numpy linalg.norm function, where axis =1 (where each column on that axis contains a full spectrum).

Preliminary analysis of spectral quality included dimensionality reduction with principal component analysis (PCA). Eigenspectra as well as averaged spectra per sample were visualized to confirm peak position in relation to previously reported spectral features of this sample type. Major peaks could be identified as reported in previously studies (**Table 2.2**).

Table 2.2: Colthup chart of molecular bonds and corresponding frequencies common in biological samples[134]

Bond	Vibrational mode	Interval (cm^{-1})	Commonly assigned biochemical component
=C-H	ν (=C-H)	~3012	Lipid (unsaturated fatty acids)
CH_3	ν_{as}(CH3)	~2960	Lipid (Methyl)
CH2 ν_{as}	$\nu_{as}(CH_2)$	2,900 – 2,938	Lipid (Methylene)
CH_3	ν_s(CH3)	~2873	Lipid (Methyl)
CH2 ν_s	$\nu_s(CH_2)$	2,831 – 2,861	Lipid (Methylene)
Fundamental stretching vibrations of double bonds			
C=O ν	ν(C=O)	1,720 – 1,760	Phospholipid esters, carbonyl ester
OCNH	80% ν(CO) 20% ν(CN)	1,635 – 1,646	Protein (Amide I)
H-O-H	γ(H-O-H)	1645	Water
O-C-N-H	60% γ(N–H) 30% ν(C–N)	1,525 – 1,565	Protein (Amide II)

	10% ν(C–C)		
C-H	ν_{as} C-H	~1468	Lipid (bending from Methyl)
COO-	ν_{as} COO-	1400-1390	Lipids, Carbohydrates (Fatty acids, polysaccharides)
C-H	ν_{s} C-H	~1468	Lipid (bending from CH_2)
Fingerprinting region			
CH3 d		1,430 – 1,470	Lipid, protein
CCH3 ν_s	γ_s (CH_3)	1,370 – 1,388	Phospholipid, fatty acid, triglyceride
OCNH		1350-1200	Protein (Amide III)
CO ν		1,210 – 1,275	DNA and RNA ribose
PO_3^{-2} ν_{as}		1,000 – 1,080	DNA and RNA ribose

ν = stretching; γ = wagging, twisting and rocking; as = assymetric; s = symmetric

2.2.4.2 Analysis of spectral features

Since chemical alterations can have an impact in peak position and peak shape, we used the integration of the area below a band of interest as a more robust measure of the molecular component of interest, since it would account for all these subtle changes. In practice this was calculated by numerical integration approximation using a custom numerical integration function written by Jonathan Grant-Peters. In all subsequent referrals to lipid content and protein content as measured by μFTIR in this thesis, this refers to the integration within the spectral regions reported in **Table 2.3**.

Table 2.3: Integration intervals per major macromolecular group

Macromolecule	Integration region (cm^{-1})
Lipids	2,800 – 3,000
Proteins	1,500 – 1,700
Nucleic Acids	1,000 – 1,300

The protein secondary structure composition of α-helix, β-sheet, β-turns and random coils were quantified by μFTIR using the method described by Goormatigh et al[135]:

$$\alpha - helix\ (\%) = -148.10 + 0.98 * A^{1654} + 1.49 * A^{1588} + 0.66 * A^{1546}$$

$$\beta - sheet\ (\%) = 35.53 + 0.56 * A^{1657} + 0.60 * A^{1629}$$

$$\beta - turn(\%) = -14.25 + 0.36 * A^{1678} + 0.85 * A^{1501}$$

$$random\ (\%) = -40.65 + 1.08 * A^{1523} + 0.62 * A^{1679}$$

For statistical comparisons of protein composition and pseudo-quantitative results from integration, a linear mixed model to control for scanning batch and sample effects was used. This accounted for the lack of independence of each spectrum. Features such as sex, age, cause of death and time from death to freezing did not correlate with the changes observed.

$$\int_{n\ cm^{-1}}^{m\ cm^{-1}} spectra \sim (\mathrm{Region} + \mathrm{Sex} + \mathrm{Age} + \mathrm{DiseaseLength} + \mathrm{TODF}) + (1|\mathrm{Patient/Sample})$$

Lipid alterations were also quantified by normalising bands of interest by the total CH_3 content (Table 2.4).

Table 2.4: Method for estimation of lipid oxidation, saturation and unsaturation

	Molecular bonds of interest	Equation	Reference
Lipid oxidation	C=O/CH_3	$\frac{\int_{1,700\ cm-1}^{1,760\ cm-1} Spectrum}{\int_{2,950\ cm-1}^{2,975\ cm-1} Spectrum}$	[136]
Lipid saturation	CH_2/CH_3	$\frac{\int_{2,900\ cm-1}^{2,945\ cm-1} Spectrum}{\int_{2,950\ cm-1}^{2,975\ cm-1} Spectrum}$	[137]
Lipid unsaturation	C=C-H/CH_3	$\frac{\int_{3,000\ cm-1}^{3,020\ cm-1} Spectrum}{\int_{2,950\ cm-1}^{2,975\ cm-1} Spectrum}$	[137]

The statistical comparison was then performed using analysis of variance (ANOVA). This analysis was performed in R 3.6.2.

2.2.4.3 Spectral clustering

In order to classify tissue spectral features, we used uniform manifold approximation and projection (UMAP) to cluster all brain and spinal cord spectra, respectively. Community detection was performed with the Leiden algorithm[138]. Spectra for each cluster were averaged for identification of cluster-distinguishing features. Cluster classification was then spatially visualized in order to correlate these results with histology.

2.3 Single-nuclear RNA sequencing

2.3.1 Nuclear isolation, hashing and sequencing

Prior to nuclear isolation, two nuclear isolation methods were compared – firstly a previously reported method which included dissociation, filtration and debris

removal[131], and secondly a sucrose gradient method based on the use a of a commercially available nuclear isolation kit (Sigma-aldrich #NUC201). The latter was found to have optimal cell output and reduced debris, and was the method taken to the next step of testing. Nuclear labelling was then tested using a Fortessa X20 flow cytometry system to ensure nuclear hashing was viable.

Using the established protocol, 2-6 sections of 10 µm thickness were scraped from a slide and dissociated in a dounce homogeneiser and nuclei were isolated based on a sucrose gradient, using the nuclear isolation kit according to the manufacturer's recommendations. RNAse inhibitors (Invitrogen, #10777019, 4 U/µl reaction) were added to all solutions prior to each step and samples and isolated nuclei were kept on dry ice or ice in order to preserve RNA quality.

Due to the time sensitivity of this protocol, nuclear isolation and preparation was performed in parallel for multiple samples with Dr. Calliope Dendrou and Dr. Victor Yeung, totaling 6 samples per batch. This shortened processing time, thus preserving RNA quality.

Multiplexing was performed by washing and hashing nuclei from each sample with one of six unique TotalSeq^{TM}-A anti-nuclear pore complex antibodies (Biolegend, #A0451-456), which are DNA-barcoded antibodies which target the nuclear pore complex thus uniquely labeling nuclei from distinct samples[19].

Multiplexed nuclei from 6 samples were then counted by Dr. Hubert Slawinski from the Oxford Genomics Centre (OGC, Wellcome Centre for Human Genetics) using an

automated system, pooled and then approximately 50,000 cells were loaded per channel. All subsequent steps were performed by the OGC, in accordance wit the manufacturer's protocols. Following a pilot single nuclear sequencing was run without multiplexing and two multiplexing batches were run on separate occasions, each including nuclei from both case and control samples.

2.3.2 Data analysis

2.3.2.1 CellRanger

Our snRNAseq data and the publicly available raw data included in this analysis[130] were processed in CellRanger v 3.1.0 using the *Homo sapiens* (human) genome assembly GRCh38 genome as the reference.

2.3.2.2 Sample demultiplexing

Sample demultiplexing was performed by Dr. Fabiola Curión and Dr. Charlotte Rich-Griffin where, in order to obtain the maximum number of nuclei, DemuxEM[19] was used for antibody demultiplexing and Vireo[139] and Freemuxlet[140] were used for genetic demultiplexing based on reference SNPs. For this, a reference from the 1,000 genomes project with 7.4 million common variants was used. CellSNP[141] was used to produce input for Vireo and Dsc-pileup was used to produce input for Freemuxlet.

Cell demultiplexing classification was determined when two or more demultiplexing methods provided the same classification. Based on records of antibodies paired with each patient sample, demuxEM results were used to match genetically demultiplexed samples back to patient clinical data.

2.3.2.3 snRNA-seq pre-processing, integration and clustering

Preliminary quality control was performed by the Oxford Genomics Centre. Subsequent steps were performed as summarised in **Fig 2.2**. Sample quality control, batch correction and clustering were performed using the analysis pipeline written by Dr. Charlotte Rich-Griffin, Dr. Fabiola Curion and Dr. Tom Thomas. Previously published 10x Genomics snRNA-seq MS data sets generated by Jäkel *et al*[131] and Schirmer *et al*[130] were included in this analysis.

Following inspection of quality indicators such as total counts, number of genes, mitochondrial content, ribosomal content and doublet scores estimated using Scrublet[142]; the samples were filtered excluding: nuclei with <200 genes; all nuclei that did not satisfy two conditions simultaneously: >10% mitochondrial content and <700 total UMI counts; cells with a doublet score >0.25. Genes present in less than 3 nuclei were excluded. A total of 86,234 nuclei remained following initial filtration of the pooled data set from all studies.

Data pre-processing such as filtration, log1p transformation, data normalization, PCA, neighbor computing, UMAP, as well as downstream visualization steps were performed using the SCANPY library[143].

For integration, the optimal batch correction method identified was Harmony[144] when compared to BBKNN and Scanorama[18]. This decision based on local inverse Simpson index (LISI) score and visual inspection of UMAP with batch corrected data.

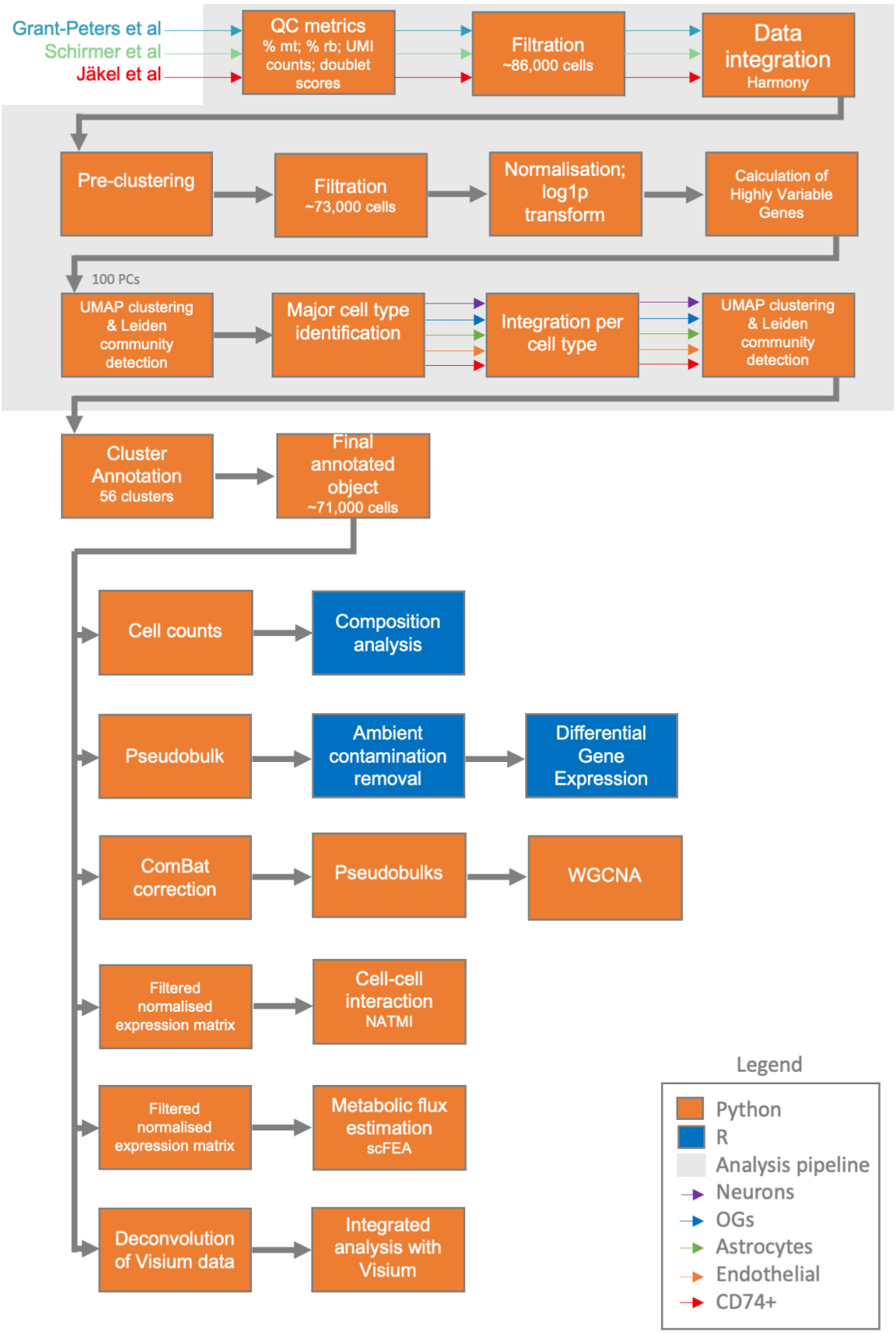


Figure 2.2: Flowchart summarising snRNAseq analysis

Using a plot of cumulative explained variance in relation to principal component (PC) number, the number of PCs used for downstream dimensionality reduction was 100 PCs. Using this input, UMAP[145] plots were generated and community detection of clusters was performed using the Leiden algorithm[138]. Low resolution cluster identification of broad cell types (neurons, oligodendrocytes, astrocytes, endothelia, microglia and leukocytes/microglia) was performed by inspection of the expression of marker genes and top differentially expressed genes. Using the identified barcodes in each of these major cell groups, each major cell group was integrated separately with Harmony and subclustered further.

The quality of the clusters was assessed by doublet scores, comparison of total counts, mitochondrial and ribosomal content within each cluster, as well as inspection of top differential genes. Low quality sub-clusters were identified in the neuronal and in oligodendrocyte cell types. These clusters were marked by at least two of the following features: (a) low total UMI counts; (b) lack of distinguishing markers; (c) large contributions from mitochondrial genes (d) high expression levels of markers from multiple major cell type groups simultaneously (e) higher doublet scores than average. These clusters were classified as low quality and of poor value for downstream analysis and were thus excluded.

In the CD74+ leukocyte/microglia, astrocyte and endothelial compartments certain clusters were identified which had high expression of markers from the

oligodendrocyte compartment (e.g. PLP1 and MBP). Whilst these clysters could represent doublets, they were retained for initial analyses as a prior study has suggested that such clusters could contain phagocytosing cells which had collected apoptotic oligodendrocytes or myelin debris along with relevant RNA[130]. Therefore, these clusters were retained to investigate their relative presence in cases and controls, hypothesizing that if genuinely phagocytosing they would show an increased abundance in MS patient samples given the demyelination occurring in these.

Following nuclear exclusion, the barcodes of remaining nuclei (n=71,078) were retrieved and the integration and clustering was performed again with the more stringent data set, resulting in a total of 54 clusters being identified (**Table 2.5**). This cell classification was used as the basis for downstream analyses. These analyses were all performed in Python 3.7.4 and R 3.6.2.

Table 2.5: List of all cluster levels and identifying markers

Major cell type	Layer 1	Layer 2	Layer 3
Neurons GABRB2 SNAP25	Excitatory neurons SATB2 RBFOX3	FEZF2+	FEZF2_TLE4hi_CDH24hi_SCUBE1
			FEZF2_SERMA3Ehi
			FEZF2_SEMA3Elow_LRRK1
			FEZF2_TLE4hi_ SERMA3Ehi_SCUBE1neg
		CUX2+	CUX2_LAMP5_PALMDneg_CNGB1neg_FREM 3neg
			CUX2_LAMP5_YWHAHhi
			CUX2_PALMDpos_ FREM3low_EPHA6hi
			CUX2_PALMDpos_FREM3pos
			CUX2_RORB_TXKhi
			CUX2_RORB_PRSS12
		RORB+	RORB_ARSJhi
			RORB_TOXhiARHGAP15hi
			RORB_GABRG1_MET
			RORB_YWHAHlow_METneg
			RORBhi_PRSS12_MME
		THEMIS+	THEMIS_GRIK3
			THEMIS_NTNG2
	Inhibitory neurons GAD1	VIP+	VIP_ABI3BP
			VIP_HGFAC
			VIPhi_EREG

			VIPhi_THSD7Bhi_PCDH18
			VIP+_FLT1_SCTR
		ID2+	ID2hi_LAMP5hi_NOS1
			ID2hi_LAMP5hi_NRG1hi
			ID2+_CXCL14hi
			ID2+_CXCL14hi_PAX6
		SST+	SST+
			SST+_FBN2+
		PVALB+	PVALB+
			PVALB+_UNC5+
OGs PLP1 MBP	**OGs** OPALINhi_PLXDC2hi		
	OGs OPALINlow_SLC5A11hi_RBFOXhi		
	OGs OPALINlow_SLC5A11low_RBFOXlow		
	OGs OPALINhi_FRYhi		
	OPCs PDGFRApos_PLP1neg		
	Early myelinating OPCs PDGFRApos_PLP1pos_BCAShi		
Astrocytes AQP4 GFAP	**Protoplasmic astrocytes** HILPDAhi_CHI3L1hi		
	Protoplasmic astrocytes SLC1A2hi_NRXN1hi		
	Reactive/Fibrous astrocytes CD44hi_S100Bpos_FOShi_VIMpos		
	PhagoProtoAstro MERTKhi_MEGF10hi		
	Doub-phagoAstro PLP1hi_MBPhi		
Endothelia LAMA2 CLDN5	**Pericytes** PDGFRBhi_NOTCH3hi		
	Fibroblasts COL1A1pos_COL15A1		
	Venous endothelial TSHZ2hi_VWFhi		
	Arterial endothelial BMX_SEMA3Ghi_EFNB2hi		
	Smooth muscle ACTA2_ CNN1		
	Erythrocyte Prog HBBhi_GYPAhi		
	Capillaries MFSD2Apos_TFRCpos_SLC16A1pos		
Microglia + immune cells CD74	**Doub-phago** PLP1hi_MBPhi		
	Perivascular Macrophages LYVE1pos		

	MicroMacro CPMhi_LPLpos		
	MicroMacro FTLhi_RPS_RPLhi		
	Homeostatic Microglia P2RY12hi_CX3CR1hi		
	Plasma cells IGHG1hi_IGHG3hi		
	T-cells CD8Ahi_CD4pos_IL7Rhi		
	Dendritic cells FCER1Apos_CLEC10Apos_HLADRB1pos		

2.3.2.4 Composition analysis

Cluster counts were converted to centered log-ratios (CLRs) using the ALDEx2 R package. The median CLR for each cluster-sample combination was used for PCA which was performed using prcomp (R version 3.6.2) with default parameters.

ANOVA was then used to identify confounding features contributing to top principal components; meta experiment (that is, the study which generated each data set) was identified as a contributing factor to several cell types and was therefore included in the subsequent composition model as a covariate.

Composition analysis was performed at the level of major cell type and the high-resolution clusters to identify differences in cell type composition when comparing cases to controls. Cell counts (prefiltered using the same criteria as PCA) were modelled using the glmQLFit function from edgeR with disease status and meta_experiement included as covariates. Differential abundance testing was performed using the quasi-likelihood F-test, and p-values were corrected using Benjamini-Hochberg adjustment.

Unlike our study and the Schirmer data set, the Jäkel study only sampled white matter areas and therefore could not be included in this analysis due to resulting skewing cluster frequencies given the lack of gray matter. A false discover rate cutoff (FDR)<0.1 was used. This analysis was performed by adapting published code[146].

2.3.2.5 Pseudobulk

In order to facilitate computational analysis, aggregated pseudobulk (Pb) data was used for certain analysis methods. In this approach, the sum of the counts of all single-cells from each donor for each cluster are used for comparisons. This is more computationally efficient and has been shown to have similar outcome to single-cell methods[147]. This was run by Melissa Grant-Peters in R 3.6.2 using code written by Charlotte Rich-Griffin.

2.3.2.6 Removal of ambient contamination

Removal of ambient contamination was performed at Pb level using the DropletUtils[148] library. Using raw Cell Ranger matrices the ambient contamination was estimated based on empty droplets, and the proportion of maximum contamination in each Pb was estimated for each cluster of that sample. For each cluster Pb this was weighted according to the representation of that cluster in that patient.

For each Pb, genes with an estimated proportion of maximum contamination >90% were excluded (circa 5-10 genes per Pb). In cases where samples were present in more than one sequencing run, the ambient contamination was calculated using the raw matrix from the run which had the largest number of outputted cells for that sample. For

sequencing runs with more than one sample, the same ambient contamination calculation was used for the background contamination estimation for all samples. This was performed in R 4.0.1.

2.3.2.6 Differential gene expression

Differential gene expression was estimated using Pbs at a major cell type and low resolution clustering level using EdgeR[149]; this compared the gene expression of cases versus controls while controlling for meta experiment. Genes simultaneously satisfying logFC> 1 and pvalue<0.05 were considered significant. This analysis was performed by Melissa Grant-Peters in R 3.6.2 using code written by Charlotte Rich-Griffin.

2.3.2.7 Gene ontology enrichment

Differential expression edgeR outputs were used for gene ontology enrichment analysis. For this, genes with a pvalue<0.05 were evaluated for pathway enrichment analysis using clusterProfiler[148] in R 4.1.0.

2.3.2.8 Weighted gene correlation network analysis (WGCNA)

PCA analysis of Pb data indicated meta experiment had a significant contribution to several PCs, therefore the data was batch corrected using ComBat[17]. Following batch correction, the cornet package implemented by the Sansom lab (University of Oxford) was used for WGCNA[150]. This analysis was performed by Charlotte Rich Griffin in R 4.0.1 and Python 3.9.1. All presented pvalues were adjusted using the Benjamini-Hochberg method.

2.3.2.9 Cell-cell interaction analysis

Cell-cell interaction analysis was performed at a single-cell level using filtered and normalised gene expression data for cases and controls separately. For this analysis clusters with indication of being doublets were excluded. Network Analysis Toolkit for Multicellular Interactions (NATMI)[151] was used with the lrc2p database as reference, since at the time of this work, NATMI had the largest database available, including other popular databases such as cellPhoneDB, and with the optional functionality of adding novel L-R pairs or interest. This was performed in python 3.7.4. Heatmaps were generated without filtration of edges, while communication networks filtered edges by specificity >0.1, where a specificity score of 0 indicates expression of this ligand-receptor pair is non-specific and 1 indicates the ligand is only expressed in a single cell type and the receptor is only expressed in a single cell type.

2.3.2.10 scFEA: Metabolic analysis

In order to estimate accumulation/depletion of metabolites and metabolic module flux we used filtered and normalized gene expression. The analysis of all single nuclei was performed jointly using single cell flux estimation analysis (scFEA)[152], which is the only computational tool available which we are aware of for estimating metabolic flux using transcriptomic data as input. Spatially resolved metabolic estimates were obtained by running scFEA on each Visium sample individually; scale bars for Visium visualization excluded the top and bottom 10% data points, as a way of controlling for outliers. Plots, quantitative comparisons and visualization were performed in python 3.7.4.

Metabolite estimation was compared for cases vs. controls within each bucket level cell annotation (astrocytes, CD74+ cells, endothelial cells, neurons, OGs). For this variance homogeneity was tested (Levene's test). Having identified that across metabolites and cells certain comparisons had homogeneous variance and others not, a non-parametric test (Mann-Witney U) was used across all comparisons in order to maintain a standardised approach. P-values were adjusted using Benjamini-Hochberg correction.

2.4 Spatial transcriptomics: 10x Genomics Visium

2.4.1 Sample preparation

Brain samples were sectioned with 10 µm thickness on a Bright 5040 cryostat in the Oxford Department of Physiology, Anatomy and Genetics with sample temperature at -13°C and chamber temperature -21°C. In order to obtain sections with a 6mm x 6mm area, containable within the active area of the Visium slides, the tissue samples were scored using a custom blade made by the Old Road Campus Research Building workshop. Regions of interest in control samples included, where possible, the transition between white matter and grey. For case samples, regions of interest spanned across NAWM and lesion areas, and where possible included grey matter as well. Where there were indications of immune cell infiltration (e.g.: hypercellular regions), these regions were prioritized. Sections were placed in the active area of the slide and stored at -80°C for < 5 days until use.

2.4.2 Optimisation of permeabilisation conditions

Using the Visium Tissue Optimisation kit (10x, protocol CG000238 Rev A) sample permeabilisation times were compared following the manufacturer's instructions; permeabilisation duration of 18 minutes was identified as optimal both for brain and spinal cord. These permeabilization conditions were used hereafter for all protocols. Fluorescence imaging of the tissue optimization protocol was performed in the Old Road Campus Research Building microscopy facilities with the support from Rhodri Wilson. Images were collected on a Nikon Eclipse system at 10X magnification.

2.4.3 Gene expression protocol

We performed the gene expression protocol according to protocol version CG000239 Rev A. Briefly, we stained samples with H&E and imaged on a Zeiss RoboPalm Laser Capture Microdissection system at 10X magnification. This was followed by tissue permeabilisation (18 minutes). In all subsequent steps, the protocol was performed with support from Dr. Victor Yeung for optimal processing time and preservation of RNA quality. The thermocycler used for all steps conducted by the Dendrou lab was a Thermo Fisher Scientific Applied Biosystems Verity.

Reverse transcription was performed, followed by second strand synthesis and cDNA denaturation in 0.08 M solution of potassium hydroxide (Millipore Sigma, #P4494). The solution from each well was transferred into individual tubes with 5 µl Tris-HCl (1M, pH 7.0) (Thermo Fisher Scientific, #AM9850G).

PCR optimal cycle number was determined using a QuantStudio system, where 1 µl of each sample was amplified using KAPA SYBR FAST qPCR Master Mix with low ROX (Sigma Aldrich, #KK4600). The qPCR protocol recommended by 10x Genomics was followed and the number of cycles at 25% of the RFU plateau for each sample was determined and ranged from 17 to 20 cycles. This cycle number was then used for PCR amplification of the remaining sample volume. Samples were then stored at -20°C and handed over to Dr. Hubert Slawinski from the OGC for all subsequent processing and sequencing, which was performed in accordance with the manufacturer's instructions.

2.4.4 Data analysis

2.4.4.1 Histological image tiling and alignment

Histological images collected on the Zeiss RoboPalm LCM system were tiled using FIJI package of ImageJ version 1.0.0-rc-69/1.52p; white balance correction was performed using a macro written by Vytas Bindokas; Oct 2006, Univ. of Chicago with modifications by Patrice Mascalchi. The images were then loaded and aligned in relation to fiducial frames for spatial gene expression mapping using Loupe Browser, where the number of active spots covered by tissue was manually determined. Dr. Charlotte Rich-Griffin used scanpy[143] pre-processing of the Visium data. Following QC inspection, Visium data were filtered based on total gene counts (brain> 500; 500<SC<15,000).

2.4.4.2 Deconvolution using snRNA-seq data

The SCVI deconvolution tool destvi and was used to deconvolute 10x Visium data with snRNAseq annotated data excluding doublet populations[153]. Mitochondrial genes were excluded from the model and default parameters were used.

This method assigned cell type proportions to each 10x Visium spot based on input snRNAseq clustering annotation and was performed at a low cluster resolution as well as high cluster resolution. Separate models were used to train for brain and SC data, respectively, although using the same snRNAseq data as reference. This was performed using scripts written by Dr. Charlotte Rich-Griffin.

2.4.4.3 Spatial co-localisation network analysis

Spatial co-localisation network analysis was performed using the deconvoluted low resolution spatial data. In this analysis the top three cell types with most likely contribution to a given spot were assumed to be co-localised in that spot. Co-localisation relationships were scored in an adjacency matrix. This adjacency matrix was used as input for network generation using the NetworkX library on python 3.7.4. Cell types with co-localisation relationships <1.5% of total were not included in the network for ease of visualization.

2.5 Spatial proteomics

2.5.1 Sample preparation

Samples were sectioned for proteomic analysis onto 1.0 PEN membrane slides (Zeiss, #415190-9041- 000). These were stained with H&E staining according previously reported spatial proteomics methods[132]. In short, samples were fixed in 70% denatured alcohol, rehydrated in water and stained with haematoxylin. Tissue was subsequently washed in distilled water (Invitrogen™, #10977035) and stained with bluing buffer followed by another wash. Finally, the samples were stained with eosin, washed a final time and dehydrated in increasing concentrations of denatured alcohol. All samples were stored in the -80°C until next processing step.

2.5.2 Laser capture microdissection

Using a Zeiss RoboPalm LCM system, a total area of 100,000 μm^2 was processed in each tube; microdissection was performed in 9 subareas (each area approximately 11,111 μm^2), since laser catapulting of tissue is more effective and precise with smaller tissue sections. For each tube, all 9 subareas were captured in 200 µl PCR tube cap with 20 µl of Radioimmunoprecipitation assay buffer (RIPA; Pierce, #89900). The number and location of captured areas was determined based on µFTIR and Visium data; 89 samples were captured in total. Following collection, the samples were frozen immediately on dry ice and stored at -80°C for <3 days until proteomic preparation.

2.5.3 Sample preparation for proteomic analysis

The sample processing for spatial proteomics was performed as previously reported[132]. Briefly, the RIPA buffer containing the microdissected samples was thawed for 30 minutes followed by centrifugation. To ensure maximum protein capture, the caps were rinsed, followed by a second centrifugation. Benzonase was added and left to incubate at room temperature for 30 minutes, followed by dithithreitol (DTT) at a final concentration of 5mM was added and tubes were incubated for 30 minutes. Iodoacetamide (IAA) at a final of concentration 20 mM was added and incubated for 30 minutes. Sera-Mag-A and Sera-Mag-B Bead mix was added and mixed, followed by acetonitrile (ACN; 70% final concentration), tubes were gently tap mixed and incubated for 18 minutes at room temperature for protein to bind to beads.

Samples were placed on magnets for 2 minutes and supernatant was discarded. Samples were rinsed twice with 70% ethanol for 30 seconds on magnet. Beads were rinsed with 180 µl 100% ACN for 15 seconds on the magnet. A solution of trypsin in 50 mM ammonium bicarbonate was added and beads were resuspended. Samples were digested overnight at 37°C.

Following digestion, samples were bath sonicated and 180 µl of ACN was added and tap mixed. This was incubated for 18 minutes off the magnet and for 2 minutes on the magnet. All supernatant was discarded and resuspended in 5µl of 2% ACN, incubated on magnet for 5 minutes and carefully transferred to vials containing 500 nL 1% formic

acid. Finally, samples were stored at -20°C and handed over to Dr. Roman Fischer's group (Target Discovery Institute, Oxford) as a mass spectrometry submission.

2.5.4 Data analysis

MaxQuant pre-processed mass spectrometry data were received and downstream analysis was performed in R 3.6.2 using Bioconductor DEP package. Initial filtration steps filtered both samples and proteins: samples with <20% of proteins present were excluded; proteins represented in <65% of samples were excluded. The 65% threshold was determined based on plotting the number of samples each protein was represented in vs. the mean intensity of that protein across samples. Following filtration, the data were log transformed and median-centred. Knn, MinProb, QRILC and random forest imputation methods were compared. Of these, random forest had the better performance with a more conservative imputation approach.

Hierarchical clustering was performed to categorise samples. In order to ensure reliability and minimize random effects from the imputation processes and hierarchical clustering, imputation and clustering were performed iteratively 10 and 100 times, with cluster classification being recorded for each sample was recorded for each iteration. Our assessment determined that cluster classification was reliable, with most samples clustered in the same group between 97-100% of iterations. A small number of samples had a lower reproducibility rate, but the rate of same classification was still 66%. In light of this, all subsequent analysis was performed using 10 iterations, due to demonstrable efficacy of the smaller number of iterations whilst having a lower computational cost.

Cluster number cutoffs were assessed separately for brain (k=4) and SC (k=5) samples to partition the major detected clusters.

Following this, differential expression of proteins was calculated in a pair-wise fashion for each cluster, with proteins with log fold change >1 and $p<0.05$ being considered statistically significant. Differentially expressed proteins (DEPs) were used for heatmap cluster visualization. Paired comparisons were performed for WM and WM lesion clusters and visualized in volcano plots. DEPs for each pair-wise comparison were analysed using gene set enrichment analysis (GSEA), of which the top 10 pathway hits were reported here.

Chapter 3: Macromolecular analysis of SPMS

3.1 Introduction

For several decades, vibrational spectroscopy methods, such as FTIR spectroscopy, have contributed valuable insights into biochemical changes related to pathology. The label-free nature of this method lends itself well to high resolution characterisation of complex biological samples with minimal sample processing for an unbiased profiling approach. Vibrational spectroscopy probes the inherent absorption patterns of molecules in a given sample by measuring their vibrational modes as a result of excitation by incident light[154]; in the case of FTIR spectroscopy, the wavelength of the incident light is within the infrared range. For each vibrational mode, all atoms vibrate at certain characteristic frequencies given in wavenumbers units (cm^{-1}) which are recorded with a detector. The collection of absorbance measurements within a wavelength range makes up an FTIR spectrum.

Generally, vibrational modes can be used to identify a corresponding specific functional group under the assumption that the rest of the molecule is stationary. This is leveraged in the analysis of a captured spectrum: specific spectral features can be used to identify the presence of a corresponding functional group. In most cases a molecule will have a unique spectrum which can make it identifiable within a sample – this is the IR fingerprint of a molecule[154].

Extensive spectra/structure correlation tables – also known as Colthup charts – have been compiled for systematic assignment of absorption bands to their corresponding functional group. Using this resource, the composition of complex multi-molecular samples can be identified or, in cases of particularly complex samples, approximated[134, 155]. This makes spectroscopy a unique method for profiling inherently complex samples such as biological samples – particularly given that it can capture in high resolution biologically relevant macromolecules such as lipid, protein and nucleic acid content.

In spite of how well developed and mature the spectroscopy field is in the context of the physical sciences, spectral interpretation for biochemically complex samples can still be challenging due to the wealth of information captured in a spectrum frequently resulting in overlapping spectral features. Changes in biochemical composition, therefore, not only manifest as the presence or absence of an absorbance peak, but can have subtle qualitative effects on peak position, width and height, convoluting spectral decomposition and molecule identification. In light of the spectral interpretation challenges, analysis of biological samples often involves a relative comparison of two or more groups in relation to each other and identification of alterations in features (**Fig 3.1**). More recently, the fingerprinting property of FTIR spectra has been harnessed for categorization of glioma and amyotrophic lateral sclerosis samples based on machine learning approaches for diagnostic purposes[156, 157].

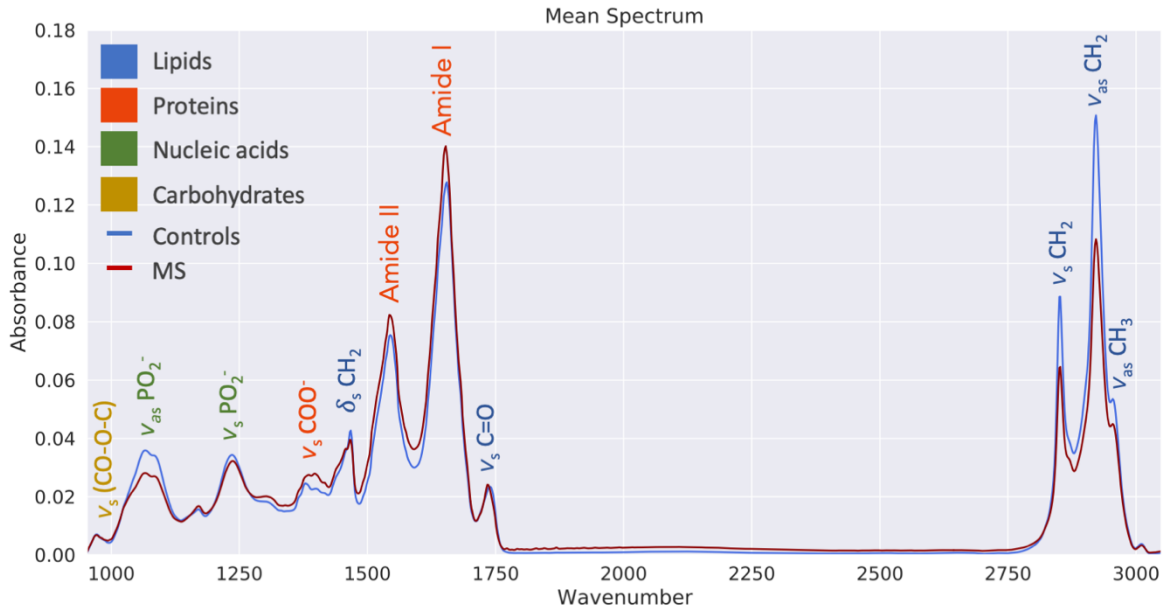


Figure 3.1: Example of mean FTIR spectra comparison for cases and controls with assignment of major absorbance peaks. This was the methodological approach most commonly used in the study of disease with spectroscopy methods.

Newer FTIR approaches for the analysis of tissue biochemical composition have not, however, been applied to MS as yet. Older studies collected spectra based on histological classification – lesion, white matter, or grey matter – and averaged the spectra from each category. These averages were then compared; the most prominent spectral change observed in MS lesions was the reduced lipid content, visible in the 2,800-3,000 cm^{-1} range, which was attributed to the process of demyelination in lesions[158, 159]. A more subtle alteration in lipid content was also noted: an increase in the C=O stretching peak at ~1,700-1,760 cm^{-1} (also known as carbonyl ester region). This was first assigned to be the C=O stretching bands of phospholipids – where the band results from two closely neighbouring peaks, one hydrogen bonded and the other non-hydrogen bonded. Its widening profile and increased peak in MS was suggested to relate to the disruption of bilayer packing caused by cholesterol, protein and the unsaturation of acyl chains, resulting in the presence of a wide range of C=O populations with varying degrees of hydration cumulatively contributing to this peak[158]. A second study interpreted this peak specifically as an indication of oxidative processes [159]. Supporting this interpretation, this spectral feature has been observed in *in vitro* cell studies in the context of lipid oxidation products and degradation of acyl chains[160, 161]. The third spectral change reported in MS are alterations in the amide 1,500-1,700 cm^{-1}; in particular in the amide I/amide II ratio which has been reported to be 1.4-2.5:1 when compared to controls. This has been hypothesized to relate to non-protein contributions

such as DNA or water in the amide I peak[158]. A broadening of the amide I peak has also been reported and suggested to relate to protein oxidation[159].

The inclusion of both RRMS and SPMS patients without endeavouring to characterise the lesion subtypes via other methods assumes that biochemical changes in both MS subtypes are the same. Since these studies were published, not only have MS subtypes been characterised more thoroughly, but the heterogeneous range of lesion classifications/stages has also become evident[66, 105]. Apart from these developments in the understanding of MS pathophysiology, technological changes have also emerged in the field of spectroscopy which significantly improve the applicability of this method to histological characterisation – firstly with increased spectral resolution and secondly with the coupling of spectroscopy systems with a microscope, known as microspectroscopy. This method enables the spatially resolved acquisition of spectral data in high resolution, generating maps of various biologically relevant macromolecules.

To our knowledge, our study presents the first example of the application of spatially-resolved FTIR spectroscopy to MS and control post-mortem brain and spinal cord tissue. Here we have used synchrotron-based μFTIR (sμFTIR) for biochemical characterisation of lesions and perilesional/NAWM areas – this approach acquires data from small areas but at high spectral resolution. We have also used conventional μFTIR (cμFTIR) equipped with an FPA detector, which enables detection of several spectra simultaneously to scan larger spatial areas for spatial characterisation of the transition at lesion borders. The latter dataset was then partitioned using unsupervised machine

learning in order to assess whether the known lesion heterogeneity was evident at a biochemical level.

Importantly, in our study design we have included perilesional and NAWM areas surrounding lesions which have not been included in previous analyses. By doing this, we intend to outline the transition from lesions to surrounding tissue (with cμFTIR) but also identify whether there are biochemical alterations in the perilesion which occur in the absence of demyelination (with sμFTIR).

3.2 Results

3.2.1 Synchrotron-based μFTIR profiling of biochemical changes within the SPMS-affected brain

SPMS lesions and NAWM in the human brain were biochemically profiled using synchrotron-based μFTIR (sμFTIR); a total of n=953 spectra were collected, where each spectrum scanned a 15 x 15 μm^2 area with overlap to total a resolution of approximately 5 x 5 μm^2 such that the resolution was comparable to cμFTIR. Thus, total scanned area was 4,765 μm^2 across all samples, with an average scanned area of 953 μm^2 per sample.

3.2.1.1 Principal component analysis of the brain sμFTIR spectra

Following quality control (as detailed in Methods section 2.2.4.1), the second derivative of each spectrum collected by sμFTIR from the MS brain samples was averaged and plotted along with the standard deviation (**Fig 3.2A**). Indicative of good

spectral quality, this plot showed prominent features characteristic of biological - and in particular CNS – samples including high amplitude signals in region corresponding to lipids (2,800-3,000 cm^{-1}) as well as the region corresponding to proteins (1,500-1,700 cm^{-1}). To identify the spectral contributors of this data set, principal component analysis (PCA) was used. The cumulative explained variance (CEV) required >250 principal components to explain 99% of the variation in the data, indicating high complexity (**Fig 3.2B**). However, the first principal component (PC1) explained a relatively large proportion of the overall variation (explained variance, EV=60.12%) of the data. The eigen-spectra for the top 3 PCs were plotted to evaluate the main biochemical features of each PC; the main feature in PC1's eigen-spectrum was high amplitude in the 2,800-3,000 cm^{-1} range, corresponding to lipids (**Fig 3.2C**). PC2 and PC3 (EV=8.37% and 3.74%, respectively) also had contributions in the lipid region. PC2 also had a strong signal at the carbonyl ester C=O region (~1,740 cm^{-1}), whilst PC3 also had a significant contribution in the amide I region (1,500-1,700 cm^{-1}) (**Fig 3.2C**). Scatter plot visualization of the spectral scores for the top 3 PCs revealed that spectra corresponding to lesional brain tissue had a smaller spread (suggestive of less variation in lipid content), while spectra corresponding to NAWM were more varied (**Fig 3.2D**). In particular, lesional spectra tended to have a negative PC1 and PC2 score, indicating lower lipid and carbonyl ester content. The PC1 and PC3 scatter plot projection segregated the lesional and NAWM spectra, suggesting that the main contributing features of these PCs, in the lipid and amide I regions, are the most important hallmarks

for distinguishing these tissue areas. Interestingly, PC2 scores were highest for a subset of the NAWM spectra and negative for other spectra, suggesting that NAWM can be further subdivided based on carbonyl ester content.

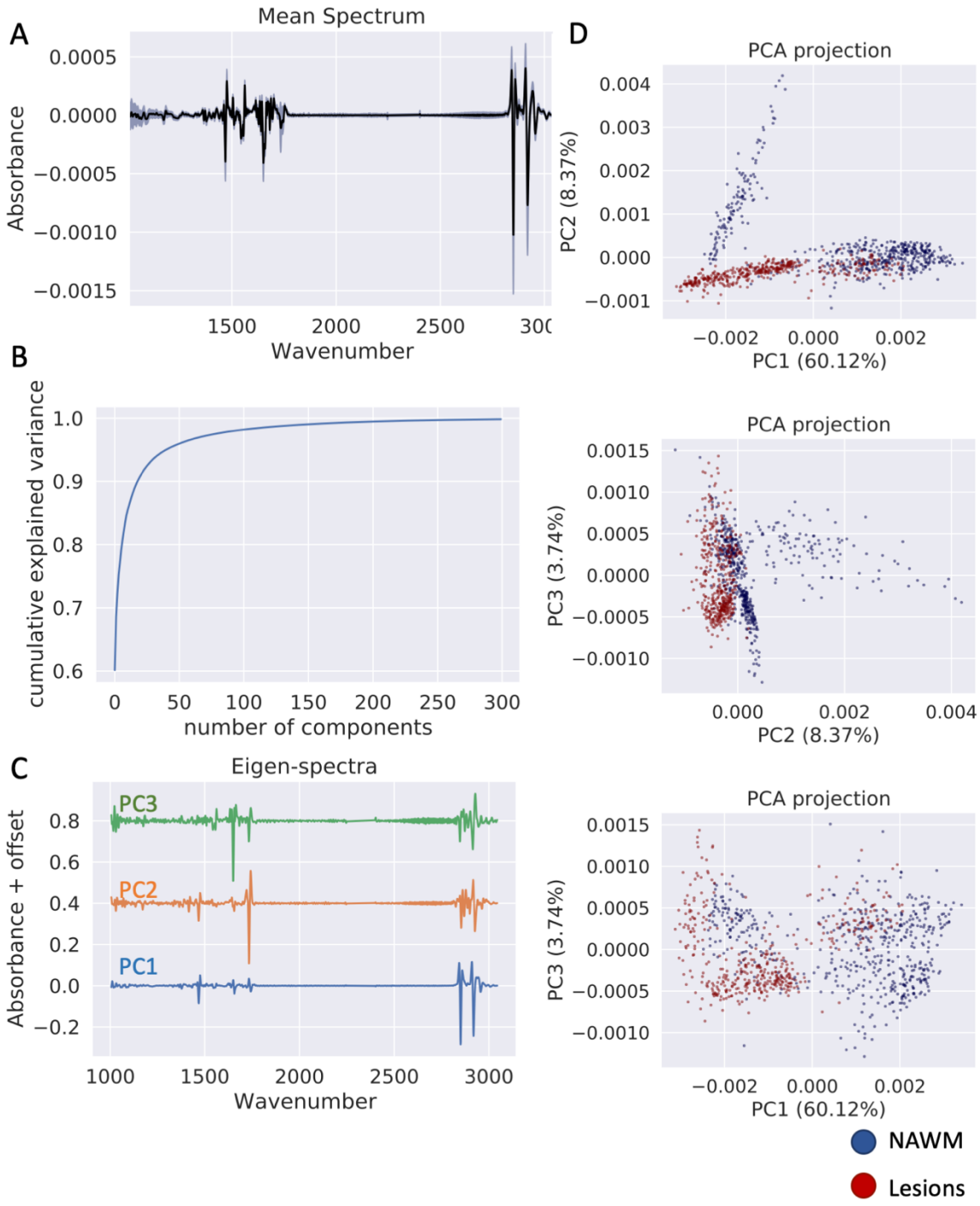


Figure 3.2: Principal component analysis of sμFTIR brain samples. (A) Second derivative of mean spectrum with standard deviation. (B) Cumulative explained variance of the first 300 PCs. Over 200 PCs are required to achieve CEV > 99%. (C) Eigen-spectra for top 3 PCs. PC1 has distinguishable peaks at CH2 and CH3 symmetric and asymmetric (2,800-3,000 cm-1), bending vibrations (1,470 cm-1); PC2 is marked by a high amplitude ester carbonyl peak (~1,740 cm-1); PC3 is marked by amide peaks (1,500-1,700 cm-1). (D) Scatter plot visualisation of top 3 PCs. PC1 and PC3 distinguish WM from lesions with fair accuracy; lesion spectra are marked by negative PC1 and PC2 scores.

3.2.1.2 Contrasting the macromolecular features of MS brain NAWM vs. lesions

In order to determine compare the biochemical profile of NAWM and lesions, the content of key macromolecules of these two groups were compared. This was performed by integrating the area below the region of interest in the spectral curve. As would be expected based on histological characterisation of myelination, lipid content in NAWM was estimated to be ~1.6 times higher on average than the lipid content in lesion ($p<0.0005$) (Fig 3.3). The carbonyl ester (1700-1,750 cm^{-1}) content was also higher in NAWM to an analogous degree, being ~1.5 higher than the lesions ($p<0.0005$). Notably, consistent with the PC2 scores, there was a large range in the NAWM carbonyl ester content (~0.0-0.8 absorbance·cm^{-1}) . This was represented as two subpopulations of spectra where part had > 0.2 absorbance·cm^{-1} and the second had carbonyl ester content <0.2 absorbance·cm^{-1}. Lipid oxidation, saturation and unsaturation were estimated based on biochemical alterations (see methods section 2.2.4.2): lipid oxidation in NAWM was ~0.8 times lower than in the lesion, whilst lipid saturation and unsaturation were ~1.15 and 0.8 times that found in lesions, respectively (all $p<0.0005$).

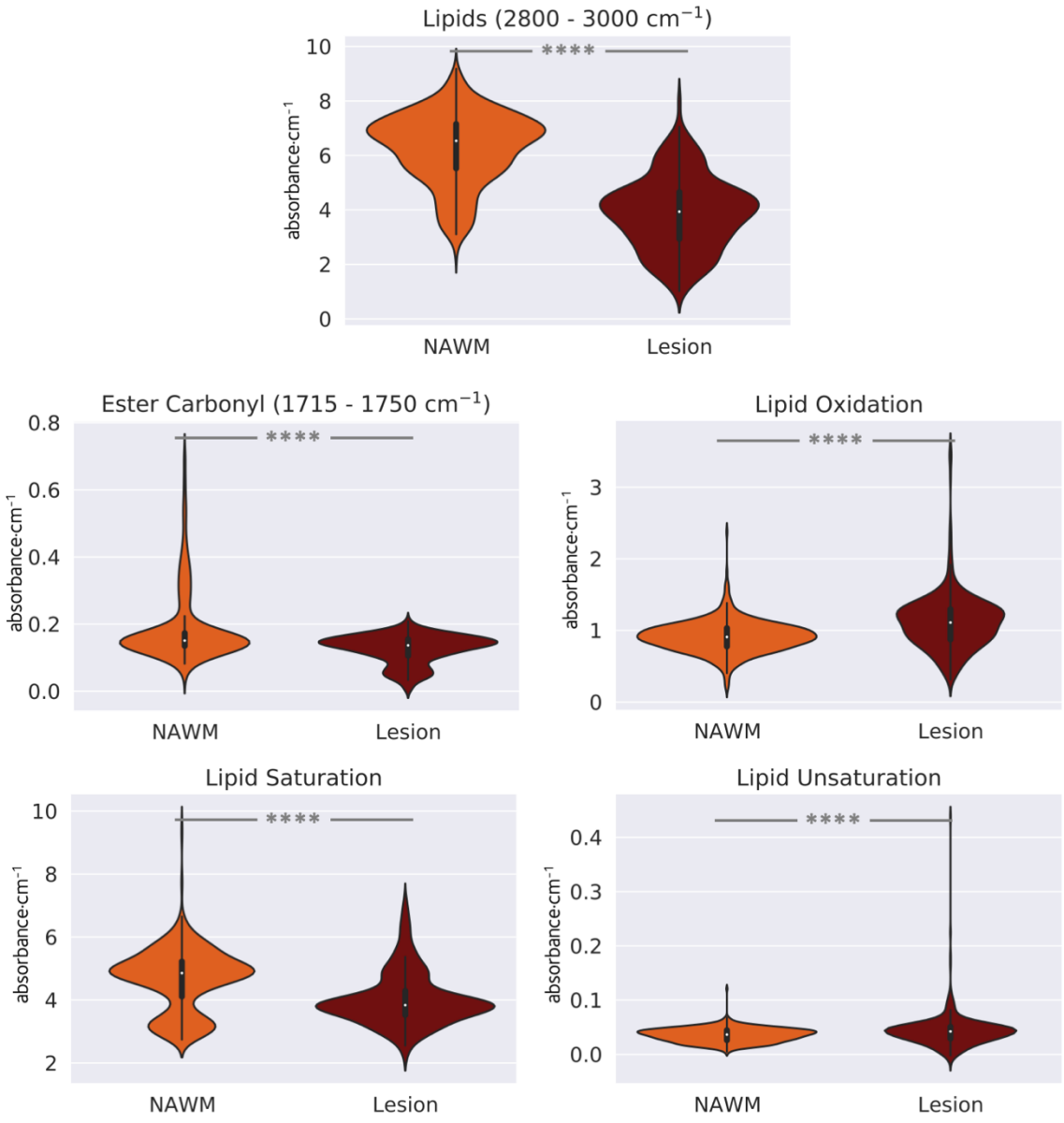


Figure 3.3: Contrasting lipid content and state comparison across NAWM and lesion spectra collected from MS-affected brains with sμFTIR. NAWM was marked by higher lipid content, ester carbonyl and lipid saturation, while in contrast the lesion was marked by higher lipid oxidation and lipid unsaturation. All statistical comparisons were statistically significant. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

Similarly to lipids, differences in mean protein content were estimated by integrating the area below spectra within the 1,500-1,700cm^{-1} region. The overall protein content of the NAWM was ~0.9 times that of lesions ($p<0.0005$) (**Fig 3.4**). Further to assessing the overall protein content, FTIR spectra can also be used to estimate the contribution of protein secondary structure, including α-helix, β-turn and β-sheet content which can be indicative of protein misfolding and accumulation, both of which are common features of neurodegenerative disorders. The high spectral resolution of sμFTIR data makes it particularly suitable for this estimation. NAWM had a ~10% higher α-helix content than NAWM, but 9% lower β-sheets, 4.34% lower β-turns and 1.87% less random folding (all $p<0.0005$) than lesions.

3.2.2 Conventional μFTIR profiling of biochemical changes reveals lesion heterogeneity in SPMS-affected brain tissue

In order to spatially characterise large areas of tissue spanning lesional and perilesional regions, cμFTIR was used scanning a 1,367 μm x 292 μm area per sample. This approach had half the spectral resolution of sμFTIR (one data point per 1.928 cm^{-1}, compared to one data point per 0.964 cm^{-1}), but enables scanning of larger areas.

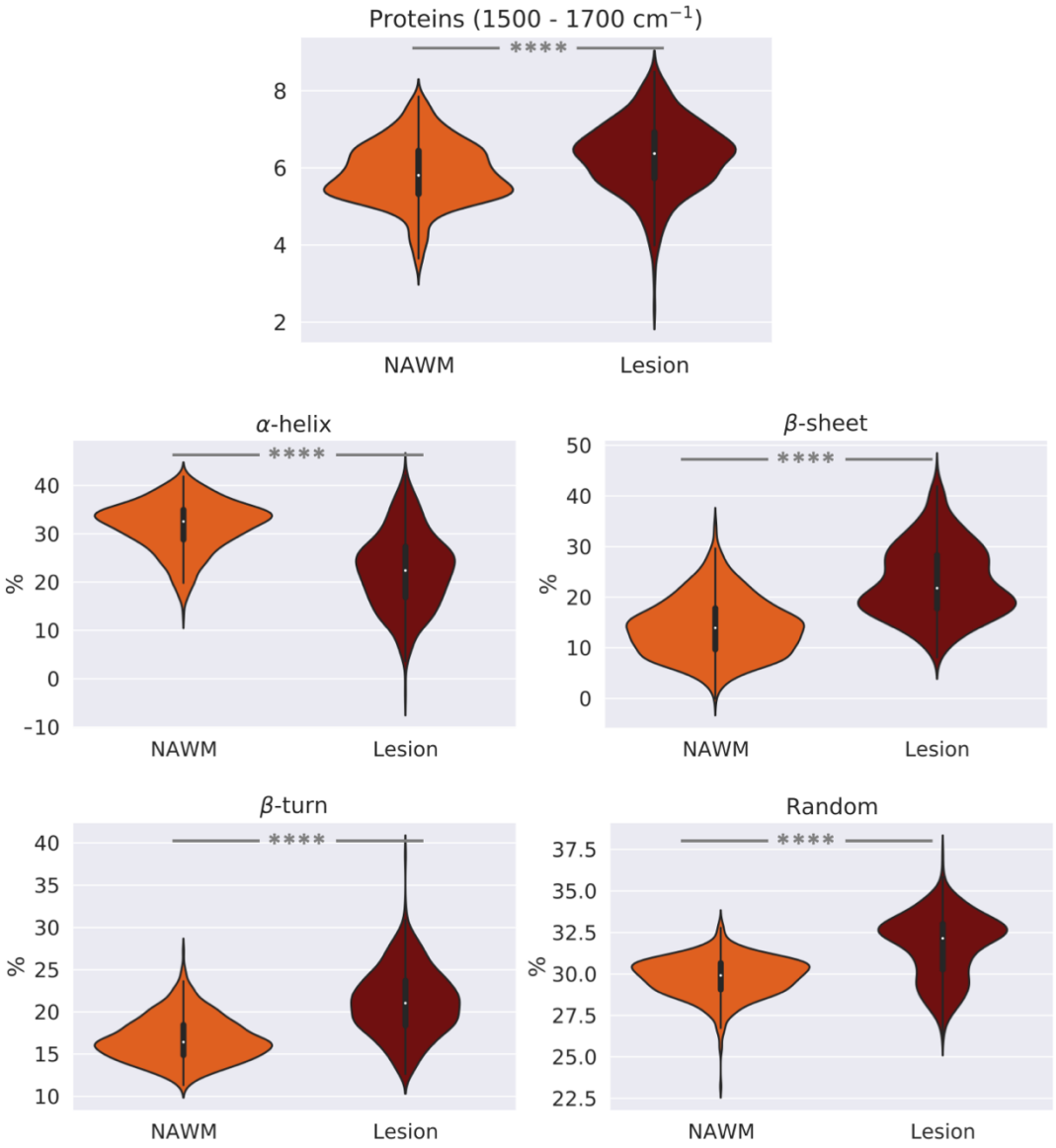


Figure 3.4: Comparison of protein content and protein secondary structure contribution to NAWM and lesion. This was estimated based on sμFTIR data collected from MS-affected brains and indicates significant protein differences across classifications. The NAWM was marked by lower total protein content, higher $\boldsymbol{\alpha}$-helix and lower $\boldsymbol{\beta}$-sheet, $\boldsymbol{\beta}$-turn and random folding content than the lesion. All differences were statistically significant. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

3.2.2.1 Principal component analysis of the cμFTIR spectra

A total of 163,840 spectra (16,384 spectra in a 320 x 1,280 μm region per sample) were collected of which 144,000 remained following QC. Similarly to the sμFTIR data, the mean second derivative and standard deviation of the spectra confirmed the high quality of the data set following pre-processing (**Fig 3.5A**) and PCA demonstrated data complexity as again >250 PCs were required for CEV >99% (**Fig 3.5B**). The eigen-spectrum for PC1 (EV=65.77%) showed variation in the lipid region as the primary contributor to the variation of the data. PC2 (EV=8.41%) and PC3 (EV=6.19%) had high amplitude contributions in the amide region (1,500-1,700 cm^{-1}) (**Fig 3.5C**). Notably, among the top 3 PCs, scatter plot distributions indicated that no single PC could fully segregate the SPMS from the control spectra (**Fig 3.5D**). We posited that this was because data capture from larger tissue areas in both the cases and controls also captures greater spectral diversity as lesional, perilesional, and NAWM areas were scanned in the cases and WM and GM transitions were scanned in the controls.

3.2.2.2 Contrasting the macromolecular features of SPMS and control brain tissue cμFTIR spectra

The integrated area below the spectral curves in the lipid and amide regions for cases and controls were compared as a quantitative metric of biochemical composition. When assessing the lipid content of cases relative to controls, SPMS samples had a lower mean lipid content, 0.77 times that of controls (**Fig 3.6**).

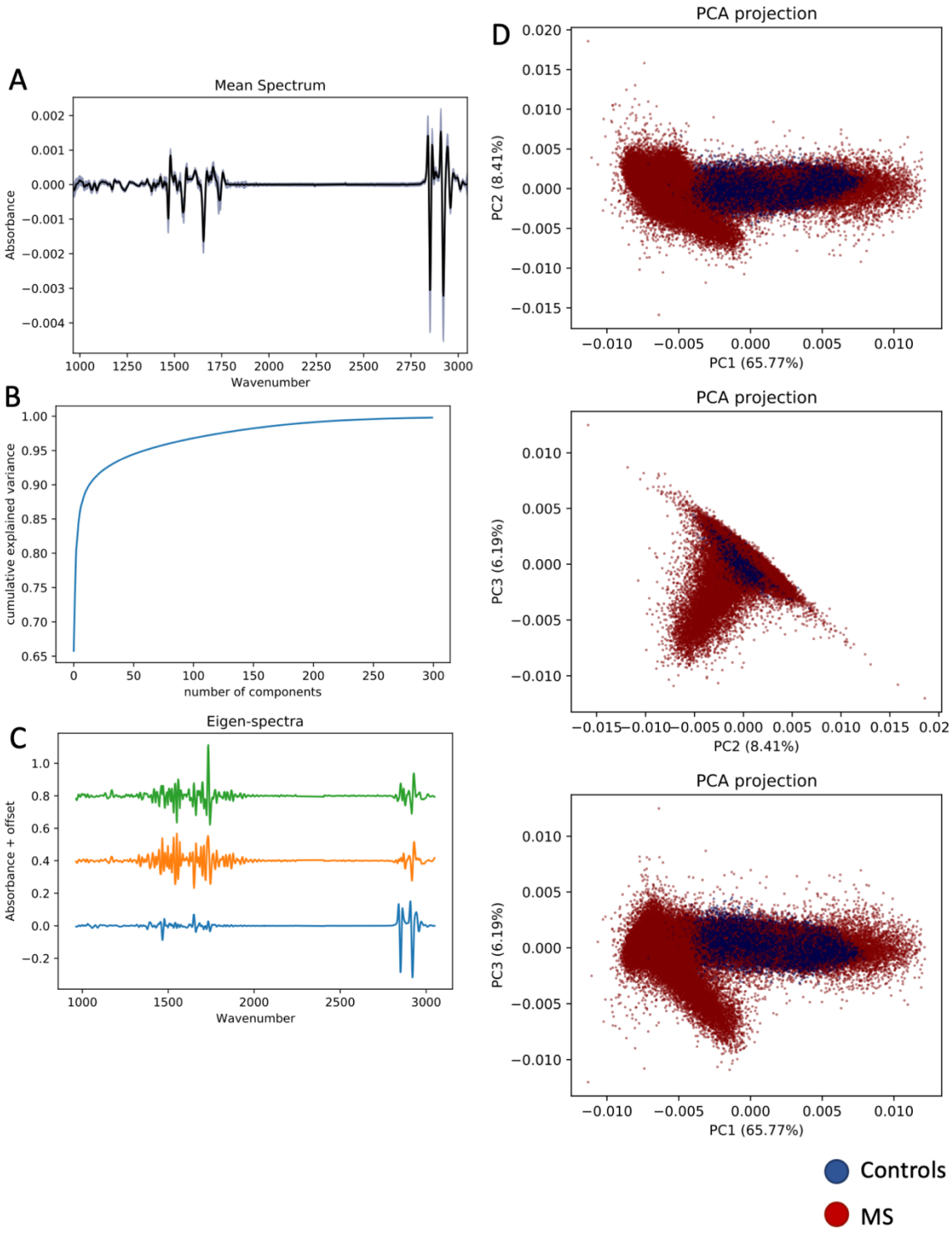


Figure 3.5: Principal component analysis of cμFTIR brain samples. (A) Second derivative of mean spectrum with standard deviation. (B) Cumulative explained variance of the first 300 PCs. Over 250 PCs are required to achieve CEV > 99%. (C) Eigen-spectra for top 3 PCs. PC1 has distinguishable peaks at CH_2 and CH_3 symmetric and asymmetric (2,800-3,000 cm^{-1}), bending vibrations (1,470 cm^{-1}); PC2 is marked by a high amplitude peaks in the amide region (1,500-1,700 cm^{-1}). (D) Scatter plot visualisation of top 3 PCs cannot clearly distinguish cases from controls. Controls have a smaller variation in PC scores, while MS has greater variation in all three top PCs.

There was also a trend towards cases having a lower lipid saturation level (0.87 times that of controls), and higher lipid oxidation was observed among cases (1.17 times that of controls). The carbonyl ester content was highly variable within cases (range of -0.02–1.3 absorbance·cm^{-1}), but more homogeneous across controls (range of 0–0.3 absorbance·cm^{-1}). The SPMS versus control contrasts relating to the lipid features were not statistically significant. However, the violin plots showed bimodal or trimodal distributions of the different lipid features, indicating a further underlying structure in the spectral data that was not well-captured by the case vs control contrasts. This could relate to the diversity of the tissue areas scanned for both the cases and controls. For example, the violin plot of the control lipid content is suggestive of at least two spectral subpopulations (**Fig 3.6**), which could correspond to WM and GM, with the former having a higher lipid content than the latter, given its enrichment in lipid-high myelinated neuronal axons as opposed to neuronal cell bodies which mainly localise to the GM.

Comparisons between cases and controls were not statistically significant for the protein-related features in this data set. There was a weak trend towards the mean protein content of cases being higher than that of controls by 1.1 times, whilst mean differences in protein secondary structure features were largely negligible (**Fig 3.7**). However, the overall range of these features was much greater in cases compared to controls, suggesting the likely existence of areas within the SPMS brain tissue displaying more extreme protein-related features.

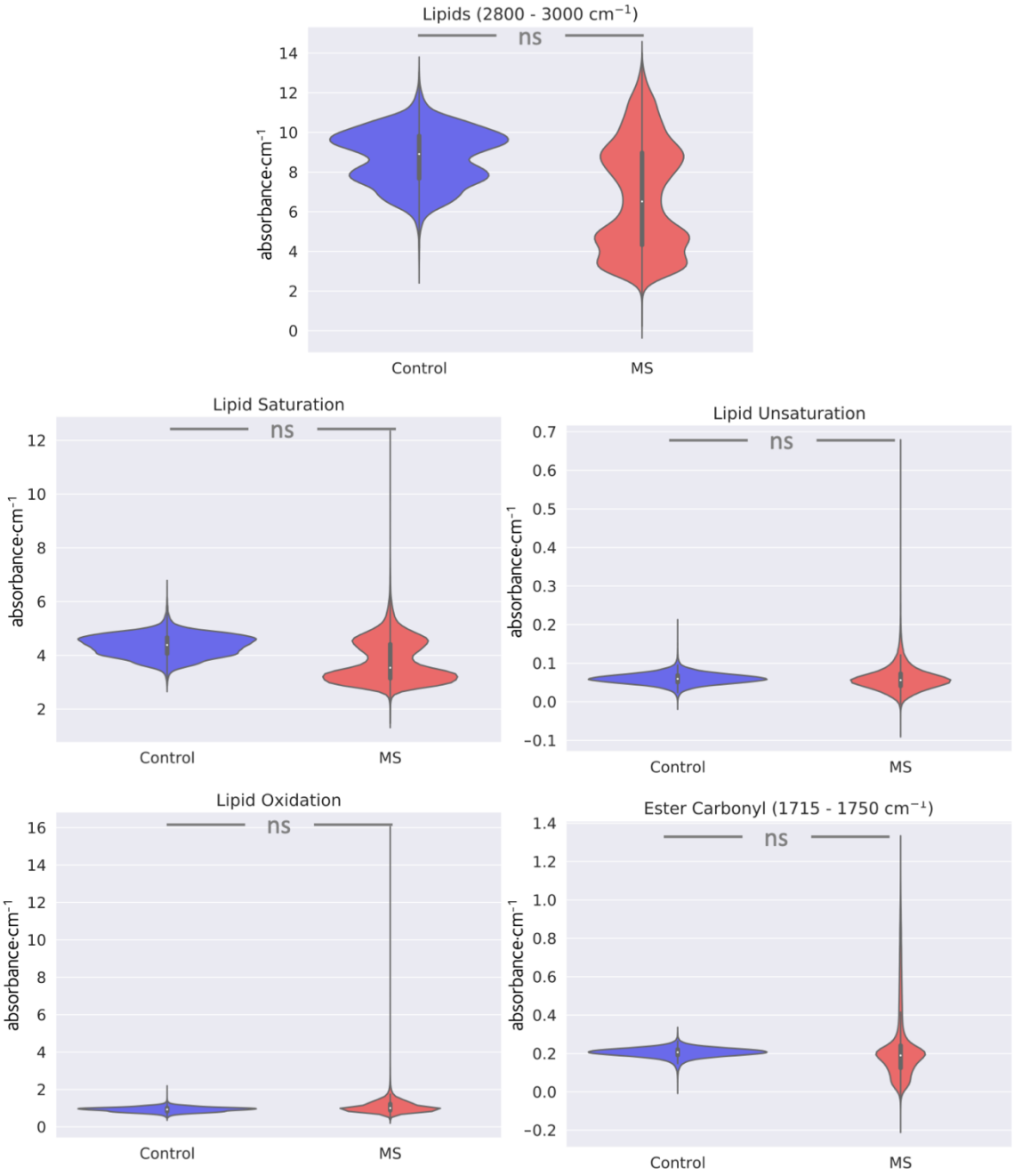


Figure 3.6: Comparison of lipid content and state in cases and controls reveals trends. Control samples are marked by higher mean lipid, lipid saturation and ester carbonyl content, while cases are marked by higher lipid oxidation. These differences were not statistically significant. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

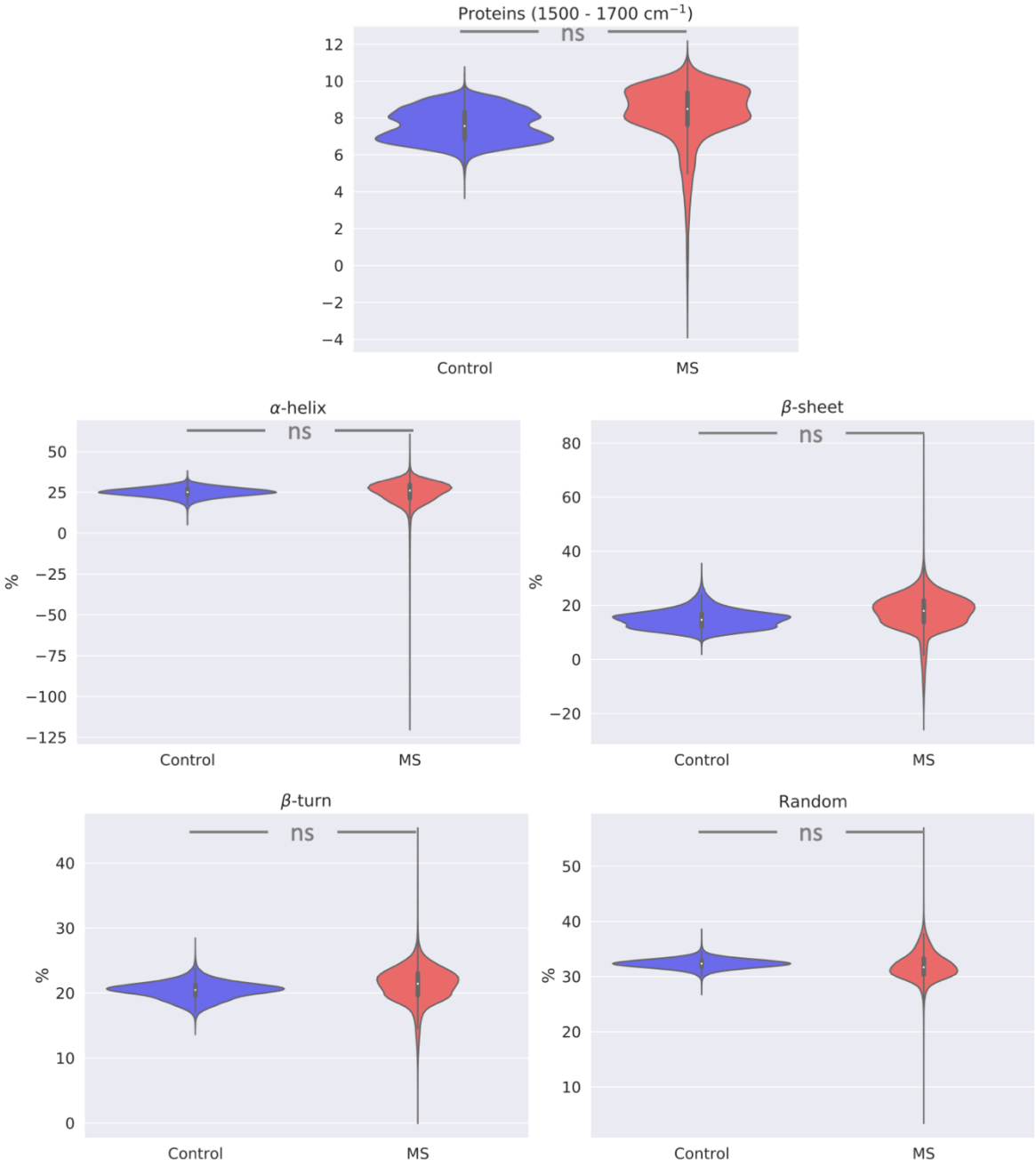


Figure 3.7: Comparison of protein content and protein secondary structure comparison with cμFTIR spectra from MS and control brains. The mean total protein content was higher in cases when compared to controls; similarly $\boldsymbol{\beta}$-sheet and $\boldsymbol{\beta}$-turn content was higher in cases than controls, although these differences were not statistically significant. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

A commonly assessed metric in spectroscopy analysis is the relationship between macromolecular features. Aiming to address this quantifiably, the correlations between several macromolecular types were visualised as scatter plots, with fitted linear models where appropriate and Pearson's R correlation calculated to help outline macromolecular relationships (Fig 3.8). In the control data the relationship between lipid and protein content showed a linear negative correlation (R=-0.912). In SPMS data, however, the relationship between the lipids and proteins was not linear and more akin to a second order polynomial relationship, suggesting a more complex relationship between these macromolecules in this context (Fig 3.8A). In controls, lipid oxidation and protein content were linearly positively correlated (R=0.757); for the cases the nature of this relationship was more variable across patients, which compromised the linear fit (R=-0.135; Fig 3.8B). For the relationship between saturation and unsaturation of lipids (Fig 3.8C) clear correlations in the case spectra were not discernible, but there was a substantially greater spread in the data compared to the controls, suggesting areas with more extreme macromolecular characteristics.). Protein α-helix and β-sheet content in controls had a relatively strong anti-correlation (R=-0.753) which was not observable in cases (R=-0.259), since these had significantly greater spread (Fig 3.8D).

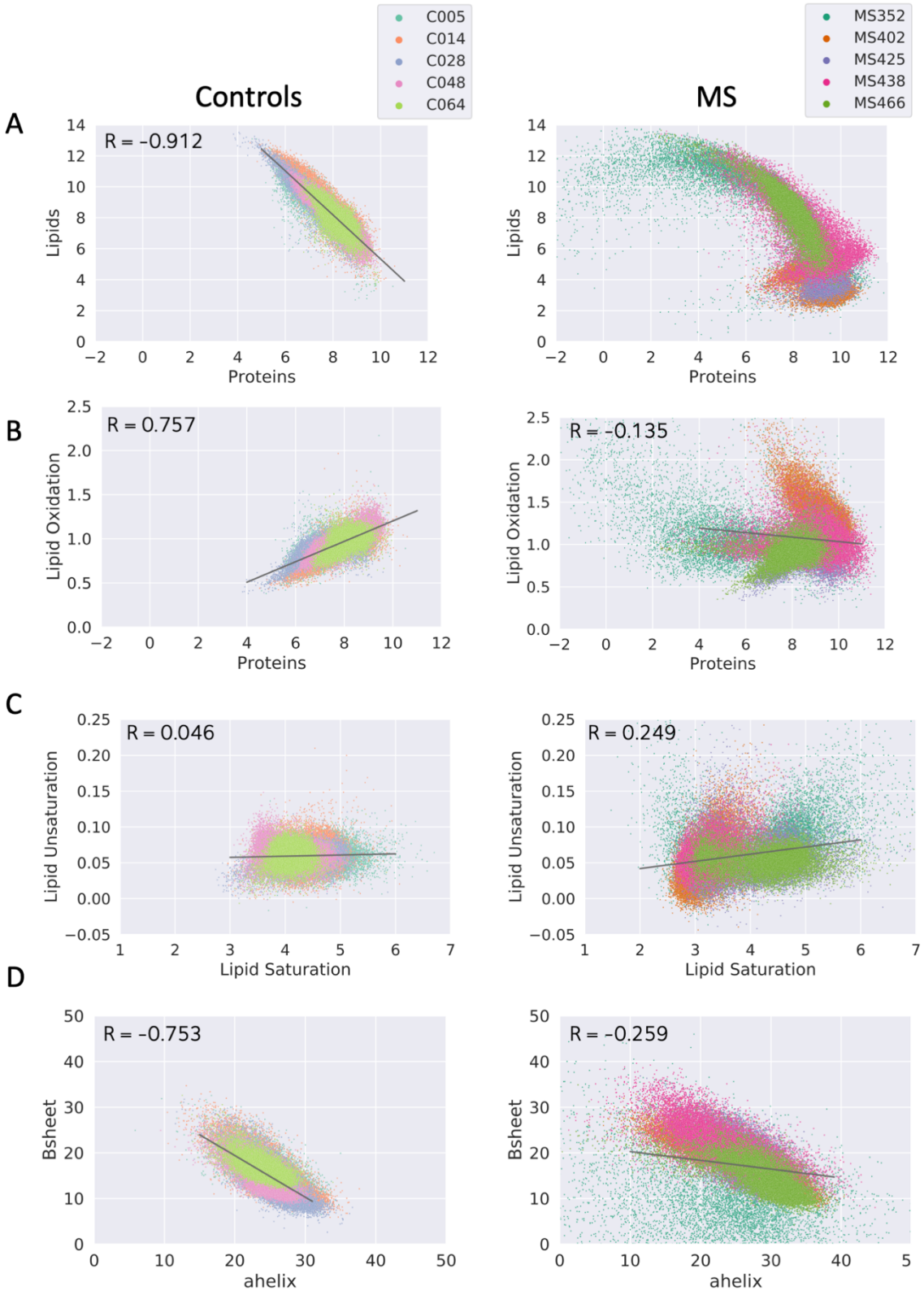


Figure 3.8: Evidence of changes in macromolecular relationships in MS when compared to controls. (A) Visualisation of protein and lipid content relationship in cases and controls reveals that, while in controls the relationship appears to be a linear anti-correlation, in MS there is a loss of linearity and increased heterogeneity. (B) Visualisation of relationship between estimated lipid oxidation and protein content in controls reveals a

linear correlation. In 4 out of 5 MS samples this relationship can no longer be observed. Notably, MS402 which has a distinct anti-correlation profile is marked by chronic features such as excessive demyelination while MS466 is the cluster with least evidence of severe demyelination, indicating the relationship between protein and lipid oxidation content may shift from a correlation to an anti-correlation with chronicity. (C) The relationship between estimated lipid saturation and lipid unsaturation is not a clear linear correlation in controls or cases, indicating an alternative interpretation of these spectral features may be warranted. (D) A clear anti-correlation can be observed between $\boldsymbol{\alpha}$-helix and $\boldsymbol{\beta}$-sheet content which was mostly undisturbed when comparing cases and controls, although cases had substantially more variation. R values indicated the calculated Pearson's R correlation.

3.2.2.3 Spatial mapping of macromolecular content distribution

Using the integration data (from section 3.2.2.2), biochemical maps were generated to spatially inspect the composition and heterogeneity across samples. Brain tissue features identified histologically through ORO+H staining, such as WM and GM in the control samples could be readily distinguished through spatial mapping of lipid and total amide content. The spatial distribution of more specific macromolecular features revealed a greater homogeneity (**Fig 3.9**), consistent with the distributions observed in the relevant feature violin plots and with the feature correlation plots (section 3.2.2.2). The SPMS cases, however, displayed more complexity (**Fig 3.10**). Non-lesional high lipid content regions in cases had higher lipid unsaturation and nucleic acids; they also had lower amide, ester carbonyl and lipid oxidation. Notably, lesions in certain samples had foci of high lipid, low amide, high carbonyl ester, higher oxidation and saturation whilst having lower nucleic acid content.

3.2.2.4 Unsupervised clustering of cμFTIR data reveals multiple SPMS brain tissue biochemical profiles

The PCA, macromolecular content contrast, correlations and spatial mapping indicated a greater complexity and underlying variation in SPMS than was captured in case vs. control comparisons. In order to partition these data in an unbiased manner, unsupervised UMAP was used for dimensionality reduction, followed by community detection with the leiden algorithm.

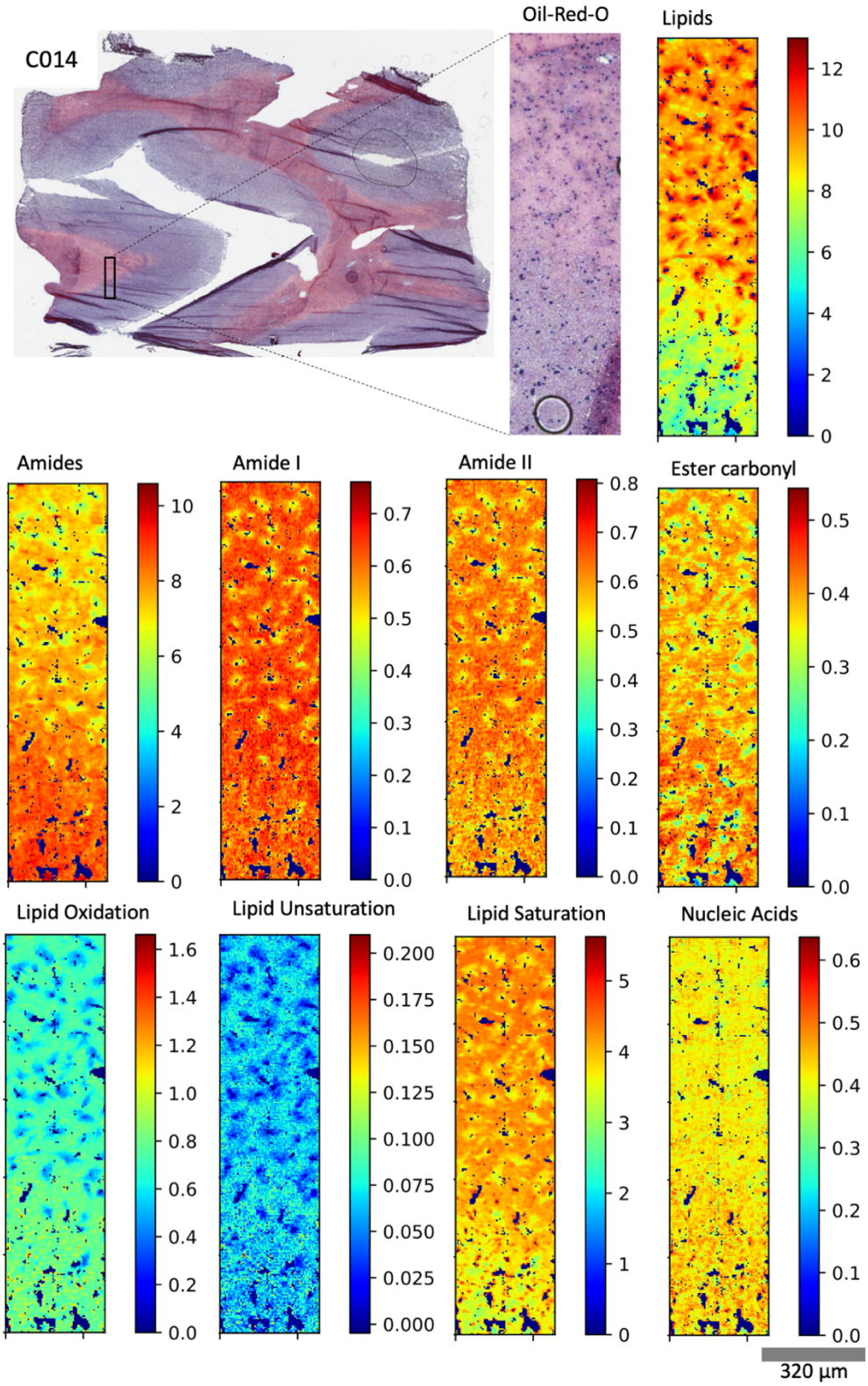


Figure 3.9: Spatial visualisation of biochemical distribution in a representative control brain section, C014 reveals features identified in histological characterisation. Total lipid, amide and lipid saturation visualisation outline WM and GM transitions. In contras with the GM, the WM is characterised by high lipid content, low total amide content, low lipid unsaturation and high lipid saturation.

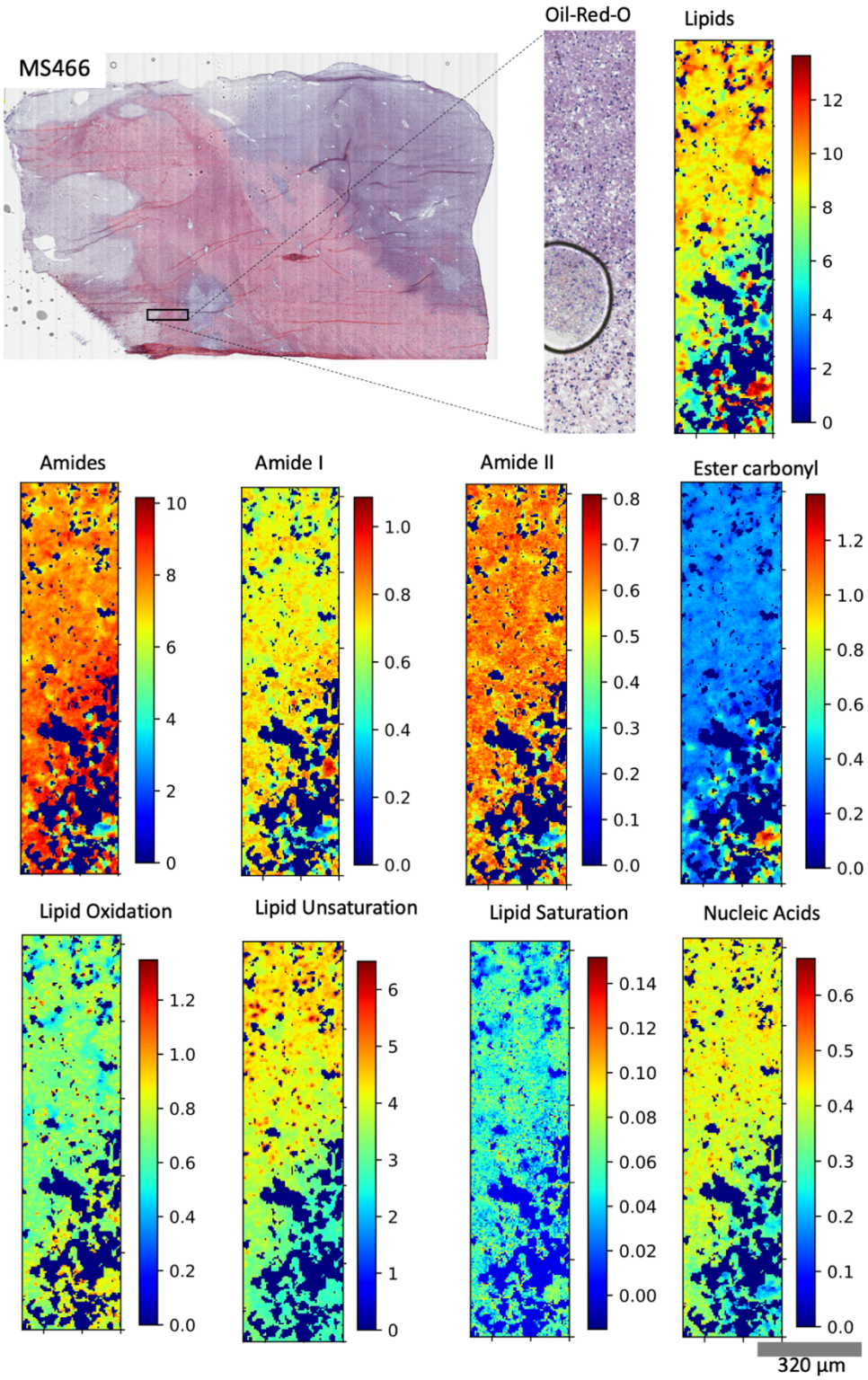


Figure 3.10: Spatial visualisation of biochemical distribution in a representative MS brain section, MS466 reveals features identified in histological characterisation. MS 466 is characterised by clearly outlined lesions, moderate demyelination and some regions of immune cell infiltration. Total lipid and amide visualisation outline lesion to perilesion transitions. This sample is marked by a focus of particularly high lipid content, low amide content, high ester carbonyl and high oxidation. The perilesion, apart from high amide content is also marked by higher lipid unsaturation and nucleic acid content than the lesion.

This is the first time the UMAP algorithm has been used on a biological μFTIR data set, and the first time the leiden algorithm has been used for data partitioning of any μFTIR data set. This approach enables the joint analysis of large data sets – this data set contained n=144,000 spectra – at a lower computational and time cost than other previously popular approaches such as t-distributed stochastic neighbour embedding (t-SNE)[145]. The approach here presented is also uniquely compatible with the SCANPY python environment, and enables streamlined analysis with detailed clinical information for each sample.

The data were clustered using as input the entire spectra and, in parallel, only the 2,800-3,000 cm^{-1} lipid region. Consistently with the indication from the PCA that lipids explain the greatest proportion of variation in the data, both methods yielded similar results (Fig 11A,B). Notably, the majority of spectra were grouped in a single large cluster with substantial internal variation, suggesting that changes in lipid content – likely due to demyelination and changes in cellular content in lesional areas in MS patients, and to the natural variation in GM and WM in controls – occur gradually along a continuum from low to high lipid content.

However, given that the correlation of lipid and protein features in the SPMS patient spectra was non-linear, we reasoned that clustering based on the protein region data might enable spectral segregation based on more subtle feature variation that may be masked by the variation in absorbance of the lipid region (Fig 3.11C).

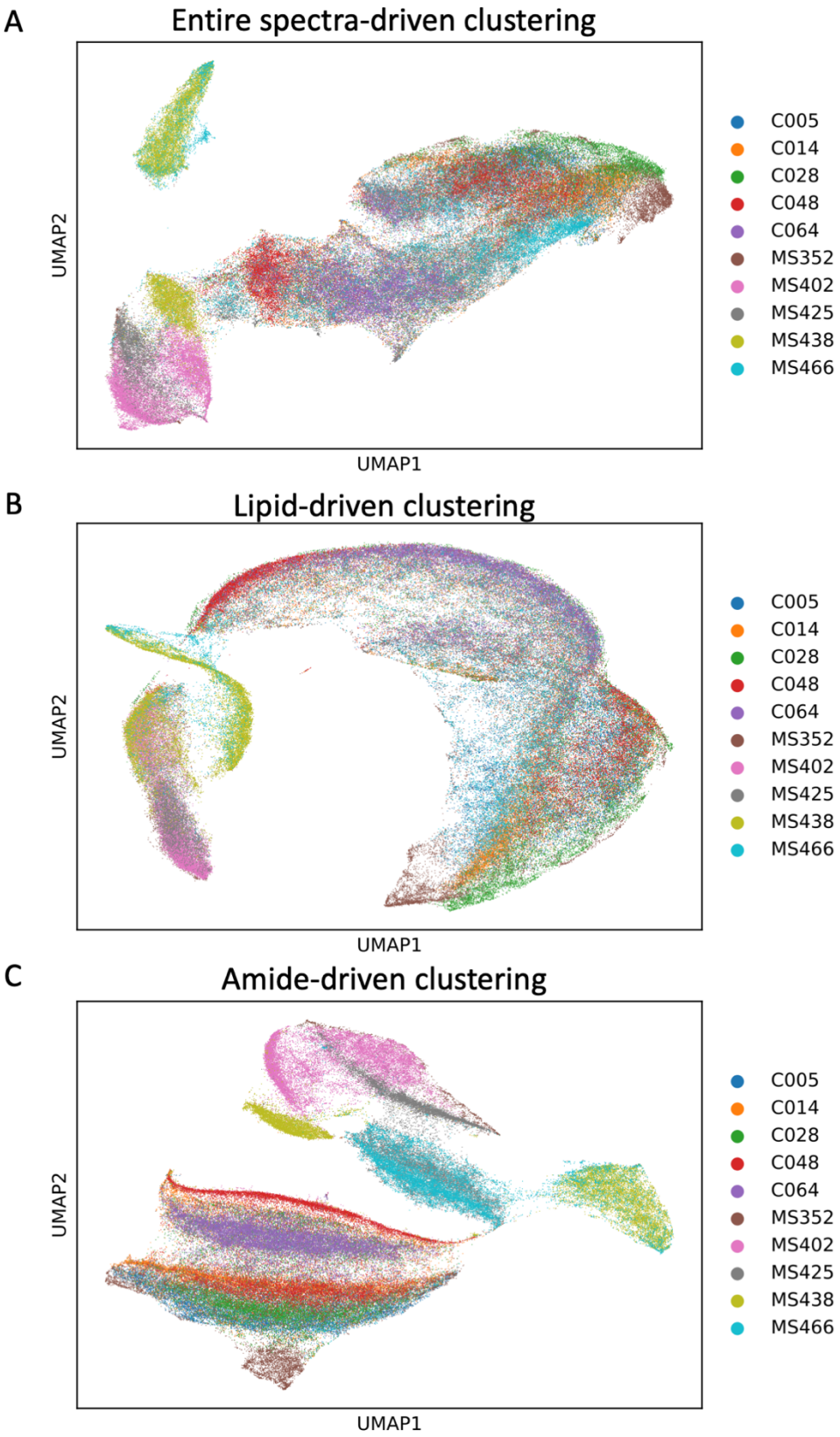


Figure 3.11: Comparison of clustering with various cluster-driving inputs. The lipid variations are dominant in clustering driven by entire spectra, with more continuous changes within a main large population. The amide region has more discrete variation captured in the architecture of the reduced dimension UMAP visualisation, making leiden community detection more easily reproducible.

Using the 1,500-1,700cm^{-1} region data as input, seven clusters were identified – two of which were predominantly present in the control samples (**Fig 3.12A**). These clusters were distributed across all controls and present in some cases; (**Fig 3.12B**). The other five clusters were each present in SPMS samples – three of these could be observed in more than one sample, and two clusters were sample-specific. The clusters were annotated based on a consideration of their key spectral features and on their spatial mapping (see section 3.2.2.5 below) relative to tissue areas characterised by histological staining.

Inspection of the mean spectra for each of the seven clusters revealed the breadth of variation in protein, lipid and in features of the fingerprinting region (1,000-1,470 cm^{-1}) across clusters (**Fig 3.12C**). The WM cluster had the lowest amide content, and lesional cluster "Les2" had the highest amide at ~1.45 times that of the WM (**Fig 3.13**). Lesional cluster "Les4" was particularly distinct: the amide II peak had visible contributions from several smaller but distinct peaks, which were absent from other clusters.

With respect to the lipid content, the WM cluster had the highest content (maximum absorbance peak ~1.2 times higher than the GM cluster and ~3.2 times higher than lesional cluster Les2, which had the lowest lipid content). All lesional clusters had a lower lipid content than WM and GM with the exception of cluster "Les1", which had an intermediary level between GM and WM.

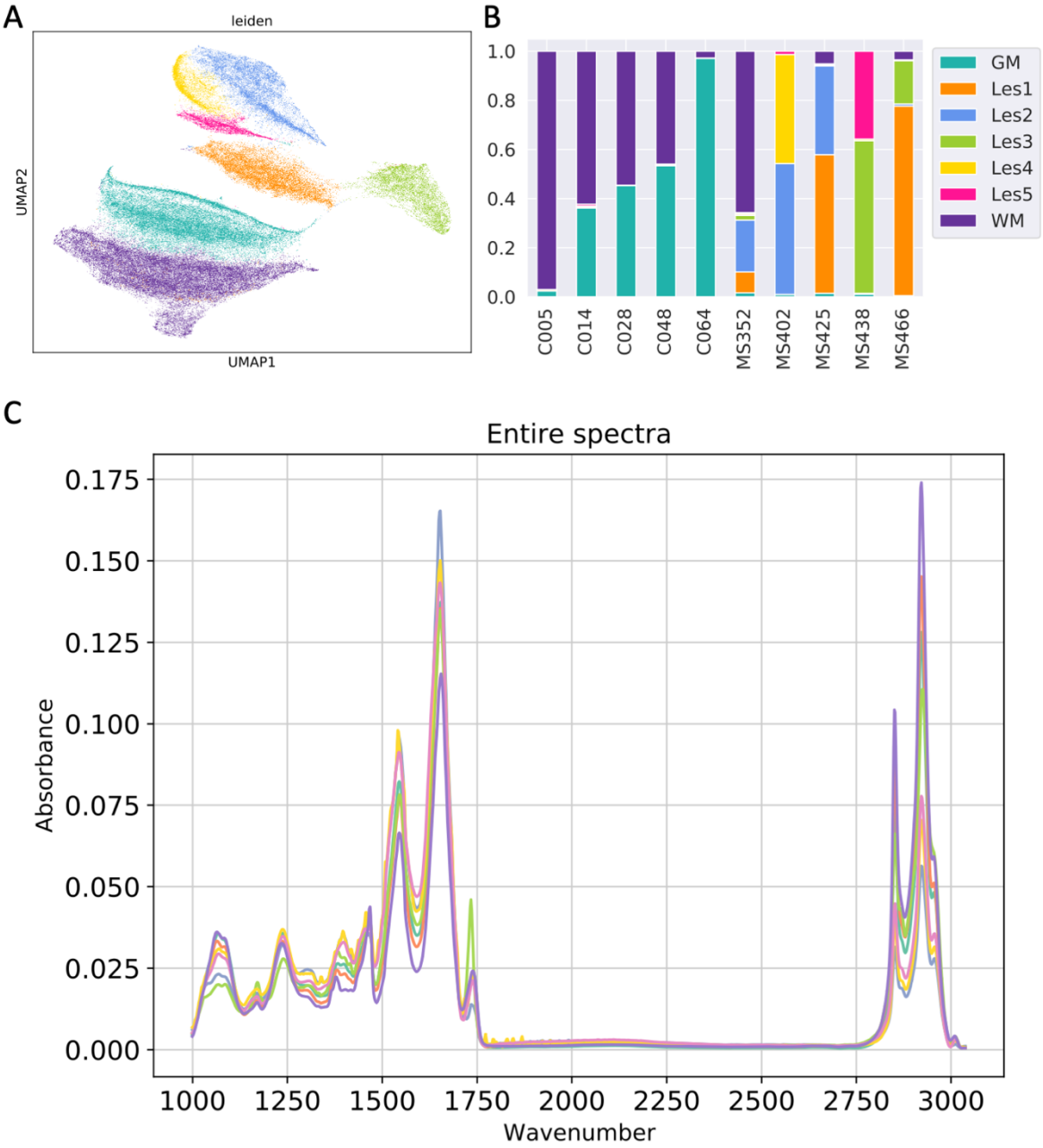


Figure 3.12: Amide driven clustering of brain cμFTIR data reveals homogeneity across control samples, as well as heterogeneity in MS. (A) UMAP dimensionality reduction paired with leiden community detection were used to cluster spectra, clearly distinguishing 7 clusters. (B) Cluster distribution across samples indicates WM and GM clusters are predominantly represented in controls; The other five clusters are distributed across cases and 3 of these are observed in more than one case. (C) Mean spectra of clusters reveal substantial variation in lipid, protein, carbonyl ester and nucleic acid content across clusters.

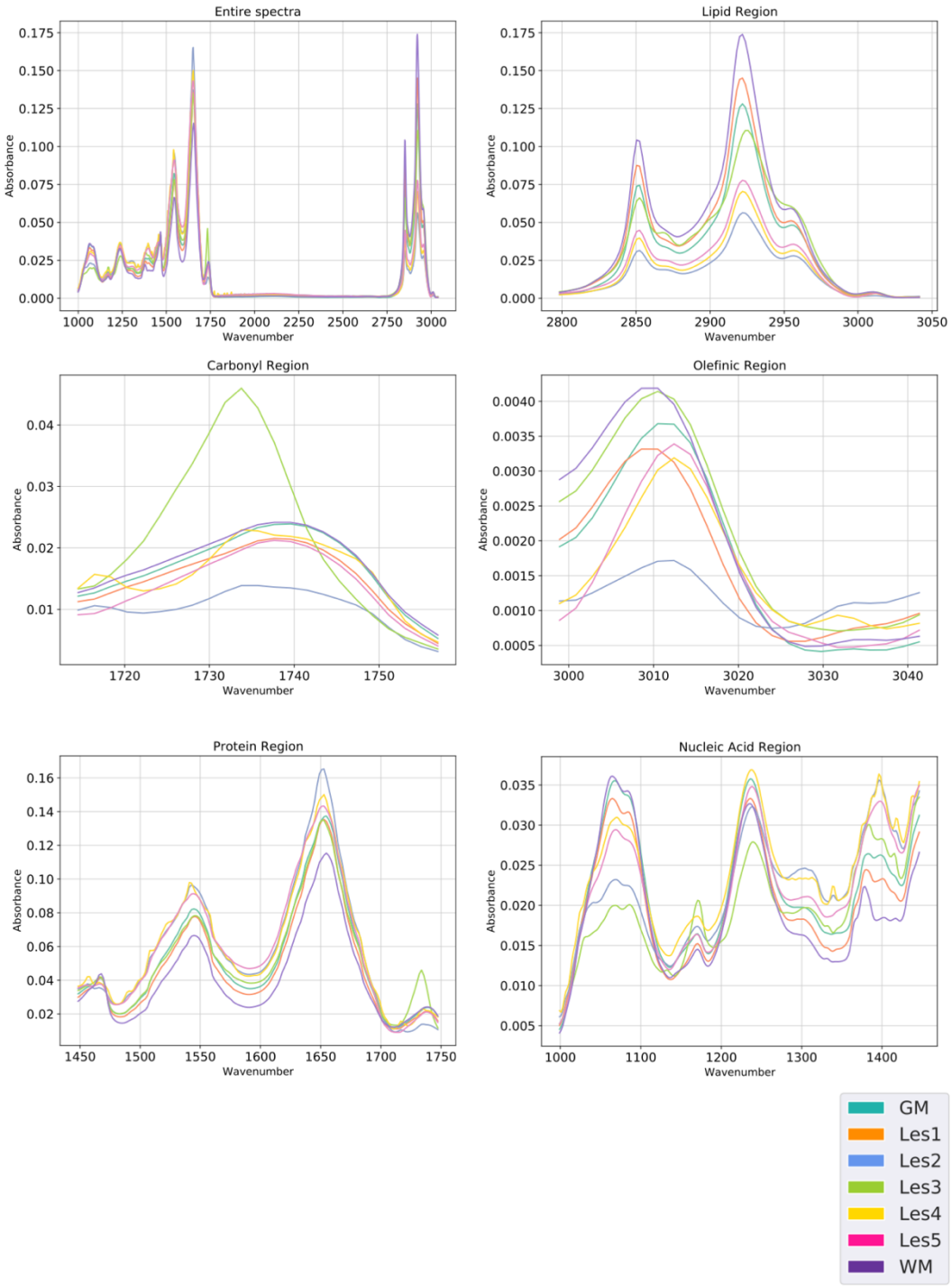


Figure 3.13: Visual inspection of qualitative spectral changes in brain cμFTIR clusters reveals changes in the lipid region and carbonyl ester region. Les3 has a unique lipid spectral profile, with a distinct absorbance peak at ~2860 cm^{-1}. Similarly to SC cμFTIR data, three distinct spectral profiles can be observed in the carbonyl ester region. WM, GM and some Les spectra have a single wide spectrum with a maximum absorbance at ~1,740 cm^{-1}. Les2 and Les4 have a small absorbance peak at ~1,715 cm -1 and a second higher peak at ~1,733 cm^{-1}. Les3 is marked by the highest absorbance peak positioned at ~1,733 cm^{-1}.

The lipid spectral profile for lesional cluster "Les3" displayed not only differences in peak amplitude relative to other clusters, but also shifts in the wavenumber at which peak maxima occurred. For example, characteristic of this cluster was the appearance of a peak at ~2,870 cm^{-1}, a shift of the CH_{2as} peak towards higher wavenumbers, and a higher v_{as} CH_3 content. Cluster Les3 also had a distinct profile in the carbonyl ester region, with a sharp peak at ~1,733 cm^{-1}, while the corresponding peak for the control GM and WM clusters, for example, was closer to 1,740 cm^{-1}. In the case of the carbonyl ester region, for instance, this may indicate changes in the number of hydrogen bonds (explored in detail in section 3.3.2). The carbonyl ester region of Les3 also had a peak absorption 1.88 times higher than the control cluster peaks. Lesional clusters Les4 and Les2 also had distinct profiles in this region, with a reduced absorption at 1,740 cm^{-1} and peaks with higher absorbance at 1,733 cm^{-1} and ~1,717cm^{-1} than neighbouring regions.

Complementary to the inspection of the mean spectra for each cluster, the integrated area below the spectra in regions of interest was also calculated and compared across clusters. These data are summarised in **table 3.1** and **table 3.2** for lipids and proteins, respectively. Highest and lowest values are highlighted in red and blue, respectively, for ease of interpretation.

Table 3.1: Summary of lipid features of each cluster

Clusters	Total spectra	Lipids	Carbonyl ester	Lipid Oxidation	Lipid Saturation	Lipid Unsaturation
GM	38178	7.563± 4.69x10-^{3}	0.206± 1.49x10^{-4}	1.014± 5.68 x10^{-4}	4.059± 1.67 x10^{-3}	0.062± 8.20 x10^{-5}
Les1	19409	8.453± 8.71 x10^{-3}	0.196± 2.38 x10^{-4}	0.895± 9.64 x10^{-4}	4.483± 3.45 x10^{-3}	0.055± 1.30 x10^{-4}
Les2	15768	3.677± 7.16 x10^{-3}	0.079± 3.83 x10^{-4}	1.144± 2.63 x10^{-3}	3.176± 3.05 x10^{-3}	0.049± 3.03 x10^{-4}
Les3	11024	7.913± 1.92 x10^{-2}	0.536± 1.94 x10^{-3}	1.010± 1.67 x10^{-3}	3.137± 2.65 x10^{-3}	0.056± 1.67 x10^{-4}
Les4	7176	4.422± 7.69 x10^{-3}	0.132± 5.74 x10^{-4}	1.466± 3.07 x10^{-3}	3.440± 3.42 x10^{-3}	0.072± 3.82 x10^{-4}
Les5	5477	5.023± 8.88 x10^{-3}	0.211± 5.96 x10^{-4}	1.127± 4.42 x10^{-3}	3.487± 5.23 x10^{-3}	0.068± 3.71 x10^{-4}
WM	47778	9.966± 4.09 x10^{-3}	0.209± 1.35 x10^{-4}	0.885± 1.20 x10^{-3}	4.658± 1.64 x10-3	0.062± 1.01 x10^{-4}

Table 3.2: Summary of protein features per cluster

Clusters	Proteins	α-helix	β-sheet	β-turns	Random
GM	8.348± 2.89x10^{-3}	23.494± 1.30x10^{-2}	17.153± 1.42x10^{-2}	21.080± 5.85x10^{-3}	32.589± 4.43x10^{-3}
Les1	7.955± 4.28x10^{-3}	28.885± 2.48x10^{-2}	14.761± 2.71x10^{-2}	20.084± 1.13x10^{-2}	30.829± 9.05x10^{-3}
Les2	9.465± 5.44x10^{-3}	27.844± 3.71x10^{-2}	18.593± 3.99x10^{-2}	22.798± 1.46x10^{-2}	32.895± 1.49x10^{-2}
Les3	8.023± 1.43x10^{-2}	24.223± 6.31x10^{-2}	19.708± 3.84x10^{-2}	21.320± 2.13x10^{-2}	30.351± 1.53x10^{-2}
Les4	8.728± 8.11x10^{-3}	19.770± 4.14x10^{-2}	21.848± 3.26x10^{-2}	23.277± 1.78x10^{-2}	35.976± 2.14x10^{-2}
Les5	9.483± 1.10x10^{-2}	20.335± 5.69x10^{-2}	25.091± 5.18x10^{-2}	22.617± 3.02x10^{-2}	32.561± 1.92x10^{-2}
WM	6.596± 5.61x10^{-3}	24.875± 4.31x10^{-2}	11.993± 2.45x10^{-2}	19.467± 9.84x10^{-3}	31.902± 7.26x10^{-3}

Integrating the area under the lipid region confirmed that the clusters with the highest lipid content were WM (9.96 absorbance·cm^{-1}), Les1 (8.45 absorbance·cm^{-1}) and GM (7.56 absorbance·cm-1), and the lowest lipid content was in Les2 (3.67 absorbance·cm^{-1}), Les4 (4.42 absorbance·cm^{-1}) and Les5 (5.02 absorbance·cm^{-1}) (**Fig 3.14**). Les3 had a particularly wide range of lipid and carbonyl ester contributions, and was the only cluster to have carbonyl content higher than the control tissue. Les2 and Les4 had the highest lipid oxidation levels when compared to all other clusters (1.14 absorbance·cm^{-1} and 1.47 absorbance·cm^{-1}, respectively). Lipid saturation was lower in Les2, Les3, Les4 and Les5 when compared to the other clusters, with values <3.5 absorbance·cm^{-1}. Lipid unsaturation was more diverse in Les2, Les4 and Les5 as indicated by the wider distribution of these clusters, although the impact on overall mean lipid unsaturation was small. All pair-wise comparisons were statistically significant (padj<0.0001) with the exception of Les1 and Les3 lipid oxidation, Les1 and WM lipid unsaturation and Les5 and WM lipid saturation which were non-significant.

The mean protein content of the Les2, Les4 and Les5 lesional clusters were the highest (>8.7 absorbance·cm-1) (**Fig 3.15**). Les3 had a wide distribution of protein content, indicating heterogeneity of this feature within that cluster. WM, GM and Les1 had the lowest protein content (<8.3 absorbance·cm-1). The protein folding secondary structure features were also estimated for each cluster. The highest α-helix content was observed in Les1 and Les2 (28.9% and 27.8%, respectively), and the lowest was in Les4 and Les5 (19.8% and 20.3%).

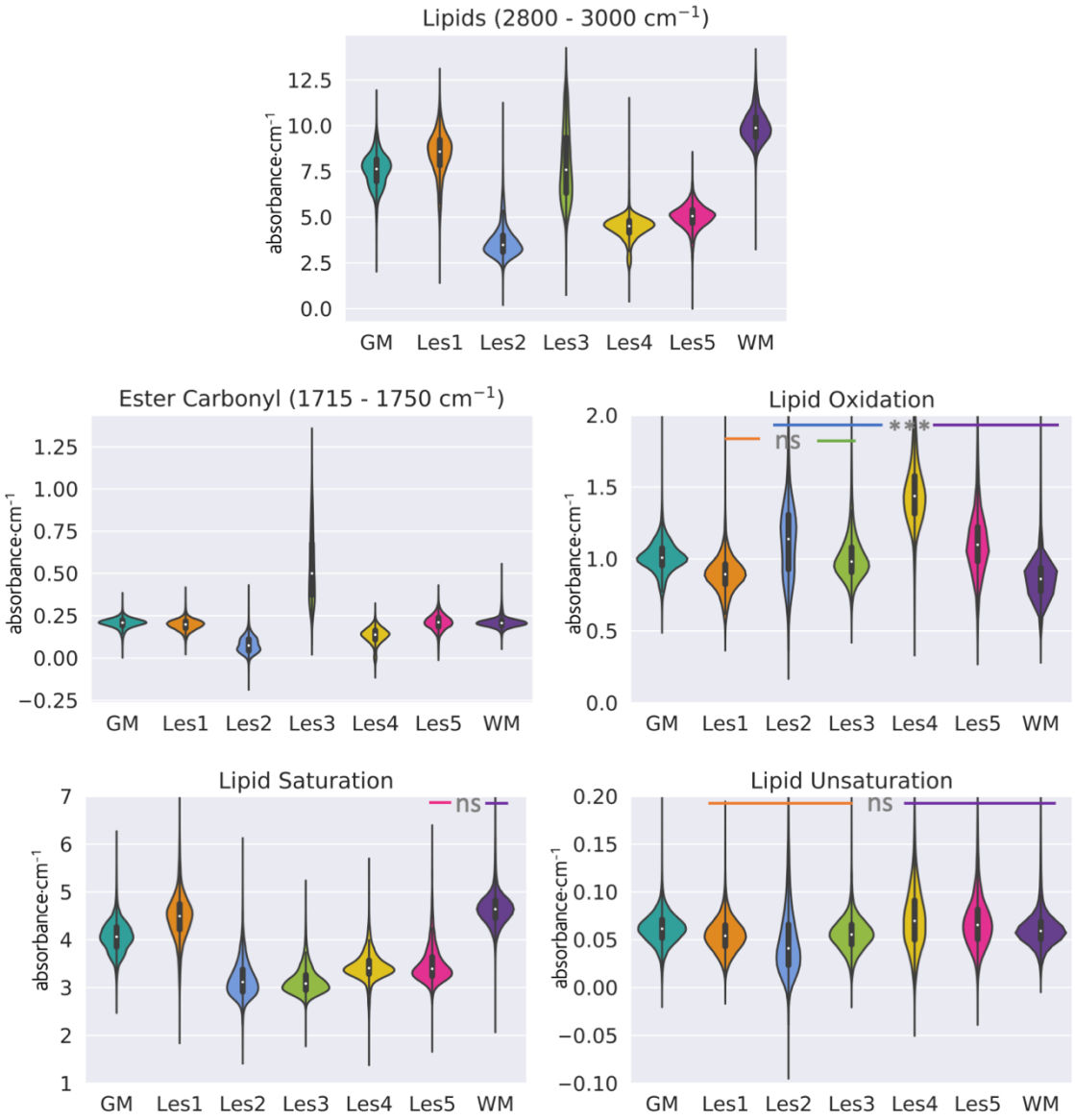


Figure 3.14: Comparison of lipid content and state across cμFTIR clusters reveals wide range of lipid variation across clusters. Differences in lipid and ester carbonyl content are significantly different across all pairwise comparisons of clusters. Lipid oxidation, lipid saturation and lipid unsaturation differences are statistically significant in most pair-wise comparisons, unless otherwise indicated. Les2, Les4 and Les5 have lower lipid content, and are also marked by higher lipid oxidation estimates. Les2 and Les4 also have decreased ester carbonyl content. Les3 is characterised by significant intra-cluster heterogeneity for lipid and carbonyl ester content – this is the cluster with the highest carbonyl ester content. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

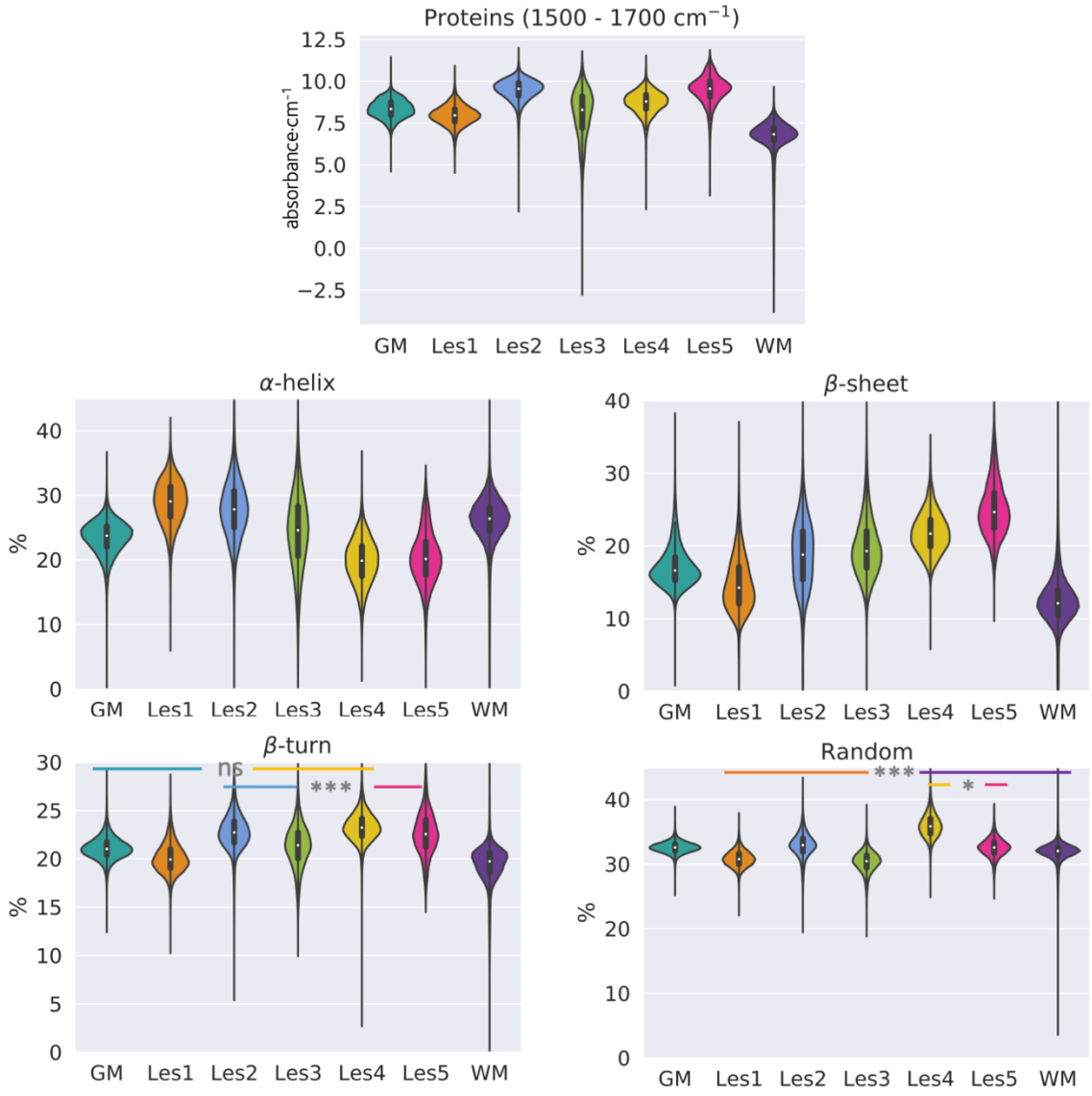


Figure 3.15: Comparison of protein content and protein secondary structure composition across cμFTIR clusters indicates significant proteomic changes in SPMS. Total protein content was higher in Les2, Les3, Les4 and Les5 clusters, also characterised by lipid loss. Lesional clusters also had notable changes in $\boldsymbol{\alpha}$-helix content: Les1, the most WM-like lesional cluster in lipid content, and Les2, the lesional cluster with most severe demyelination, had a higher $\boldsymbol{\alpha}$-helix content when compared to WM while Les4 and Les5 had substantially lower $\boldsymbol{\alpha}$-helix content. All Les clusters had an increased $\boldsymbol{\beta}$-sheet content when compared to the WM cluster. Les2, Les3, Les4 and Les5 also had an increase in $\boldsymbol{\beta}$-turns and Les2 and Les4 had an increase in random folding. All cluster pair-wise comparisons were statistically significant unless otherwise indicated. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

Les5 had the highest β-sheet content (25.1%), followed by Les4 (21.8%), and the GM and WM clusters had the lowest (17.1% and 12%, respectively). β-turn content was highest in Les2 (22.8%) and Les5 (22.6%) and lowest in WM (19.5%) and Les1 (20.1%). Finally, Les4 had the highest random folding content (36%) and Les3 the lowest (30.3%). All pair-wise comparisons were statistically significant (padj<0.0001) except for β-turn comparison between GM and Les4 clusters.

3.2.2.5 Spatial mapping of cμFTIR spectral clusters recapitulates histological features

The cluster classification of each spectrum was colour coded and mapped back into the conFiguration at time of collection to inspect the relative spatial distribution of each cluster (Fig 3.16). Using as a reference ORO histological characterisation, the two clusters prevalent in control samples were determined to correspond to the WM and GM. The WM cluster was also observed in case MS352, suggesting the presence of a region biochemically similar to control WM within this brain tissue section. The Les1 cluster mapped to MS425 and MS466 in areas of minimal damage observable through ORO – this is consistent with the finding that this cluster has the highest lipid content among lesional clusters. Les2, characterised by the lowest lipid content was present in the MS352, MS402 and MS425 case samples in areas with evidence of severe tissue damage and reduced tissue integrity. Les3 was present in MS466 and MS438 in areas characterised by some damage, albeit not as severe as Les2 based on the histological staining.

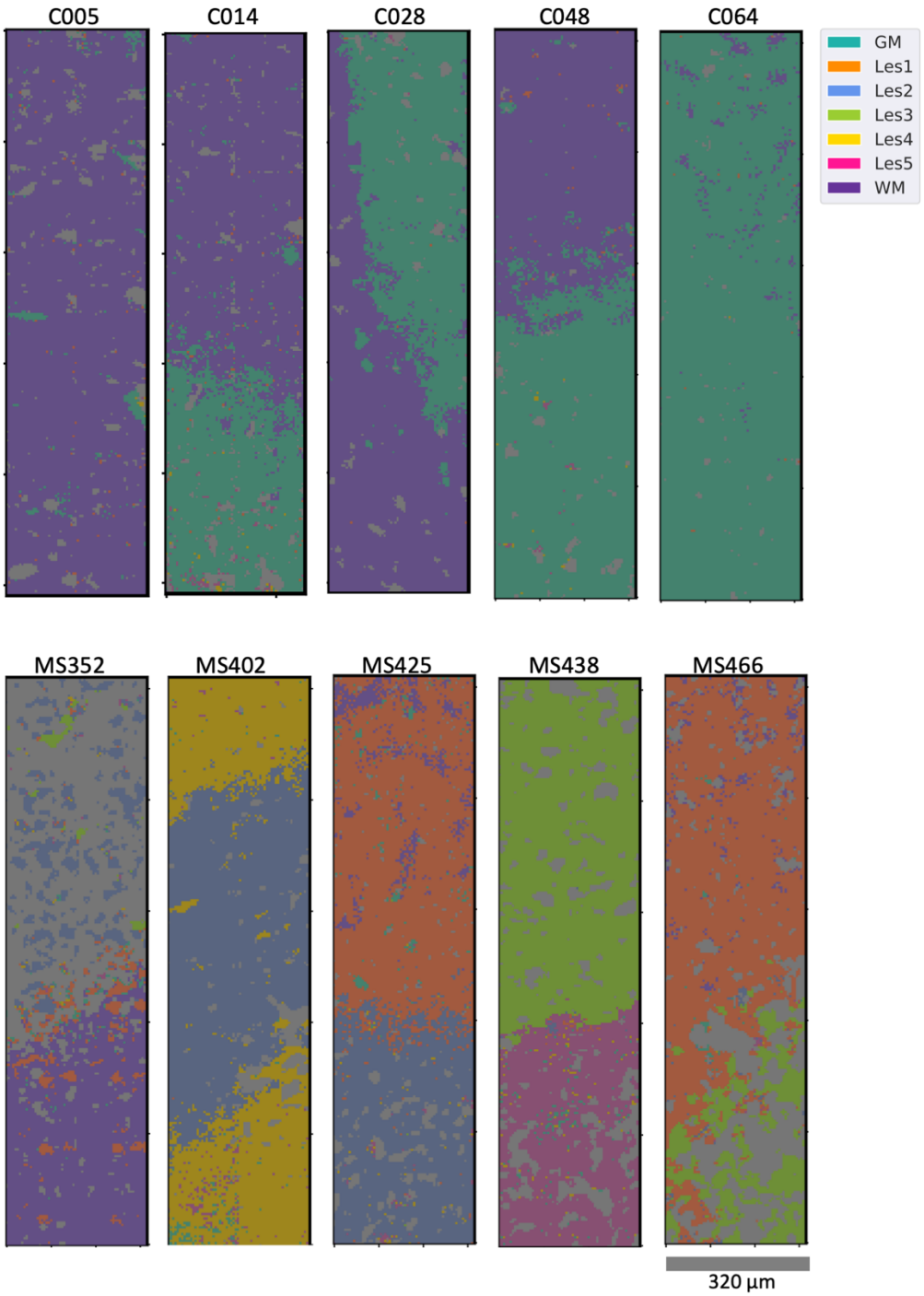


Figure 3.16: Cluster classification maps of all brain samples based on cµFTIR data outline histological features. Cluster classification maps of controls have GM/WM classification borders consistent with those observed in ORO histological characterisation. WM cluster appears in cases MS352 and MS425. Les1, Les2 and Les3 clusters are recapitulated in more than one sample, indicating a distinct biochemical profile which is not sample-specific. MS352 and MS425 are marked by stark differences in myelination in the transition from lesion to perilesion creating a cliff-like lipid content transition; MS402 is characterised by overall severe loss of lipid content and MS438 and MS466 have more moderate differences between the two main clusters characterising each sample, all of these seemingly having a more slant-like lesion transition.

The Les4 and Les5 clusters were mostly sample-specific clusters: Les4 mapped to MS402 in the area surrounding the most severe damage; Les5 mapped to MS438 and had a lower lipid content than the neighbouring Les3. Notably, this suggests that the Les3 cluster may represent a perilesional intermediate area; however, in sample MS466 Les3 had the lowest lipid content and thus this cluster cannot be definitively characterised as lesional or perilesional without a further assessment of cellular representation and gene expression analysis.

3.2.3 Synchrotron-based µFTIR-based molecular profiling of spinal cord in SPMS and controls

Although MS spinal cord (SC) samples cannot be easily obtained from MS tissue banks in the UK, we were able to separately obtain and profile samples from two SPMS patients (one with an active lesion and one with a chronic lesion), and two controls using sµFTIR. A total of 769 spectra were collected from the patient samples and 339 from the controls.

3.2.3.1 Principal component analysis of the spinal cord sµFTIR spectra distinguishes cases and controls

Similarly to brain tissue spectra, the mean second derivative of the pre-processed SC spectra was plotted with the standard deviation as an assessment of spectral quality (Fig 3.17A).

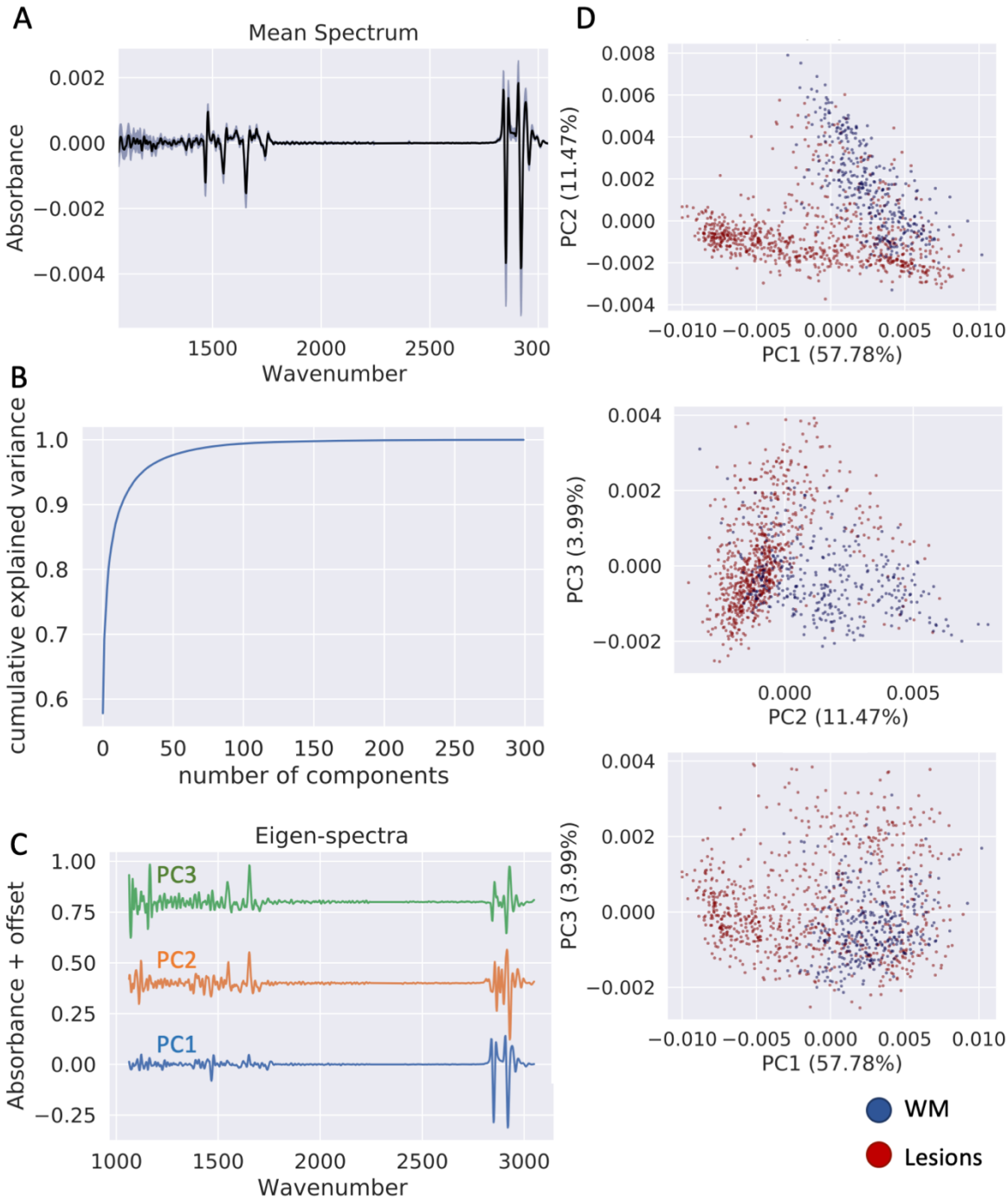


Figure 3.17: Principal component analysis of sμFTIR SC samples. (A) Second derivative of mean spectrum with standard deviation. (B) Cumulative explained variance of the first 300 PCs. Over 50 PCs are required to achieve CEV > 99%. (C) Eigen-spectra for top 3 PCs highlight the importance of major lipid peaks (2,800-3,000 cm^{-1}) and amide peaks (1,500-1,700 cm^{-1}) for sample characterisation. (D) Scatter plot visualisation of top 3 PCs. PC1 and PC2 distinguish WM from lesions with fair accuracy.

Some features where identified in the fingerprinting region (1,000-1,470 cm^{-1}) which may be associated to noise/beamline changes during data collection, the remaining spectrum which is particularly of interest to this analysis appeared to have good quality. PCA was performed to identify the strongest features explaining the variation in the data. More than 50 PCs were required to achieve CEV>99% (**Fig 3.17B**). The main hallmarks of the PC1 eigen-spectra (EV=57.78%) were two high-amplitude features corresponding the CH_2 and CH_3 symmetrical and asymmetrical vibrations in the 2,800-3,000 cm^{-1} lipid range (**Fig 3.17C**). PC2 and PC3 each explained 11.47 and 3.99% of the variance of the data, respectively. High-amplitude contributions in the lipid region were also visible in PC2 and PC3, as were signals corresponding to the amide I and amide II peaks. A scatter plot of the PC1 and PC2 scores could largely distinguish the spectra belonging to cases vs. controls (**Fig 3.17D**).

3.2.3.2 SμFTIR detects significant macromolecular differences when comparing active and chronic SPMS spinal cord lesions to control WM

By integrating the area below the region of interest in the spectral curve, the mean lipid content was estimated (**Fig 3.18**), and was found to be ~1.2 times higher in the WM than the active lesion and 2.0 times higher than the chronic lesion. Notably, all comparisons in this data set were not statistically significant. In order to control for beamline variation, similarly to brain data, a linear model was used to control for each scanning sequence to control for beam-dependent variation – this, combined with the small starting number of individuals represented, resulted in non-significant contrasts.

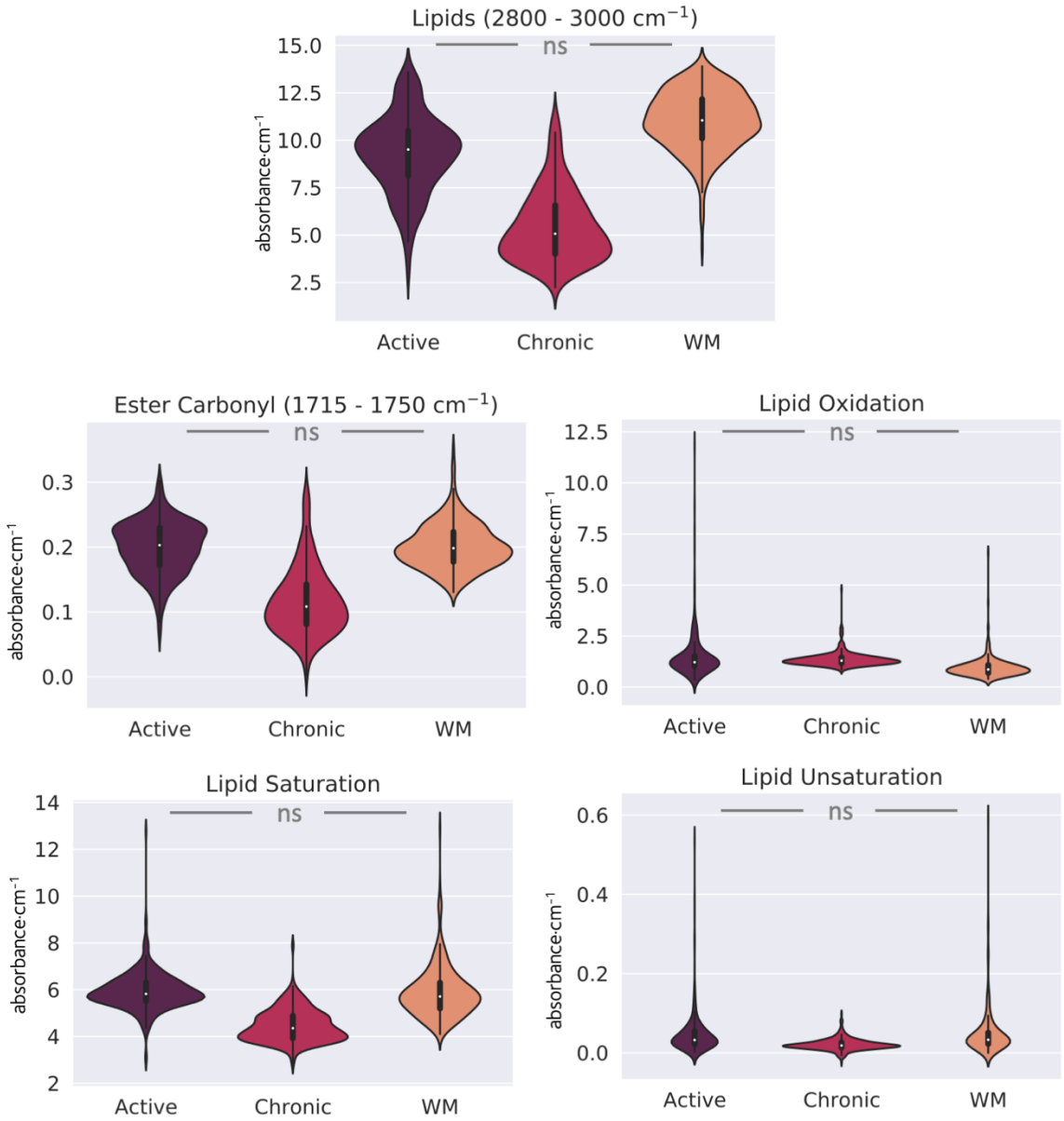


Figure 3.18: Comparison of lipid content and state across SC active lesions, chronic lesions and control WM. This was accomplished by integrating the area below sμFTIR spectra, revealing statistically significant differences across all pair-wise comparisons of tissue types. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

In spite of lack of significance, several trends could be observed which are consistent with MS pathophysiology. The mean carbonyl ester (1,700-1,750 cm^{-1}) level was ~1.7 times higher in WM when compared to the content of the chronic lesion; the WM and the active lesion mean carbonyl ester content did not differ, although the distribution in the active lesion was more variable. Changes to lipid saturation and lipid unsaturation were most pronounced when comparing WM to the chronic lesion (1.3 and 2.3 higher in the WM, respectively). The mean lipid oxidation in the WM was estimated to be ~0.65 times of that in the active lesion and 0.70 in the chronic lesion. Overall, the active lesion biochemical profile showed greater variation, while the chronic lesion was more homogeneous.

The highest protein content was that of the chronic lesion (1.6 times higher than the WM), while the active lesion was 1.1 times higher (**Fig 3.19**). Changes in protein folding secondary structure features in the chronic lesion were less pronounced with respect to the α-helix content, but more pronounced for the other folding features relative to the active lesion. The chronic lesion had 2.8% less α-helix features than the WM, 7.4% more β-sheet, 3.4% more β-turns and 1.8% more random folding. The active lesion had 3.3% less alpha-helix features than the WM, 2.2% more β-sheet, 1.5% more β-turns and 0.5% more random folding.

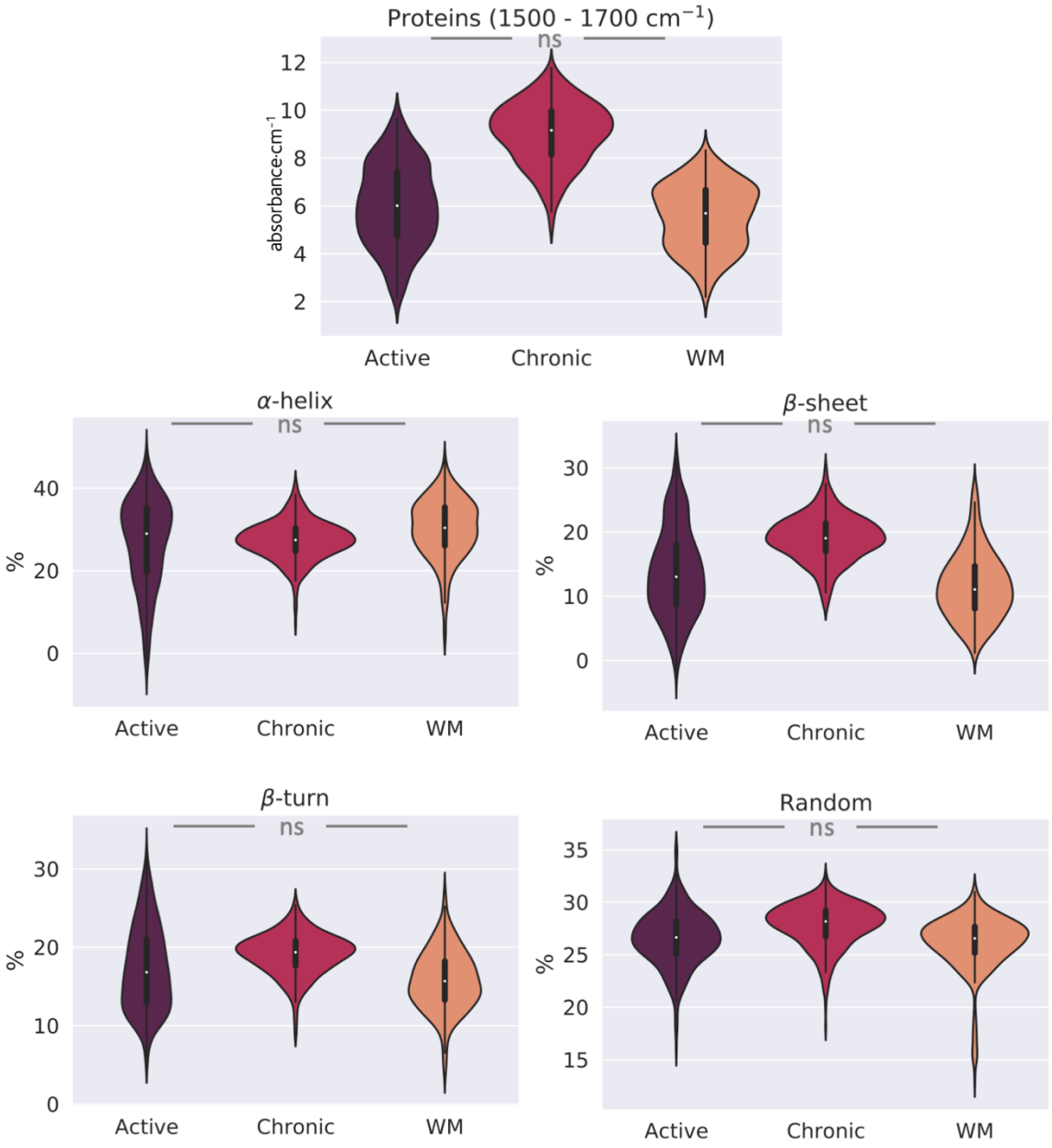


Figure 3.19: Comparison of protein and protein secondary structure content across SC active lesions, chronic lesions and control WM. This analysis revealed statistically significant differences across all pair-wise comparisons of tissue types. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

3.2.4 CµFTIR of the human spinal cord

As a complement to the sµFTIR biochemical analysis of the available SC sections, cµFTIR was used to scan 40,987 control spectra and 102,400 case spectra.

3.2.4.1 Principal component analysis of the spinal cord cµFTIR spectra does not distinguish cases and controls

The mean second derivative of the spectra and standard deviation were plotted to assess spectral quality following QC (**Fig 3.20A**). The PCA revealed that >250 PCs were required to achieve 99% CEV (**Fig 3.20B**), again indicating high-complexity data. Scatter plot visualisation of PC scores for the top 3 PCs did not clearly distinguish control and MS spectra – reminiscent of the equivalent brain tissue data - although certain trends could be observed, such as controls tending to have higher PC1 scores than SPMS, while SPMS tended to have higher PC2 scores (**Fig 3.20C,D**). Similar to the sµFTIR results, eigen-spectra for PC1 had spectral contributions with high amplitude in the lipid peaks, and PC2 had high amplitude contributions in the protein range.

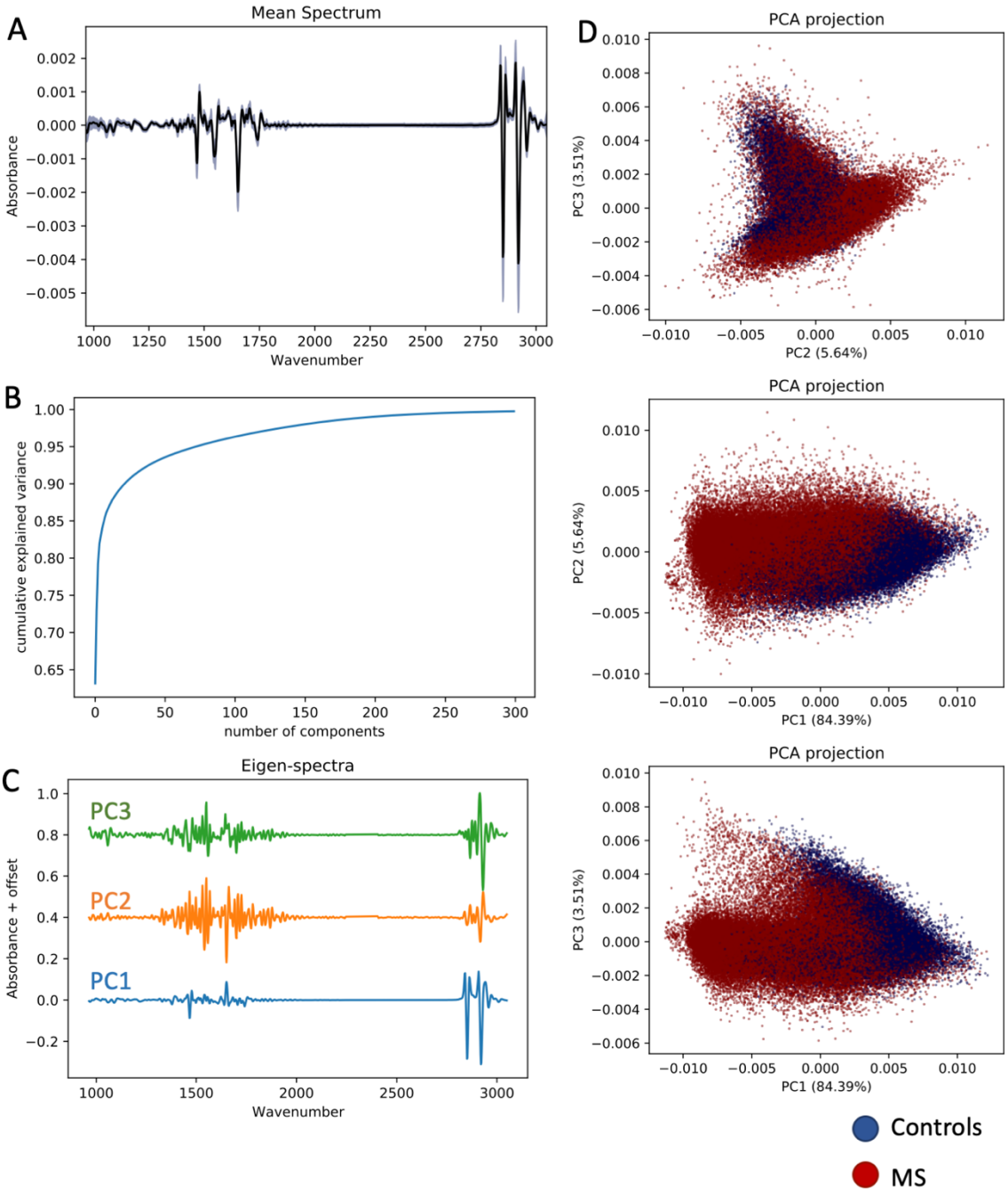


Figure 3.20: Principal component analysis of cμFTIR SC. (A) Second derivative of mean spectrum plotted with standard deviation. (B) Cumulative explained variance for top 300 PCs reveals high complexity of the data set, requiring >250 PCs to achieve CEV > 99%. (C) Eigen-spectra for top three PCs reveal major contribution of CH_2 and CH_3 symmetrical and asymmetrical vibration peaks to PC1. PC2 and PC3 have high amplitude contributions in the 1,400-1,700 cm^{-1} region. (D) Scatter plot visualisation of top 3 PCs indicate cμFTIR data does not clearly distinguish MS from control spectra, although clear trends exist such as higher prevalence of low PC1 and high PC2 scores for MS spectra.

3.2.4.2 Contrasting the macromolecular features of SPMS and control SC tissue cμFTIR spectra

Integrating the area below the lipid region demonstrated a trend towards a lower mean SPMS lipid content that was 0.7 times that of the controls (Fig 3.21). There was also a weak trend towards a decreased lipid saturation and carbonyl ester levels (0.9 and 0.6 times that of controls, respectively) and an increased lipid oxidation (1.1 times that of controls). However, none of these trends were statistically significant – potentially partly due to the low number of samples. Similarly to brain characterisation, the violin plots of the individual spectral features suggested the presence of at least two subpopulations within the SPMS spectra.

Similarly to the lipid region, integration of the area below the protein region did not reveal significant differences between cases and controls. The mean SPMS protein content was ~1.2 times that of controls (Fig 3.22). Of the protein folding secondary structure features, the β-sheet features contributed 1.7 times more to MS cases than controls. β-turns were also slightly more frequent in MS by 1.22 times. Of these features, the β-sheet had the most varying distribution within both case and control populations based on an inspection of violin plots.

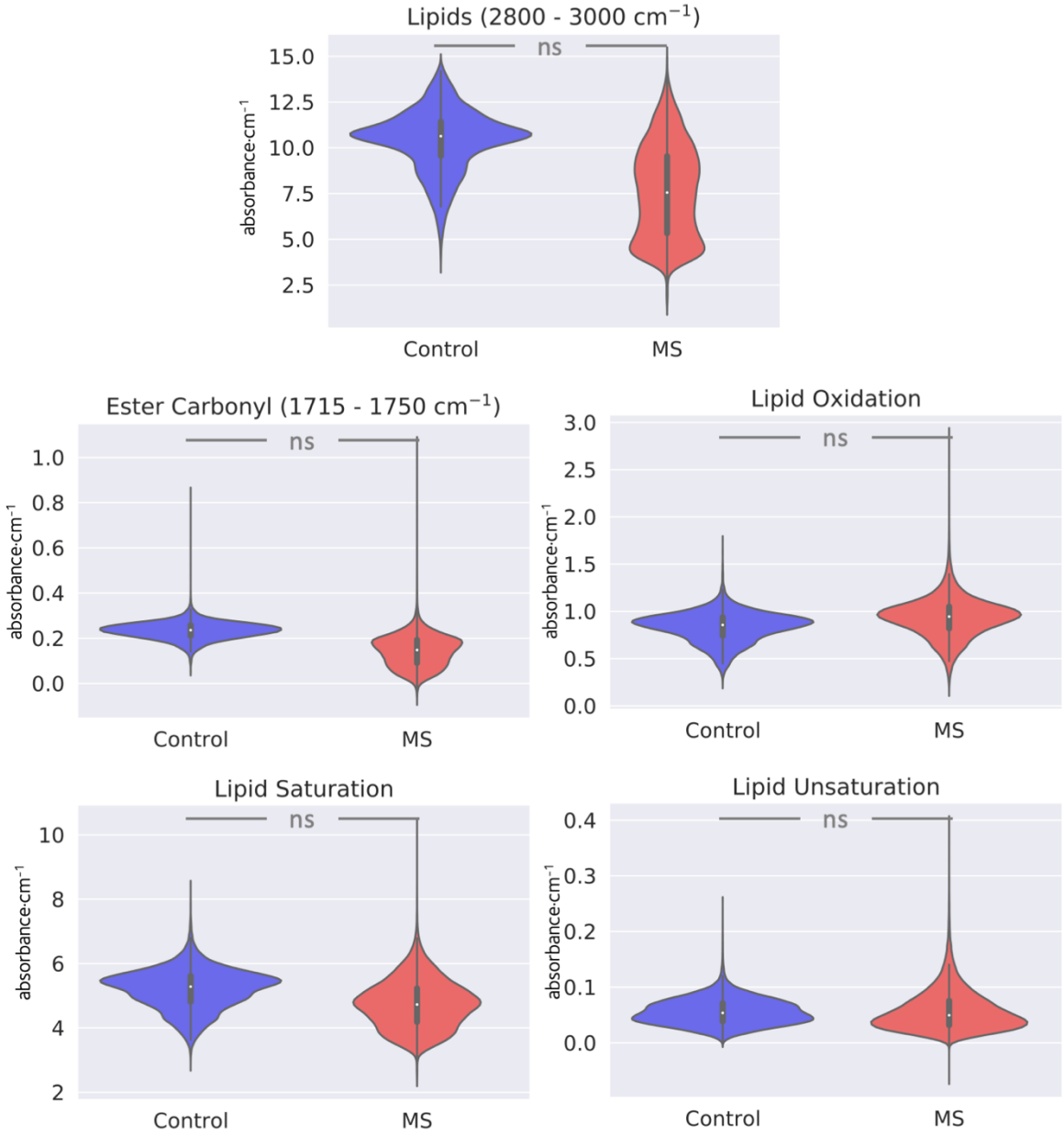


Figure 3.21: Comparison of lipid content and state in control and MS cμFTIR control and MS SC reveals no statistically significant differences. Violin plot visualisation indicates intra-group heterogeneity is present in both MS and controls. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001

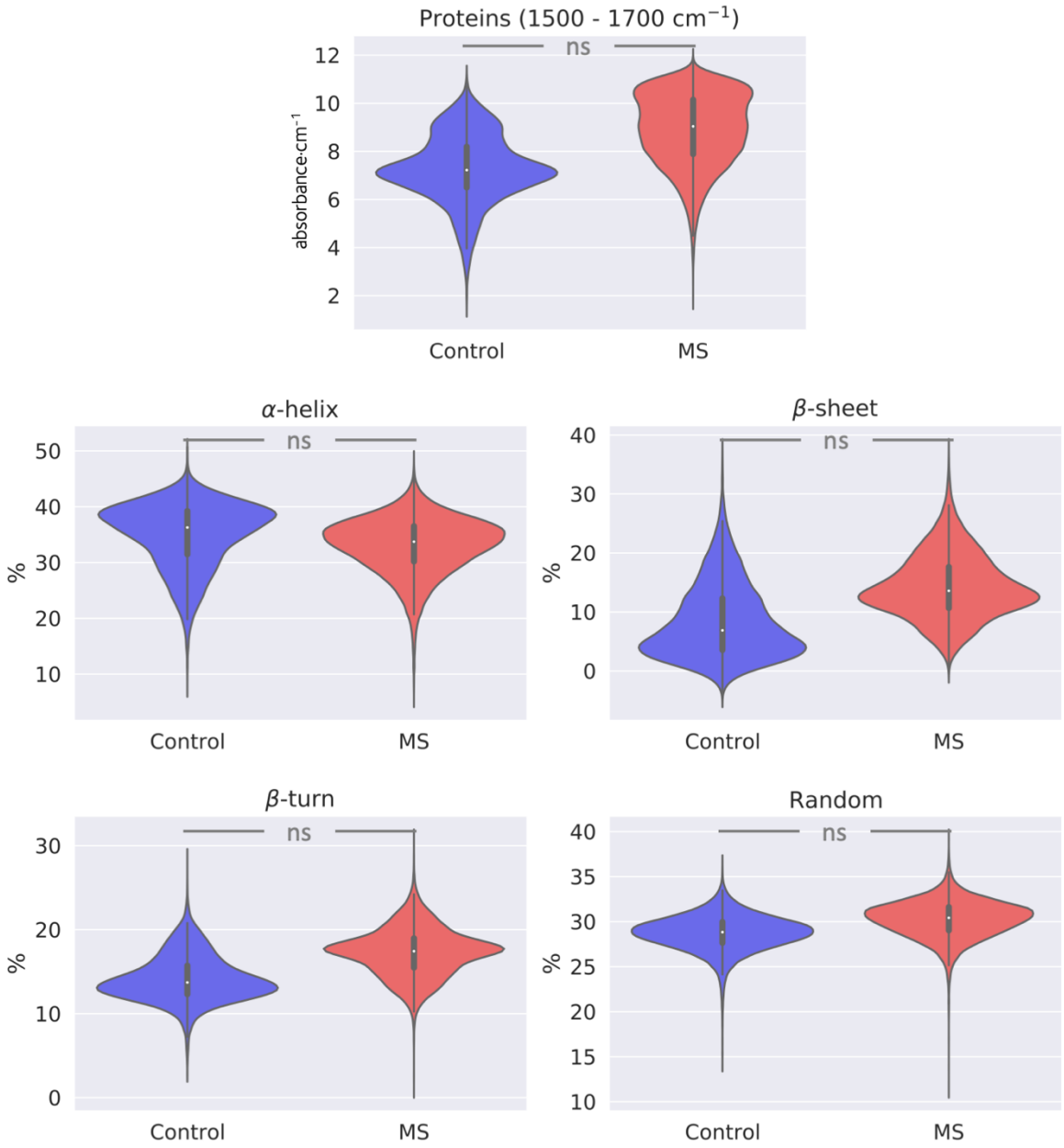


Figure 3.22: Comparison of protein content and protein secondary structure in cμFTIR control and MS SC reveals no statistically significant differences. Violin plot visualisation indicates intra-group heterogeneity is present in both MS and controls. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

3.2.4.3 Spatial mapping of macromolecular content distribution

In order to characterise the macromolecular content relative to its spatial distribution within the SC tissue sections, the integrations calculated for each feature of interest were mapped spatially in the original configuration at the time of data collection. In the controls a clear distinction could be observed in molecular composition which corresponded to WM and GM tissue types. This was most clear when mapping total lipid and total amide content, where the WM had high lipid content, high lipid saturation, low lipid unsaturation (**Fig 3.23**).

For SPMS patient MS376, the sample analysed was histologically characterised by hypercellularity and presence of myelin laden cell types, and featured substantial variation in all macromolecular maps (**Fig 3.24**). Low lipid regions were characterised by high amide content, low carbonyl ester, high lipid oxidation, lower lipid saturation and also a lower nucleic acid (1,045-1,080+1,210-1,260 cm^{-1}) content. In contrast, regions high in lipid had a lower amide content, higher carbonyl ester, lower lipid oxidation and higher lipid saturation and higher nucleic acid content. Interestingly, regions transitioning from low to high lipid content had high amide II levels but not as high amide I levels and were the regions particularly marked by lipid oxidation. The co-variance in amide levels and oxidation levels may, for instance, be linked to processes of oxidative amide synthesis[162].

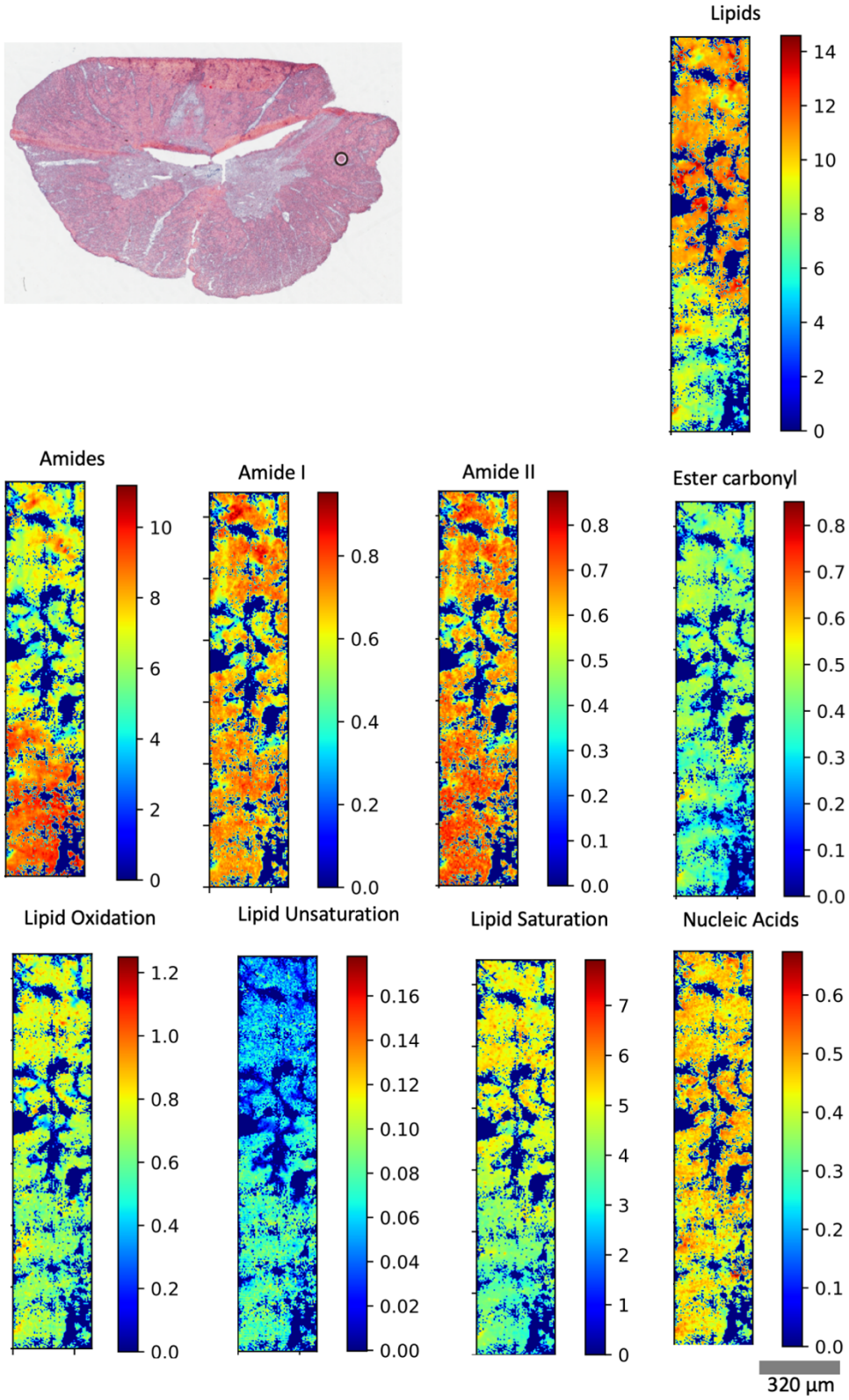


Figure 3.23: Spatial visualisation of biochemical distribution in C085 control SC reveals features identified in histological characterisation. Total lipid and amide visualisation outline WM and GM regions. In contrast with the GM, the WM is characterised by high lipid content, low total amide content, low lipid unsaturation and high lipid saturation.

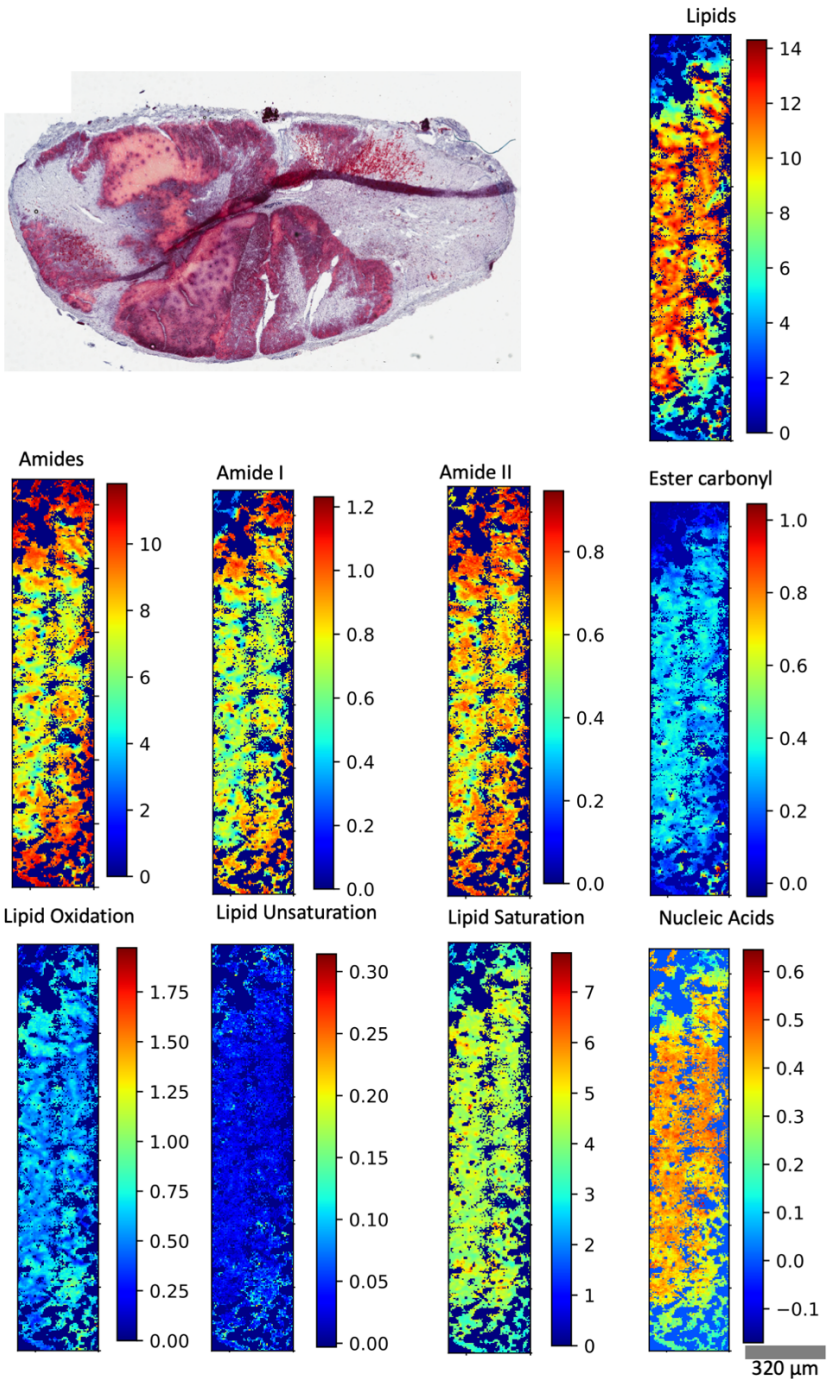


Figure 3.24: Spatial visualisation of biochemical distribution in MS376 MS SC reveals features identified in histological characterisation. MS376 is histologically characterised by active immune cell infiltration (indicated by hypercellularity), myelin laden phagocytosing cells and demyelinated lesions. Demyelinated areas are characterised by lower lipid content, higher amide, amide I and amide II content, low ester carbonyl and high lipid oxidation scores. The myelinated area is characterised by the prevalence of higher lipid content, low amide and amide I specifically, as well as high ester carbonyl, lower lipid oxidation as well as high nucleic acid content. Notably, the regions of high lipid oxidation scores can be mapped most clearly to the transition between the demyelinated and still myelinated region.

The second SPMS sample, from patient MS432, was histologically characterised by severe atrophy, low lipid content and the absence of common architectural features expected in SC sections, and has therefore been classified as a chronic inactive lesion (Fig 3.25). The biochemical spatial maps show a mostly homogeneous region with low lipid content, low lipid oxidation, saturation and unsaturation, low nucleic acid content and high protein content. A single distinguished area has high lipid and amide content, and surrounding this region is an area with higher carbonyl ester content with a distinctive pattern recapitulated in the lipid saturation map but not in the lipid oxidation map.

3.2.4.4 Unsupervised clustering with UMAP

Taking an analogous approach to that used for the brain tissue data, the data from the protein region of the spectra were used to drive clustering using UMAP for dimensionality reduction and leiden community detection, enabling the identification of six clusters (Fig 3.26a). The cluster architecture here was significantly distinct from that of the brain data, which may relate to the smaller number of samples. Cluster annotation was informed by the biochemical profiling of each cluster and their spatial mapping relative to the histological staining (see section 3.2.4.5 below), as performed for the brain tissue clusters. Likely lesional "Les" clusters were labelled numerically based on cluster size, and importantly do not necessarily have a biochemical profile corresponding to that of the brain tissue clusters with the same annotation; the brain and SC lesional clusters should thus be considered distinctly.

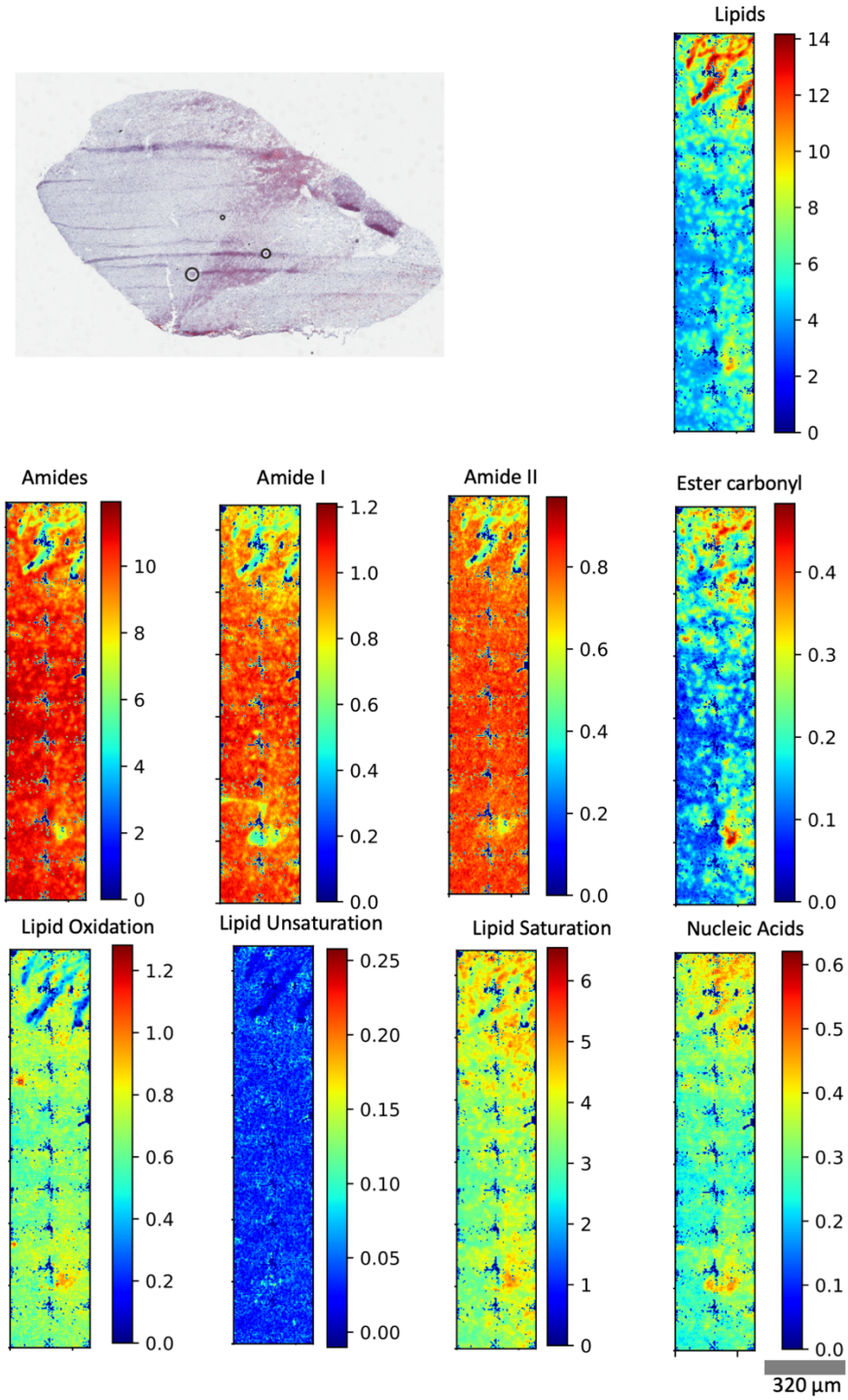


Figure 3.25: Spatial visualisation of biochemical distribution in MS432 MS SC reveals features identified in histological characterisation. MS432 is histologically characterised by severe demyelination, spinal cord atrophy and overall loss of stereotypical SC architectural features. A small region with seemingly intact myelin content has similar biochemical profile to WM in C085, while the majority of the scanned area is marked by low lipid content, high amide content, low ester carbonyl and small changes in lipid oxidation, lipid saturation and nucleic acid content. Notably, there were no clear regions of differential lipid oxidation content in the transition from myelinated to demyelinated regions.

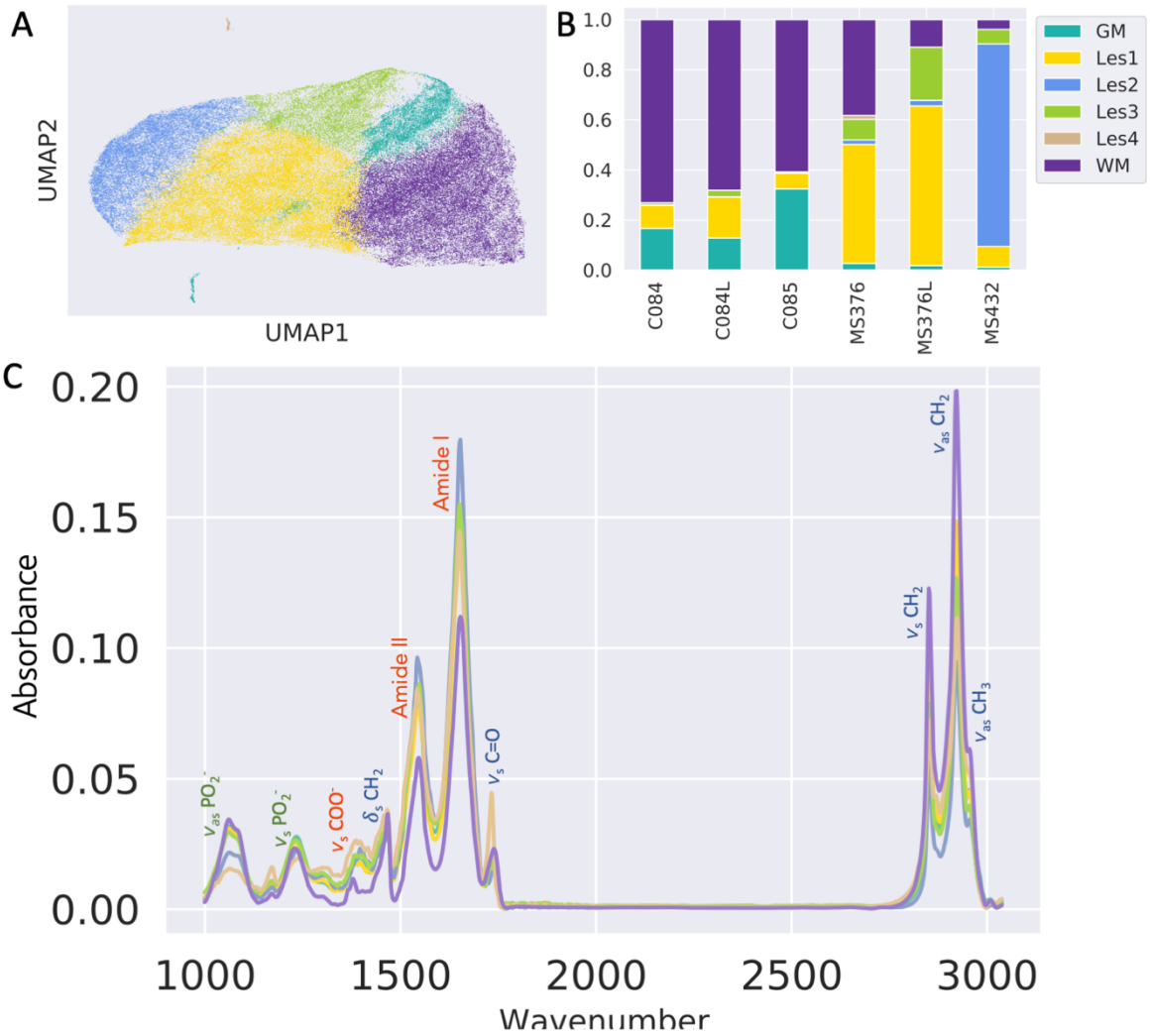


Figure 3.26: Amide-driven unsupervised clustering of cμFTIR spectra reveal a continuous biochemical gradient from WM across GM, and lesion types. (A) Leiden community detection was used to partition data and characterise the biochemical gradient from spectra characterising healthy SC to MS-affected SC. (B) Cluster distribution across samples indicate WM and GM are predominantly represented in control samples, while Les clusters are predominantly – but not exclusively – represented in MS. Les2 is mostly represented in MS342, indicating a distinct biochemical profile characterises this sample. (C) Mean spectra for each cluster reveals WM spectrum is characterised by having the highest lipid content, a moderate carbonyl ester absorbance (~1,740 cm^{-1}), low amide content and high nucleic acid content. Les2, in contrast, has the lowest lipid and highest amide content. Les4 is uniquely characterised by a particularly high absorbance in the carbonyl ester region
(~1740 cm^{-1}).

Inspection of the mean spectra for each of the six clusters showed that the WM cluster was characterised by high lipid content and low protein content and was present in most scanned areas, contributing mostly to controls but also to MS376 (Fig 3.26b). The GM, Les1, Les3 and Les4 clusters had lower lipid content than the WM and were also characterised by increasing protein content. Les4 had a distinct spectral profile with a high absorbance peak in the carbonyl ester region, a shift in the peak in the olefinic region (3,000-3,020 cm^{-1}) and particularly low absorbance in the nucleic acid peaks (Fig 3.26c). Several distinguishing features can in fact be observed for each cluster in the carbonyl ester region: Les1 and Les2 had a shift in absorbance to lower wavenumbers (1,733 cm^{-1}) and have a small absorbance peak at 1,715 cm^{-1} (Fig 3.27). Les3 had two apparent absorbance peaks in this region – one at 1,730 cm^{-1} and a second at 1,737 cm^{-1}; this may indicate intra-cluster heterogeneity or the presence of two distinct absorbance peaks due to hydrogen binding. Lastly, Les2 was sample-specific, being found only in the sample from patient MS432, and mapped to regions of chronic damage. This cluster had the highest protein content and lowest lipid content.

Complementary to the inspection of the mean spectra for each cluster, the integrated area below the spectra in regions of interest was also calculated and compared across clusters. For the lipid region integration (Fig 3.28), WM and Les1 had the highest mean lipid content (11.24 absorbance·cm^{-1} and 8.17 absorbance·cm^{-1}, respectively), while Les2 had the lowest (5.46 absorbance·cm^{-1}).

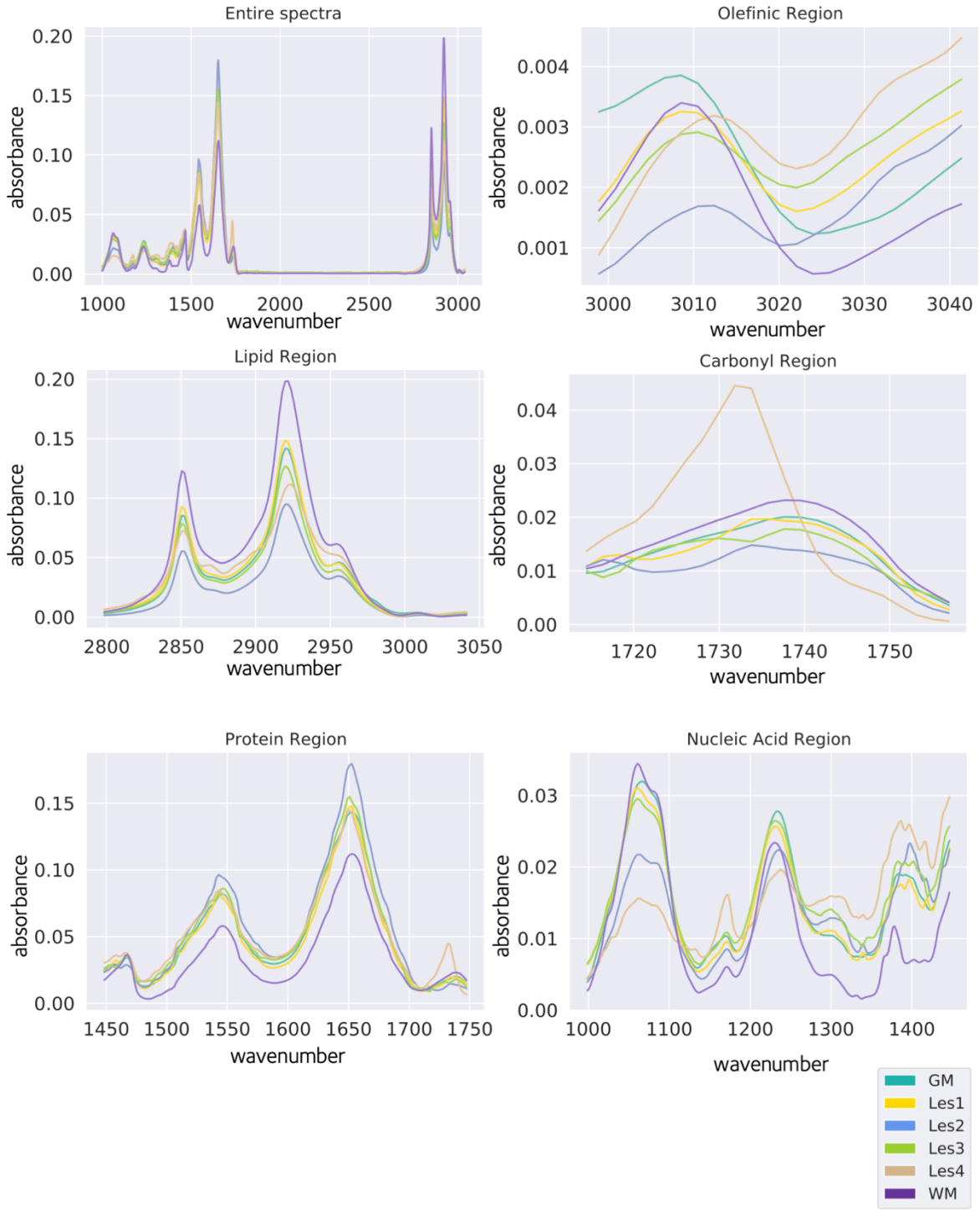


Figure 3.27: Visualisation of qualitative spectral alterations in cμFTIR SC clusters. Several qualitative differences in lipid, protein and nucleic acid regions are evident. Most notably and clearly distinguishable are significant qualitative alterations in the carbonyl ester region. Les2 is marked by a shift in maximal absorbance position from ~1,740 cm^{-1} (position of maxima for WM and GM) to ~1,733 cm^{-1} . Les1 and Les2 also have a small absorbance peak in the 1,733 cm-1 position as well as a novel small peak at 1,715 cm^{-1}. Les3 is marked by two peaks at ~1727 cm^{-1} and ~1737 cm^{-1}.

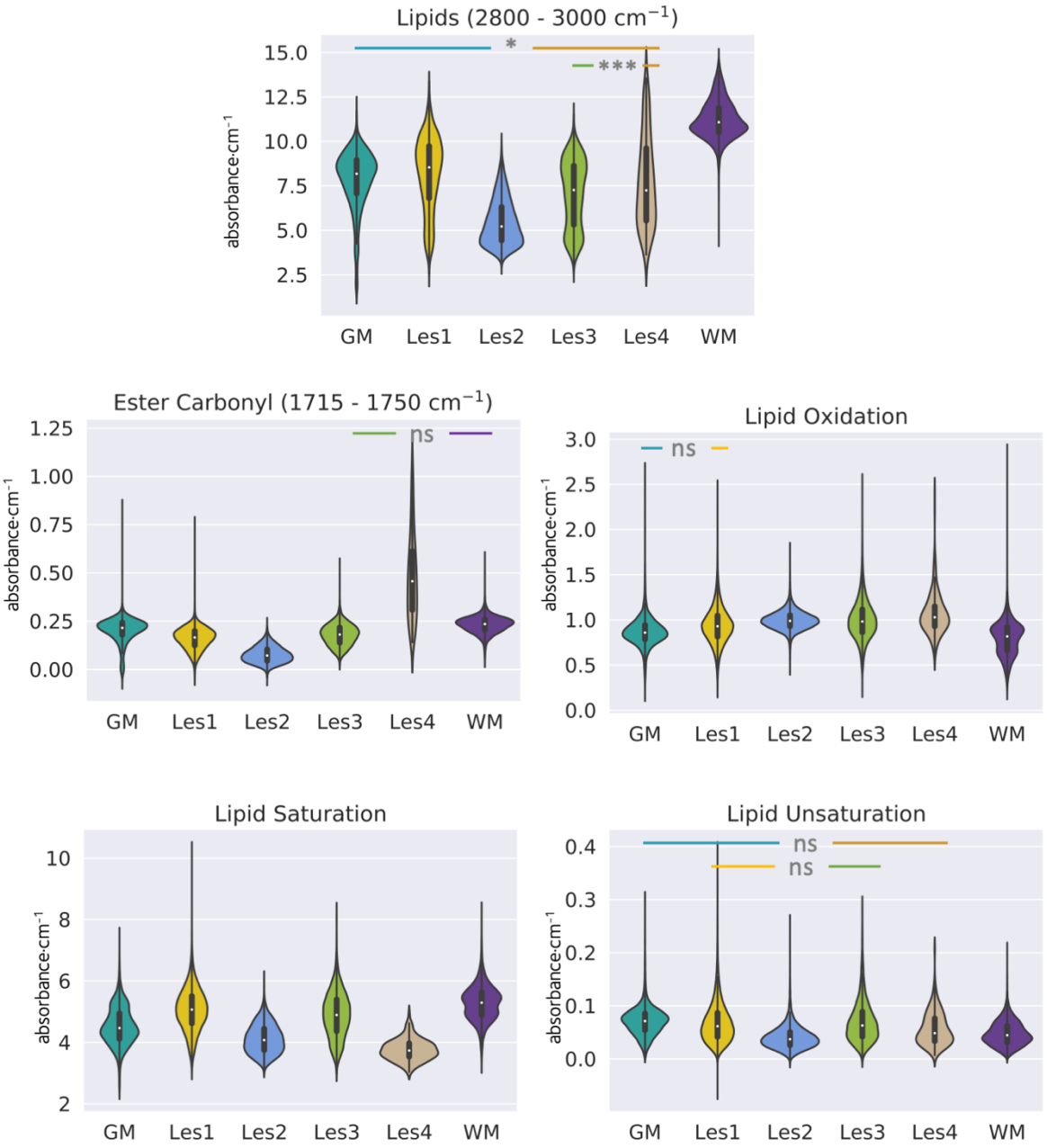


Figure 3.28: Comparisons of lipid content and state across clusters from cµFTIR SC data. All pairwise comparisons are statistically significant, unless otherwise indicated. WM, Les1 and GM clusters are marked by higher mean lipid content, while Les2, Les3 and Les4 are characterised by decrease in lipid content, increase in lipid oxidation; Les2 and Les4 are also marked by a decrease in lipid saturation. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

The cluster with lowest lipid oxidation levels was WM, and the two clusters with highest lipid oxidation were Les4 and Les3, which had 1.34 and 1.25 times the lipid oxidation of WM. Les4 had the highest carbonyl ester content (0.48), while Les2 had the lowest at 0.07 (6.8-fold difference). The pair-wise comparison differences of lipid features were statistically significant, except for the Les3 and WM comparison of carbonyl ester content, GM and Les1 lipid oxidation, and the GM/Les4 and Les1/Les3 lipid unsaturation comparisons.

The protein region Les2 had the highest protein content (10.35 absorbance·cm^{-1}), while the WM had the lowest (6.64 absorbance·cm^{-1}; 1.55-fold difference) (**Fig 3.29**); these clusters also had the highest α-helix content (34.86% and 36.2%, respectively), while Les4 and GM had the lowest (23.54% and 26.73%, respectively). WM had the lowest secondary structure contribution for β-sheets (6.61%), β-turns (13.26%) and random: folding features (28.15%). Conversely, Les4 had the highest contribution for these features (β-sheet: 22.19%; β-turn: 20.44% ; random: 31.58%). The protein-related comparison differences across these clusters were statistically significant (padj<0.0001), except for the Les1/Les4 and Les2/Les4 comparisons of total protein content. These data are summarised in **Table 3.3** and **Table 3.4** for lipid and protein-associated features, respectively.

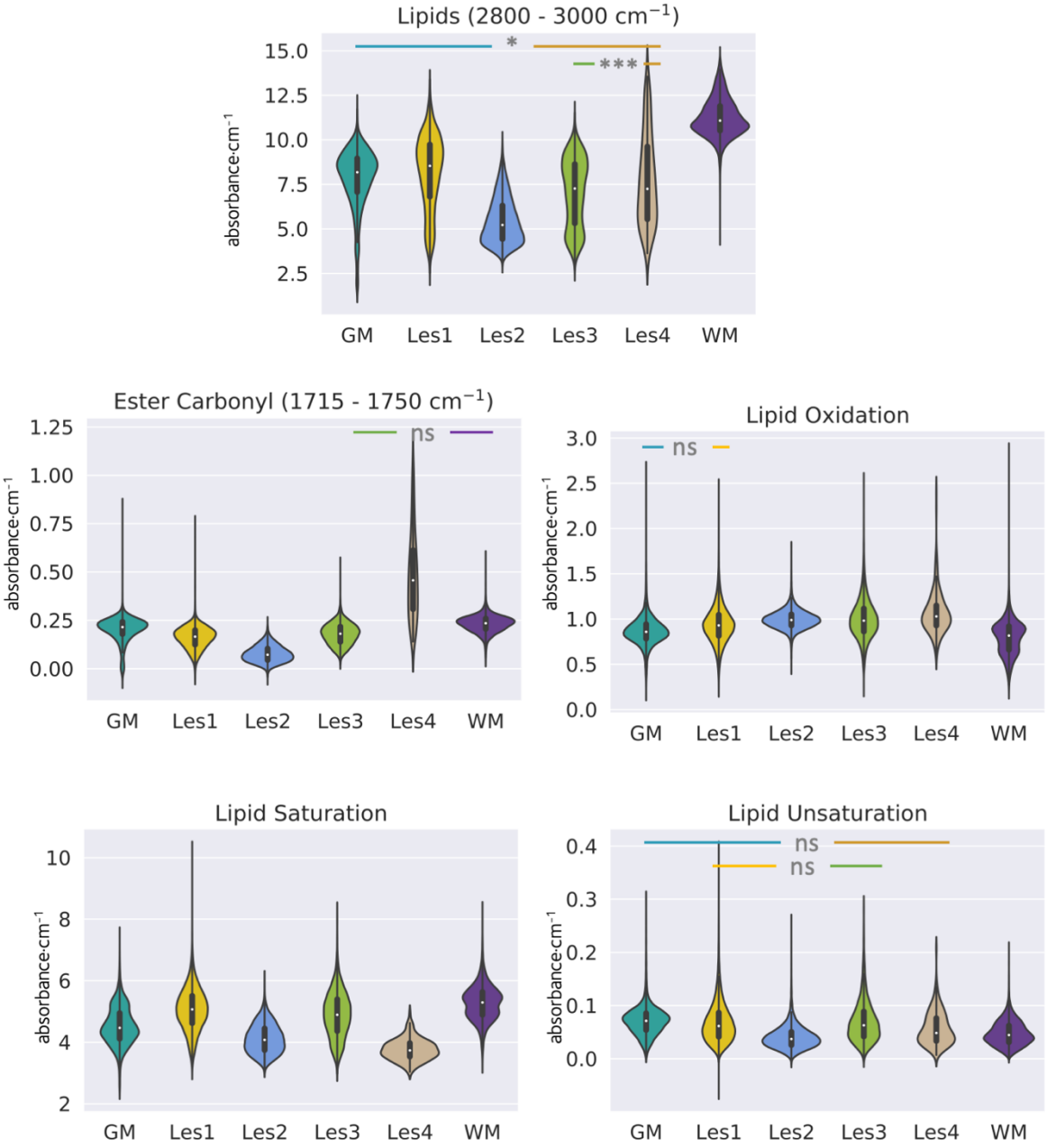


Figure 3.29: Comparisons of protein and protein secondary structure across clusters from cμFTIR SC data. Lesions have an overall increase in protein content and changes in protein folding features. Les clusters are marked by a decrease in α-helix content when compared to WM; this is most notable in Les3 and Les4 which have intermediary demyelination, while Les1 (higher myelin content among Les clusters) and Les2 (lowest myelin content among Les clusters) have more discrete changes. All Les clusters are marked by a higher β-sheet, β-turn and random folding content - this is most notable in samples with a decrease in α-helix content. All pairwise comparisons are statistically significant except where indicated otherwise. ns – non-significant; * padj<0.05; ** padj<0.01; *** padj<0.001; **** padj<0.0001.

Table 3.3: Summary of lipid-associated features of spinal cord clusters

Clusters	Total nuclei	Lipids	Carbonyl ester	Lipid Oxidation	Lipid Saturation	Lipid Unsaturation
GM	8837	7.817± 1.76 x10^{-2}	0.202± 6.80 x10^{-4}	0.865 ± 1.58 x10^{-3}	4.529± 6.53 x10^{-3}	0.071± 2.78 x10^{-4}
Les1	30643	8.177± 1.19 x10^{-2}	0.161± 3.19 x10^{-4}	0.937 ± 1.14 x10^{-3}	5.060± 3.91 x10^{-3}	0.068± 2.12 x10^{-4}
Les2	16478	5.462± 9.96 x10^{-3}	0.076 ± 3.28 x10^{-4}	0.994± 7.88 x10^{-4}	4.112± 3.67 x10^{-3}	0.040± 1.62 x10^{-4}
Les3	9124	7.034± 1.99 x10^{-2}	0.179 ± 5.70 x10^{-4}	0.998 ± 2.24 x10^{-3}	4.885± 7.66 x10^{-3}	0.070± 4.09 x10^{-4}
Les4	199	7.778± 1.79 x10^{-1}	0.481± 1.59 x10^{-2}	1.069± 1.67 x10^{-2}	3.761± 2.45 x10^{-2}	0.057± 2.19 x10^{-3}
WM	31909	11.240± 5.91 x10^{-3}	0.232± 2.38 x10^{-4}	0.794 ± 1.03 x10^{-3}	5.277± 3.10 x10-3	0.048± 1.26 x10^{-4}

Table 3.4: Summary of protein-associated features of spinal cord clusters

Clusters	Proteins	α-helix	β-sheet	β-turns	Random
GM	9.111 ± 8.16x10^{-3}	26.737± 5.67x10^{-2}	17.388± 6.08x10^{-2}	17.928± 2.50x10^{-2}	30.270± 1.82x10^{-2}
Les1	8.453± 6.34x10^{-3}	33.783± 2.86x10^{-2}	14.811 ± 3.36x10^{-2}	16.917± 1.70x10^{-2}	30.417± 1.08x10^{-2}
Les2	10.354± 5.08x10^{-3}	34.858± 2.23x10^{-2}	12.947 ± 1.92x10^{-2}	17.670± 8.43x10^{-3}	31.370± 8.64x10^{-3}
Les3	9.372± 9.64x10^{-3}	29.712± 5.11x10^{-2}	16.309± 5.44x10^{-2}	19.469± 2.66x10^{-2}	29.459± 2.00x10^{-2}
Les4	8.303± 1.14x10^{-1}	23.547± 3.46x10^{-1}	22.193± 2.76x10^{-1}	20.443± 1.48x10^{-1}	31.590± 1.04x10^{-1}
WM	6.644± 6.04x10^{-3}	36.216± 2.53x10^{-2}	6.608± 2.61x10^{-2}	13.265± 1.18x10^{-2}	28.151± 1.05x10^{-2}

3.2.4.5 Spatial mapping of cμFTIR of active spinal cord lesion shows no clear lesion borders

Each spectrum was colour coded based on their cluster classification and mapped in the position of spectral collection (**Fig 3.30**). In the controls, grey matter and white matter patterns were distributed as expected based on histology. In the cases, there were no clearly identifiable stereotypical architectural features associated with SC anatomy, such as the butterfly-shaped GM in the centre of the SC section.

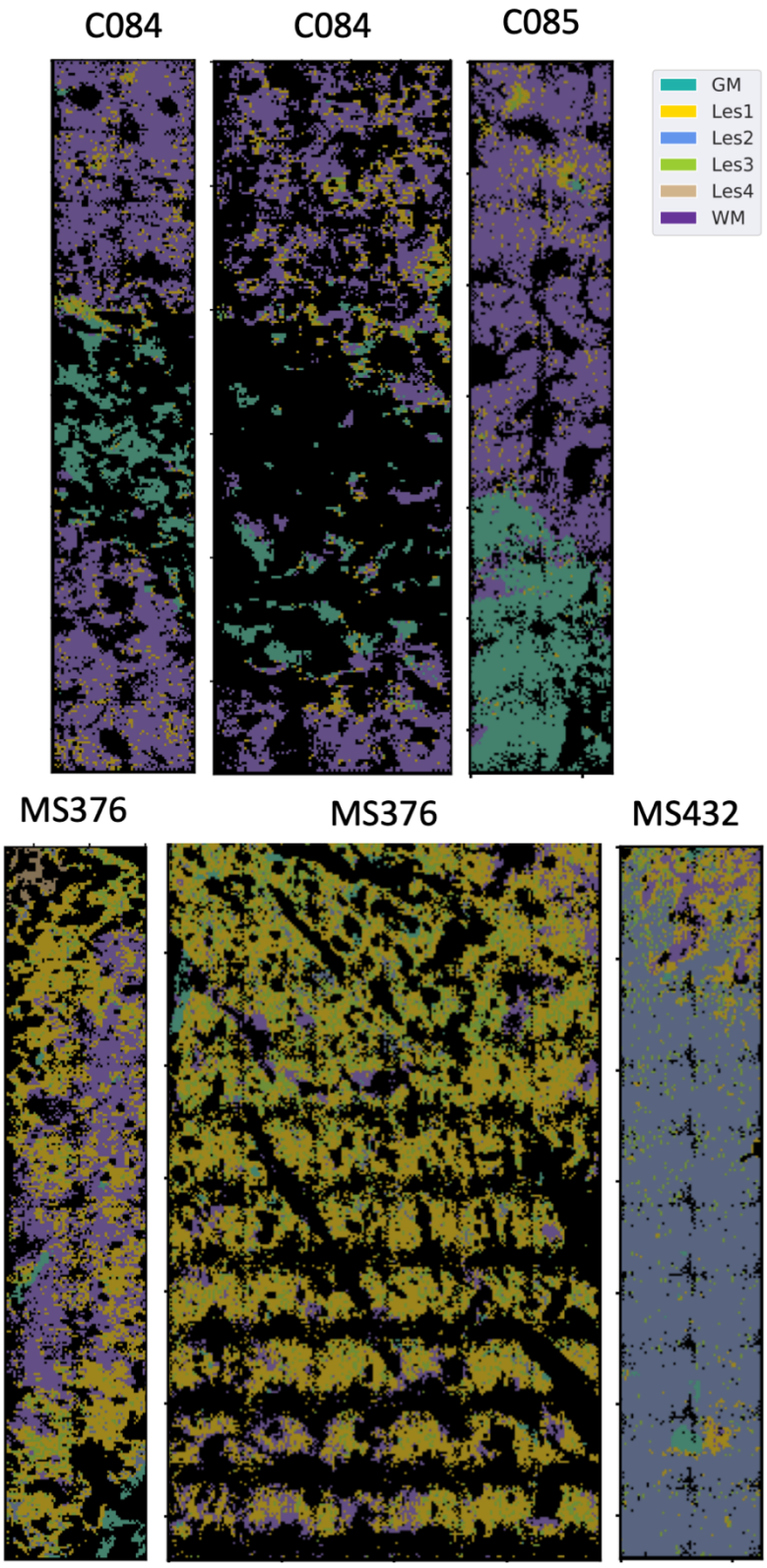


Figure 3.30: Spatial mapping of cluster classification reveals histologically identified architectural features. The chronic MS lesion in MS432 is marked by homogeneity, while the active MS lesion in MS376 is marked by heterogeneous composition and complex distribution.

MS376 had sites classified as WM which transitioned to lesional. The small Les4 cluster, characterised by high carbonyl ester content and protein secondary structure changes was localised to a small site in the MS376 map. The most abundantly present cluster classification in this section was Les1 with interspersed areas classified as WM and GM with no clear borders of lesion focus. Consistently with the chronic nature of the lesion, the section from sample MS432 was mostly characterised by the presence of the low lipid cluster Les2, with a small region remaining that was classified as WM.

3.3 Discussion

3.3.1 Outlining the relationship between lipid and protein content in SPMS pathology

In line with early spectroscopy analyses of MS, these μFTIR analyses demonstrate that changes in the lipid and protein content in the CNS tissue of SPMS patients are defining biochemical features of pathology[158, 159]. Consistently, damaged tissue areas with a pronounced decrease in lipid content showed an increase in protein content. Beyond this, we have been able to outline that relationship more mathematically than previous work – determining that this was not simply a linear correlation in the context of pathology and was better explained as a second-degree polynomial relationship – indicating an underlying biological complexity in the variation of these macromolecules in patients. Moreover, the broader changes in lipid and protein content were also accompanied by changes in lipid state, with less lipid saturation and more lipid oxidation

generally occurring in regions with a lower lipid content, as well as by alterations in protein folding secondary structure. We have also observed that these alterations are spatially complex, with focal areas in brain samples which were biochemically distinct from surrounding lesions, and in particular in the active MS376 sample there were no clear lesion borders. We observed both lesions with well outlined borders of demyelination, such as MS352, as well as lesions with more gradient-like differences in lipid content such as MS466.

3.3.2 High-resolution lipid biochemistry profiling in the SPMS CNS

The prominent role of lipid content in the changes of the MS-affected CNS is indisputable. Traditional histological characterisation of MS lesions clearly features visualisation of lipid using LFB and ORO staining given the demyelination that occurs in the disease, and μFTIR analysis has consistently reported alterations to lipid content in MS lesions[158, 159]. Thus, the reproducible identification of lipid-associated CH_2 and CH_3 vibrations as the main contributors to PC1 across all μFTIR analyses here reported reiterates analytically the prominent role of changes in lipid content in MS. Moreover, the breadth of variation in lipid content across MS samples was made evident in these analyses, with levels of lipid in SPMS being as much as 2.7 times less than the content in the WM in the brain and 2.2 times less than the WM in the SC. Apart from differences in relative lipid content, alterations in lipid oxidation, lipid saturation and lipid unsaturation were also identified.

In healthy physiological conditions, reactive oxygen species (ROS) produced in the process of aerobic respiration are balanced by antioxidant detoxification of radical species. An imbalance in this system – whether through an increase of ROS output or a decrease of antioxidant capacity – will cause cellular damage and can have pathological repercussions. ROS production results from a radical-induced modification of molecular oxygen or incomplete consumption of molecular oxygen. This produces a superoxide anion which is metabolized by superoxide dismutase to hydrogen peroxide which in turn is transformed into water. However, if instead the hydrogen peroxide reaches cellular regions rich in free iron it can be converted to hydroxyl radical. In the context of MS, due to the role of oligodendrocytes in iron storage, oligodendrocyte cell death will release stored iron making this event more likely to take place[163].

Previous μFTIR characterisation of MS reported an increased carbonyl ester absorbance in MS when compared to controls[158, 159], and this was suggested to relate to lipid oxidation since mechanisms such as protein carbonylation – a process by which a C=O carbonyl group is added to proteins and is increased in the context of oxidation – contribute to this peak. A subset of sample regions indeed was been identified to have this spectral feature - in particular the SC-Les4 and brain-Les3 clusters. However, sample clustering showed that this biochemical feature was not common across all SPMS cases and was frequently localised to specific tissue regions in particularly high content. Since these previous reports used the mean MS spectra, it is possible that this feature was overrepresented in the mean due to a small subset of highly localised

spectra with particularly high absorbance, such as is the case with the small SC-Les4 cluster localised to a small region of the MS376 tissue section (Fig 3.24), or the small region of the MS466 tissue section which had a particularly high carbonyl ester content (Fig 3.10).

Interestingly, however, in the clusters where this feature was present, unlike the control WM and GM, the carbonyl ester peak position was at wavenumber 1,733 cm^{-1}. The absorbance peak position in this region is known to be highly susceptible to hydrogen bond number and strength; the shift from 1,740 cm^{-1} to 1,733 cm^{-1} may thus indicate an increase in the number of hydrogen bonds[164, 165]. A change in number of hydrogen bonds would be consistent with oxidative events, since once formed, hydroxyl radicals cannot be metabolized and can only be consumed by removing hydrogens from neighbouring molecules, thus forming unsaturated lipids. Although the evidence of a potential increase, rather than decrease, in hydrogen bonds is intriguing, potentially indicating overcompensation of anti-oxidative mechanisms.

The appearance of a peak in this position has been reported to be characteristic of the formation and decay of a COOH mode upon reduction[166], which could indicate an increase in reduction in response to oxidation. Thus, the use of the integrated area of this region for indicating the presence or absence of oxidative processes while useful is not all-encompassing, also requiring inspection of the spectral distribution in order to assess whether carbonyl ester peak may be indicative of oxidative or reductive processes in response to oxidation.

Carbonyl ester maxima in the 1,733 cm^{-1} position are also present in a second spectral profile observed in SC-Les1, SC-Les2, Brain-Les2 and Brain-Les4, which had local maximum absorbance at 1,733 cm^{-1} as well as an absorbance peak prominently featured at ~1,715 cm^{-1} which was not present in other clusters. Three out of four of these clusters had markedly low total lipid content. The appearance of this absorbance peak is consistent with the aftermath of lipid oxidation, particularly given the evidence of advanced demyelination in the clusters with this feature. The 1,715 cm^{-1} region is known to have contributions from small polyunsaturated molecules which are a product of oxidation. Among the molecules that might have hydrogen bonds removed for this process to occur are polyunsaturated fatty acids which, following hydrogen removal, have oxygen added instead, as lipid oxidation occurs.

Notably, among the molecules represented in this spectral region are α,β-unsaturated aldehydes which are a class of molecules produced in the lipid peroxidation process, of which 4-hydroxynonenal (HNE) has been shown to be present in several neurodegenerative conditions, such as AD, PD, diffuse Lewy body disease, and is considered to be indicative of oxidative stress[167]. The impact of the presence of these cleaved unsaturated lipid molecules is not limited to lipids and can also modify DNA, RNA and proteins, thus warranting further investigation of alterations resulting from oxidation.

In order to quantify changes in lipid saturation/unsaturation, a previously reported approach based on the ratio of CH_2/CH_3 as a measure of lipid saturation and

Olefinic/CH_3 as lipid unsaturation was used[137]. The SPMS affected CNS was found to have overall lower lipid saturation and higher lipid unsaturation in sμFTIR analysis; this difference was more pronounced in lesions with a chronic/advanced demyelination profile. A decrease in lipid saturation would be consistent with the observation of increased oxidation in lesions, since removal of hydrogen bonds is a key step preceding the addition of an oxygen bond to lipid molecules. However, the absence of a clear correlation between the estimated saturation and unsaturation is a confounding factor for this interpretation (**Fig 3.8**). An alternative interpretation for this specific spectral ratio is that the CH_2/CH_3 ratio is a measure of lipid length[136], in which case the reduced representation of this ratio in cases could instead indicate the presence of shorter lipids. This would be consistent with the findings of increased absorbance in the ~1,715 cm^{-1} region associated with the presence of cleaved lipids in clusters with advanced demyelination.

The lipid unsaturation interpretation of the Olefinic/CH_3 ratio relates to the fact that the olefinic C=C-H group at 3,012 cm^{-1} is a product of free radicals reacting with each other; this reaction would also decrease the absorption of C-H bonds at 2,920 cm^{-1} [137]. However, the applicability of this metric to SPMS may be confounded by the process of demyelination also impacting the absorption at 2,920 cm^{-1}. Therefore, while the lipid unsaturation metric is inconclusive within the context of MS, the CH_2/CH_3 ratio does indicate alterations in lipid metabolism and both the interpretation of decreased lipid saturation and length are indicative of ongoing oxidative stress, which is more

pronounced in Brain-Les2, Brain-Les3, Brain-Les4 and Brain-Les5, with Brain-Les2, Brain-Les4 and Brain-Les5 being the clusters with highest oxidation scores, hence supporting these findings.

3.3.2 High-resolution protein biochemistry profiling in the SPMS CNS

Further to extensive differences in lipid biochemistry in the SPMS CNS compared to controls, we also consistently observed an increase in total protein content which was negatively correlated with lipid content. While there are other macromolecular contributors to the 1,500-1,700 cm^{-1} region used for this estimation, such as water and NO[158], which may have some effect on this outcome, an increase in the absolute total protein content is functionally plausible in the pathophysiological context of MS.

An increase in total proteins may relate to hypercellularity which can occur as a consequence of immune cell infiltration or, alternatively, in more chronic lesions, due to recruitment of CNS-resident astrocytes and microglia in response to longer term damage. Due to the demyelination process within the area, the overall lipid content of the region would decrease; while hypercellularity would include cells with intact lipid content, the lipid contribution from these infiltrating cells may be small when compared to the lipid loss due to demyelination. Thus, hypercellularity would be mostly reflected in the protein content. The variable nature of the cells within lesions and hence the overall lesional architecture could contribute to the lack of linearity in the lipid/protein relationship observed in the SPMS cases (Fig 3.8a). Alternatively, given the substantial evidence of oxidation, and increase in protein carbonylation may also contribute to the

detected increase in total protein content in MS. While moderately carbonylated proteins are degraded, heavily carbonylated proteins have been reported to form degradation-resistant aggregates which accumulate and contribute to neurodegenerative processes[168, 169]. The accumulation of protein aggregates would inevitably result in a higher protein content as measured by μFTIR. Lastly, oxidative processes have a role in amide synthesis[162]. The observation of co-localised increased amide II and lipid oxidation simultaneously in MS376 may be indicative of this process.

In addition to changes in protein content, the high-resolution sμFTIR data showed a decrease in α-helix protein folding structures and an increase in other protein secondary structures in the SPMS CNS when compared to controls, particularly in β-sheet content. Using the clustered cμFTIR data it was also possible to identify these alterations to protein secondary structure. In particular, lesional clusters with an apparent intermediate degree of demyelination (as inferred by the lipid content) had a notable decrease in α-helix content; an increase in β-sheet content could also be observed which was highest in lesional clusters with intermediate demyelination (SC-Les3, SC-Les4, Brain-Les4 and Brain-Les5). These proteomic alterations correlated with high lipid oxidation. A decrease in α-helix and increase in β-sheet content have been previously reported in the brain affected by neurodegenerative conditions with protein folding involvement, such as AD and PD[170, 171] but have not, to our knowledge, been previously reported in SPMS.

These findings could be explained by the occurrence of a difference in overall protein composition in response to processes triggered during neurodegeneration, and these proteins may have an inherently higher incidence of β-sheet folding. A potential candidate for this is collagen, which contributes to the amide region[172] – already identified to have higher absorbance in MS when compared to controls – and which is associated with CNS response to injury[173]. Spectroscopic analysis of collagen has already indicated that certain collagen types, including type IV and VI have increased β-sheet and random folding content.

Other protein candidates which may be overrepresented, for instance, include neurotrophic factors promoting neurorescue, and this may be consistent with these protein folding features being prominently observed in tissue areas with an apparent intermediate degree of demyelination, where neuroreparative mechanisms may be at play. Exploration of this hypothesis is reliant on either improved access to protein folding features in databases jointly with proteomic profiling, or may benefit from protein folding prediction tools such as AlphaFold[174].

Extrapolating from the recognised inherent involvement of protein misfolding in PD and AD, which is characterised by an increase in β-sheet content, the identification of an excess of this folding feature could be interpreted as a result of protein aggregates. Protein aggregates are not a characteristic feature of MS to the extent that they are in PD and AD, but a lower level of protein accumulation due to shared

neurodegenerative processes is plausible, particularly in late disease which has a neurodegenerative-like pathophysiology.

Rather than reflecting the accumulation of proteins with an inherently higher incidence of β-sheet folding, the profiles observed may also represent the accumulation of proteins that are selectively vulnerable to carbonylation processes[175] – while the overall effects of protein carbonylation in protein folding and function are protein- and carbonylation-site-dependent, certain amino acids or folding structures may be more susceptible to this type of modification. This hypothesis is consistent with reports of increased protein carbonylation in WM and GM of brains affected by MS [176]. Alternatively, this effect may be an indirect consequence of increased carbonylation, which has also been reported to induce dysfunction of protein folding processes due to the carbonylation of specific endoplasmic reticulum chaperones[177, 178].

To begin to address such hypotheses, and to provide a more mechanistic context for the biochemical profiles observed in the SPMS-affected CNS with respect to cellular and molecular pathway contributions, we next characterised the SPMS and control tissue through transcriptomic and proteomic approaches, as detailed in the subsequent chapters of this thesis.

3.3.3 Improved instrumentation and updated μFTIR analysis yielded valuable biological findings

This analysis successfully demonstrates the complimentary applicability of synchrotron-based v. conventional FTIR technology – the first ideal for disease characterisation, whilst the latter contributing to inquiry of spatial features and machine-learning spectral classification, jointly providing valuable insights into the biochemical nature of SPMS pathology. We also demonstrate the benefit of machine learning classification of spectral data and present a novel methodological approach and computational tool in python for this analysis. This comes at a time where the need for improved computational resources for maximising utility of spectroscopic data efficiently is becoming increasingly evident in the field, both in the context of diagnosis[156, 179] as well as novel applications in surgery[180, 181]. In subsequent chapters we continue to build on this data exploring integration with other high-resolution molecular profiling methods.

Chapter 4: Single-nuclear transcriptomic analysis of SPMS

4.1 Introduction

Histological analysis has been a valuable tool in profiling the cellular composition of MS lesions and has been the main approach used for lesion classification. The use of such analyses – which include staining with dyes and/or with antibodies against key markers – have also helped to provide some insights into the mechanisms of pathology, but have been limited due to restrictions in terms of the numbers of markers that can be assayed at any one time. The advent of omics technologies, however, have the power to enable more extensive interrogation of the cellular transcriptome and proteome – capturing hundreds to thousands of genes from individual cells or nuclei, and hence capturing dynamic cellular phenotypes, cell activity and cell-cell interactions in the context of both health and disease. Importantly, the lack of labels means this approach lends itself well for hypothesis generation as well as data reusability for separate specific queries. Among the omics technologies, single-nuclear RNA sequencing (snRNAseq)[182] has been particularly valuable for the study of conditions such as MS, since it can be applied to snap frozen tissue, thus uncoupling sample collection time from processing and sequencing[183] and enabling the use of archival, post-mortem frozen tissue.

The first application of snRNAseq to MS focused on the brain WM, and in particular oligodendrocyte (OG) characterisation[131]. A total of 6 OG clusters were reported to be differentially distributed when comparing cases and controls[184]. Oligo2

(LURAP1L-AS1high), oligo3 (MAGhigh) and oligo5 (KLK6high) were reportedly overrepresented in MS, whilst oligo1 (CDH20high), oligo4 (MAGlow, CDH20low) and oligo6 (OPALINhigh) were underrepresented. Differential frequency of these clusters was also reported across control WM, NAWM, active, chronic active, chronic inactive and remyelinating lesions, and the study showed a higher frequency of oligodendrocyte progenitor cells (OPCs) and committed oligodendrocyte progenitor cells (COPs) in the control WM than other groups. However, these observations were reported without statistical validation.

Notably, a second snRNAseq study did not find an analogous set of OG subpopulations such that these initial findings have not been confirmed. This second study applied a more unbiased sampling approach whereby both brain WM and GM were analysed, agnostic to lesion classification[130]. This study reported compelling evidence for the selective vulnerability of CUX2 neurons in MS. Astrocyte reactivity was also distinct in cortical and subcortical MS lesions, in particular astrocytes in the GM had high *SLC1A*, while WM astrocytes expressed *CD44*, for example. Transcriptomic changes in phagocytosing cells were additionally reported, including an increase in the number of *PLP1* transcripts in the nuclei of cells phagocytosing myelin, with *in vitro* data suggesting that the presence of these transcripts can at least in part be explained by the phagocytosis process.

In more recent work, 7T MRI brain imaging was used to identify regions of interest, with compositional differences in cell types characterised by snRNAseq focusing

on the demyelinated cortex, periplaque WM, chronic active WM and lesion core regions. Two main proinflammatory phenotypes were identified – 'microglia inflamed in MS' and 'astrocytes inflamed in MS'. Notably, C1q expression was found to be a critical mediator in the former, which was demonstrated through the use of a knockout EAE model and pharmacological intervention blocking C1q[185]. In a different approach which involved the use of scRNAseq of fresh samples as validation for findings from the EAE animal model, Wheeler *et al.* identified an astrocyte cluster which had a supressed expression of transcription factor *NRF2* – which controls cellular responses limiting oxidative stress and inflammation; this role was seemingly compensated by *MAFG* expression - which promotes DNA methylation to repress antioxidant and anti-inflammatory transcriptional programs in cooperation with MAT2alpha.

Whilst these studies have collectively shown how snRNAseq can provide further insights into MS pathology, there is much variation across studies with respect to cell subpopulation identification, annotation, and representation across samples which complicates the interpretation of cell type and pathway relevance to pathology. Moreover, there is much variation with respect to the investigation of whether or not apparent differences in cell abundance and in gene expression are statistically robust, which again hinder the interpretation of findings. To address these caveats in the existing literature, we performed snRNAseq from 6 SPMS patients and 5 controls, employing sample multiplexing to reduce potential batch effects derived from sample processing variation, and recovering 14,821 high quality nuclei. We then integrated our

data with those available from the two other main snRNAseq MS and control studies[130, 131]. We hypothesised that by performing a meta-analysis of our and publicly available snRNAseq, we would be able to better annotate the nuclear clusters, and identify more statistically robust differences in nuclear abundance and gene expression between cases and controls, thereby providing further mechanistic insights into brain health and disease.

4.2 Results

4.2.1 Meta-analysis of snRNAseq MS and control data uncovers high-resolution differences in cell type abundance and gene expression

To perform the meta-analysis of our demultiplexed snRNAseq data and of snRNAseq data from the two further studies, the data were jointly QCed, integrated with batch correction and clustered (as described in Methods section 2.3.2.3). This integration yielded 71,078 high-quality nuclei, from a total of 18 case and 19 control samples. The nuclei were classified into 5 major cell types: OGs, neurons, astrocytes, endothelial cells and immune cells, which included both CNS-resident microglia as well as infiltrating immune cells (**Fig 4.1**). The two most abundant cell types were oligodendrocytes (n=29,605) and neurons (n=27,038). Performing an initial characterisation of the abundance of these broad cell types showed that, as expected, MS cases had a significantly higher immune cell content (p=0.0140) (**Fig 4.2**).

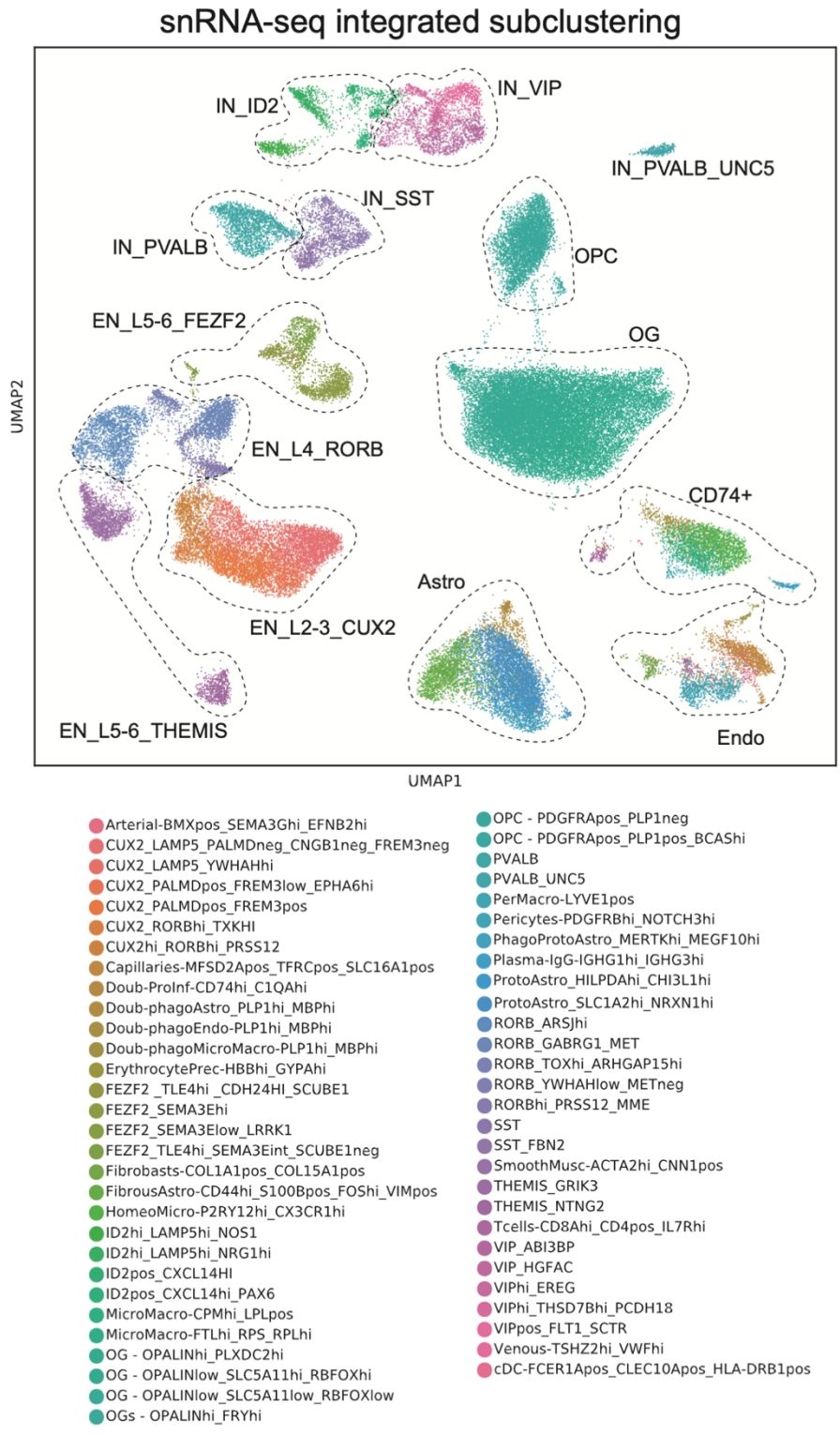


Figure 4.1: A total of 56 clusters were identified in snRNA-seq final cluster annotation of integrated dataset (n = 71,078). Major cell types include oligodendrocytes, astrocytes, endothelial cells, immune cells and neurons, of which 8 subtypes in total were identified. Cluster annotation was performed with a combination of known cell-type markers as well as genes differentially expressed within clusters. Astro – Astrocytes; EN – Excitatory neurons; IN – Inhibitory neurons; Endo – Endothelial cells; OG – Oligodendrocytes; OPC – Oligodendrocyte progenitor cells.

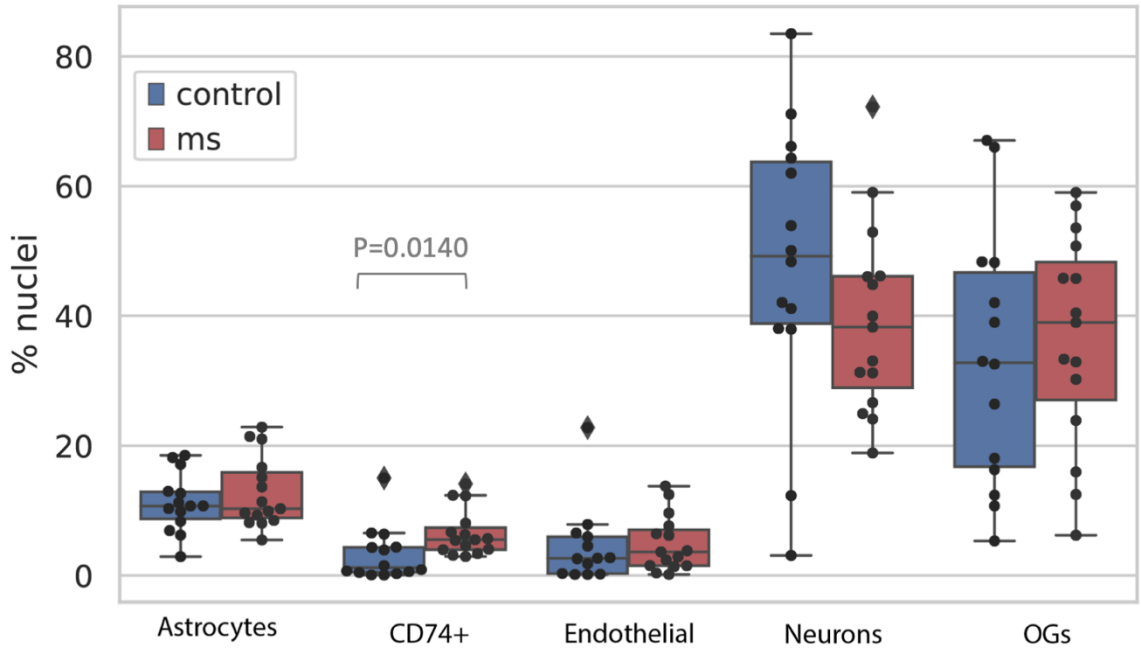


Figure 4.2: Differential abundance comparison for major cell types. CD74+ immune cells were the only cell type statistically significantly differentially abundant (pval = 0.0140; FDR = 0.0698) and was more represented in MS than in controls.

Subsequent analyses, detailed in the sections below, were also performed with higher resolution cell subsets, and included differential abundance (DA) and differential gene expression (DGE) analyses. As cell abundance, DA and DGE can be inter-linked, in order to visualise the relationship between abundance and DA with the number of differentially expressed genes (DEGs) for each high-resolution cluster, a scatter plot was generated (**Fig 4.3**). In general, clusters with a larger number of differentially expressed genes (>20) had a log_2FC <0.5 in DA, whilst clusters with a differential abundance of log_2FC >0.5 had DEGs<10.

The functional relevance of DEGs was further characterised through the use of gene ontology (GO) enrichment analysis to help identify key biological pathways, whilst WGCNA was employed to determine if functional gene expression modules correlating with disease could be identified. Lastly, given that metabolic dysfunction has been reported in MS[186, 187] and was further suggested by our biochemical profiling by FTIR microscpectroscopy (Chapter 3), metabolite flux analysis was also performed. For simplicity, the key results of these analyses are described below for each of the 5 major cell types identified.

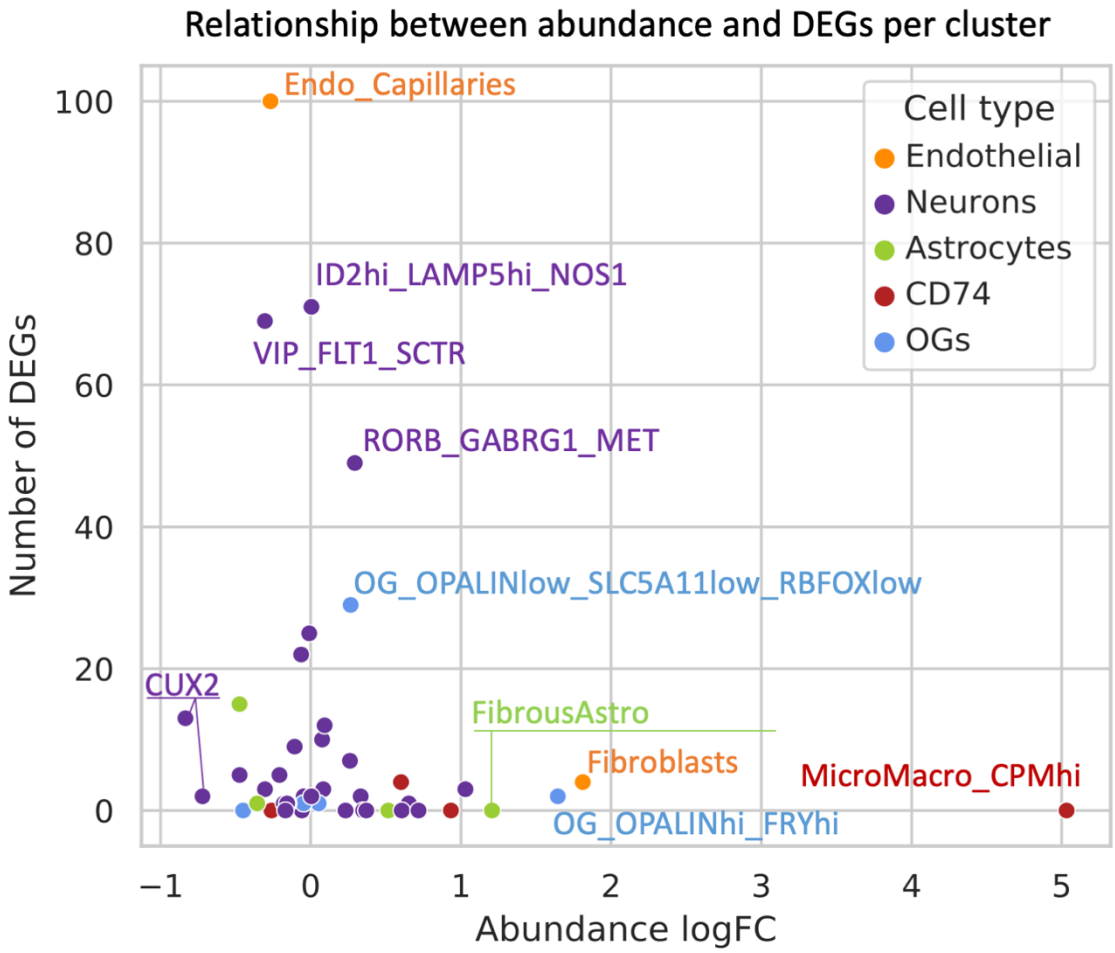


Figure 4.3: Visualisation of the relationship between differential abundance and differential gene expression indicates a non-linear relationship between both cluster features. Clusters characterised by an increase in abundance in MS were MicroMacro_CPMhi_LPLpos (5.03 logFC; DEGs 0), Fibroblasts (1.81 logFC; DEGs 4), OG_OPALINhi_FRYhi (1.64 logFC, DEGs 2), FibrousAstro (1.206 logFC, DEGs 0) and FEZF2_TLE4hi_CDH24hi_SCUBE1 (1.03 logFC, DEGs 3). The two clusters with greater negative log fold change are CUX2_PALMDpos_FREM3pos (-0.83 logFC, DEGs 13) and CUX2_RORBhi_TXKhi (-0.72 logFC, DEGs 2). Clusters with greater differential gene expression had significantly smaller differences in abundance. The clusters with highest DEGs were Capillaries (100 DEGs), ID2hi_LAMP5hi_NOS1, VIP_FLT1_SCTR, RORB_GABRG1_MET and OG_OPALINlow_SLC5A11low_RBFOXlow.

4.2.1.1 Oligodendrocytes

Among OGs, two distinct OPC clusters were identified, PDGFRA^{+} PLP1^{-} and BCAS1^{+} OPCs, and the latter may correspond to a previously reported myelinating OPC subpopulation observed in MS[188] (**Fig 4.4A**).

The main large cluster of OGs was subclustered further based on a gradient of expression of the key marker genes *OPALIN*, *PLXDC2*, *SLC5A11*, *RBFOX* and *FRY*.

The OPALINhi FRYhi subcluster was the only one with statistically significant DA when comparing cases to controls (p=0.0018) and was more abundant in MS cases (**Fig 4.4B**), although there was some trend towards an increase in the OPALINlow_SLC5A11low_RBFOXlow cluster and a decrease OPCs in cases when compared to controls.

DGE analysis indicated OPALINlow SLC5A11low RBFOXlow OG cluster as having the most DEGs when comparing cases and controls (**Fig 4.4C**), but these genes did not fall into known pathways as identified by GO enrichment analysis. A few DEGs of potential functional interest could be discerned across the different OG clusters. The OPALINhi_PLXDC2hi cluster had higher *MIF* levels in cases; this gene is involved in the negative regulation of the p53 pathway[189], which may trigger apoptosis, response to DNA damage and negative regulation of cell cycle arrest, and *MIF* can also be involved in intramolecular oxidoreductase activity and cytokine receptor binding[190]. Cases had a higher *SLC22A17* expression in OPALINlow_SLC5A11hi_RBFOXhi OGs, which is involved in iron ion transport and homeostasis[191].

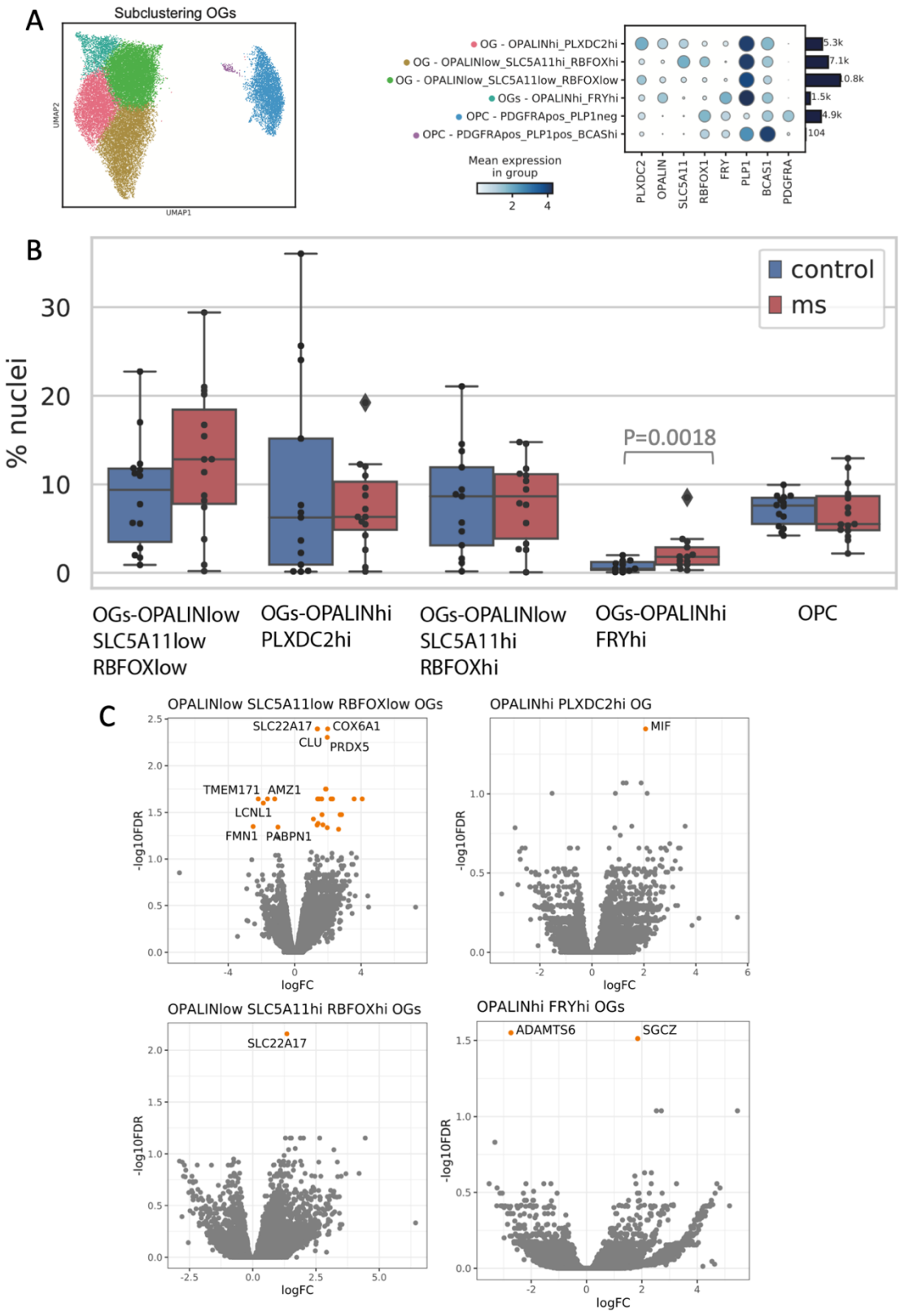


Figure 4.4: Oligodendrocytes subclustering and characterisation. (A) Among OGs three main groupings were identified which corresponded to OPCs (PDGFRA^{+} PLP1^{-} and PDGFRA+ PLP1^{+} BCAShi) and mature OGs. The mature OG cluster was subdivided based on gradient gene expression changes of markers OPALIN, SLC5A11, RBFOX, PLXDC2 and

FRY. (B) OPALINhi FRYhi OGs were more abundant in MS than controls and this difference was statistically significant (pval = 0.0018; FDR = 0.0090) (C) DGE analysis identified OG OPALINlow SLC5A11 low RBFOXlow cluster had the most differentially expressed genes, although no known GO biological process or GO molecular function pathways were associated with these genes in Go enrich analysis (D) OPALINhi PLXDC2hi OGs had upregulated MIF in MS (E) OPALINlow SLC5A11hi RBFOXhi OGs had upregulated SLC22A17 in MS (F) OPALINhi FRYhi OGs had upregulated ADAMTS6 in controls and SGCZ in cases.

WGCNA analysis of the oligodendrocyte compartment identified the pale turquoise module to have a statistically significant correlation with disorder (OR: 0.42; p=0.02) (**Fig 4.5**). This module was characterised by biological process pathways relating to protein folding, protein refolding, chaperone-mediated protein complex assembly, 'de novo' protein folding and mitochondrial electron transport (**Fig 4.6**). Cellular compartment pathways identified included chaperone complex, respiratory chain complex, respirasome, mitochondrial respiratory chain complex I and NADH dehydrogenase complex (**Fig 4.7**). Molecular function pathways included heat shock protein binding, ATPase regulator activity and unfolded protein binding (**Fig 4.8**). KEGG terms which correlated with this module include neurodegenerative diseases such as PD, HD and prion disease (PrD) (**Fig 4.9**).

Metabolite analysis also points towards alteration in protein folding. Analysis of metabolite accumulation/depletion in cases and controls identified 56 metabolites which were statistically significantly different in cases when compared to controls (**Fig 4.10**). Of these, $Glc_3GlcNac_2Man_9Asn_1$ has a role in protein processing in the endoplasmic reticulum according to pathways associated with this metabolite reported in the KEGG database and had one of the log-fold changes (>1).

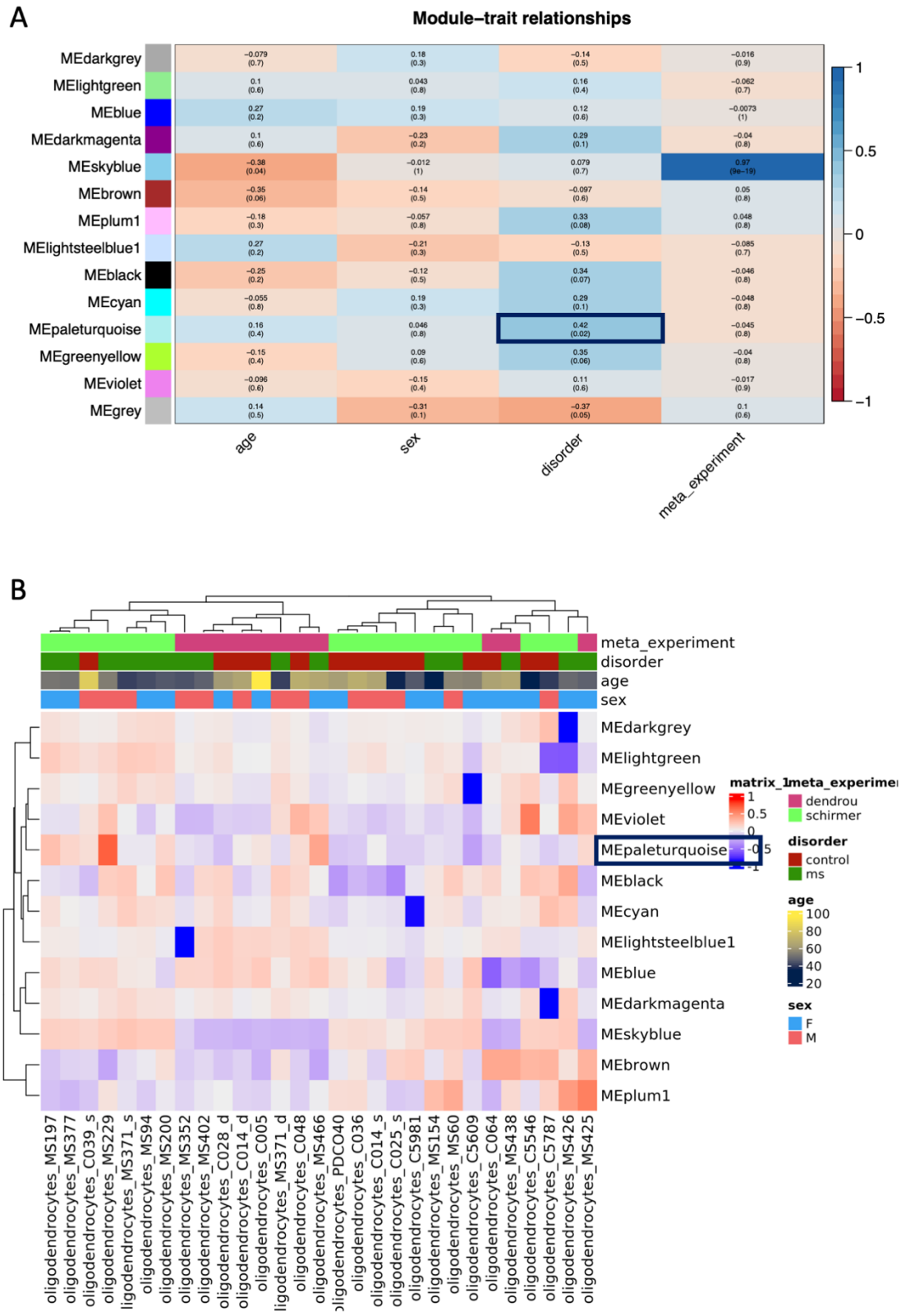


Figure 4.5: WGCNA analysis of oligodendrocytes (A) WGCNA analysis of the oligodendrocyte compartment identified the pale turquoise to be associated with MS (OR: 0.42, padj = 0.02) (B) The MS cases most strongly associated with the disease correlated pale turquoise module are MS MS425, MS466, MS229, MS377 and MS197.

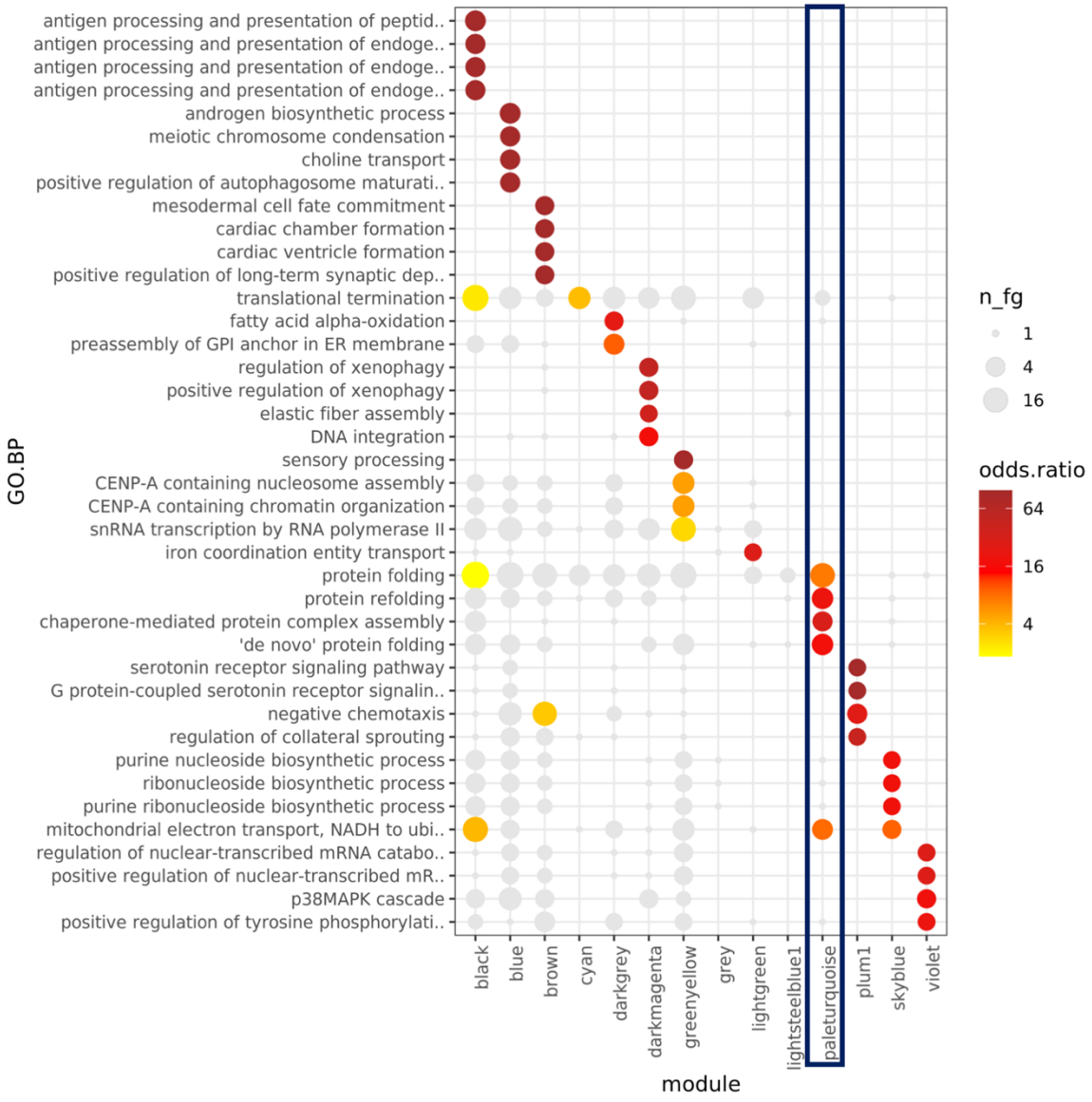


Figure 4.6: GO biological process pathways for oligodendrocytes. The pale turquoise module correlated with pathways relating to protein folding, protein refolding, chaperone mediated protein complex assembly, 'de novo' protein folding, mitochondrial electron transport NADH.

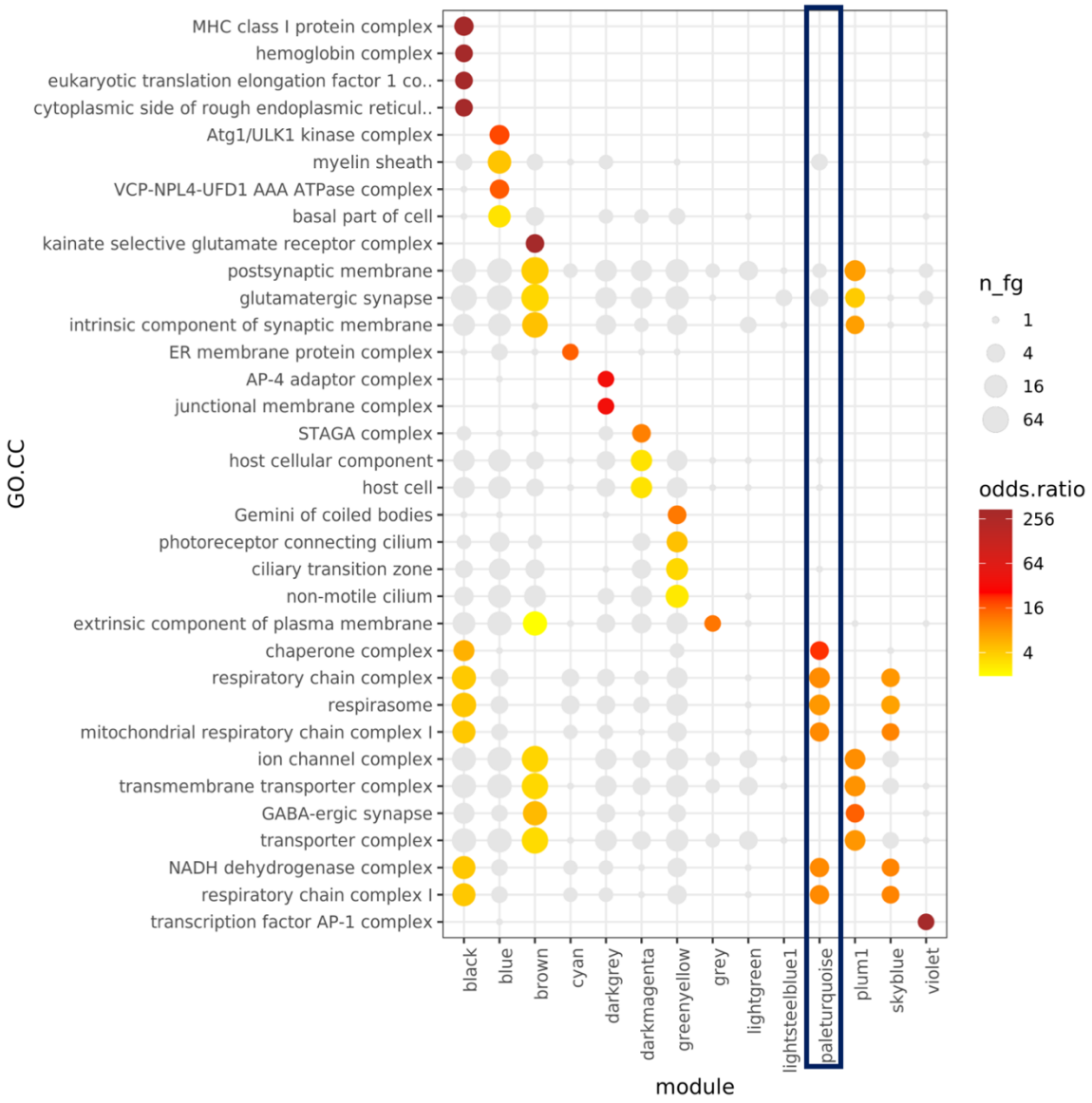


Figure 4.7: Oligodendrocyte GO cellular component pathways. The pale turquoise gene expression module is associated with pathways including the chaperone complex, respiratory chain complex, respirasome, NADH dehydrogenase complex, respiratory chain complex I.

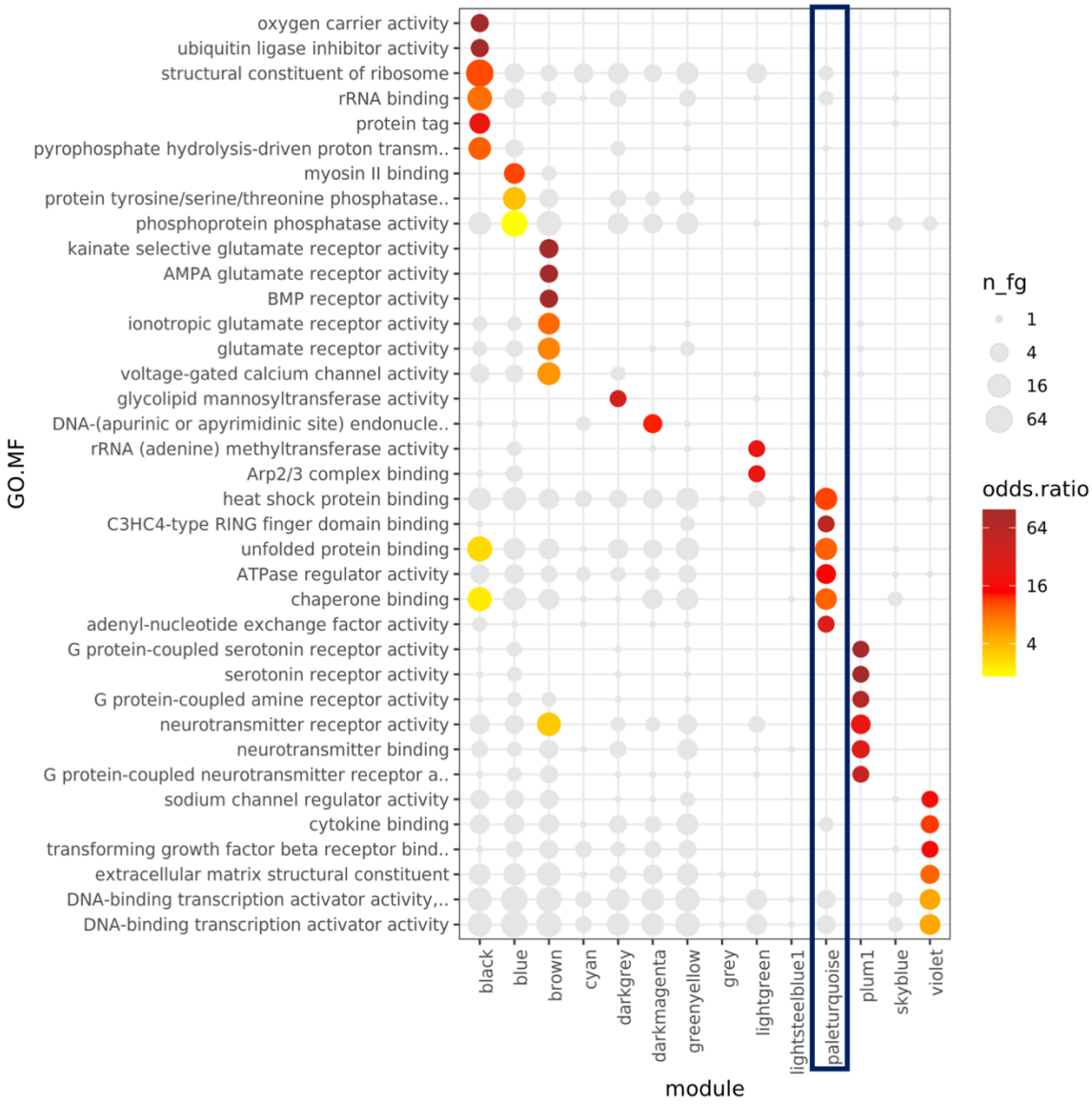


Figure 4.8: Oligodendrocyte GO molecular function pathways. The pale turquoise gene expression module correlated with heat shock protein binding, C_3HC_4 type ring finger domain binding, unfolded protein binding, ATPase regulator activity, chaperone binding and adenyl-nucleotide exchange factor activity pathways.

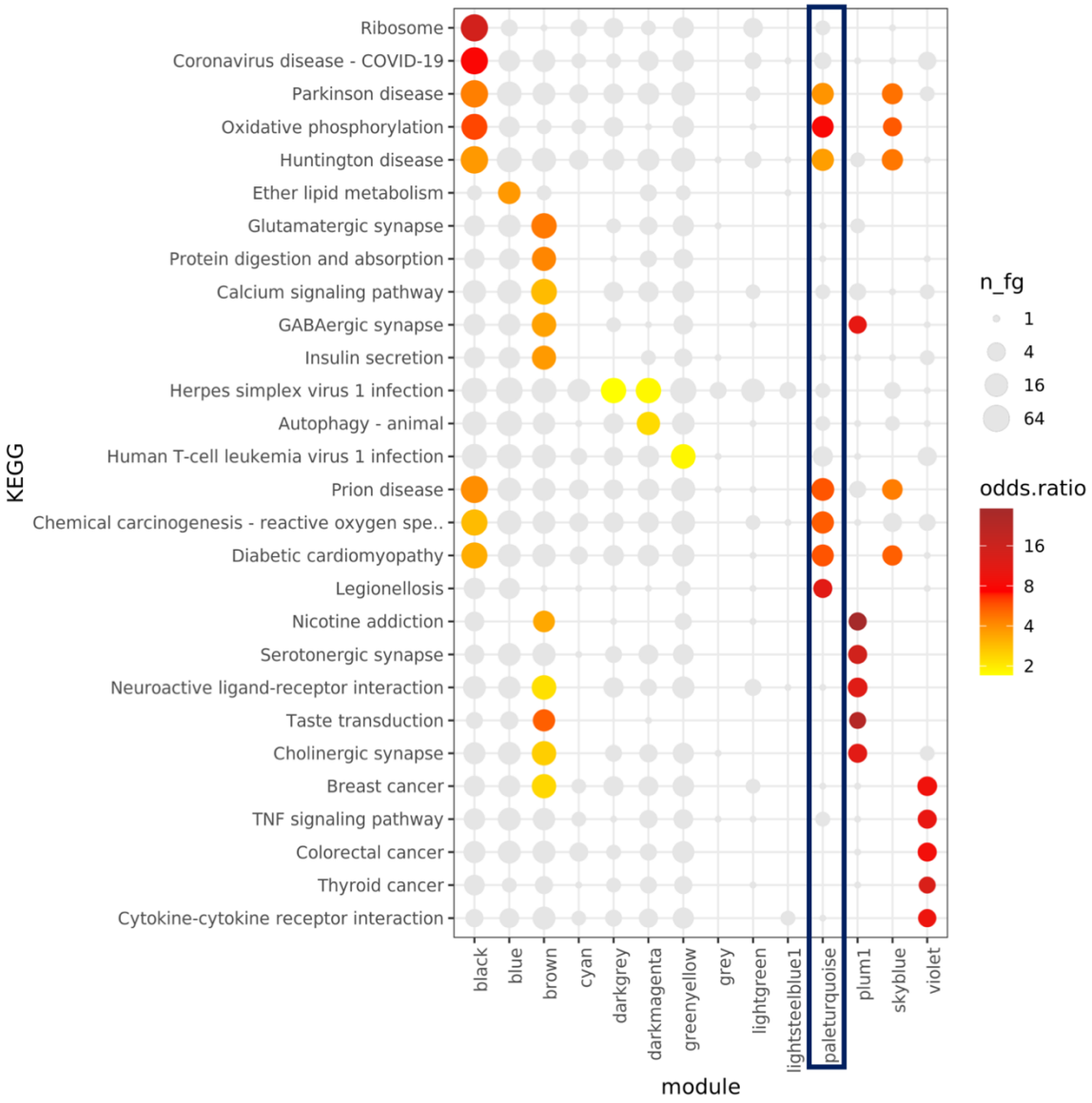


Figure 4.9: Oligodendrocyte KEGG terms. The pale turquoise gene expression module correlated with the Parkinson disease, oxidative phosphorylation, Huntington disease, prion disease, chemical carcinogenesis – reactive oxygen species, diabetic cardiomyopathy and legionellosis pathways.

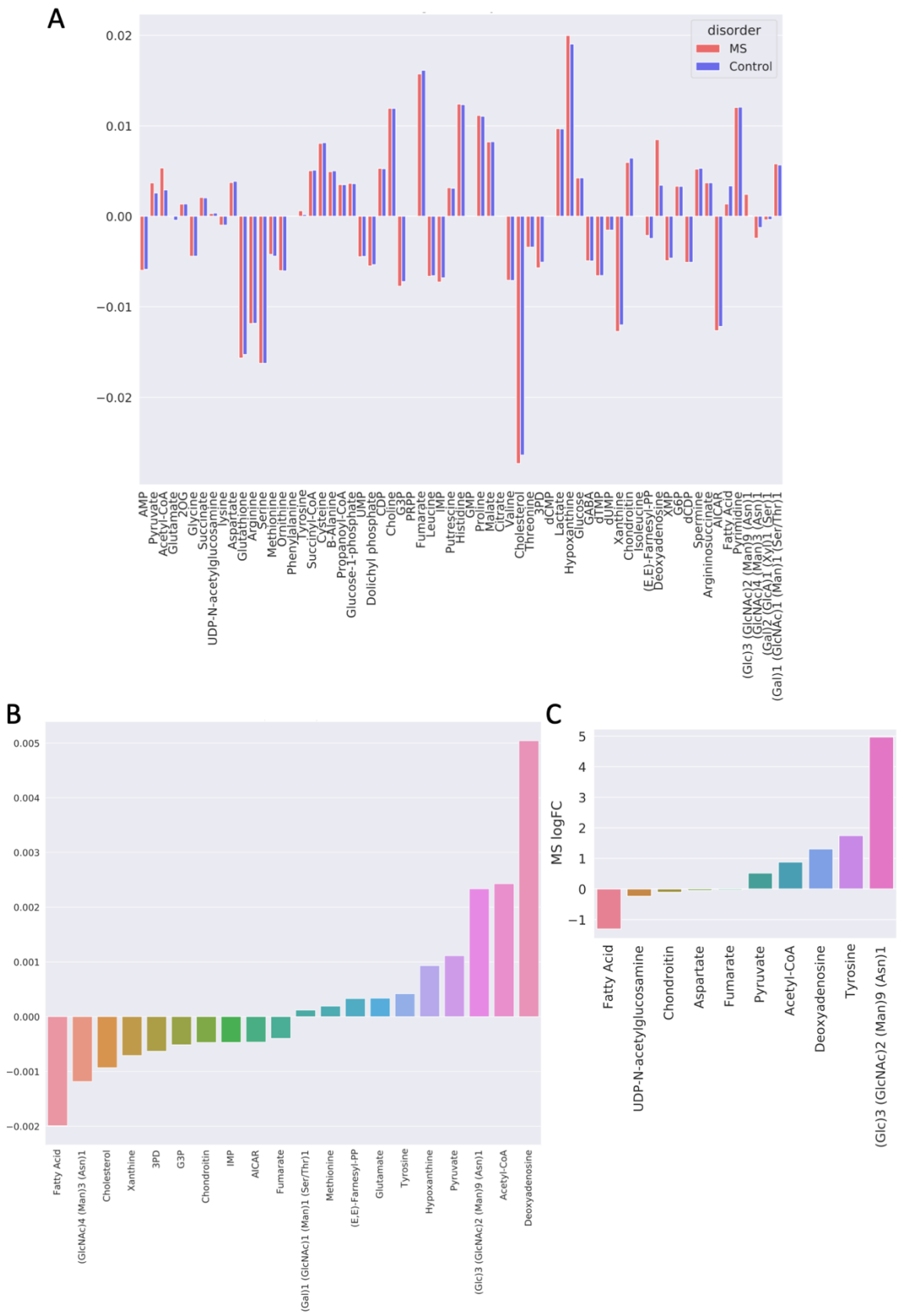


Figure 4.10: Metabolite accumulation and depletion analysis in oligodendrocytes. (A) Comparison of mean metabolite accumulation/depletion in cases and controls (B) Metabolite difference between cases and controls, top 10 most differentially depleted

and accumulated metabolites in oligodendrocytes. The most differentially depleted metabolite when comparing cases and controls is fatty acid (padj non-significant), while the most accumulated metabolites are deoxyadenosine (padj=4.782 x 10^{-67}), acetyl-CoA (padj non-significant) and $Glc_3GlcNac_2Man_9Asn_1$ (padj=0.021). (C) LogFC of accumulation and depletion when comparing cases and controls – fatty acid still is the most depleted with a logFC<-1 and $Glc_3GlcNac_2Man_9Asn_1$ is the most differentially accumulated metabolite with logFC ~5. Tyrosine (padj=4.900 x 10^{-291}) and deoxyadenosine are also accumulated and have a logFC >1 (see **table 4.1** for all statistical comparisons).

4.2.1.2 Astrocytes

Subclustering of the astrocytes revealed three protoplasmic astrocyte subpopulations (**Fig 4.11A**), the first marked by the presence of *HILPDA*, the second ("PhagoProtoAstro") by phagocytosing markers *MERTK*[192] and *MEGF10*[193], and the last by the lack of *HILPDA* and *MERTK* but presence of other protoplasmic astrocyte markers such as GFAP and AQP4. Fibrous astrocytes expressed astrocytic activation markers including *CD44*, *FOS* and *VIM*. The fibrous astrocytes were differentially abundant with higher representation in MS cases (p=0.0108) (**Fig 4.11B**).

TUSC3 was the only DEG in the astrocyte compartment when comparing cases and controls (**Fig 4.11C**). This gene is involved in magnesium ion transport and several glycosylation processes, as identified by GO enrichment analysis. WGCNA analysis of this compartment did not identify any gene expression modules which were significantly correlated with disease (**Fig 4.12**), suggesting that differences in subtype abundance may be the most prominent disease-associated feature in the astrocytic compartment. However, metabolic analyses identified 47 metabolites showing significant differences in depletion/accumulation when comparing cases to controls (**Fig 4.13**). The metabolite with largest absolute variation was cholesterol (padj=8.630×10^{-19}), followed by $Glc_3GlcNac_2Man_9Asn_1$ (padj=2.800×10^{-5}). In the context of fold changes, dCMP had the largest significant log FC (<-20, padj=3.777×10^{-26}).

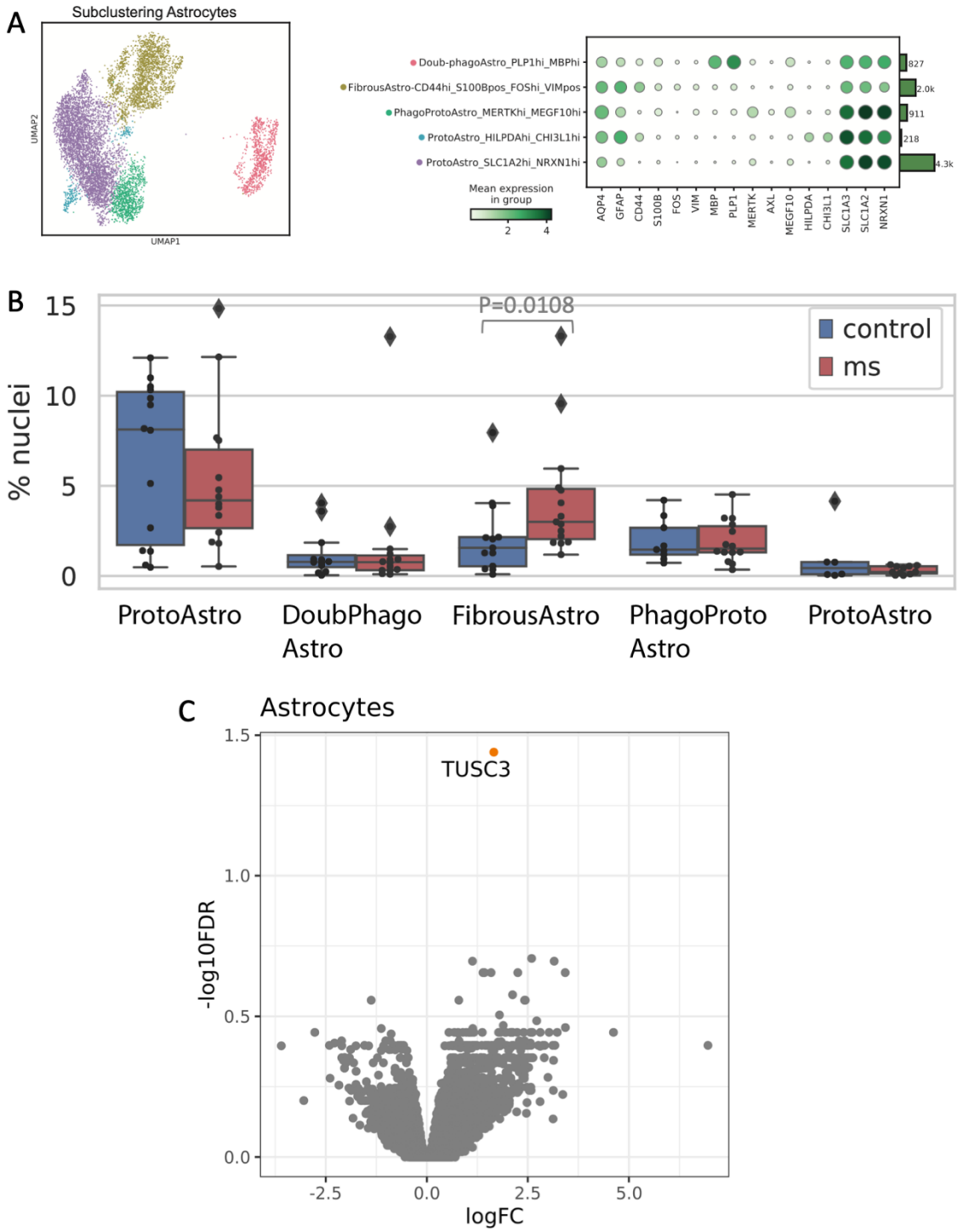


Figure 4.11: Astrocyte subclustering and characterisation (A) Protoplasmic and fibrous astrocytes were identified in astrocyte subclustering. Three subtypes of protoplasmic astrocytes were identified: PhagoProtoAstro (MERTKhi MEGF10hi), ProtoAstro (SLC1A2hi NRXN1hi), ProtoAstro (HILPDAhi CHI3L1hi). An MBPhi subcluster was identified and later determined to constitute a doublet cluster and not phagocytosing astrocytes, and was therefore not included in any subsequent analysis other than differential abundance analysis. (B) Fibrous astrocytes were more abundant in MS (p=0.0108; FDR = 0.0541) (C) DGE analysis of astrocytes identified TUSC3 to be differentially expressed in MS.

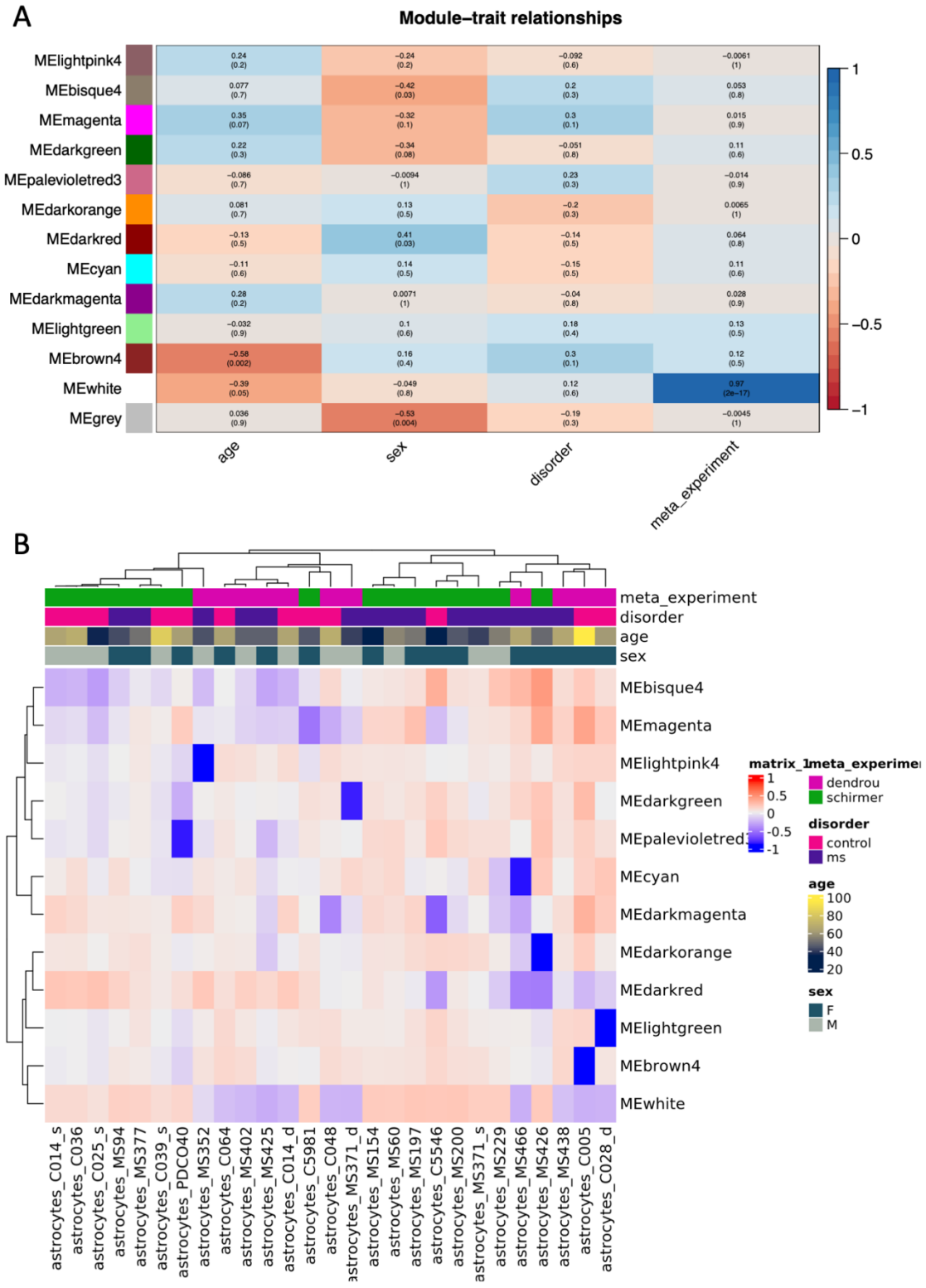


Figure 4.12: WGCNA analysis of astrocytes. (A) No WGCNA modules in the astrocyte compartment have a statistically significant correlation with MS or health. (B) Correlation of WGCNA modules in relation to samples.

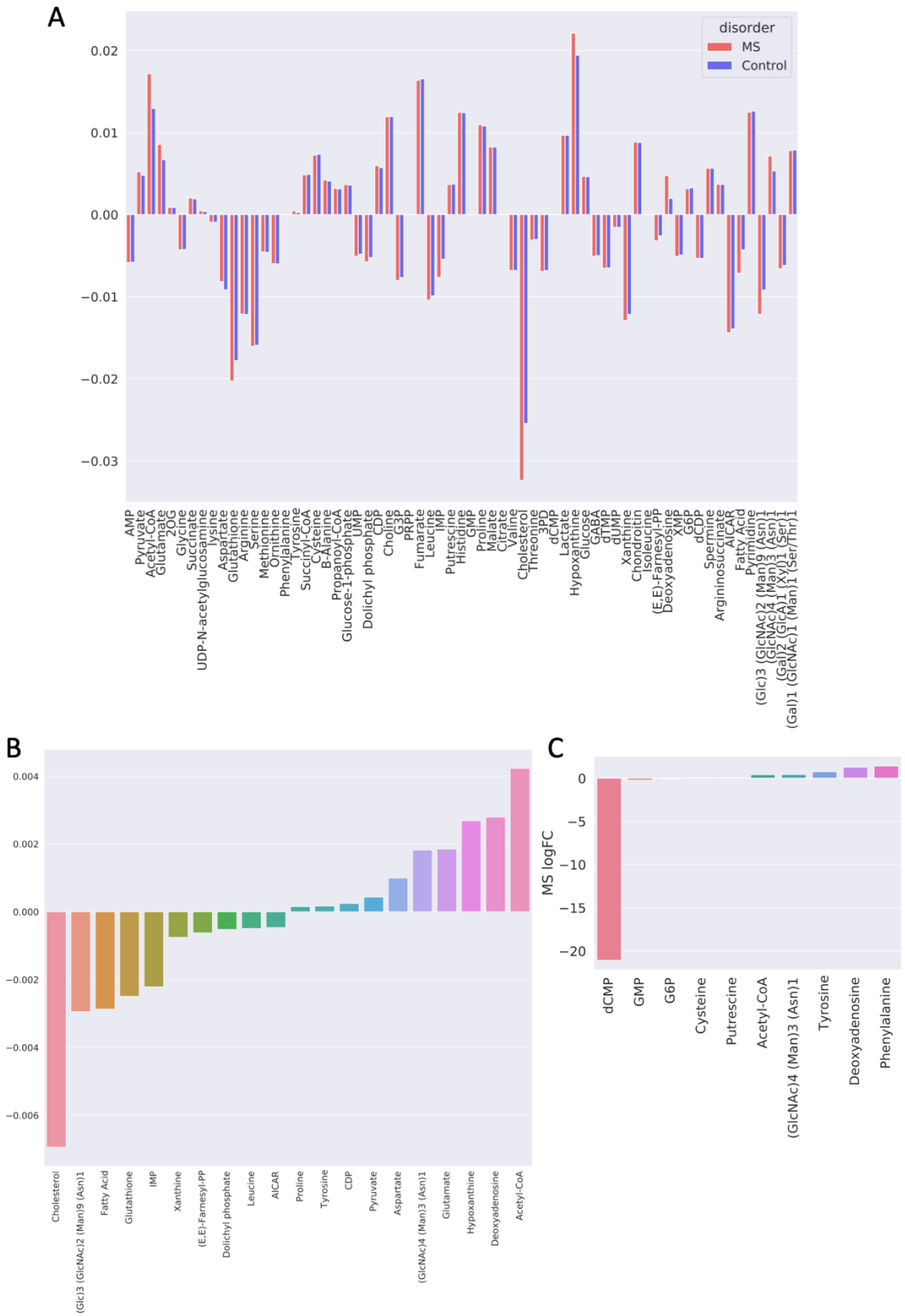


Figure 4.13: Metabolite accumulation and depletion analysis in astrocytes. (A) Comparison of mean metabolite accumulation/depletion in cases and controls (B) Metabolite difference between cases and controls, top 10 most differentially depleted

and accumulated metabolites in astrocytes. The most differentially depleted metabolite is cholesterol (padj=8.630 x 10^{-19}) followed by $Glc_3GlcNac_2Man_9Asn_1$ (padj=2.800 x 10^{-5}) and fatty acid (ns), while the most accumulated metabolites are Acetyl-CoA (ns), deoxyadenosine (padj=4.865 x 10^{-4}) and hypoxanthine (padj=4.693 x 10^{-5}). (C) LogFC of accumulation and depletion when comparing cases and controls – dCMP the most depleted metabolite (padj=3.777 x 10^{-26}), with logFC<-20 (see **table 4.1** for all statistical comparisons).

4.2.1.3 Neurons

Neurons were subclassified based on key neuronal markers as follows: four subtypes of excitatory neurons (EN) were identified using *CUX2*, *RORB*, *FEZF2* and *THEMIS*; four subtypes of inhibitory neurons (IN) were identified using *SST*, *PVALB*, *ID2* and *VIP*. These clusters were further subdivided into a total of 30 clusters, based on existing cluster classification annotation[130, 194] as well as genes expressed differentially across the identified subclusters (**Fig 4.14A**). DA analysis of the main neuronal subtypes did not identify statistically significant compositional differences, although trends could be observed towards reduced CUX2, RORB, VIP and PVALB neurons frequency in MS (**Fig 4.14B**). DA could be observed at a higher resolution: two CUX2 subclusters were significantly less abundant in MS: CUX2 PALMD^{+} FREM3^{+} (p=0.0002) and CUX2 RORBhi TXKhi (p=0.0108) (**Fig 4.14C**), and two other smaller clusters were significantly more abundant in cases when compared to controls FEZF2_TLE4hi_CDH24hi_SCUBE1 (p=0.0122) and SST_FBN2 (p=0.0076).

DGE analysis of the total neurons indicated 5 genes to be upregulated in cases (**Fig 4.15A**). Of these, *CYSLTR2* is involved in icosanoid receptor binding; *RPL9* is a ribosomal gene involved in rRNA binding and *NOS3* relates to oxidoreductase activity, monooxygenase activity, amino acid binding, actin monomer binding, NADP binding and flavin mononucleotide (FMN) binding (**Fig 4.15B-C**).

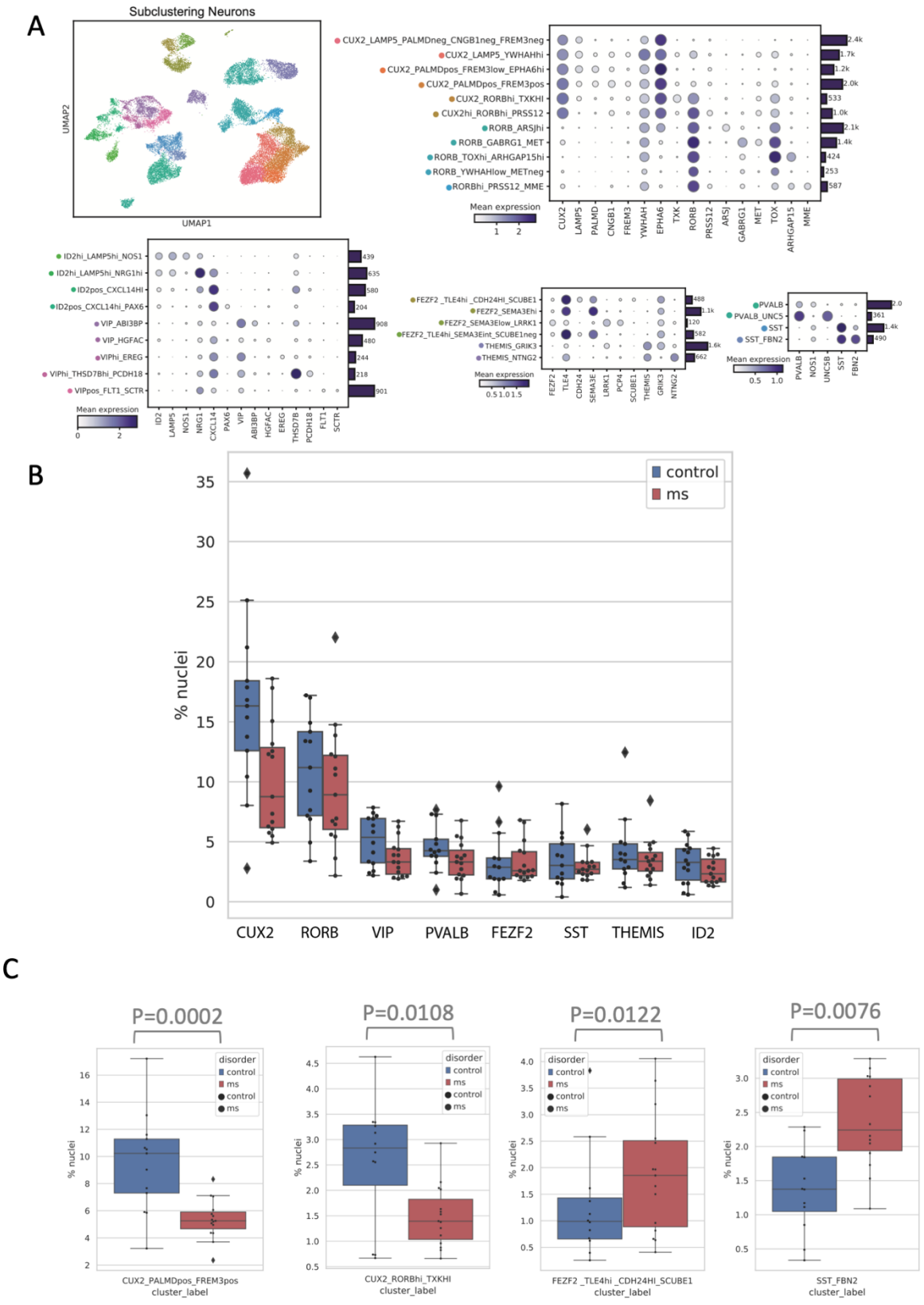


Figure 4.14: Neuronal subclustering and characterisation (A) 8 neuronal subtypes were identified, these were further subclustered into 30 high resolution subclusters. (B) No statistically significant differentially abundant neuronal subtypes were identified. (C) Four high resolution clusters were differentially abundant when comparing cases and controls.

Two were more abundant in controls – CUX2 PALMD+ FREM3+ (p=0.0002) and CUX2 RORBhi TXKhi (p=0.0108) – and two were more abundant in cases - FEZF2 TLE4hi CDH24hi SCUBE1 (p=0.0122) and SST FBN2 (p=0.0076).

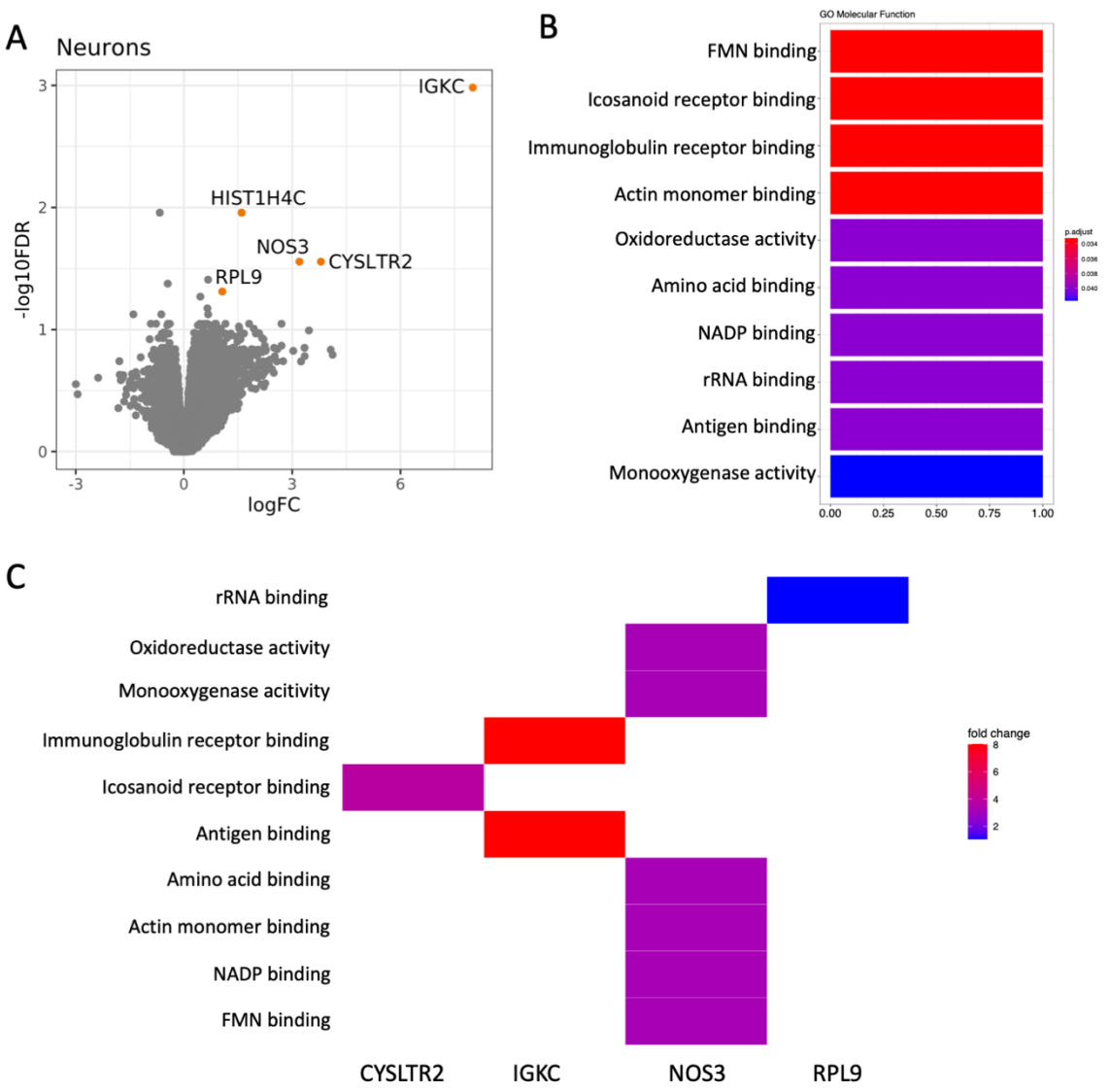


Figure 4.15: DGE analysis of neuronal compartment (A) DGE analysis identified five genes to be upregulated in MS when compared to controls. (B) Go enrich analysis identified 10 GO molecular function pathways significantly associated ($p<0.05$) with the DEGs in MS neurons: FMN binding, icosanoid receptor binding, immunoglobulin receptor binding, actin monomer binding, oxidoreductase activity, amino acid binding, NADP binding, rRNA binding, NADP binding, rRNA binding, antigen binding, monooxygenase activity (C) relationship between GO enrich relevant pathways and upregulated genes.

DGE analysis of the neuronal subtype indicated these changes to be most prominently represented in the CUX2 neurons, which showed overlapping DEGs as well as *WFS* and *PIK3R3* (**Fig 4.16**). *PIK3R3* is related to phosphatidylinositol kinase activity.

WGCNA analysis identified three gene expression modules which positively correlated with disease – turquoise (OR: 0.41; p = 0.03), salmon (OR: 0.56; p = 0.002), orange (OR: 0.52; p=0.006) and one which negatively correlated with disease, grey60 (OR: -0.4; p=0.04) (**Fig 4.17A**). The turquoise gene expression module was highly expressed in a subset of MS patients, mainly MS438, MS229, MS200, MS377, MS371_s, MS402 (**Fig 4.17B**). Enriched GO pathways included several ribosome-related pathways (e.g. ribosomal small subunit assembly, viral translation, U2 snRNP, U4 snRNP, cytosolic large ribosomal subunit, ribosomal subunit, large ribosomal subunit). Other significantly enriched pathways included cotranslational protein targeting to membrane, proteasome core complex beta-subunit complex, endoplasmic reticulum chaperone complex, ubiquitin specific protease binding, and NADH dehydrogenase activity amongst others (**Fig 4.18-20**). KEGG terms associated with this module included oxidative phosphorylation, Parkinson's disease, spliceosome, pathways of neurodegeneration, and Alzheimer disease (**Fig 4.21**), suggesting that this module has a fundamental function in neurodegeneration.

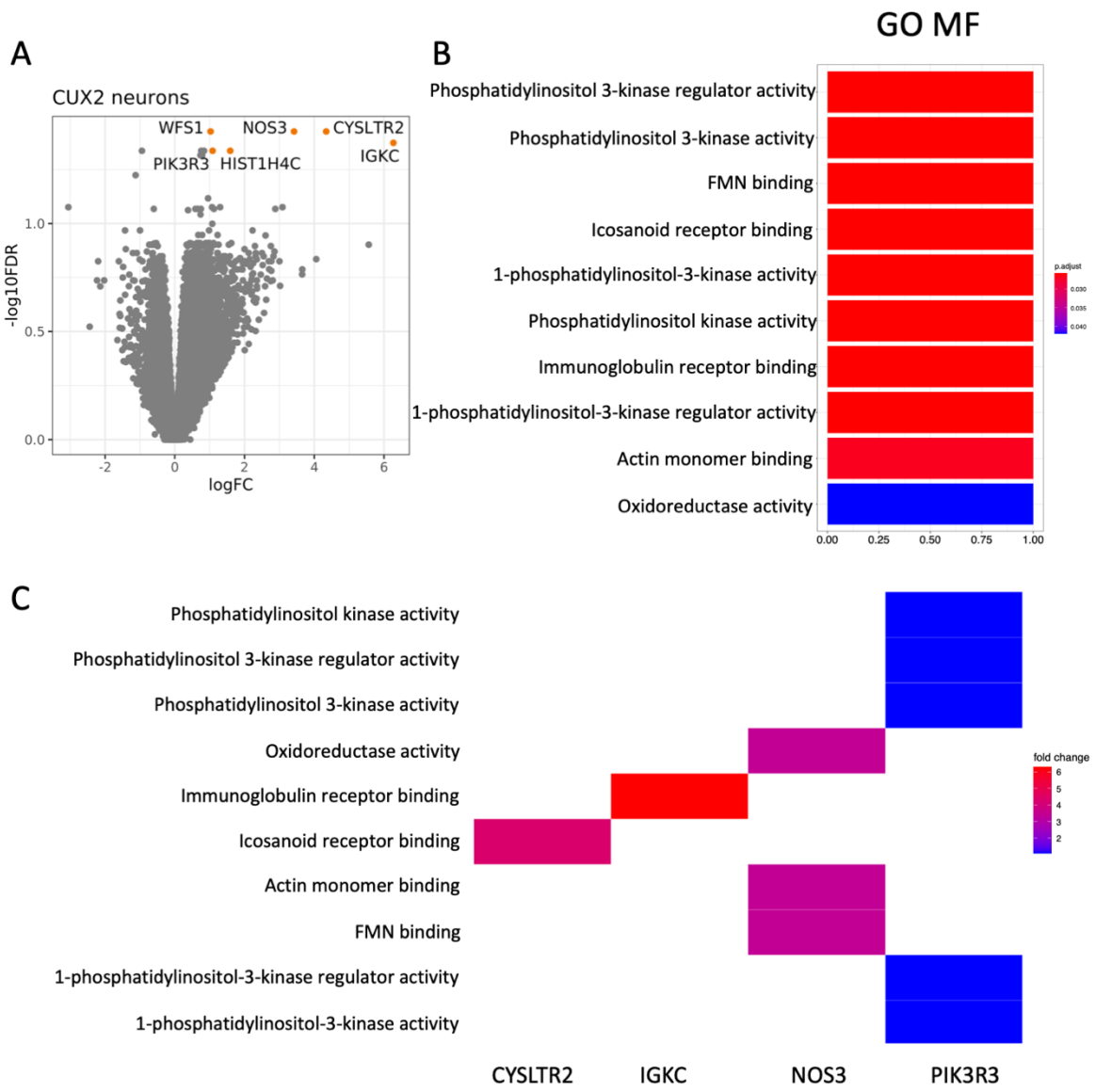


Figure 4.16: Characterisation of CUX2 neuronal DGE (A) Six DEGs were identified in MS when compared to controls (B) Top 10 GO enrich molecular function pathways significantly associated ($p<0.05$) with the DEGs in MS CUX2 neurons included pathways associated with phosphatidylinositol 3-kinase activity, FMN binding, icosanoid receptor binding, immunoglobulin receptor binding, actin monomer binding and oxidoreductase activity (C) Relationship between DEGs and GO enrich identified pathways.

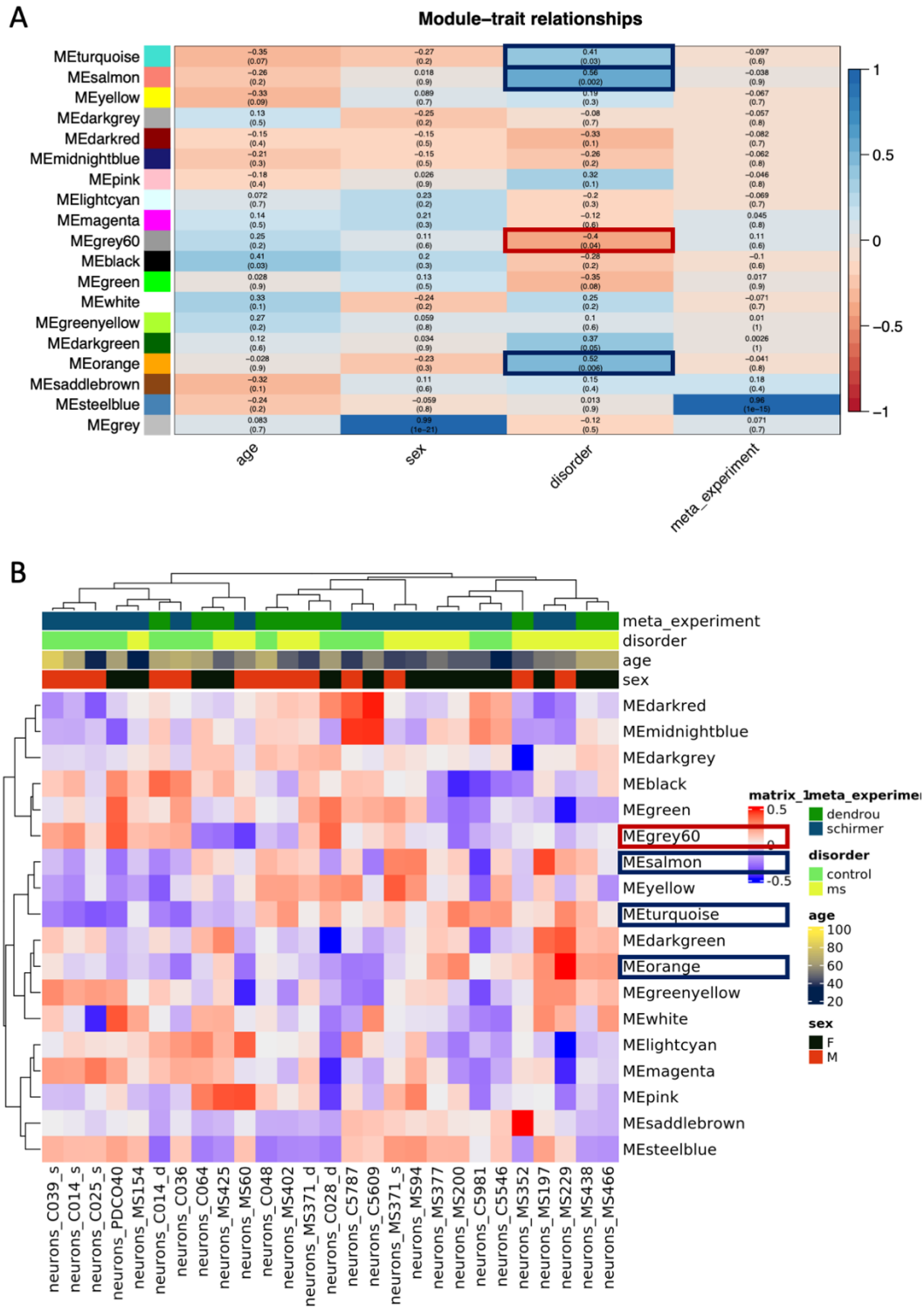


Figure 4.17: WGCNA analysis of neuronal compartment (A) Three gene expression modules were identified in the neuronal compartment which correlated with MS, turquoise(0.41, padj=0.03), salmon (0.56, padj=0.002), orange (0.52, padj=0.006), and one module correlated with health - grey60 (-0.4, padj=0.004) (B) The grey60 module had positive correlation with controls n=9, MS n=2. The salmon module positively correlated with n=4 controls and n=10 MS. The turquoise module positively correlated with n=5 controls and n=9 MS. The orange module positively correlated with n=1 controls and n=8 MS.

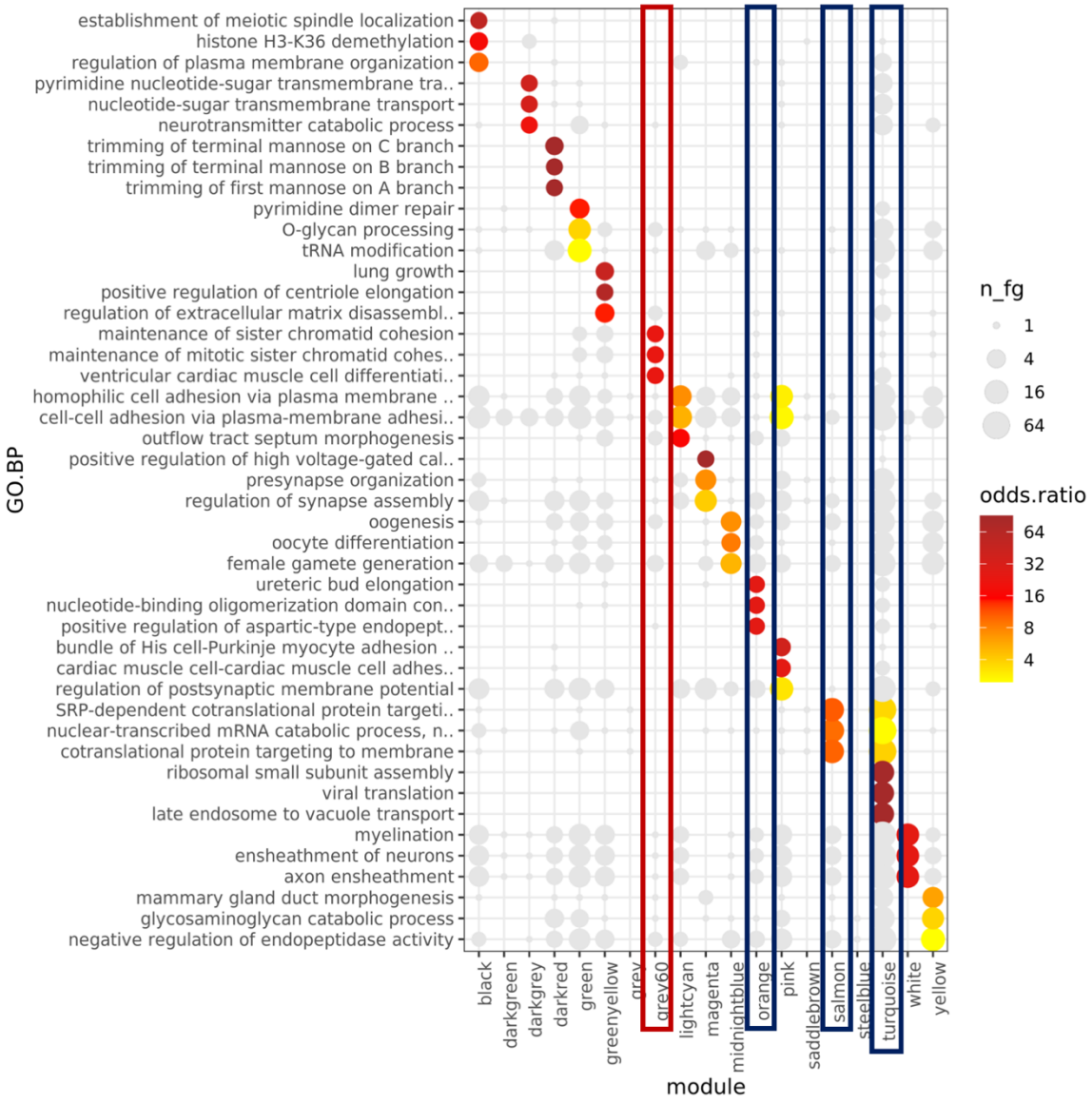


Figure 4.18: GO biological process pathway hits for WGNA neuron gene expression modules significantly correlated with MS or healthy controls. Turquoise module: SRP-dependent cotranslational protein targeting, nuclear transcribed mRNA catabolic process, cotranslational protein targeting to membrane, ribosomal small subunit assembly, viral translation, late endosome to vacuole transport. Salmon module: SRP-dependent cotranslational protein targeting, nuclear transcribed mRNA catabolic process, cotranslational protein targeting to membrane. Orange module: ureteric bud elongation, nucleotide binding oligomerization domain, positive regulation of aspartic type endopeptide. Grey60 module: maintenance of mitotic sister chromatic cohesion, ventricular cardiac muscle cell differentiation.

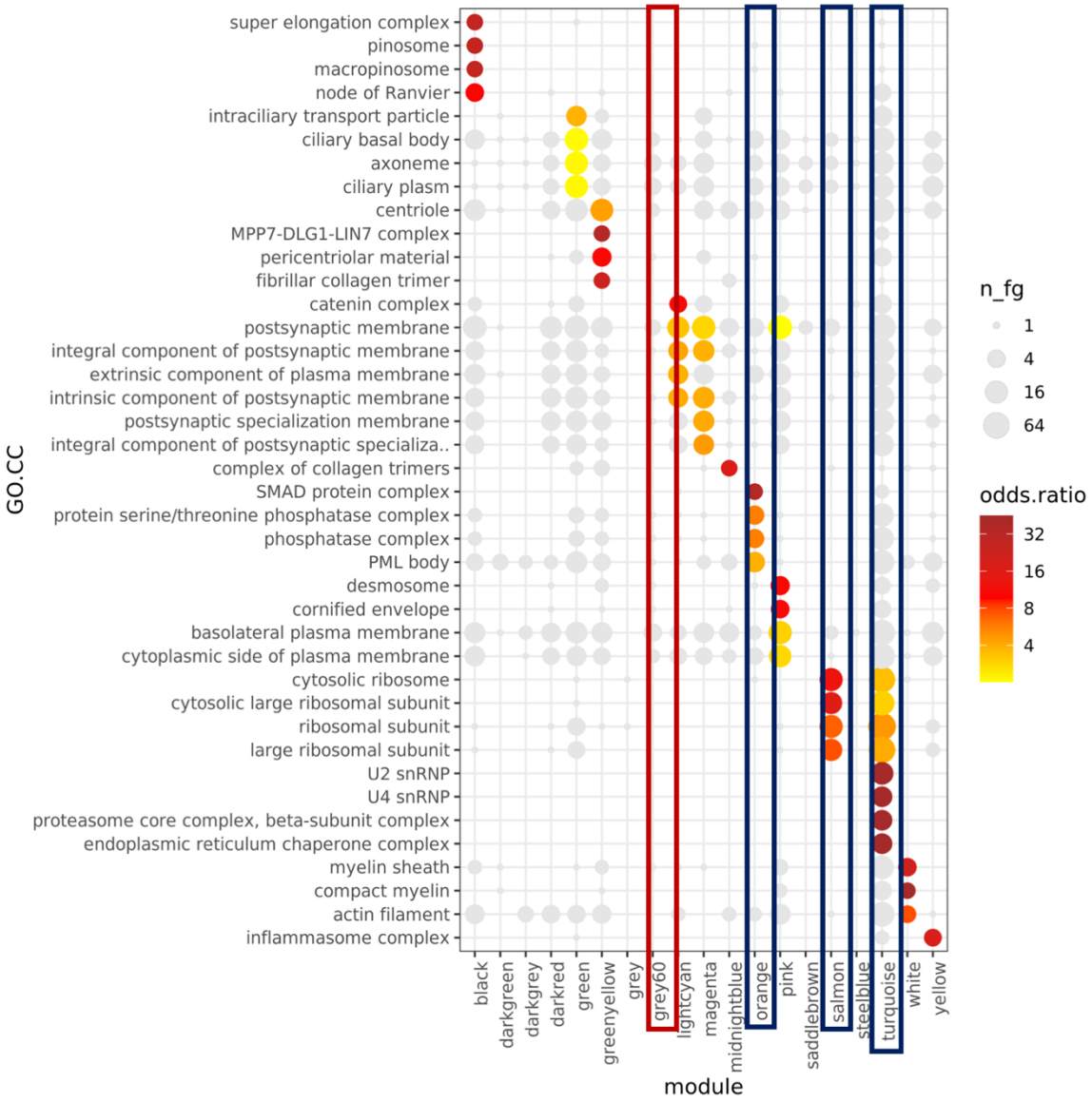


Figure 4.19: GO cellular component pathway hits for WGNA neuron gene expression modules significantly correlated with MS or healthy controls. Turquoise module: cytosolic large ribosomal subunit, U2 snRNP, U4 snRNAP, proteasome core complex, bet-subunit complex, endoplasmic reticulum chaperone complex. Salmon module: cytosolic large ribosomal subunit. Orange module: SMAD protein complex, PML body, protein serine/threonine phosphatase complex, phosphatase complex

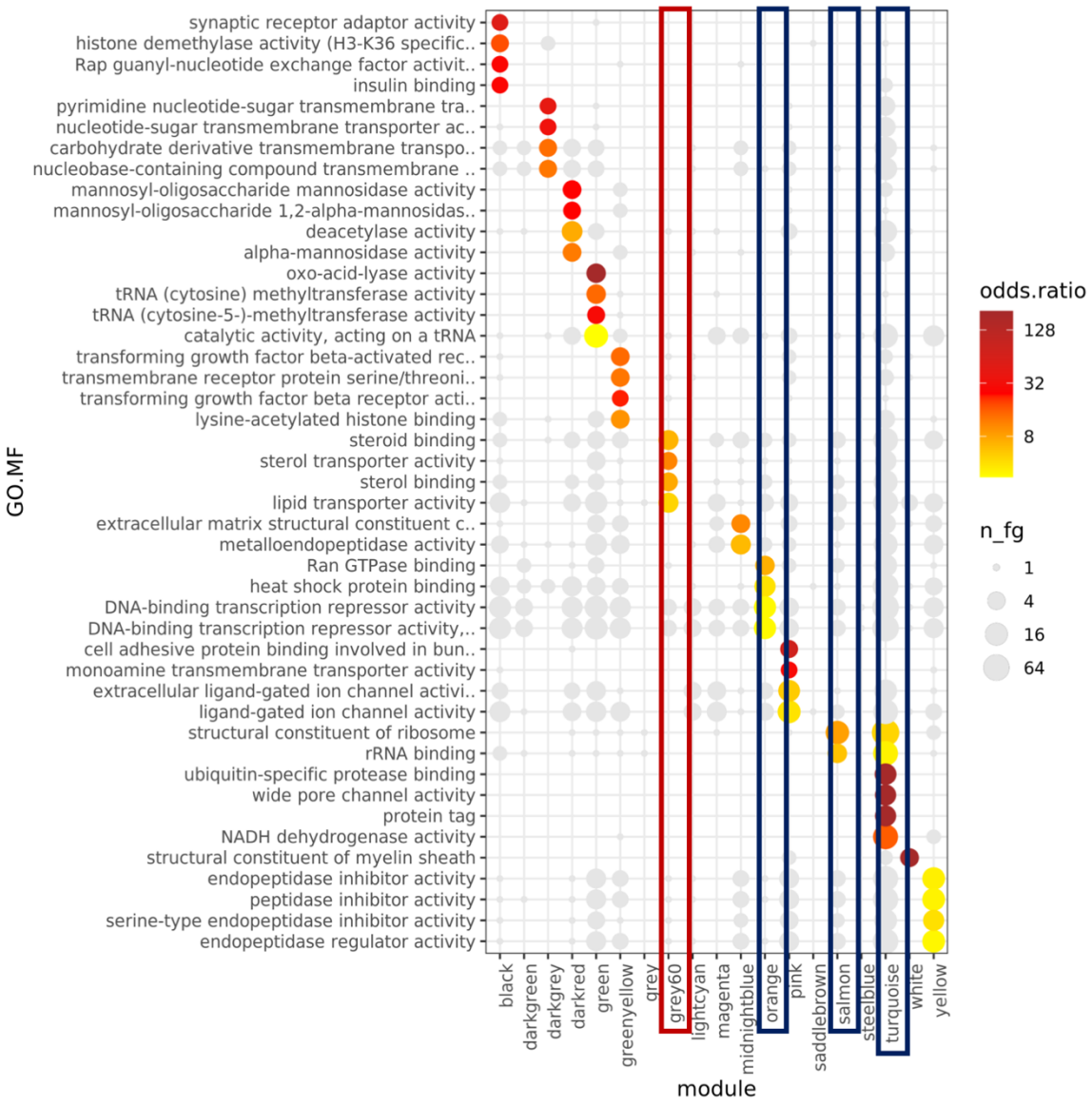


Figure 4.20: GO molecular function pathway hits for WGNA neuron gene expression modules significantly correlated with MS or healthy controls. Turquoise module: NADH dehydrogenase activity, protein tag, wide pore channel activity, ubiquitin-specific protease binding. Salmon module: structural constituent of ribosome, rRNA binding. Orange Ran GTPase binding, heat shock protein binding, DNA-binding transcription repressor. Grey60 module: steroid binding, lipid transporter activity.

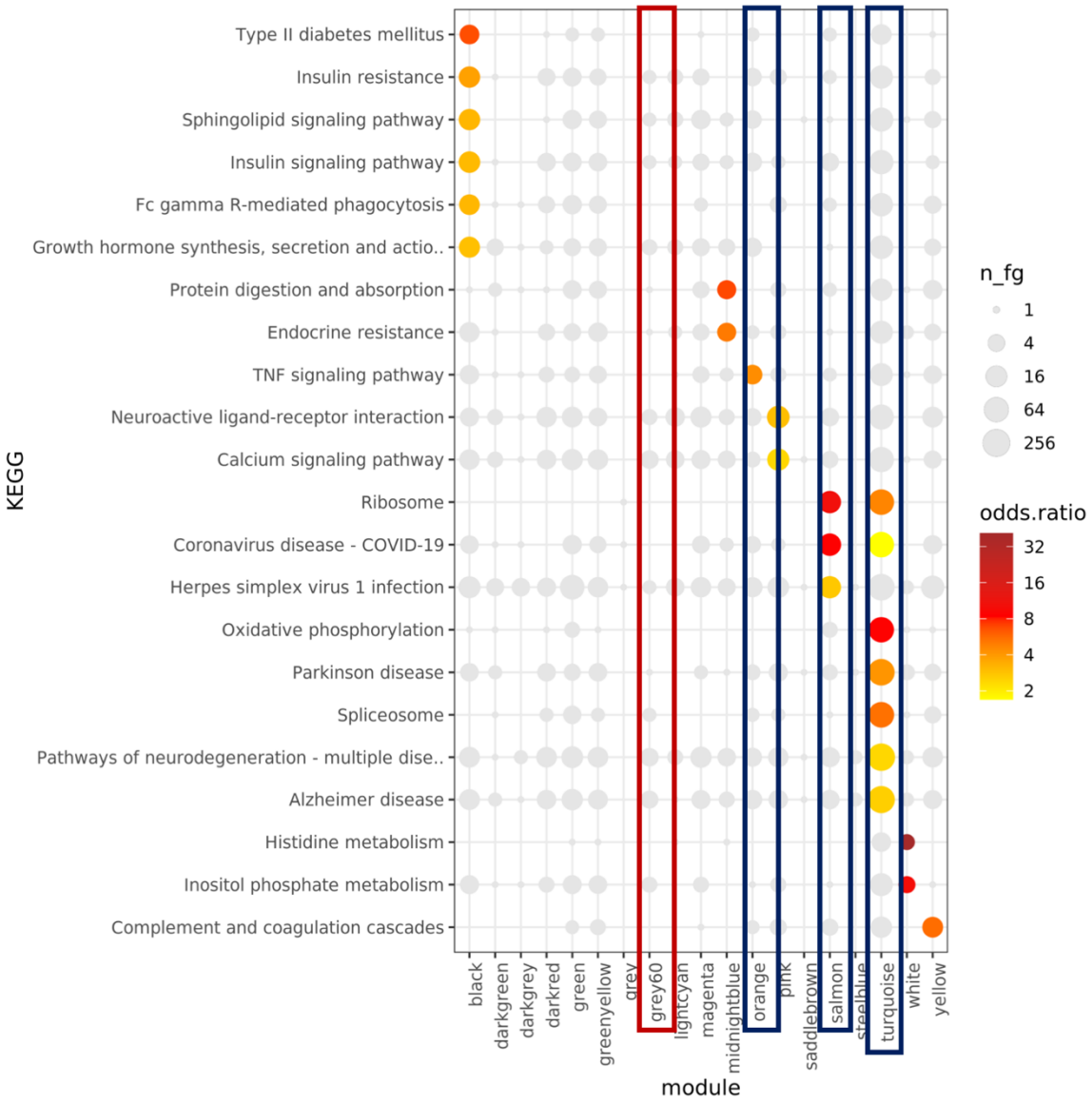


Figure 4.21: KEGG term hits for WGNA neuron gene expression modules significantly correlated with MS or healthy controls. Turquoise module: oxidative phosphorylation, PD, spliceosome, Neurodegeneration, AD. Salmon module: Ribosome, COVID19, Herpes simplex infection.

The salmon module had several similarities in enriched pathways with the turquoise module and was observable in a larger subset of MS cases, including the ones with the turquoise module as well as MS197 and MS94 (**Fig 4.17B**). The overlapping pathway hits between both modules mainly corresponded to ribosomal pathways and pathways associated with viral infections (**Fig 4.18-20**). KEGG terms included ribosome, COVID19, herpes simplex virus 1 infection (**Fig 4.21**). Intriguingly, this module was not associated with KEGG terms for Parkinson's or Alzheimer disease.

The orange gene expression module correlated with MS466, MS438, MS229, MS197, MS352, MS200, MS377 and MS425 (**Fig 4.17B**). This module was associated with GO pathways related to development and growth factor expression and KEGG term TNF signalling (**Fig 4.18-21**). Other terms related to oxidative stress such as phosphatase complex, heat shock protein binding, and DNA binding transcription repressor activity.

The grey60 module correlated mostly with controls, including C039_s, C014_s, C025_2, PDCO40, C014_d, C036, C028_d, C5609 (**Fig 4.17B**). Of the GO terms it was associated with, pathways relating to steroid and lipids were prominently enriched including steroid binding, sterol transporter activity, sterol binding, and lipid transporter activity (**Fig 4.20**).

Lastly, metabolite accumulation/depletion estimation identified 53 metabolites differentially present in MS neurons compared to controls (**Fig 4.22**).

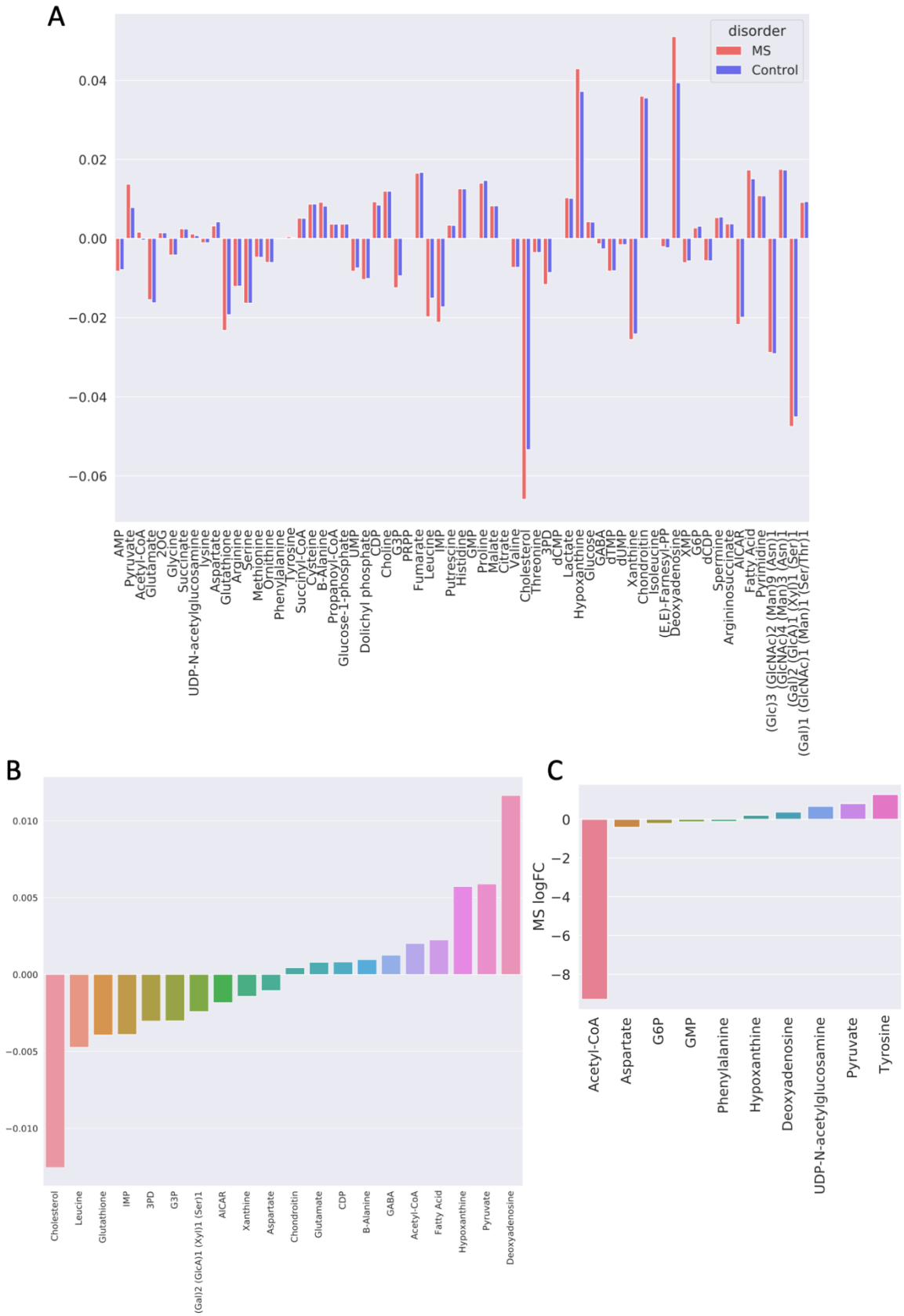


Figure 4.22: Metabolite accumulation and depletion analysis in neurons. (A) Comparison of mean metabolite accumulation/depletion in cases and controls (B) Metabolite difference between cases and controls, top 10 most differentially depleted

and accumulated metabolites in neurons. The most differentially depleted metabolite is cholesterol (ns) followed by leucine (padj=4.528 x 10^{-9}) and glutathione(padj=1.168 x 10^{-66}), while the most accumulated metabolites are Deoxyadenosine (ns), pyruvate (padj=1.640 x 10^{-84}) and hypoxanthine (padj=2.429 x 10^{-5}) (C) LogFC of accumulation and depletion when comparing cases and controls – acetyl-CoA (ns) has a logFC<-8 and Tyrosine was the most depleted metabolite (logFC>1; padj=3.764 x 10^{-39}; see **table 4.1** for all statistical comparisons).

Of these, the most depleted metabolites were leucine (padj=4.528 x 10^{-9}) and glutathione (padj=1.168 x 10^{-66}), and the most significantly accumulated ones were pyruvate (padj=1.640 x 10^{-84}) and hypoxanthine (padj=2.429 x 10^{-5}).

Notably, pyruvate has a significant role in energetics and has been reported to be altered in MS[195] – glycolysis produces pyruvate, which is in turn reduced to lactate in the cytoplasm. The difference in lactate in neurons from cases and controls is only slightly elevated in MS (logFC<1), although this difference is statistically significant (5.736 x 10^{-44}). This small difference may relate to the fact that lactate is secreted into the extracellular space when in the presence of low oxygen availability or when pyruvate formation exceeds pyruvate oxidation, such as in a state of low mitochondrial oxidative capacity[196]. Moreover, ethyl-pyruvate has been reported to have a suppressive effect on secretion of pro-inflammatory cytokines by microglia, macrophages and astrocytes, and may thus have a neuroprotective role in the context of neuroinflammation[197]. Tyrosine had the largest significant logFC >1 (padj=3.764 x 10^{-39}).

4.2.1.4 Immune cells

Subclustering of the CD74+ immune nuclei divided them into five main clusters: microglia/macrophages, plasma cells, T cells, dendritic cells and a *PLP1*+ cluster with a microglia/macrophage phenotype (Fig 4.23A).

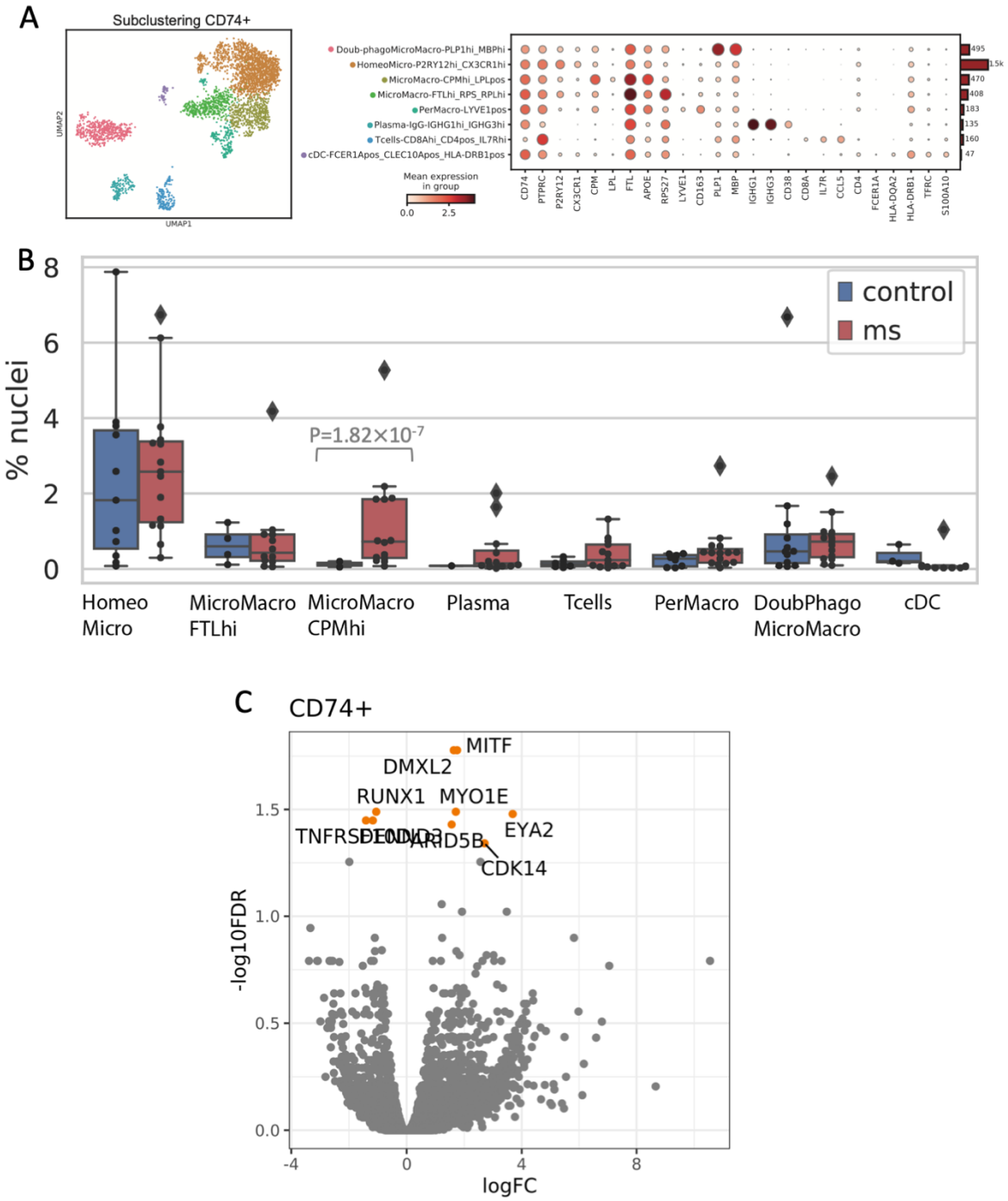


Figure 4.23: Immune cell compartment subclustering and characterisation. (A) Main immune cell types identified include Plasma cells (IGHG1hi IGHG3hi), Tcells (CD8hi CD4^{+} IL7Rhi), dendritic cells (FCER1A^{+} CLEC10A^{+} HLA-DRB1^{+}), Microglia and macrophages which were subdivided into homeostatic microglia (P2RY12hi CX3CR1hi), peripheral macrophages (LYVE1^{+}), microglia/macro (CPMhi LPL^{+} and FTLhi RPS^{+} RPLhi). Similarly to astrocyte compartment, a PLP1hi MBPhi was identified which was clustered due to previous reports of phagocytosing macrophages transporting MBP transcripts to the nucleus. (B) The MicroMacro CPMhi cluster was found to be differentially abundant when comparing cases and controls – it was significantly more abundant in cases ($p=1.82x10^{-7}$; FDR = $9.09x10^{-7}$). Due to both the high doublet score of the DoubPhagoMicroMacro

population and the lack of differential abundance in cases and controls, this cluster was excluded from all subsequent analysis. (C) DGE analysis identified three genes upregulated in controls and 6 genes upregulated in cases, but these genes were not significantly associated with any GO enrich terms.

Within the microglia/macrophages, there was further subclustering which distinguished peripheral macrophages (LYVE1+), homeostatic microglia (P2RY12hi CXCR1hi), and cells annotated as MicroMacro_CPMhi LPLhi and MicroMacro_FTLhi_RPS_RPLhi. Among the immune cell subtypes, likely due to the low overall abundance of infiltrating cells such as T and plasma cell, only the MicroMacro-CPMhi_LPLpos had significantly different abundance ($p=1.82x10^{-7}$), being more abundant in MS (**Fig 4.23B**).

DGE analysis identified *MITF*, *DMXL2*, *MYO1E, EYA2*, *ARID5B*, and *CDK14* to be upregulated in cases and *RUNX1*, *TNFRSF10D*, and *DENND3* in controls (**Fig 4.23C**). GO enrichment analysis did not identify any pathways relating to these DEGs. WGCNA analysis identified one gene expression module which was positively correlated with disease, light cyan (OR: 0.49; p=0.02) and one which was negatively correlated, turquoise (OR: -0.45; p=0.03) (**Fig 4.24A**). The light cyan module, however, was not strongly associated with any biological pathways, but the turquoise module was identified to relate to positive regulation of GTPase activity, specific granule membrane, SH3 domain binding, chemokine signalling pathway and regulation of actin cytoskeleton (**Fig 4.25-28**).

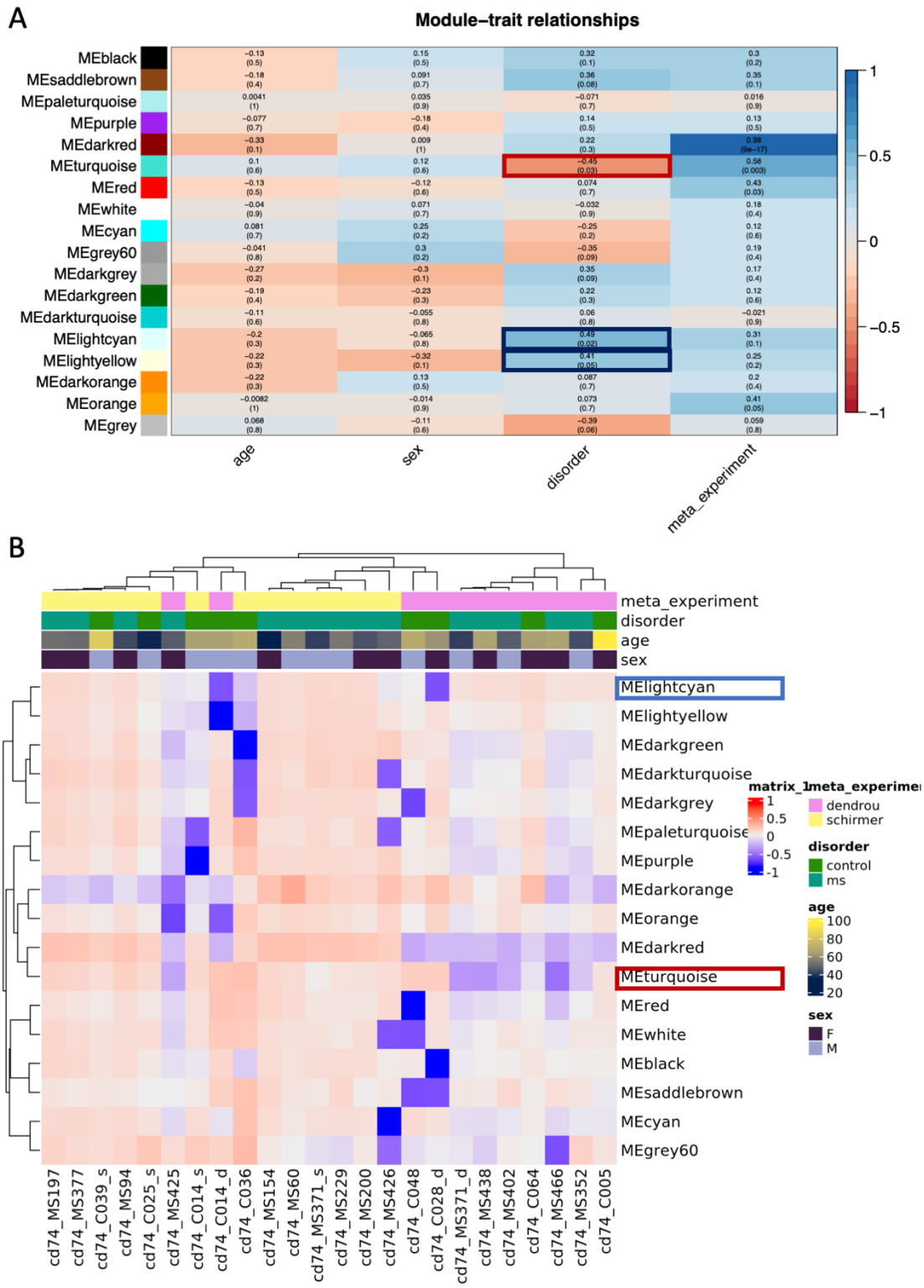


Figure 4.24: WGCNA analysis of the immune cell compartment (A) Three gene expression modules were identified which had statistically significant associations with MS, light cyan (0.49, padj = 0.02) and light yellow (0.41, padj = 0.05). The turquoise module was positively correlated with health (-0.43, padj = 0.03) (B) The light cyan module was positively correlated with n=4 controls and n=13 MS. The light yellow module was positively correlated with n=4 controls and n=12 MS. The turquoise module was positively correlated with n=8 controls and n=8 MS.

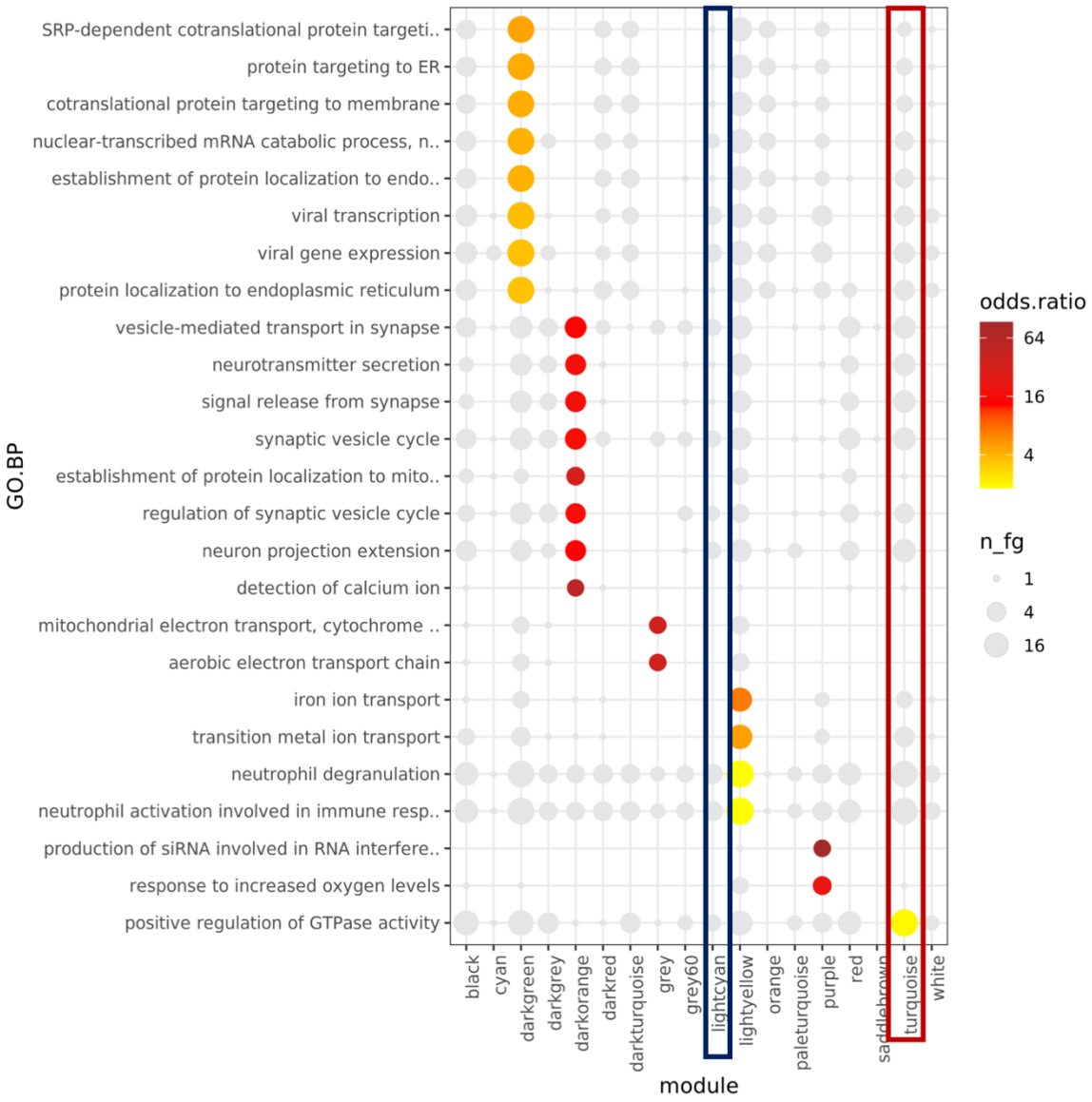


Figure 4.25: GO biological process pathway correlation with WGCNA gene expression modules of immune cell compartment. The turquoise module was marked by positive regulation of GTPase activity.

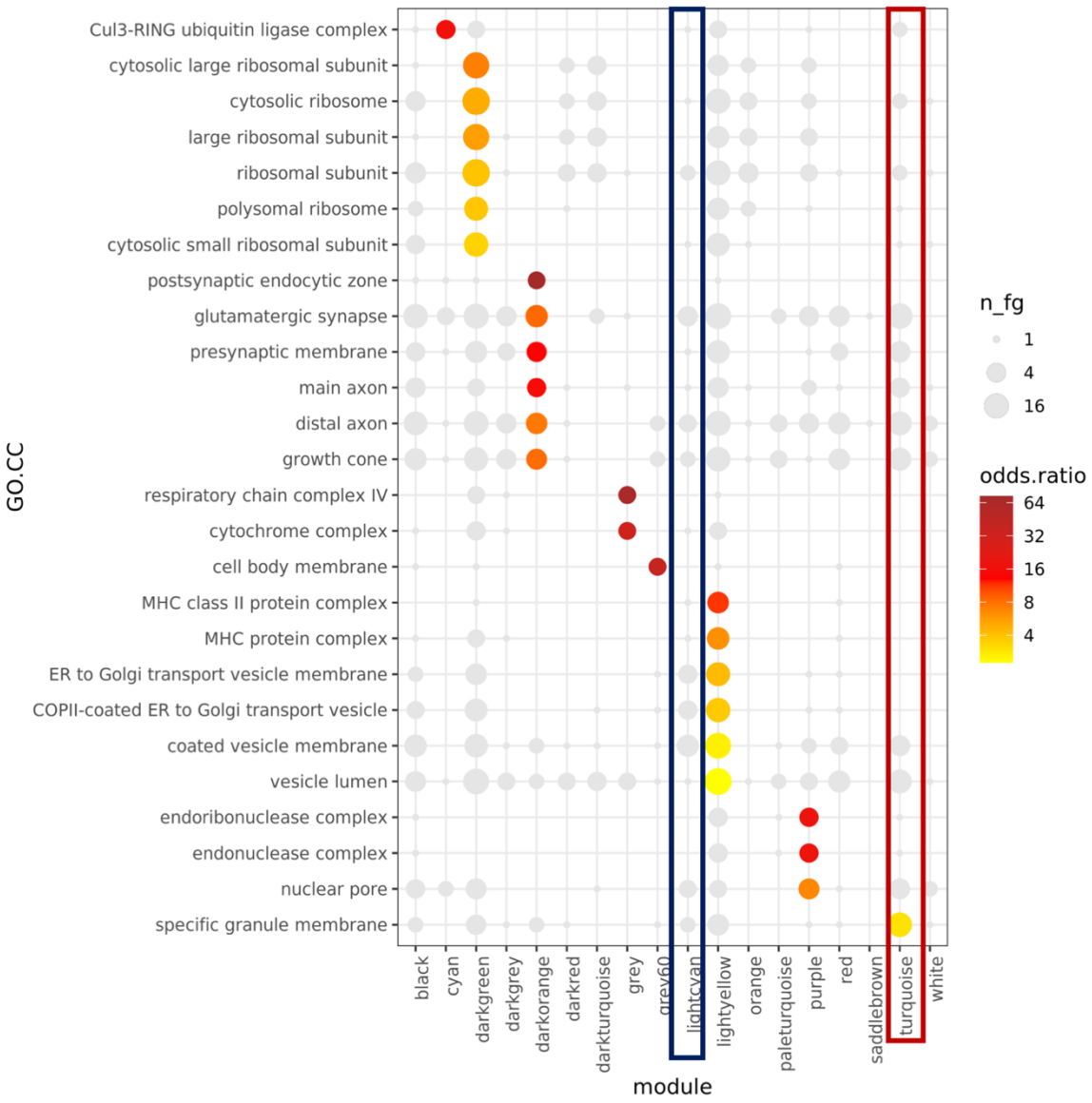


Figure 4.26: GO cellular component pathway correlation with WGCNA gene expression modules. The turquoise module was marked by specific granule membrane.

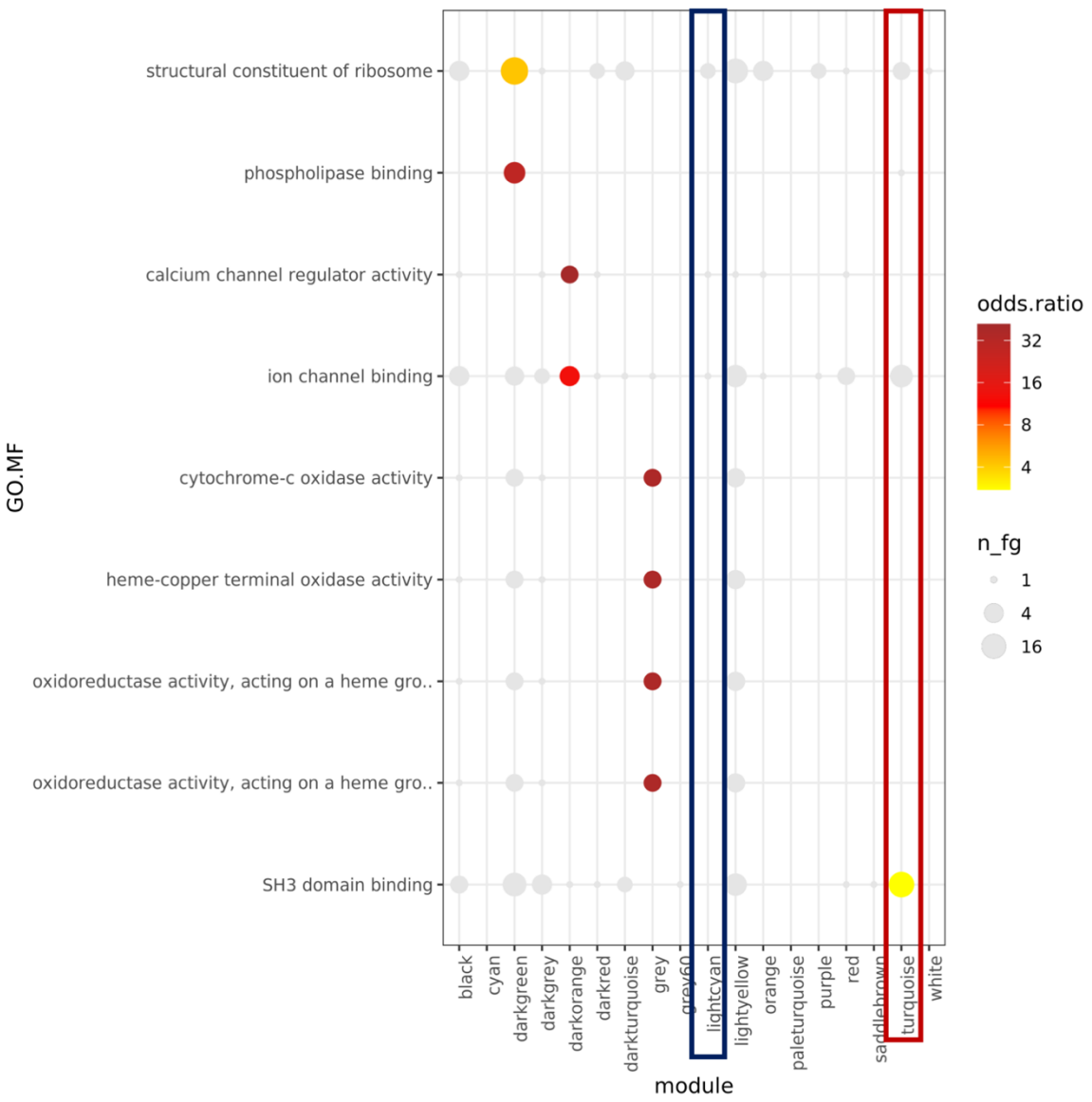


Figure 4.27: GO molecular function pathway correlation with WGCNA gene expression modules. The turquoise module was marked by SH3 domain binding.

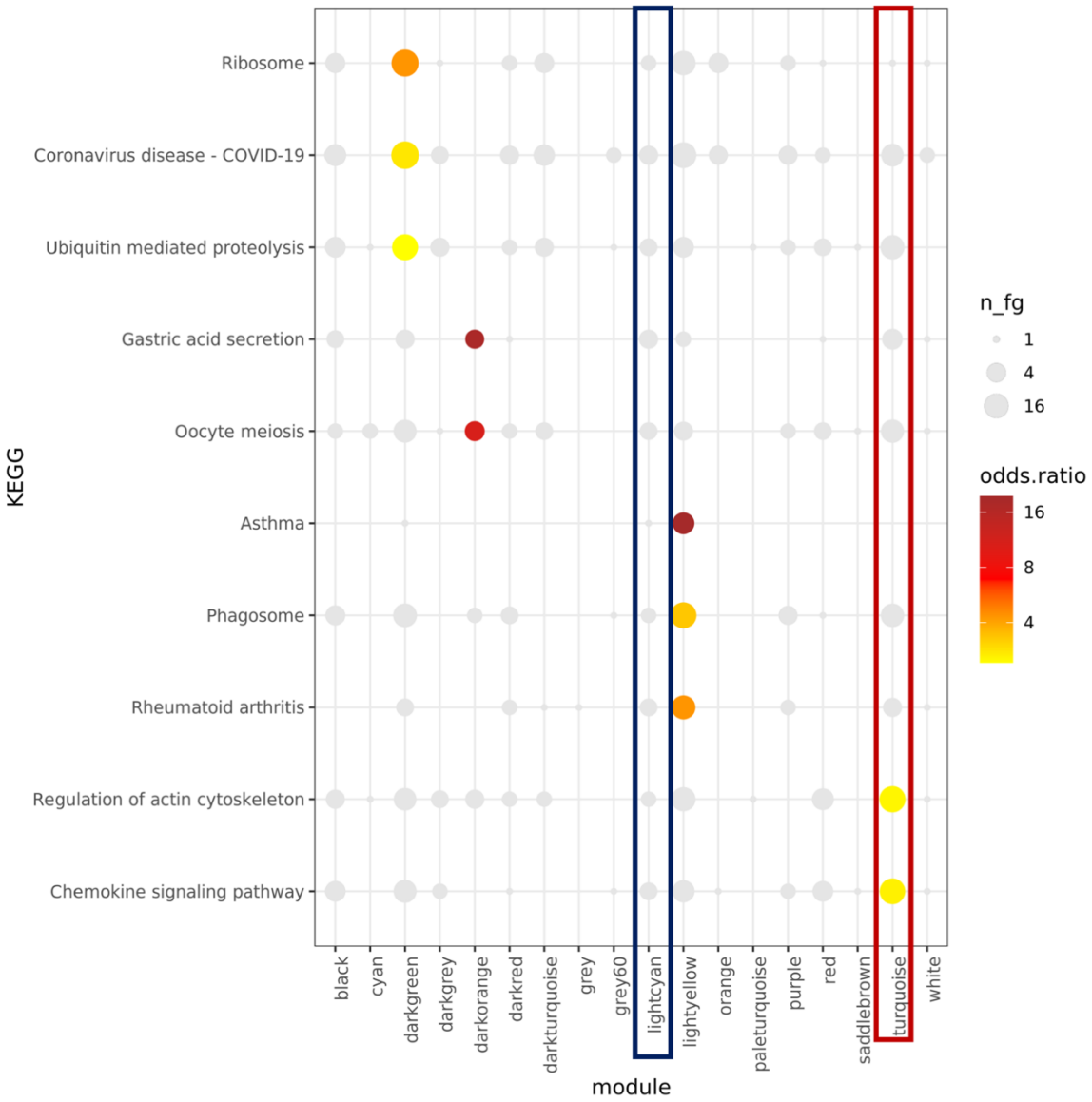


Figure 4.28: The KEGG term correlation with WGCNA gene expression modules. The turquoise was module associated with regulation of actin cytoskeleton and chemokine signalling pathway.

41 metabolites were estimated to be differentially accumulated/depleted in MS when compared to controls (**Fig 4.29**). Many of the largest absolute changes to accumulation/depletion were not statistically significant – deoxyadenosine was the most accumulated metabolite to be significant (padj=9.067 x 10^{-05}). Deoxyadenosine levels are associated with purine metabolism[152], and metabolite levels of all metabolites associated with this process have been reported to be elevated in MS patients – this has been suggested to be associated with deficits in lipophilic antioxidants[198]. Tyrosine, although smaller in absolute differences, again had the largest significant logFC compared to controls (logFC ~2, padj=0.010). Tyrosine, identified as accumulated in neurons, immune cells and OGs is involved in the leucine + valine + isoleucine metabolic supermodule[152]. Leucine and isoleucine have been reported to be significantly higher in MS and have a role in aminoacid catabolism and is implicated in immunoregulatory mechanisms associated with autoimmune disorders, including MS[199].

4.2.1.5 Endothelial cells

Within the endothelial cell subset arterial, venous, smooth muscle and capillary cell types were identified based on known vascular markers[200] (**Fig 4.30A**). A high HBB cluster was identified and classified as an erythrocyte progenitor cell; interestingly, this cluster was specific to a single donor whose cause of death was a stroke. Fibroblasts (COL1A1+ COL15A1+) and pericytes (PDGFRBhi NOTCH3hi) were also represented in this compartment.

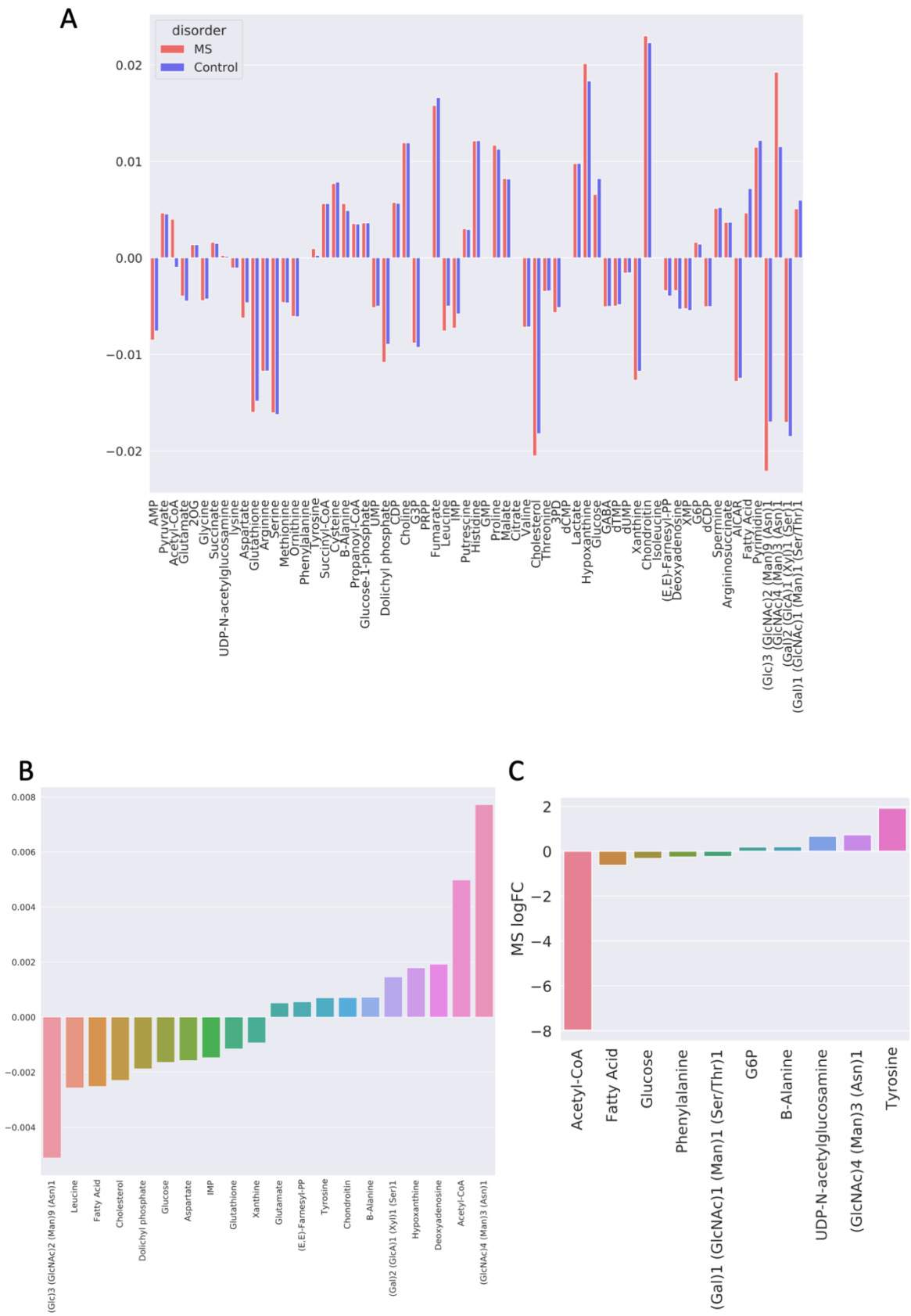


Figure 4.29: Metabolite accumulation and depletion analysis in immune cells. (A) Comparison of mean metabolite accumulation/depletion in cases and controls (B) Metabolite difference between cases and controls, top 10 most differentially depleted

and accumulated metabolites in immune cells. The most differentially depleted metabolite is $Glc_3GlcNac_2Man_9Asn_1$ (ns), while the most accumulated metabolites are $GlcNac_4Man_3Asn_1$ (ns), acetyl-CoA (ns) and deoxyadenosine (padj=9.067 x 10^{-05}) (C) Log FC of accumulation and depletion when comparing cases and controls – acetyl-CoA has a logFC~8 and is accumulated in cases and depleted in controls, due to the negative value of the control and positive value of the MS mean the ratio was converted to be negative. Tyrosine is accumulated (logFC ~2; padj=0.010, see **table 4.1** for all statistical comparisons).

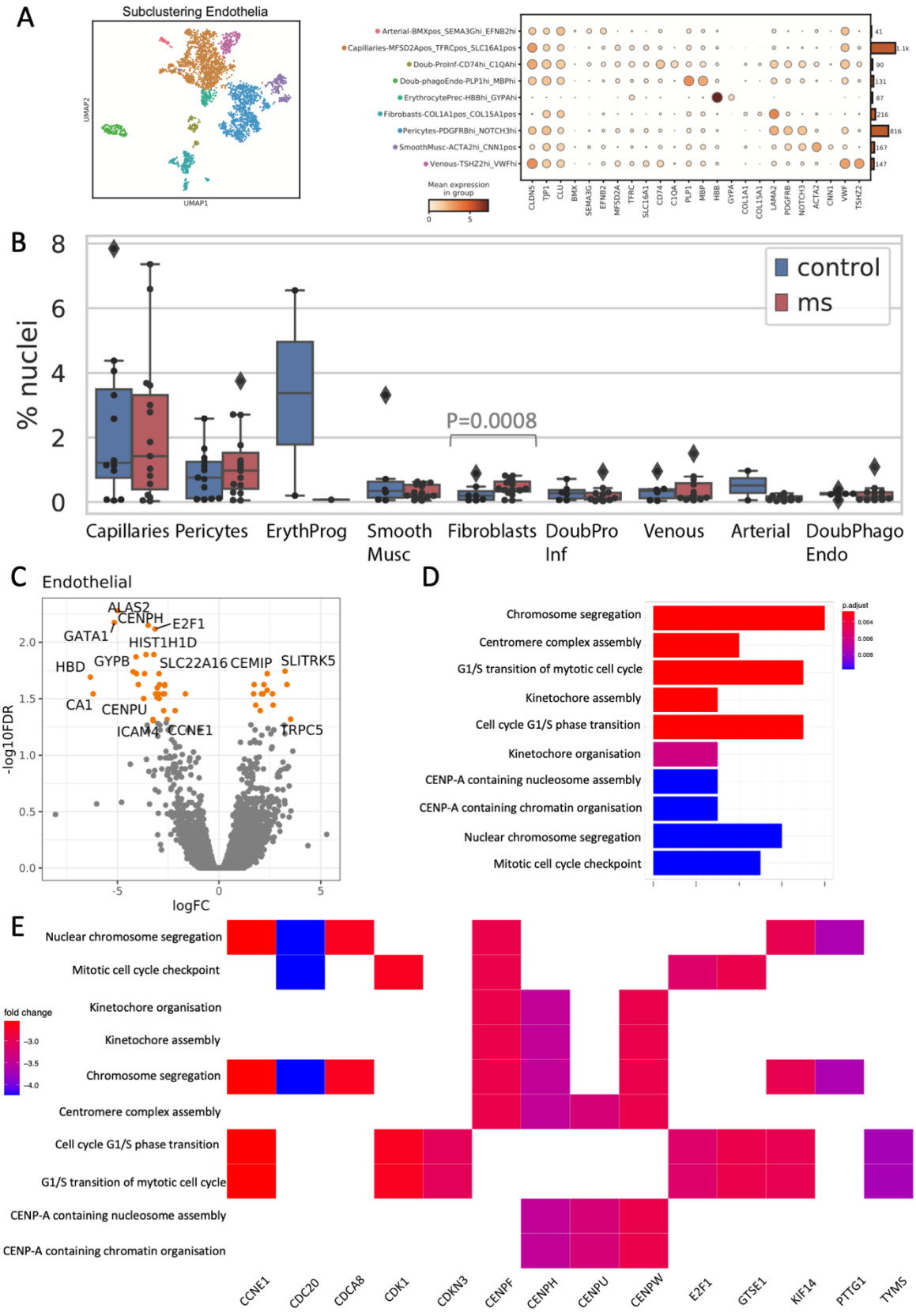


Figure 4.30: Subclustering and characterisation of endothelial cells. (A) Subclustering the endothelial compartment clusters of vascular cells included arterial (BMX^{+} SEMA3Ghi EFNB2hi), capillary (MFSD2A^{+} TFRC^{+} SLC16A1^{+}), smooth muscle (ACTA2hi CNN1^{+}) and

venous (TSHZ2hi VWFhi). Pericytes (PDGFRBhi NOTCH3hi), fibroblasts (COL1A1^{+} COL15A1^{+}) and erythrocyte precursors (HBBhi GYPAhi). Similarly to immune and astrocyte compartment two clusters were identified which could have a phagocytosing phenotype (DoubPhagoEndo-PLP1hi MBPhi and DoubProInf CD74hi C1QAhi) (B) Fibroblasts were more abundant in MS than controls (p=0.0008; FDR = 0.0023). Given the lack of differential abundance of doub clusters (indicating less likelihood of phagocytosing process taking up myelin products) and high doublet scores of these clusters, they were excluded from all subsequent analyses (C) Multiple differentially expressed genes were identified when comparing cases and controls, with the majority of DEGs being upregulated in controls. (D) GO enrich analysis was performed on DEGs and the top 10 biological process pathways with padj<0.05 included chromosome segregation, centromere complex assembly, G1/S transition of mitotic cell cycle, kinetochore assembly, cell cycle G1/S phase transition, kinetochore organisation, CENP-A containing nucleosome assembly, CENP-A containing chromatin organisation, nuclear chromosome segregation, mitotic cell checkpoint. (E) Relationship between DEGs and GO enrich pathways.

Fibroblasts were the differentially abundant cell type in this compartment, which were more abundant in MS than in controls (p=0.0008) (**Fig 4.30B**). DGE analysis identified over 100 DEGs; GO enrichment analysis of the genes upregulated in controls related to cell cycling/mitotic processes (e.g. nuclear chromosome segregation, mitotic cell cycle checkpoint, kinetochore organisation and assembly, centromere complex assembly, cell cycle G1/S phase transition, chromatin organisation, nucleosome assembly) (**Fig 4.30D,E**).

WGCNA analysis identified a gene expression module which positively correlated with disease, green yellow (OR: 0.44; p=0.02) (**Fig 4.31**). GO terms associated with this module included nucleoside monophosphate phosphorylation, ADP biosynthetic process, adenylate kinase activity, and nucleoside monophosphate kinase activity (**Fig 4.32-34**). KEGG terms associated with this module were thiamine metabolism, biosynthesis of cofactors and purine metabolism (**Fig 4.35**).

A total of 51 metabolites were identified to be differentially accumulated/depleted in MS endothelial cells compared to controls (**Fig 4.36**). Of these, the most depleted ones were cholesterol (padj=4.782 x 10^{-67}) and inosine monophophate (IMP; padj=1.205 x 10^{-4}), while the most significantly accumulated was glucose (padj=9.845 x 10^{-5}). The presence of nucleoside monophosphate phosphorylation among GO terms associated with disease alongside inosine monophosphate depletion is interesting, since inosine is a nucleoside – potentially indicating increased phosphorylation and activity of this metabolite in spite of overall depletion. Inosine has been tested in a small cohort as a

treatment for MS, which increased the Th2 to Th1cytokinesin circulating cells and improved the Kurtzke expanded disability status scale of treated patients[201]. The largest logFC in metabolites was (E,E)-Farnesyl-PP (logFC>10; padj=5.554 x 10^{-5}).

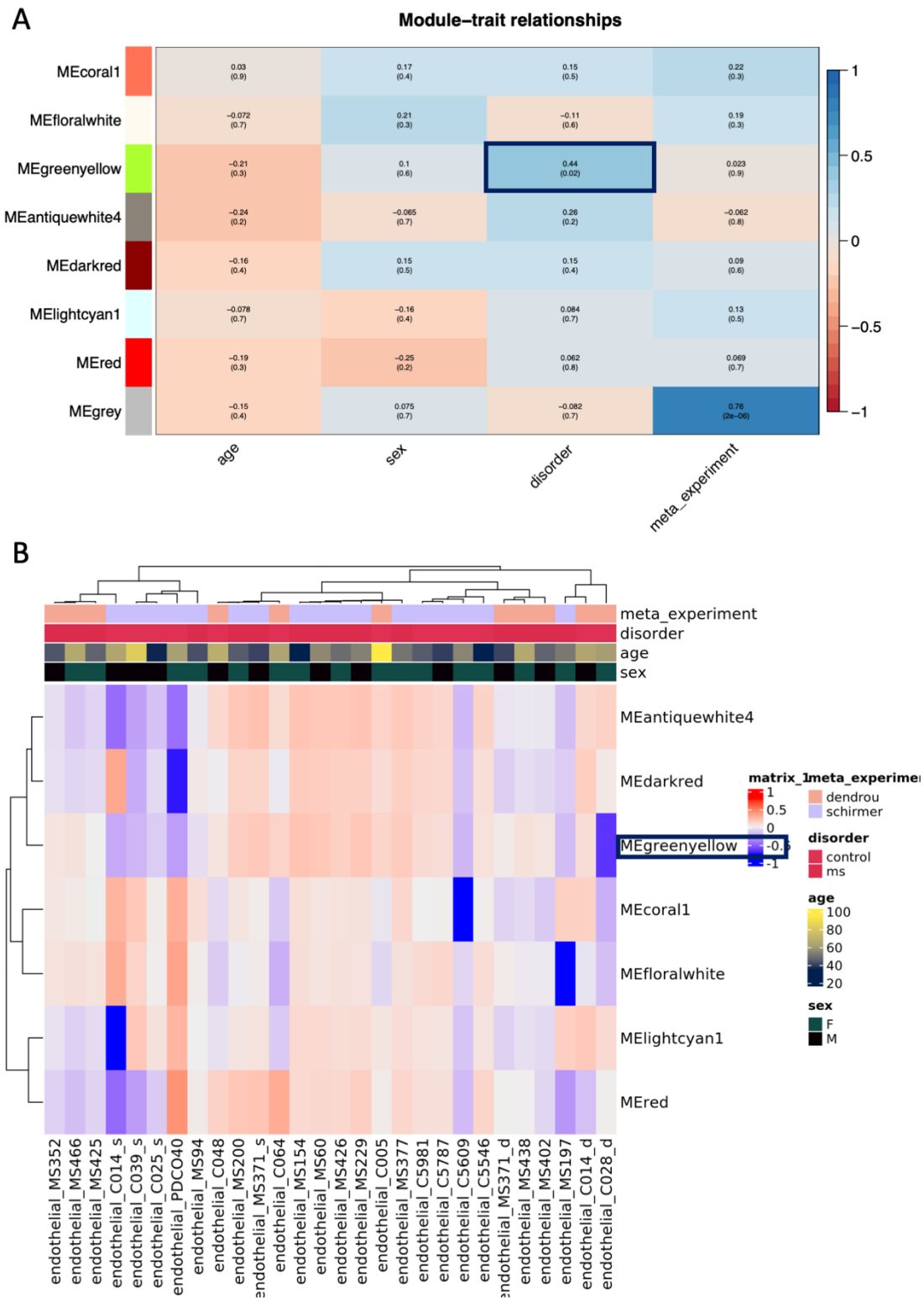


Figure 4.31: WGCNA analysis of endothelial compartment. (A) The green yellow gene expression module positively correlates with MS (0.44, padj = 0.02) (B) The green yellow module is positively correlated with n=7 controls and n=12 MS.

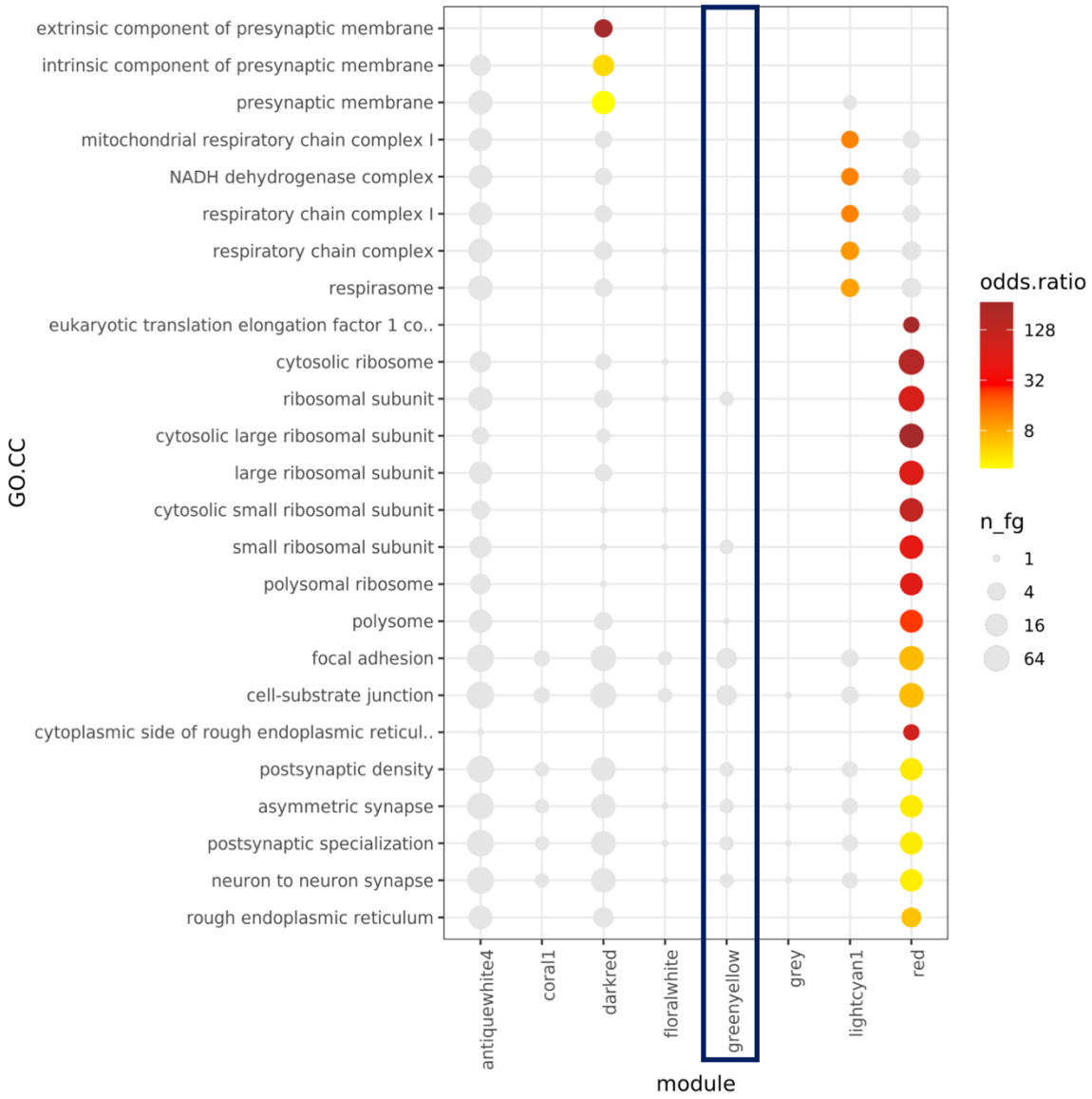


Figure 4.33: The green yellow WGCNA gene expression module for endothelial cells did not correlate with any GO cellular component pathways.

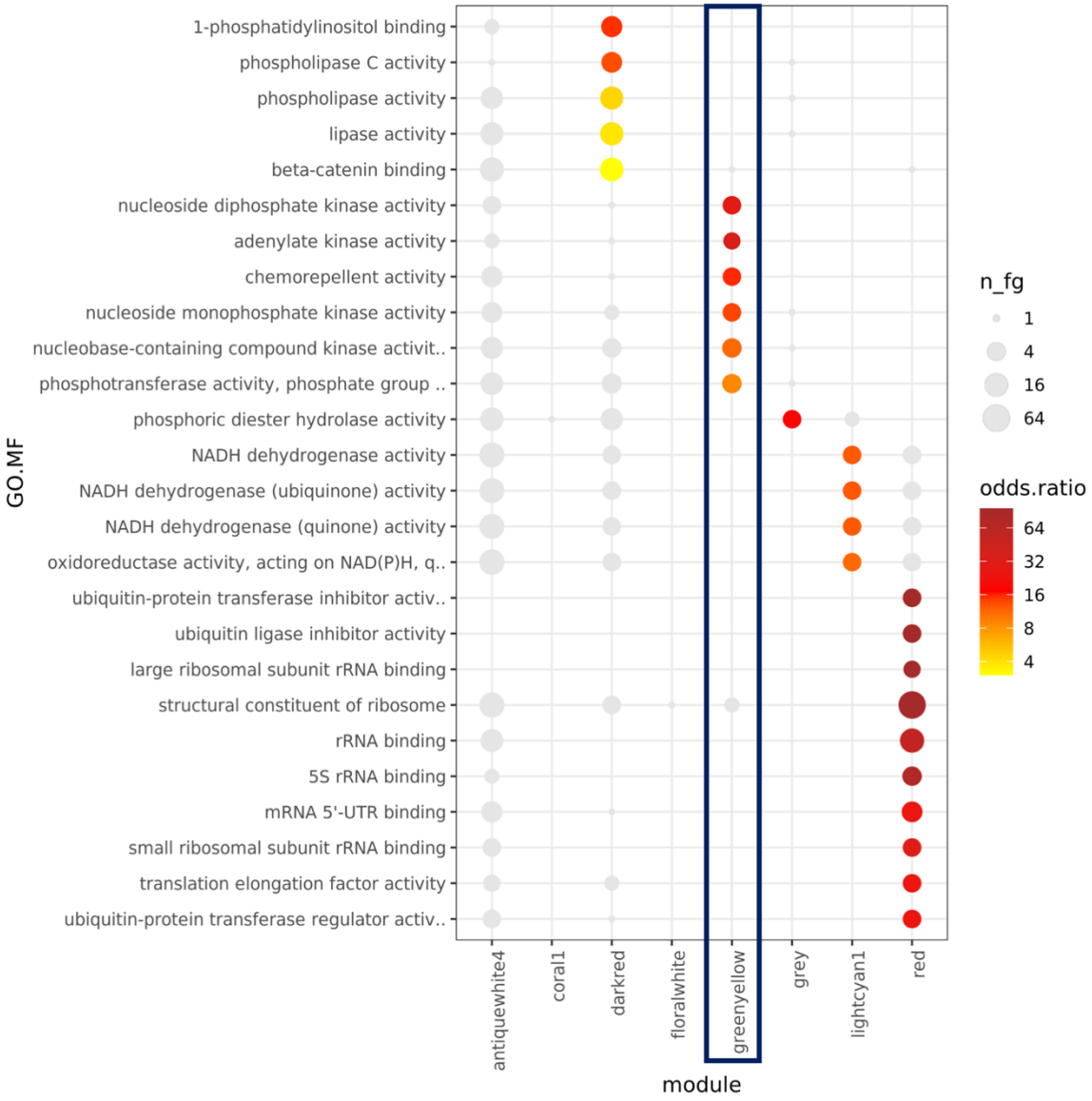


Figure 4.34: Endothelial compartment WGCNA GO biological function correlations. The green yellow gene expression module correlated with the nucleoside diphosphate/monophosphate kinase activity, adenylate kinase activity, chemorepellent activity, nucleobase-containing compound kinase activity, phosphotransferase activity phosphate group as acceptor pathways.

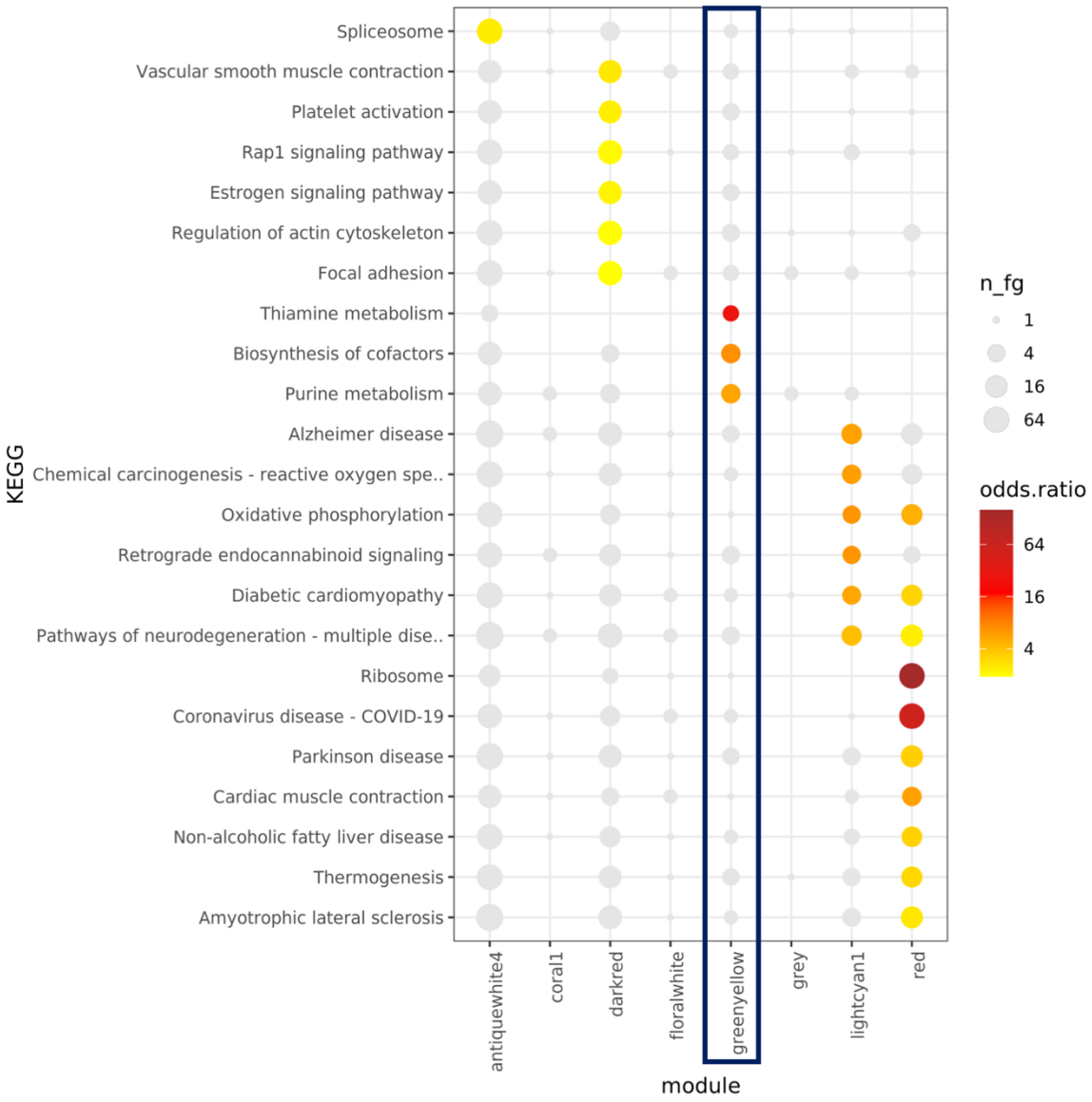


Figure 4.35: KEGG terms associated with WGCNA gene expression modules of endothelial compartment. The green yellow WGCNA gene expression module correlated with KEGG terms thiamine metabolism, purine metabolism and biosynthesis of cofactors.

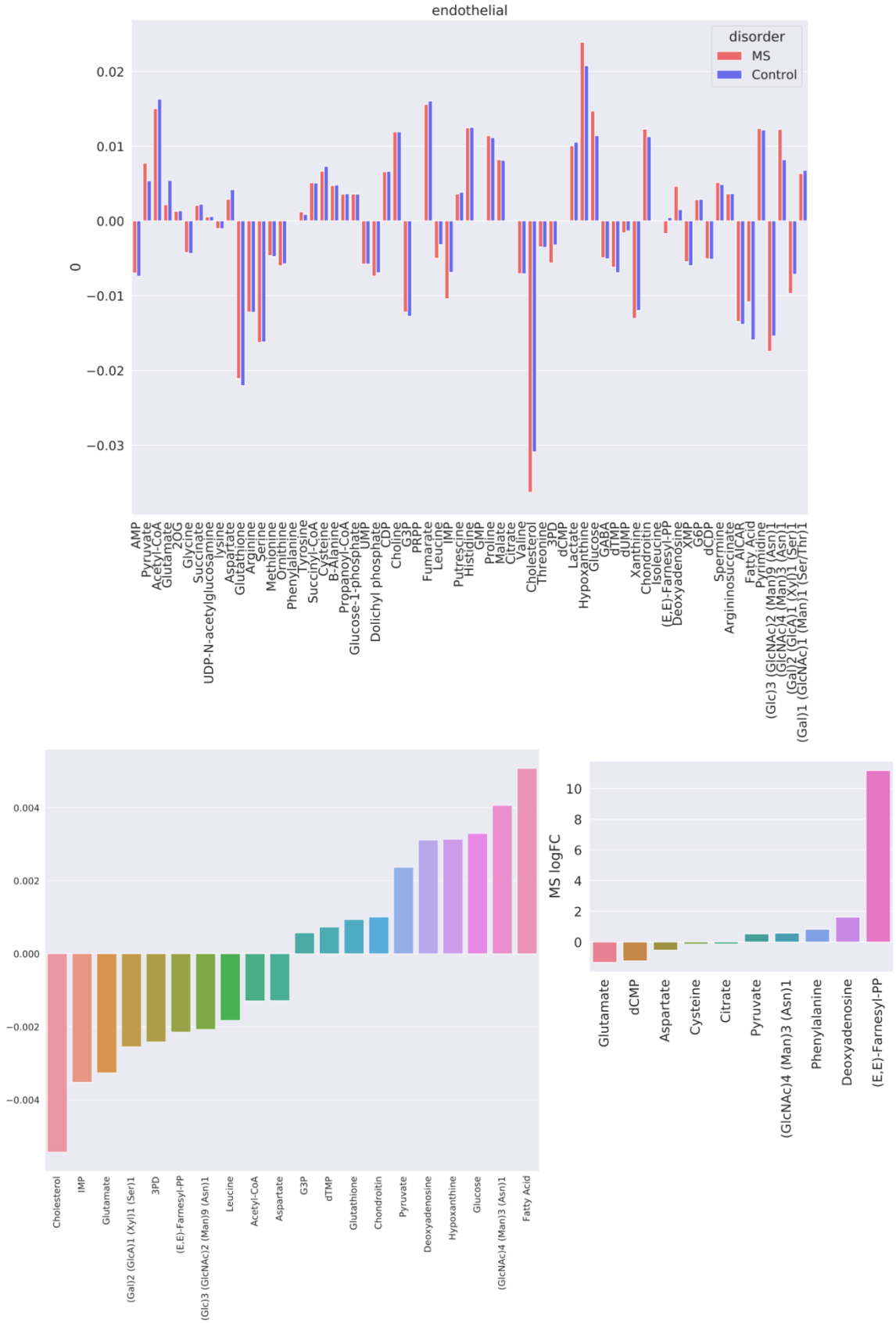


Figure 4.36: Metabolite accumulation and depletion analysis in endothelial cells. (A) Comparison of mean metabolite accumulation/depletion in cases and controls (B) Metabolite difference between cases and controls, top 10 most differentially depleted

and accumulated metabolites in endothelial cells. The most differentially depleted metabolite is cholesterol (padj=4.782 x 10^{-67}) followed by IMP (padj=1.205 x 10^{-4}) and glutamate (ns), while the most accumulated metabolites are fatty acid and $GlcNac_4Man_3Asn_1$ (ns). (C) Log FC of accumulation and depletion when comparing cases and controls – glutamate is the most depleted with a logFC<-1 and (E,E)-Farnesyl-PP is the most accumulated (logFC>10; padj=5.554 x 10^{-5}) (see **table 4.1** for all statistical comparisons).

Table 4.1: Summary of statistical comparison of metabolites accumulation/depletion for each cell type comparing cases vs. controls

Metabolite	Astrocytes (padj)	CD74+ (padj)	Endothelial (padj)	Neurons (padj)	OGs (padj)
AMP	4.27E-01	4.95E-01	1.14E-02	1.49E-02	3.89E-21
Pyruvate	2.64E-25	1.71E-01	8.24E-03	1.64E-84	3.96E-281
Acetyl-CoA	6.86E-01	6.82E-01	3.79E-01	7.88E-01	3.72E-01
Glutamate	1.72E-02	3.65E-01	1.10E-01	1.07E-02	2.33E-09
2OG	5.39E-01	9.40E-03	2.72E-25	3.49E-04	2.15E-30
Glycine	2.94E-29	1.07E-51	3.33E-107	1.04E-136	5.18E-01
Succinate	1.27E-15	1.75E-03	7.66E-07	1.41E-14	3.82E-103
UDP-N-acetylglucosamine	5.73E-05	1.11E-07	1.17E-02	5.46E-122	1.72E-08
lysine	1.77E-06	2.25E-01	3.08E-09	4.95E-02	7.95E-02
Aspartate	7.18E-04	3.26E-01	2.09E-07	1.38E-306	1.36E-08
Glutathione	2.28E-31	1.18E-06	2.54E-01	1.17E-66	6.02E-56
Arginine	3.82E-124	1.46E-02	3.98E-32	0.00E+00	4.05E-01
Serine	1.32E-05	6.47E-43	1.42E-60	1.09E-10	8.93E-01
Methionine	3.91E-02	1.28E-02	1.81E-10	4.10E-22	2.20E-62
Ornithine	2.72E-69	2.17E-10	1.14E-02	5.94E-33	0.00E+00
Phenylalanine	2.08E-127	2.23E-15	2.92E-41	9.34E-187	1.76E-03
Tyrosine	2.36E-55	1.01E-02	4.48E-03	3.76E-39	4.90E-291
Succinyl-CoA	8.66E-03	3.65E-01	6.97E-03	7.93E-06	1.78E-23
Cysteine	8.76E-01	7.98E-03	8.33E-03	8.14E-16	4.75E-04
B-Alanine	1.49E-03	1.75E-03	1.12E-01	3.95E-33	5.85E-04
Propanoyl-CoA	4.04E-01	5.12E-12	1.85E-26	1.67E-144	7.89E-12
Glucose-1-phosphate	5.14E-05	7.76E-05	4.24E-04	5.19E-18	1.65E-51
UMP	8.21E-11	8.77E-01	1.14E-01	2.65E-09	2.34E-52
Dolichyl phosphate	2.13E-05	1.79E-06	1.08E-04	2.29E-01	1.39E-16
CDP	1.01E-08	6.65E-01	9.18E-02	1.10E-08	2.27E-121
Choline	1.97E-70	1.25E-01	3.83E-01	1.67E-15	1.87E-155
G3P	7.97E-55	3.91E-03	2.60E-03	5.46E-111	0.00E+00
PRPP	2.68E-09	1.92E-02	1.51E-08	4.38E-280	5.27E-46
Fumarate	2.65E-10	1.89E-03	6.97E-03	5.91E-10	1.17E-93
Leucine	5.61E-01	1.66E-03	6.49E-25	4.53E-09	2.54E-06
IMP	3.05E-02	1.46E-01	1.20E-04	2.65E-01	8.40E-01
Putrescine	2.98E-24	1.72E-04	5.94E-33	1.18E-253	6.42E-37
Histidine	1.96E-44	4.55E-02	6.94E-11	0.00E+00	0.00E+00
GMP	8.25E-03	8.77E-01	3.35E-13	1.92E-04	1.01E-86
Proline	1.41E-22	1.60E-03	2.60E-03	1.07E-92	7.37E-39
Malate	5.37E-28	1.91E-25	8.71E-62	7.12E-04	0.00E+00
Citrate	1.51E-31	4.95E-01	2.73E-17	2.30E-01	1.84E-02

Valine	5.93E-01	9.40E-03	5.68E-02	2.10E-09	4.46E-02
Cholesterol	8.63E-19	4.60E-02	2.54E-01	8.31E-01	8.86E-12
Threonine	3.90E-01	1.13E-12	6.23E-28	1.87E-146	8.26E-07
3PD	4.27E-11	2.81E-01	1.13E-05	7.10E-66	3.33E-22
dCMP	3.78E-26	3.63E-06	1.89E-02	4.12E-01	3.71E-149
Lactate	6.86E-01	3.35E-09	1.69E-07	5.74E-44	4.48E-135
Hypoxanthine	2.69E-05	4.16E-02	1.12E-08	2.43E-05	8.59E-30
Glucose	7.60E-02	1.09E-12	9.85E-05	1.06E-08	5.32E-02
GABA	7.25E-63	7.01E-02	5.91E-06	1.11E-243	7.34E-39
dTMP	8.33E-01	1.66E-02	4.49E-10	3.94E-01	9.05E-24
dUMP	1.75E-07	3.89E-34	4.66E-26	2.12E-21	3.48E-11
Xanthine	1.09E-235	1.06E-38	7.36E-88	8.60E-10	0.00E+00
Chondroitin	2.10E-02	6.65E-01	1.51E-08	7.96E-01	3.40E-13
Isoleucine	4.38E-15	2.40E-04	5.00E-01	1.99E-45	6.94E-299
(E,E)-Farnesyl-PP	3.88E-01	3.26E-01	5.55E-05	3.71E-03	3.62E-06
Deoxyadenosine	4.86E-04	9.07E-05	6.52E-03	9.11E-02	4.78E-67
XMP	2.04E-11	1.92E-02	1.04E-06	1.31E-06	3.43E-193
G6P	9.10E-01	1.60E-03	1.66E-01	1.19E-67	1.13E-07
dCDP	3.28E-01	3.11E-01	7.54E-05	5.35E-13	2.53E-34
Spermine	8.25E-03	3.49E-03	1.70E-05	6.23E-47	6.75E-09
Argininosuccinate	6.55E-01	5.84E-01	1.04E-71	7.17E-08	0.00E+00
AICAR	3.77E-12	6.64E-01	7.22E-03	1.45E-04	6.82E-192
Fatty Acid	8.76E-01	7.00E-01	5.87E-01	6.11E-01	1.33E-01
Pyrimidine	3.14E-36	2.61E-04	5.87E-01	3.59E-35	2.08E-02
(Glc)3 (GlcNAc)2 (Man)9 (Asn)1	2.80E-05	3.26E-01	3.24E-02	8.91E-02	2.15E-02
(GlcNAc)4 (Man)3 (Asn)1	6.86E-01	3.26E-01	2.14E-01	6.53E-01	6.45E-02
(Gal)2 (GlcA)1 (Xyl)1 (Ser)1	2.59E-01	7.67E-01	4.93E-03	9.90E-01	1.89E-01
(Gal)1 (GlcNAc)1 (Man)1 (Ser/Thr)1	5.36E-03	1.27E-02	1.63E-02	5.10E-07	2.44E-54

4.2.2 Cell-cell interaction analysis

In addition to DA, DGE, WGCNA, and metabolite analyses, cell-cell interaction analyses can provide potential functional insights regarding which cells in a given tissue type or area engage with each other the most and via ligand-receptor (LR) interactions.

4.2.2.1 Top specificity cell-cell interactions

LR interactions were ranked based on specificity using the NATMI software and its associated library, where the highest specificity score indicates a ligand is only expressed in one cell type and the corresponding receptor is only expressed in one cell type. Top LR interactions were plotted in a chord plot using LR analysis from controls and cases (**Fig 4.37**). Interacting cell types present in controls and absent in cases included CUX2 and ID2 neurons, chandelier cells (PVALB_UNC5), arterial cells and smooth muscle. Interacting clusters unique to MS cases included plasma cells, PeriMacro, MicroMacro-CPM_LPL, VIP EREG, VIP FLT_SCTR, and PhagoProtoAstro. Fibroblasts had increased communication in MS, particularly communicating with the OPALINlow_SLC5A11hi_RBFOXhi OGs.

Overall, as anticipated the chord plot visualisation indicated a prominent role of the immune cell compartment in MS, with greater cellular diversity and prominent specific communication. It also highlighted the shift in the astrocyte compartment from a homeostatic state to an activated state, with specific internal communication of fibrous astrocytes gaining prominence over that of protoplasmic astrocytes (SLC1A2hi_NRXNhi), as well as the appearance of the PhagoProtoAstro cluster.

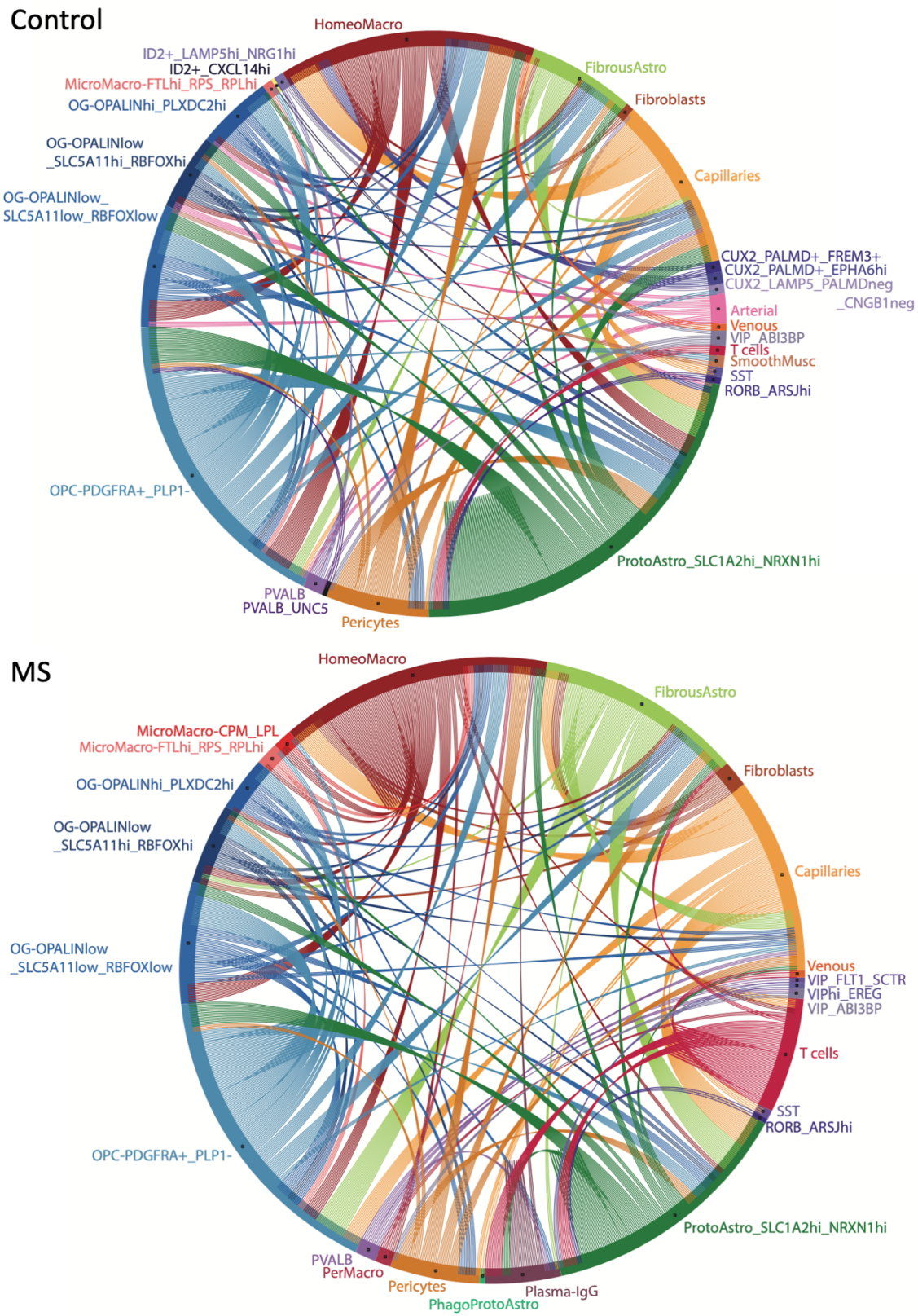


Figure 4.37: Top specificity cell-cell communication networks in cases and controls. There is increased communication of astrocytes, macrophages/microglia, fibroblasts, VIP neurons, T-cells and Plasma cells, PhagoProtoAstro in the context of MS when compared to controls. Cell types with notable decrease in communication include CUX2 neurons, chandelier cells (PVALB_UNC5), arterial cells, smooth muscle.

In spite of OG loss being prominent in MS, the overall indication was that OGs continued to have a prominent role in specific cell-cell communication processes. Compared to other cell types neurons had a less prominent role in specific communication in controls, and this was further decreased in cases.

4.2.2.2 LR interactions specific to controls

LR interactions unique to controls involved *FGF10* and *FGF17* signalling to receptors *FGFR1/2/3* (Fig 4.38). APOE and LRPAP1 ligand signalling to *LDLR* also are present. The cell type most dominantly involved in these signalling patterns was CUX2 neurons both in ligand synthesis for *FGF17* and *LRPAP1*, as well as receptor presentation of *LDLR* and *FGFR2* (Fig 4.39).

4.2.2.3 LR interactions specific to MS

To aid biological interpretation, LR pairs were divided into four groups based on the EntrezID and SwissProt gene function description of the ligands. Whilst this is not an all-encompassing classification, it was employed for aiding visualisation of interactions. Where genes could be classified in more than one group, they were assigned to one based on biological relevance to MS. For example, the *APP*, *LRP1* and *PSEN1* genes which have pleiotropic functions but are known to have a significant role in pathogenesis of AD, have been placed in the protein folding category.

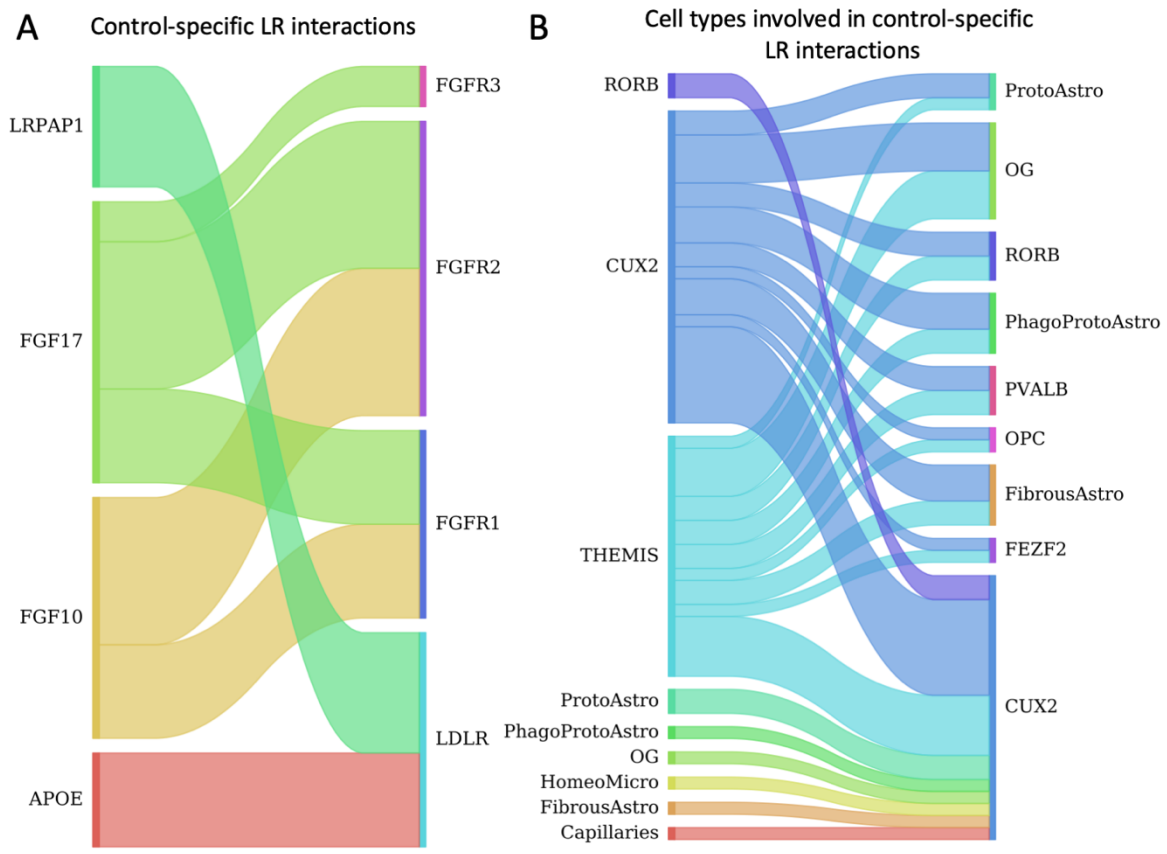


Figure 4.38: Control-specific LR interactions. A total of 7 ligand-receptor interactions were present in the controls which were not present in cases; many of these included FGF or FGF receptors. The top 3 cell types accounting for ligands were neuronal cell types and jointly accounted for 87% of ligands – CUX2 neurons alone accounted for 40% of ligands. CUX2 also accounted for 40% of receptors, although OGs and OPCs, astrocytes and neuronal cell types also expressed receptors.

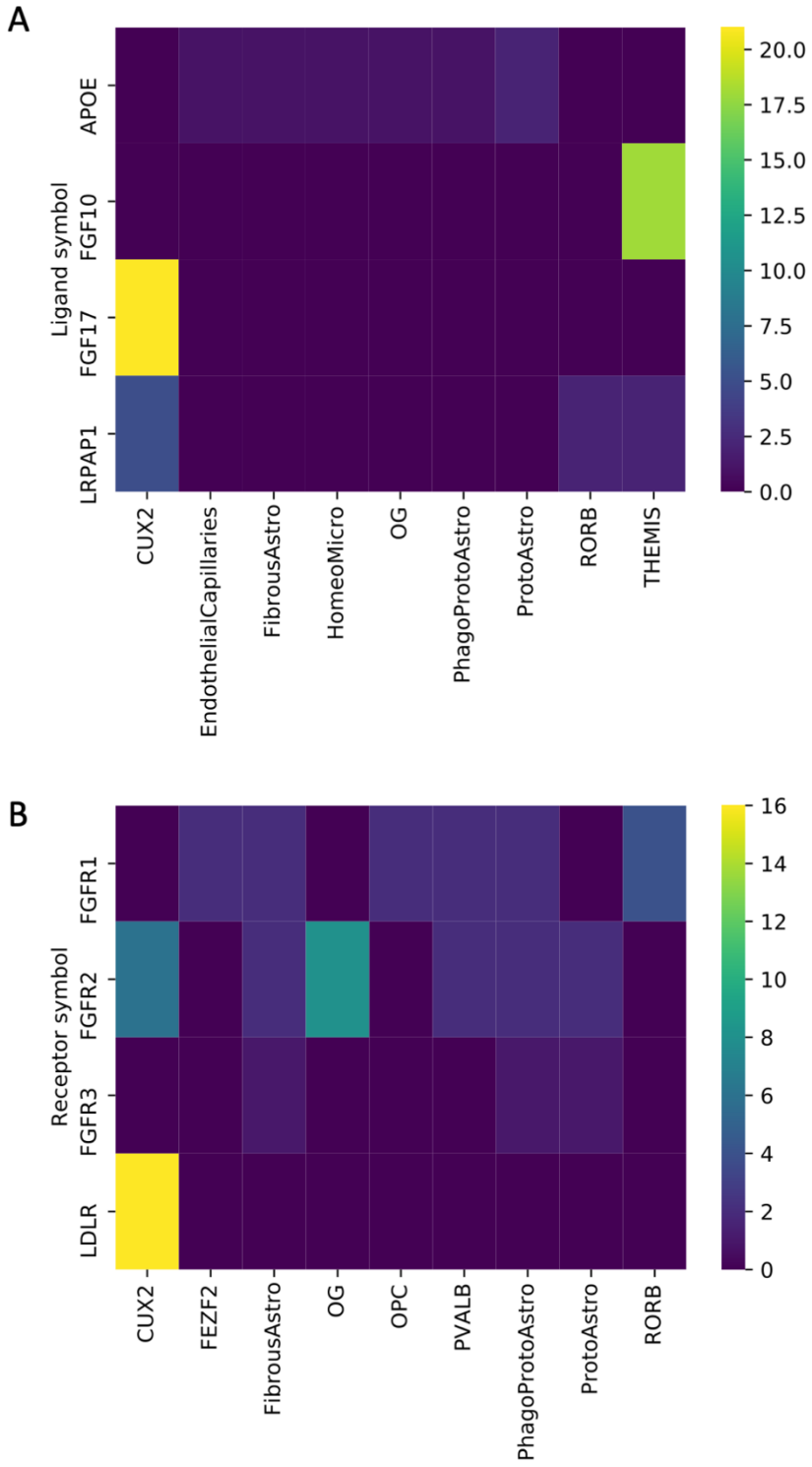


Figure 4.39: Heatmaps of cell types signalling in control-specific LR interactions. (A) CUX2 and THEMIS neurons are the most highly expressing cell types in control-specific LR communications, accounting for most FGF17 and FGF10 communication, respectively (B) CUX2 is the cell type expressing LDLR. FGFR1 and FGFR2 are lowly expressed in several cell types, with FGFR2 most highly expressed in CUX2 and OG.

Out of the 29 pairs of LR interactions associated with immune function, a large proportion of the interactions involved the complement *C3* ligand which is involved in both innate and adaptive immune responses (**Fig 4.40**). *LRP1*, involved in lipid metabolism and reportedly involved in clearance of amyloid beta plaque accumulation in AD, was found alongside *NRP1*, involved in cell migration. Several members of the integrin family were also present, including *ITGAV*, *ITGA9*, *ITGAX*, *ITGAM*, *ITGB1*, *ITGB2*. As would be expected, many of the ligand-expressing cells were of an immune cell type, such as macrophages/microglia and T cells (**Fig 4.41**). Endothelial cell types such as pericytes and capillaries also significantly contributed to the expression of these genes.

Several ligands involved in coagulation were also identified, such as *FN1*, *FGL1*, *FBLN1*, and *F13A1* (**Fig 4.42**). Integrin cell adhesion molecules were prominently represented in receptors in this group of LR interactions. The ligand-expressing cells in this group were mostly of endothelia, such as fibroblasts, pericytes, smooth muscle, capillaries; the receptor-expressing cells were mostly neuronal (**Fig 4.43**).

A total of 32 ligand-receptor pairs were identified which may be involved in protein folding (**Fig 4.44**). Among these, the two most expressed ligands were *APP*, which forms the basis of amyloid plaque formation found in the brains of patients with AD, and *CALR* which has a role in protein quality control and calcium homeostasis. Other ligands such as *DNAJB11*, *HSP90AA1*, *PSEN1* and *LRPAP1* have direct roles in protein folding. Notably, ligands *LPL*, *LIPC* and *LRPAP1* and the receptor *LRP1* have close relationships with lipid metabolism.

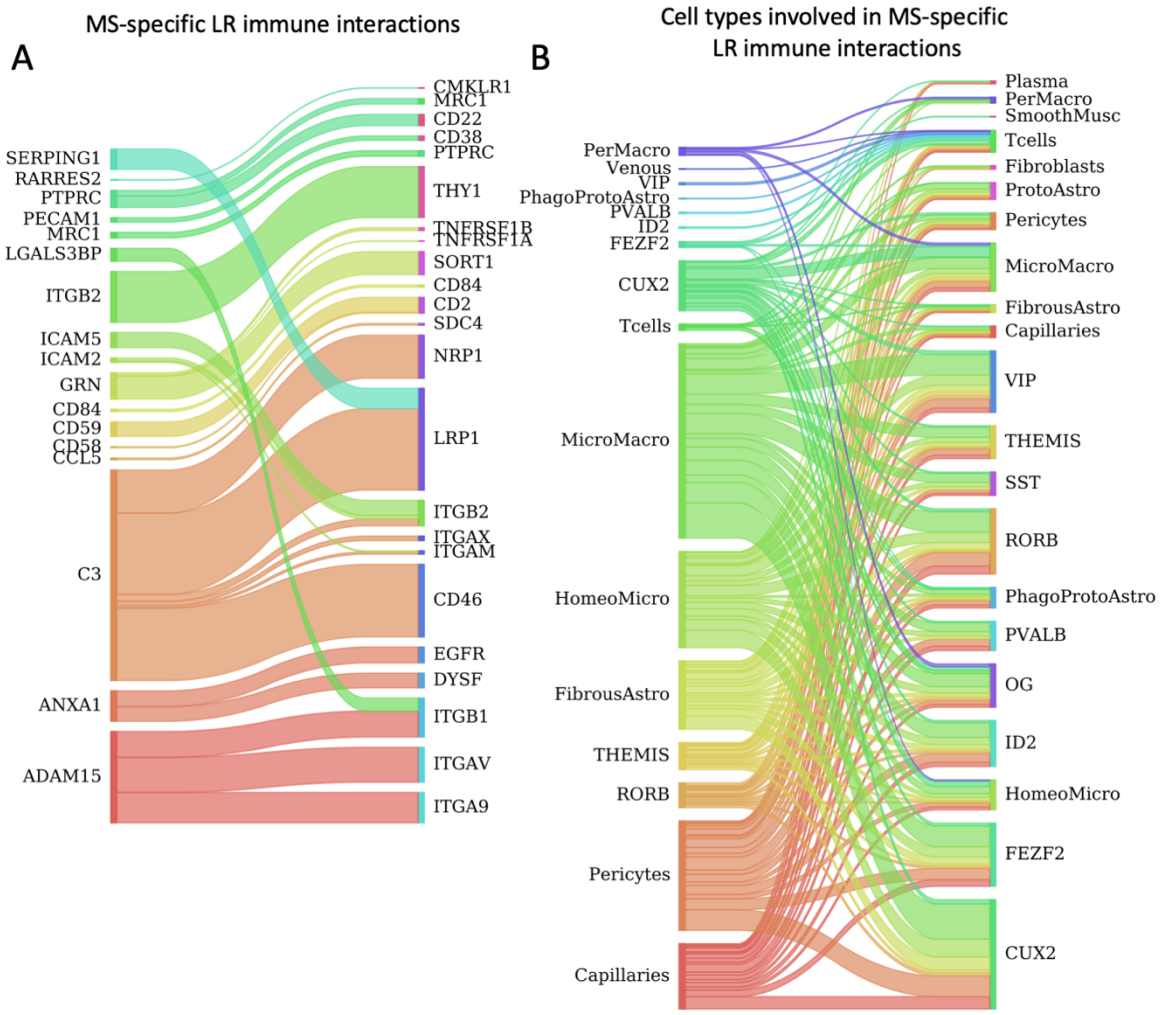


Figure 4.40: MS-specific LR interactions where the ligand has a role in immune response. (A) C3 is the most highly expressed ligand communicating to 6 distinct receptors; ADAM15 is the second highest expressed ligand and communicates with three members of the integrin family. (B) The cells most prominently involved in ligand production in these interactions are microglia/macrophages, pericytes, fibrous astrocytes and capillaries, while the receptors are distributed across a large number of cell types including immune cells and neurons.

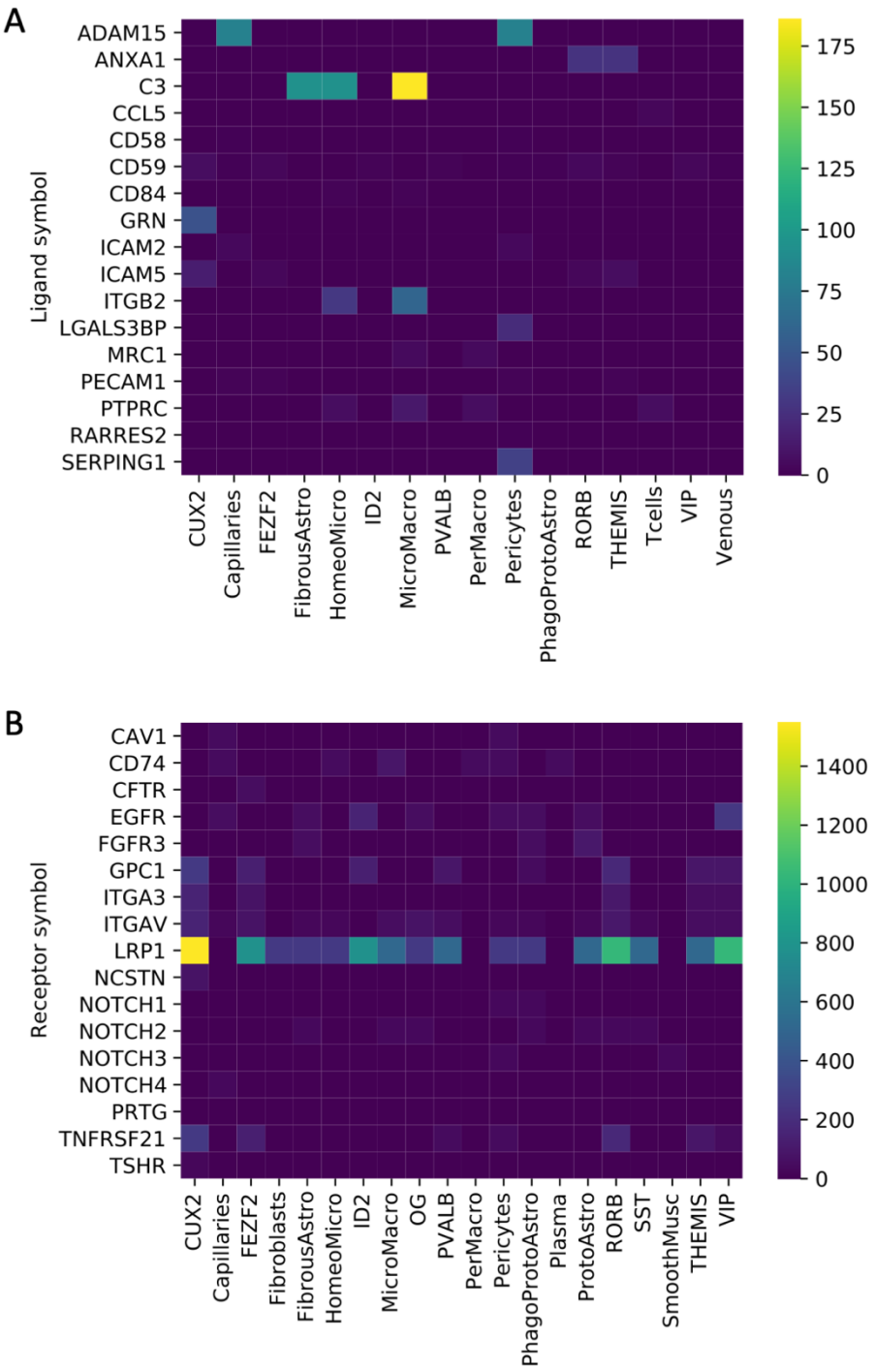


Figure 4.41: Heatmaps of cell types signalling in MS-specific LR interactions where the function of ligands is associated with immune response. (A) C3 expression is the highest and mostly occurs in microglia/macrophage, homeostatic microglia and fibrous astrocytes. ADAM15 expression is the second highest and is specific to capillaries and pericytes. (B) The most highly expressed receptor is LRP1, which is expressed in almost all cell types with the exception of capillaries, peripheral macrophages, plasma cells and smooth muscle cells. LRP1 expression is particularly high in neuronal cell types.

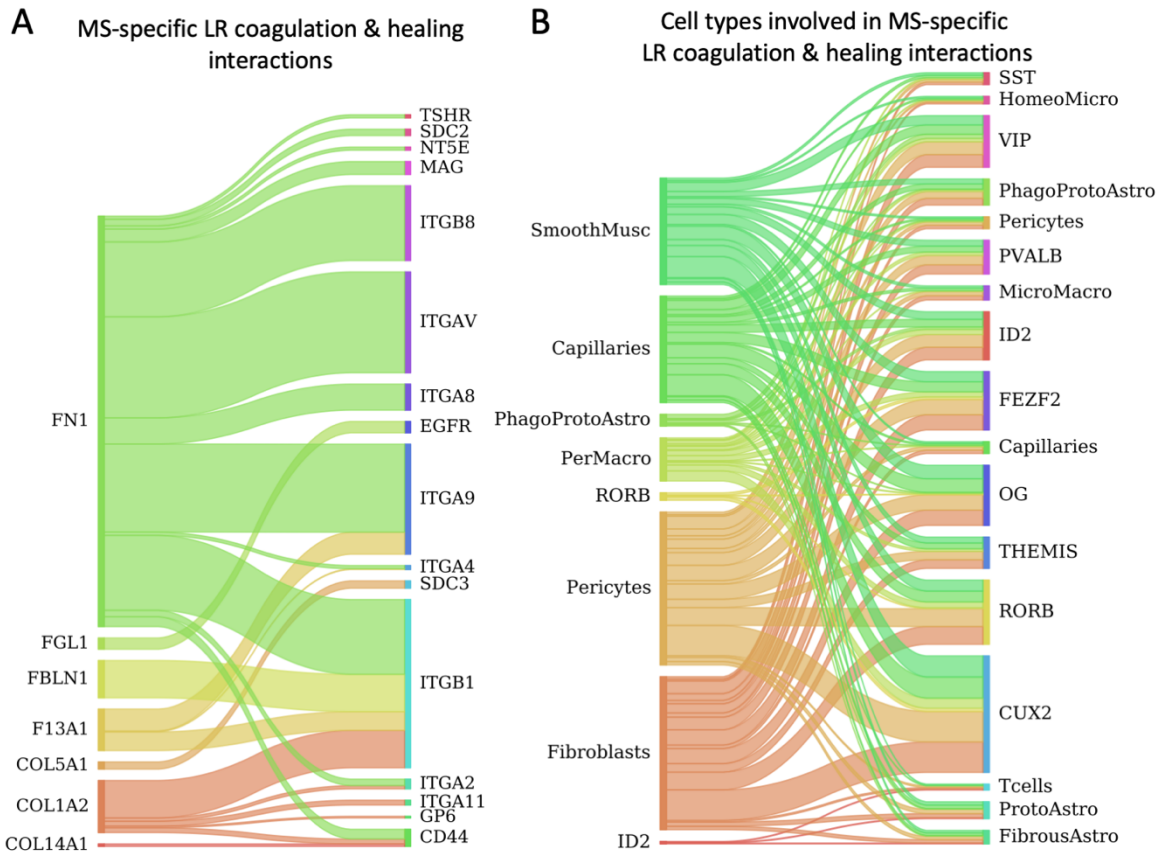


Figure 4.42: MS-specific LR interactions where the ligand is involved in coagulation and/or healing processes. (A) The most highly expressed ligand is FN1, which is communicating with 8 receptors members of the integrin families as well as several other receptors more lowly expressed. Three classes of collagen are also expressed. (B) Endothelial cells such as smooth musc, capillaries, pericytes and fibroblasts are responsible for most ligand synthesis, while receptors are present in a large number of cells, including neurons, astrocytes, OGs and immune cells.

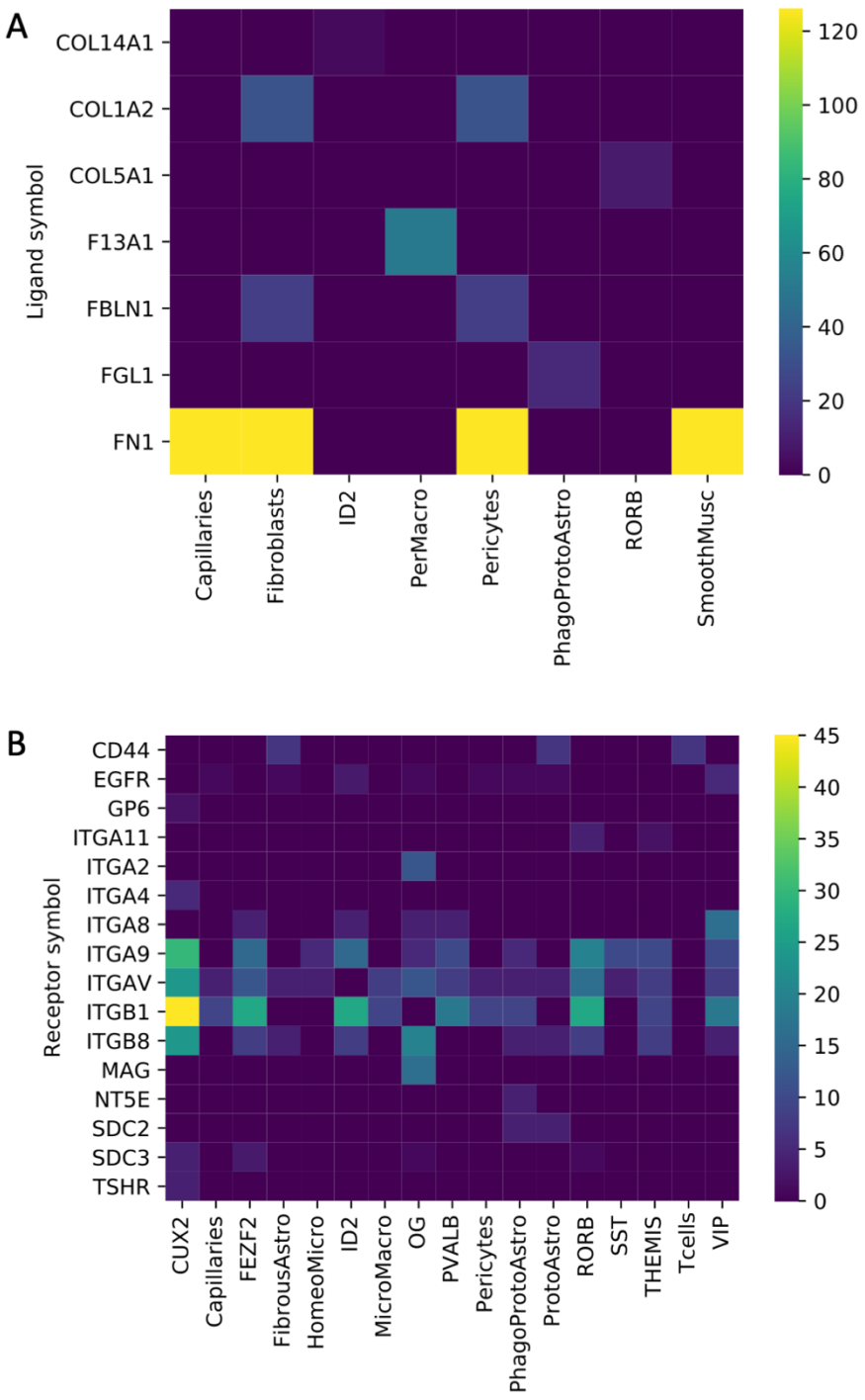


Figure 4.43: Heatmaps of cell types signalling in MS-specific LR interactions where the function of ligands is associated with coagulation and/or healing. (A) FN1 is the most highly expressed ligand, with high expression in endothelial cells – capillaries, fibroblasts, pericytes and smooth muscle cells. (B) The most highly expressed receptors are four members of the integrin family, ITGB9, ITGB1, ITGAV, ITGA9 which are expressed across multiple cell types. The mostly expressed of these is ITGB1 expression in CUX2 neurons.

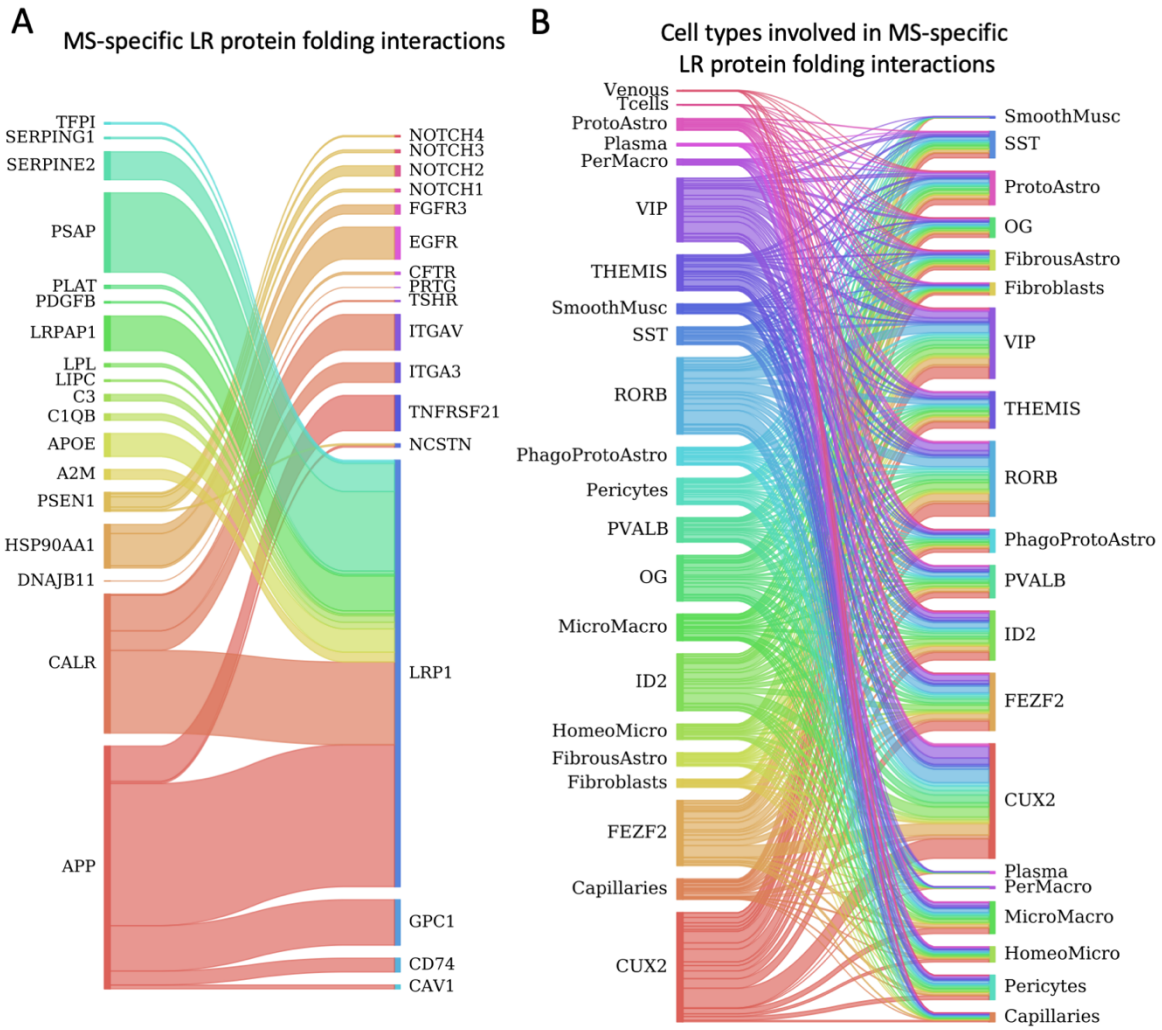


Figure 4.44: MS-specific LR interactions related to protein folding or conditions relating to aberrant protein folding. (A) APP and CALR are the two most highly expressed ligands, and both communicate highly with LRP1 – the most highly expressed receptor. (B) Both ligand and receptor expression are spread across various cell types, with highest expression both of ligands and receptors in CUX2.

Among receptors, several growth factors such as *EGFR* and *FGFR3* were identifiable as well as cell adhesion molecules *ITGAV*, *ITGA3* and *NOTCH1-NOTCH4*. Mainly neuronal subtypes expressed ligands in this group of LR interactions (**Fig 4.45**).

A total of 82 LR interactions were identified where the ligand had a role associated with development, neurite outgrowth or proliferation (**Fig 4.46**). Of these, the neural cell adhesion molecule *NCAM1* which has roles regulating neurite outgrowth, neurogenesis and cell migration was the most abundantly expressed ligand, and five members of the semaphorin protein family are also represented, which have roles in axon path guidance as inhibitors or chemorepellants. Among receptor families abundantly expressed in this subset, *ROBO1* was the most expressed and has a role in axon guidance and neuronal precursor cell migration. Several members of the netrin family of receptors were represented (*UNC5A/B/C/D*); this family mediates axon repulsion of neuronal growth cones in the developing CNS. Similarly, several receptors members of the ephrin family (EPH) were also abundantly expressed, which have also been implicated in CNS development. The neuronal cell types were the dominant communicators in this LR group (**Fig 4.47,48**).

Of the remaining LR pairs, the majority of signalling involved the ligands *GNAS* and *GNAI2*, which are involved in several cellular processes, as indicated by the wide range of receptors these ligands are interacting with (**Fig 4.49**). Most of the ligands in this group were expressed by neuronal cell types (**Fig 4.50**).

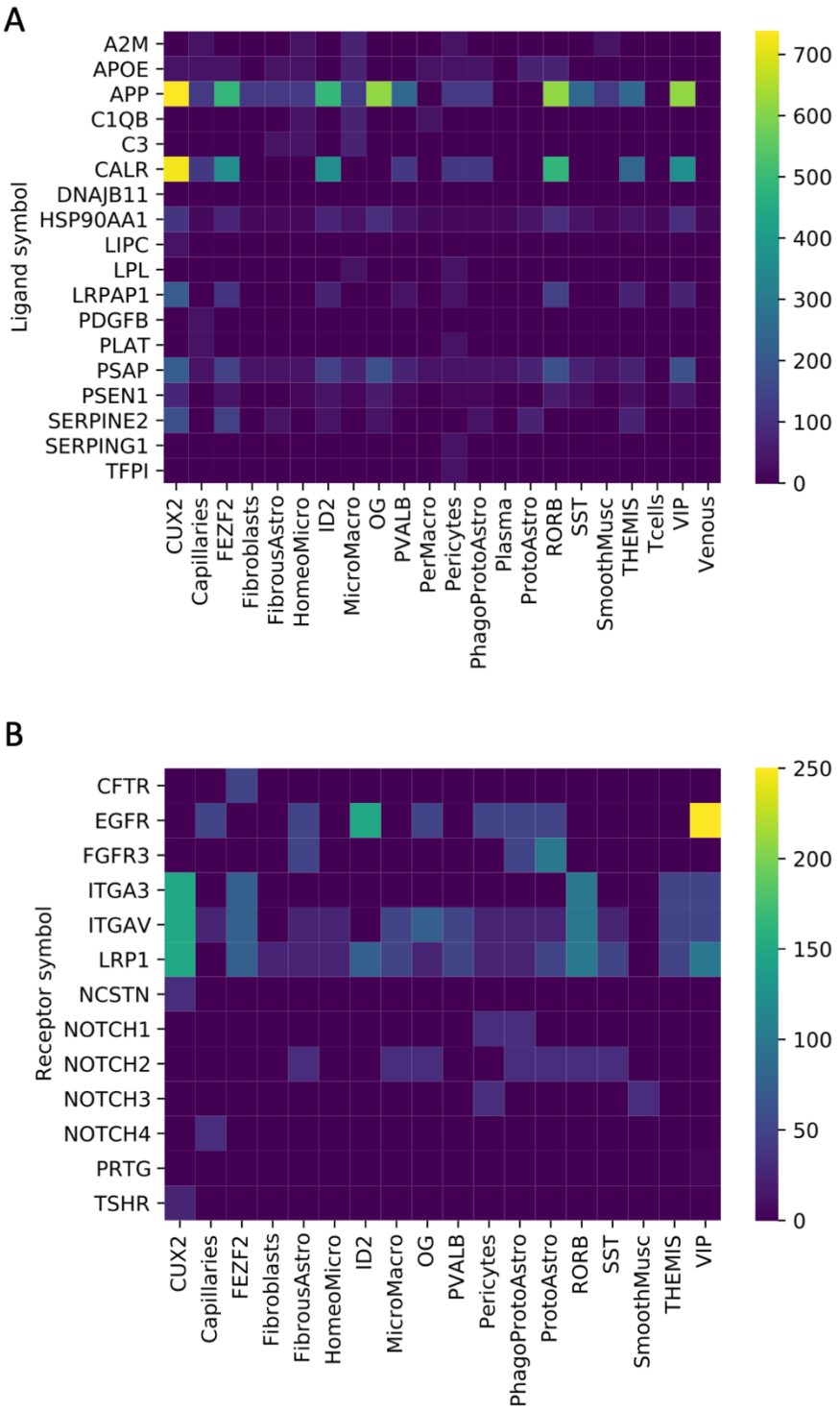


Figure 4.45: Heatmaps of cell types signalling in MS-specific LR interactions where the function of ligands is associated with protein folding. (A) APP and CALR, the most highly expressed ligands, are highly expressed specifically in neuronal cell types. This pattern is observable, albeit with lower expression, in PSAP, LRPAP1 and HSP90AA1. (B) EGFR was the most highly expressed receptor, particularly in VIP cells. ITGA3, ITGAV and LRP1 had similar expression levels in excitatory neurons, CUX2, RORB, FEZF2 and THEMIS. It was also expressed in inhibitory VIP neurons, albeit not uniformly.

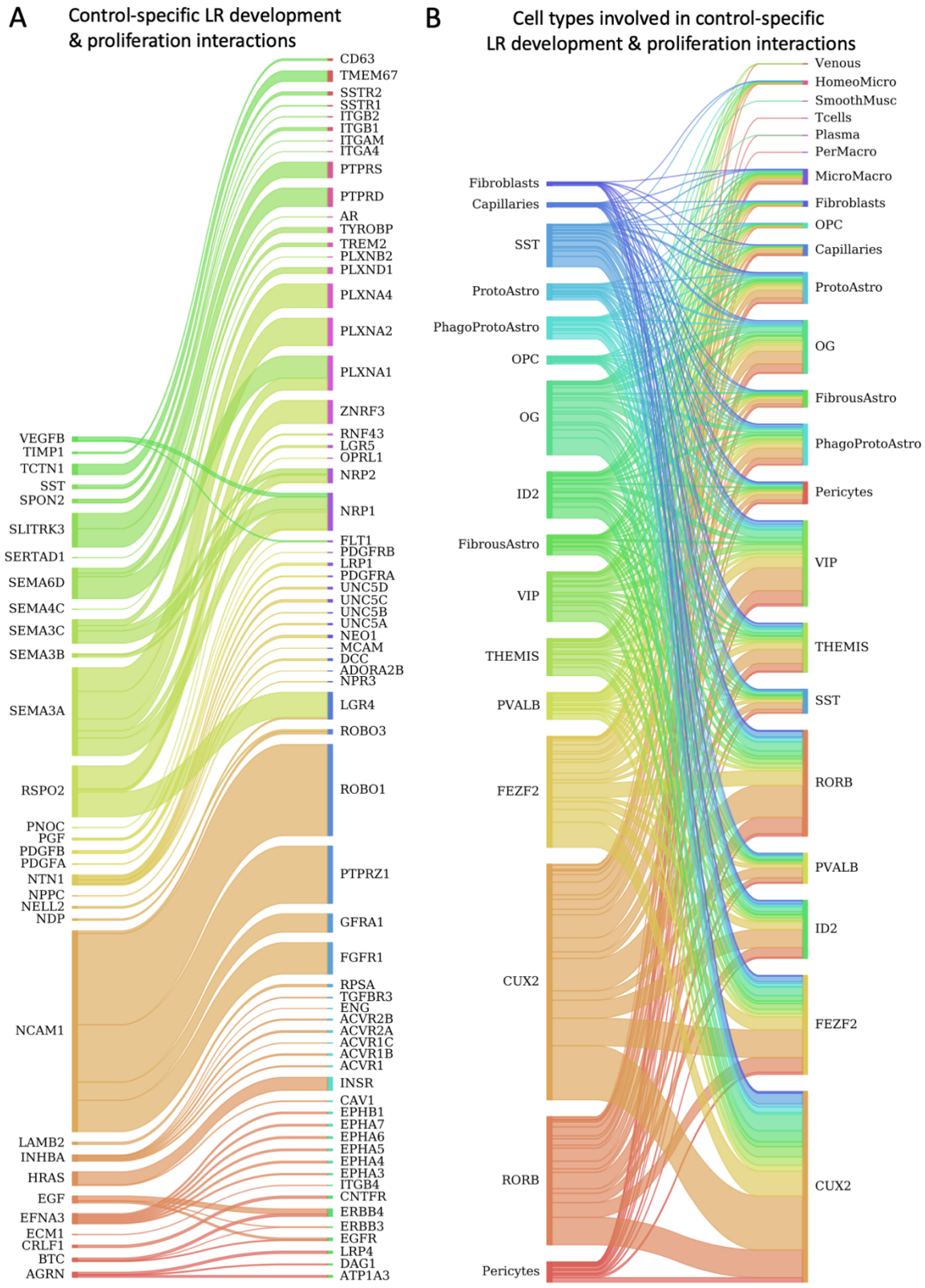


Figure 4.46: MS-specific LR interactions related to development, proliferation and neurite outgrowth. (A) The most highly expressed ligands in this group are NCAM1 and SEMA3A. A large number of receptors were expressed, among which NCAM1-communicating ROBO1 and PTPRZ1 were the most highly expressed. (B) Neurons were the cell types expressing both ligands and receptors most highly.

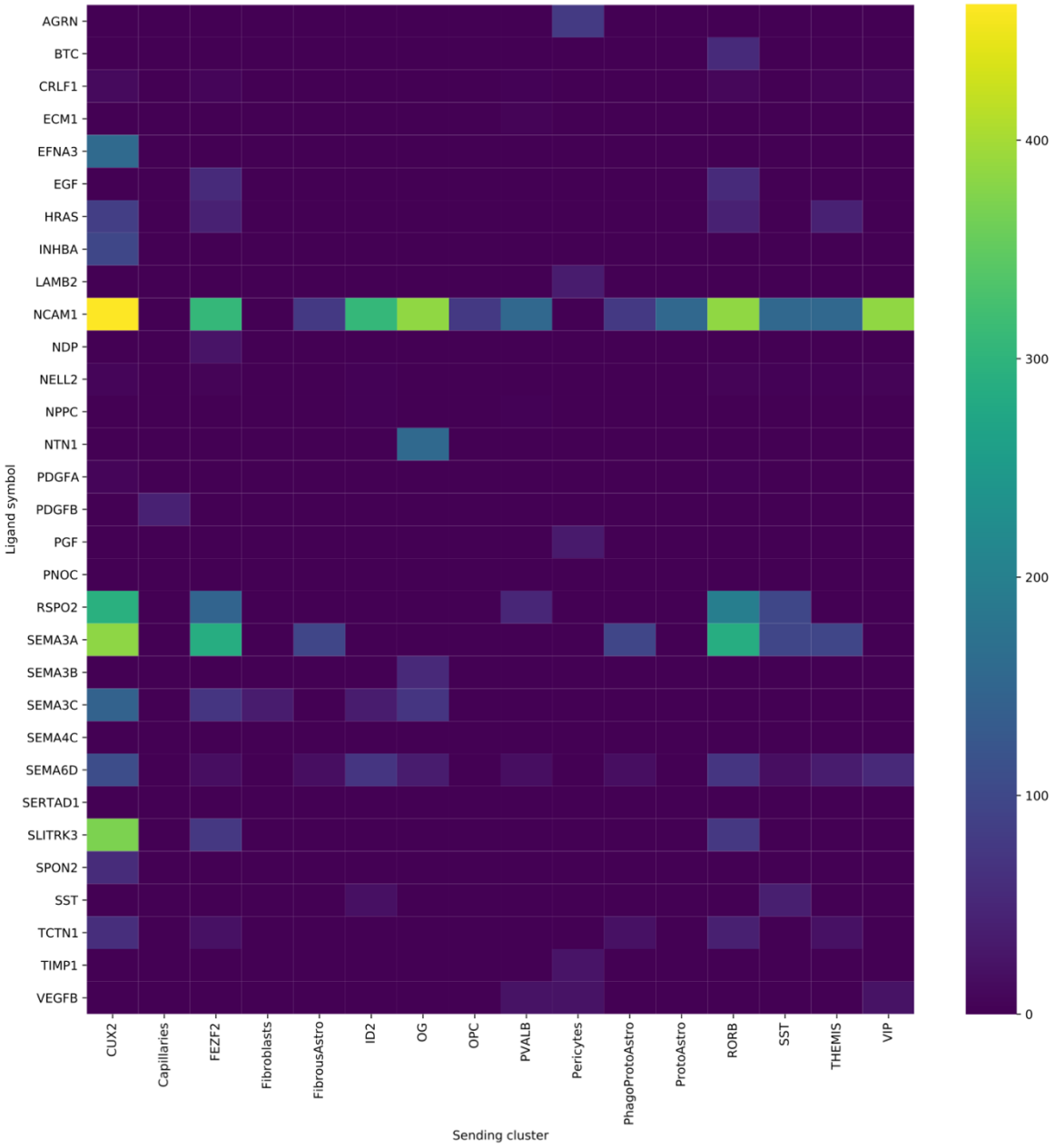


Figure 4.47: Heatmap of cell types expressing ligands in MS-specific LR interactions where the function of ligands is associated with development, proliferation or neurite outgrowth. The most highly expressed ligand, NCAM1 was most expressed in CUX2 neurons, followed by OGs, RORB neurons, VIP neurons and FEZF2 neurons. CUX2 neurons were the highest expressing cells, including SEMA3A, SLITK3, RSPO2, EFNA3, SEMA3C and SEMA6D. Some of these ligand patterns were shared with RORB and FEZF2 neurons.

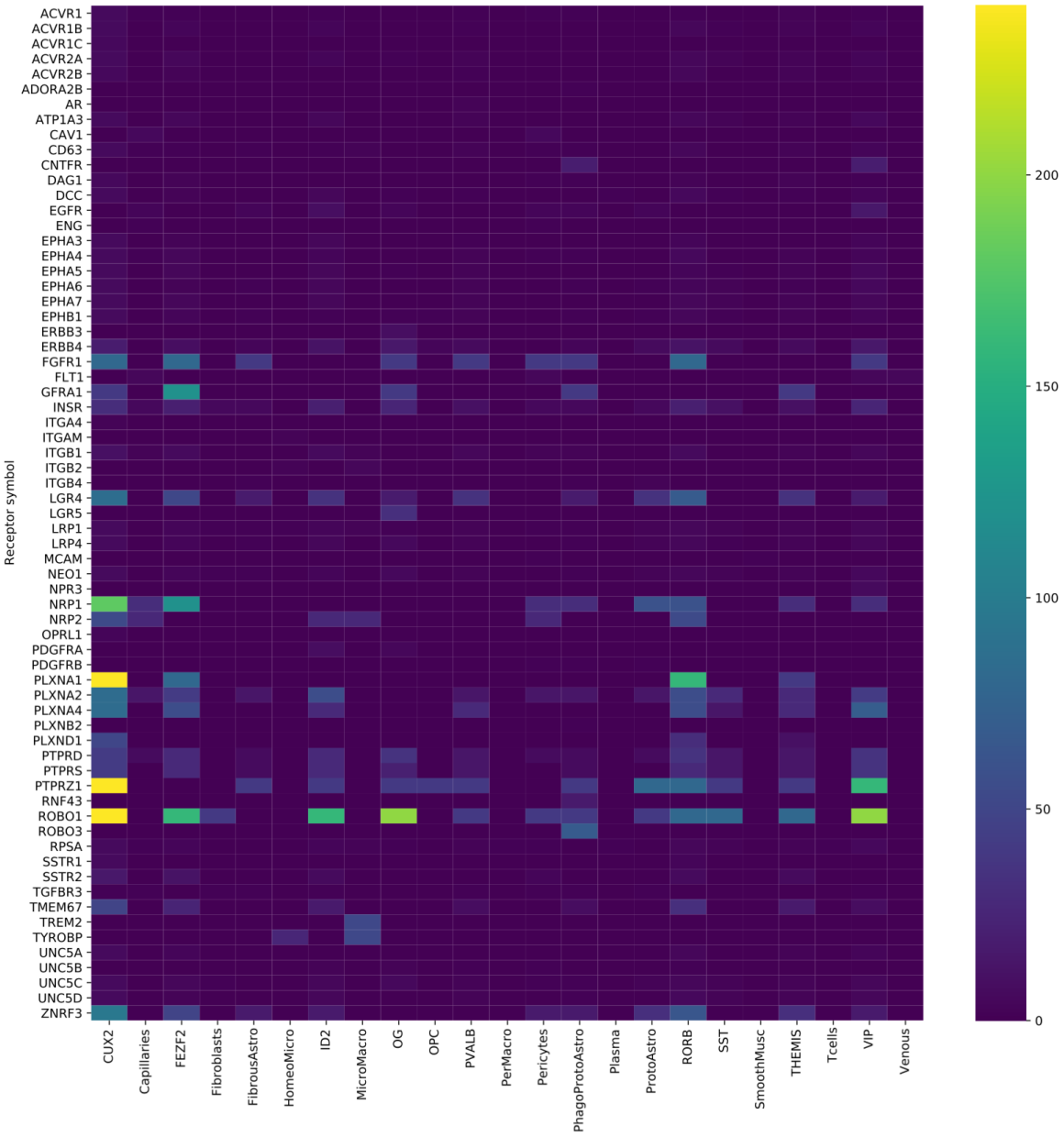


Figure 4.48: Heatmap of cell types expressing receptors in MS-specific LR interactions where the function of ligands is associated with development, proliferation or neurite outgrowth. The most highly expressed receptor is ROBO1, which is most expressed in neuronal cell types but also present in other cell types at lower levels. The neuronal cell type expressing most receptors is CUX2, followed by FEZF2 and RORB.

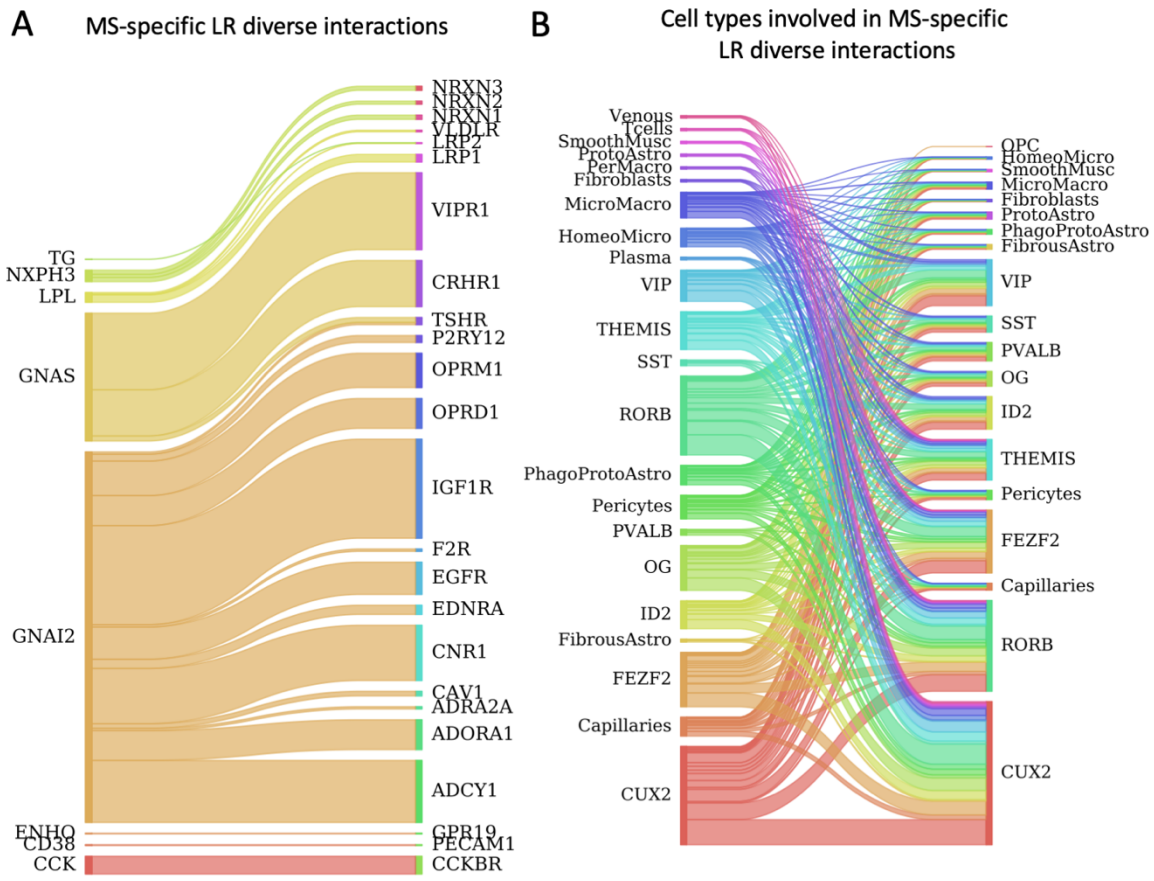


Figure 4.49: MS-specific LR interactions in diverse categories. (A) GNAS and GNAI2 are the two dominantly expressed ligands, particularly GNAI2 communicates with a large number of receptors. (B) Neurons are the most highly expressing cell type for both ligands and receptors, followed by oligodendrocytes.

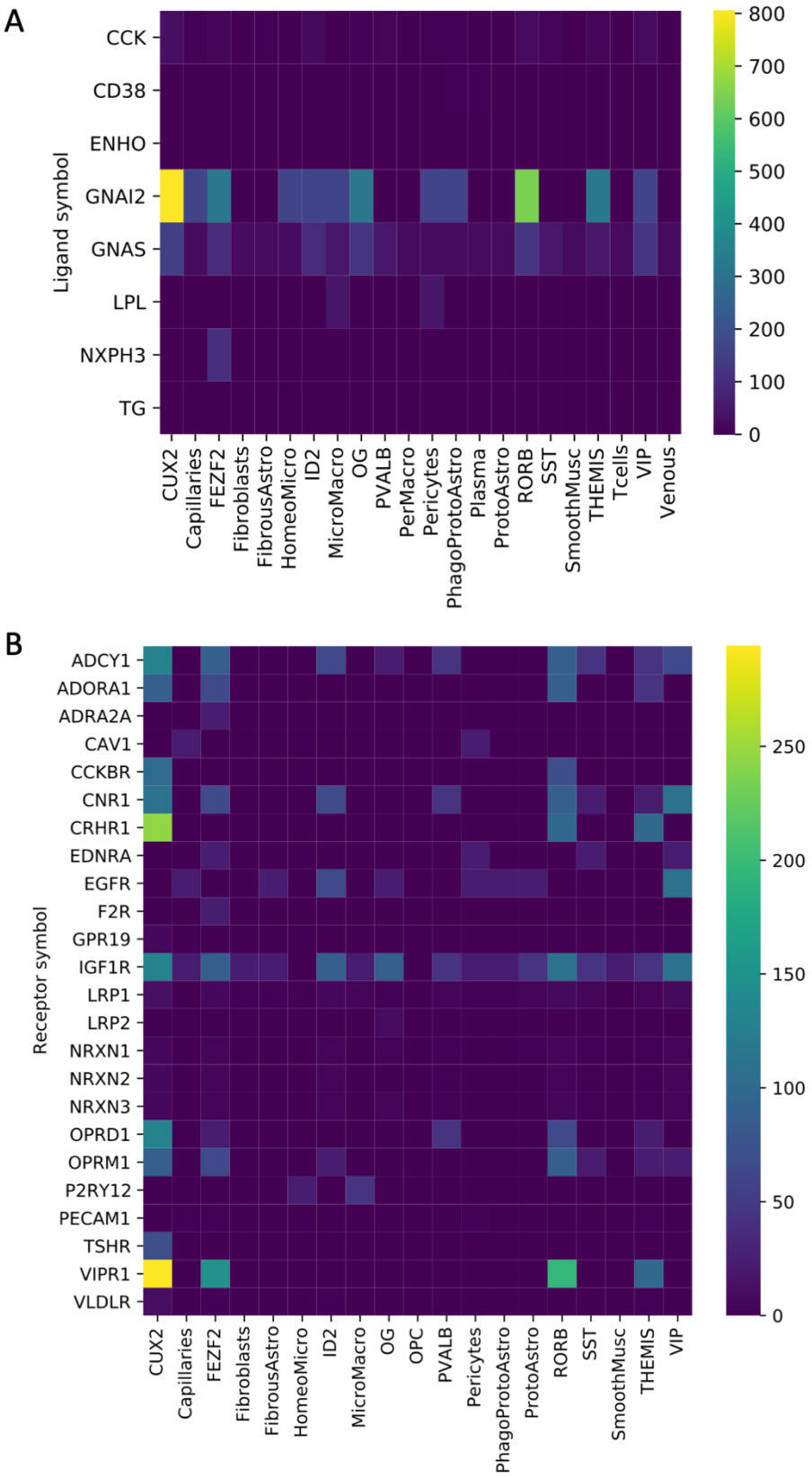


Figure 4.50: Heatmaps of cell types signalling in remaining MS-specific LR interactions. (A) GNAI2 was the most highly expressed ligand. CUX2 followed by RORB neurons were the highest expressing cells, although there was no apparent cell type specificity in expression. (B) The most highly expressed receptor was VIPR1 and its

expression was highly specific to excitatory neurons. CUX2 neurons expressed receptors in this category most highly and had similar receptors expressed to RORB and FEZF2 neurons.

4.3 Discussion

The snRNAseq meta-analysis identifies several key compositional and gene expression differences when comparing MS to controls. Clusters such as MicroMacro CPMhi LPLhi, OG OPALINhi FRYhi, fibroblasts and fibrous astrocytes had large log fold changes in abundance but subtle gene expression changes. This may reflect the fact that the clustering assigns not only cell identity, but also functional phenotypes associated with disease. Two CUX2 subclusters were identified which had the largest DA negative logFC in SPMS when compared to controls across all clusters. In light of previous evidence reporting CUX2 neurons to be selectively vulnerable in MS[130], these clusters may indicate a decrease in this neuronal population.

Overall, the cellular landscape of MS is composed of cell types which are responding to the physiological stressors of the disease by altering abundance and other cell types which maintain abundance but alter gene expression. While it is intuitive to think of an increase in abundance as relating to increased proliferation and migration, and a decrease in abundance as relating to cell death, these alterations may take place also due to sufficiently substantial changes in gene expression programmes that then become classified as a distinct cell type (e.g.: activation of macrophages, microglia and astrocytes). This reflects an increasingly appreciated conundrum in the single-cell field – that of distinguished cell type from cell state. Thus, a more cell subset-agnostic approach was used to interrogate functionality – chiefly WGCNA which profiles gene

expression modules associated with disease. In addition, metabolite and cell-cell interaction analyses were also employed to further probe functional processes.

4.3.1 Compositional and transcriptional evidence of glial scar formation in SPMS

The most significant changes to the compositional landscape of SPMS included an increased representation of fibrous astrocytes, MicroMacro_CPMhi_LPLhi, and fibroblasts. Jointly, this revealed a profile associated with a well-established CNS process: glial and fibrotic scar formation. In the context of traumatic CNS injury, glial and fibrotic scar formation is characterised by an outer rim lined with reactive astrocytes, while the central regions are populated by fibroblasts, activated macrophages and OPCs[173] (**Fig 4.51A**) – a remarkably similar cellular composition to that which is observable in MS lesions.

The gene expression profile of these over-represented clusters is also consistent with this hypothesis. Gene expression findings from the fibrous astrocyte population indicated these to be activated, with this subtype being the one expressing the highest level of activation markers *CD44*, *FOS* and *VIM*. These data are also in line with previous reports that fibrous astrocytes populate MS WM lesions more so than protoplasmic astrocytes, which tend to populate GM lesions[130].

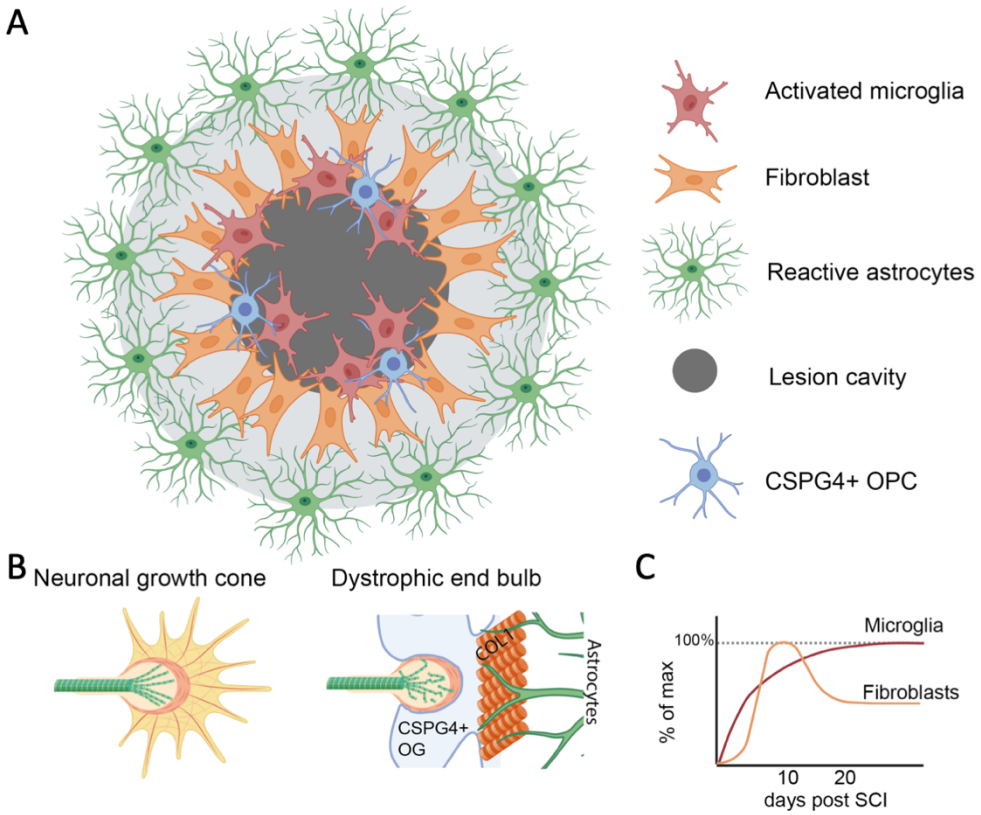


Figure 4.51: Cellular architecture and events in CNS injury (A) Schematic representation of CNS scar architecture in response to CNS injury- the outer rim of lesions is lined by interlocking GFAP+ astrocytes isolating the area of damage and CSPG4+ oligodendrocytes, the inner rim is populated by fibroblasts synthesising collagen which provides structure and insulation for the damaged area. The inner lesion cavity is populated by activated microglia/macrophages phagocytosing debris. (B) Schematic of a normal neuronal growth cone with orderly microtubule radiating structure (green) compared to dystrophic end bulb, where microtubule organisation is lost and neuron forms synaptoids with CSPG4+ OGs. Astrocytes and collagen form physical barriers preventing neurite outgrowth. (C) Cellular composition of SCIs over time – fibroblasts are rapidly recruited and peak in population in the early stages of damage, while the microglial population continues to increase and plateaus in ~20 days.

The gene expression pattern of the macrophages/microglia also corresponded to an activated phenotype. The differentially abundant macrophage/microglia cell type marked by high expression of *CPM* and *LPL* is consistent with a previously characterised transcriptomic profile of foam cells, macrophages which localize to fatty deposits on blood vessel walls where they ingest low-density lipoproteins and become laden with lipids. The expression of *LPL* and lipid content was shown to be a requirement for *CPM* expression in these cell types, since in the absence of lipid *CPM* expression was repressed *in vitro*[202]. Thus, given previous reports likening myelin-laden macrophages to foam cells[203] and the stark compositional contrast of this cell type almost exclusively represented in MS and not in controls, it is plausible that these cells are activated macrophages/microglia which are phagocytosing lipid-rich myelin debris at the site of MS lesions.

This is further supported by the association of the disease-correlating gene expression module including the KEGG term 'phagosome'. The association of this module with multiple iron ion transport pathways may further indicate that the phagocytic activity is occurring in the context of myelin debris/OG apoptosis since OGs have a central role in iron storage, and thus uptake of this debris would require iron transport pathways to be activated in phagocytosing cells.

Fibroblasts were also more abundant in cases than controls and were marked by *COL1A1* expression. In CNS injury, they have an indispensable role in response to injury by producing collagen which, along with other extracellular matrix proteins, provides

structure to the injured area and halts the influx of blood cells and toxins into the CNS by creating a physical barrier to contain the injured area. While this is a necessary step for recovery to be viable, it also limits repair in the long term, since neuronal axons cannot grow past these extracellular matrix (ECM) proteins[173, 204]. Deposition of collagen surrounding blood vessels and an increase in fibrotic gene expression in lesions has been reported in MS[205-207] and EAE[208, 209]. Therefore, these findings are again consistent with previous observations in MS and with known traits of scar formation in the CNS in response to injury.

Notably, however, while all other cell types involved in scar formation are differentially abundant in MS, OPCs were an exception. NG2/CSPG4$^+$ OPCs present at the borders of sites of damage serve as attractants for dystrophic neurons, stabilising them and forming synaptoids – synapse-like structures between dystrophic neurons and NG2+ OPCs which ultimately contribute to preventing the reestablishment of the neuronal network in that area[204] (**Fig 4.51B**). This has a neuroprotective effect, since it prevents neuronal processes from reaching areas rich in neurotoxic and neuroinflammatory factors present at sites of damage, but comes at the cost of long-term regeneration. The lack of DA in OPCs may indicate either a dysfunctional aspect of scar formation in the CNS, or may indicate that OPCs are recruited via migration rather than local proliferation. We explore this further in the spatial analysis in chapter 5.

While the exploration of scar formation as a whole in the context of neurodegeneration has been limited, its involvement in MS has been explored in the EAE

animal model in a compelling study. Extensive fibrotic scar formation following immune cell infiltration was observed and associated with proliferation and migration of fibroblasts. Fibrotic scar presence was demonstrably present several months following initial injury. Remarkably, inhibition of CNS fibroblasts decreased scarring and disability scores in the later disease stages[209].

The large number of LR interactions which relate to coagulation and healing is also surprising in this dataset given the analysis of SPMS tissue (and thus lesion chronicity) and contributes compellingly to fibrotic scar formation, particularly since tissue fibrosis is thought to result from a dysregulated wound healing response in response to continuous local injury[210]. The identification of compositional and transcriptional changes consistent with scar formation in MS in humans strengthens the arguments in favour of therapeutic interventions which target these mechanisms. Notably, these observations were made in the context of late MS, which poses a substantial therapeutic challenge.

Supporting the CNS' capacity for recovery should scar formation be regulated, is the large number of LR interactions identified which relate to neurodevelopment, including axonal guidance and development of neurite processes. This indicates the presence of a recovery response and plasticity potential in neurons even in the neurotoxic biochemical context of late MS which would directly enable lesion recovery in the absence/reduction of fibrotic scars.

While targeting fibroblast proliferation proved effective in EAE, this approach would require further investigation for humans. In spite of the fibroproliferative role of the coagulation cascade[210] and the reported proliferation of fibroblasts in EAE scar formation process[209], we have limited evidence of fibroblast proliferation. This may be partly due to the overall small number of fibroblasts captured in snRNAseq but also in the context of the entire endothelial compartment, we have identified a decrease in mitosis and proliferation in MS. Upregulated gene *CEMIP* in this compartment has a role in cell migration and may indicate that the increased abundance of fibroblasts results from migration rather than proliferation.

The reduced proliferation of cells in the endothelial compartment in SPMS contrasting with EAE may relate to the chronic and long-term nature of the lesion development in SPMS. The green yellow module identified in WGCNA analysis of the endothelial compartment relates to ADP biosynthetic processes, phosphorylation, kinase activity, thiamine metabolism, biosynthesis of cofactors and purine metabolism. Thiamine and purine metabolism have a close relationship with energetics[211-213] and energy stress, and may indicate a level of energetic stress either due to the cytotoxic environment of MS lesions or, alternatively, cellular exhaustion following extended periods of proliferation in response to insult. In order to ascertain whether this indicates that the differential fibroblast representation in MS compared to controls results from fibroblast migration or proliferation, a larger data set and further assays would be required.

4.3.1.1 Regulation of fibrotic scar formation may be affected by neuronal loss

The overwhelming contribution of FGF pathways to control-specific LR interactions and its implications in fibrotic scar formation must be considered. Evidence in the context of epithelial damage and lung damage indicate that FGF can reduce fibrotic scarring[214, 215], and LR interaction analysis in our data indicate that *FGF10* and *FGF17* signalling with FGF receptors are among LR pairs absent in MS; this may result from the fact that selectively vulnerable CUX2 neurons are significant contributors to these interactions. This may suggest increasing accumulation of scar formation with MS advancement, where with this neuronal subset increasingly damaged and/or depleted, the decreased *FGF1*0 and *FGF17* signalling may reduce the regulation of fibrotic scar formation. This would suggest a scenario where not only would scarring accumulate over time, but potentially increase in rate - contributing to limited lesion recovery and increasing disability.

4.3.1.2 NOTCH regulation in late MS

The presence of cell adhesion proteins (CAMs) such as integrins and NOTCH in MS-specific LR interactions in late MS could be valuable when considering therapeutic intervention of fibrotic scar formation. These molecules have a well-established role in immune response, managing leukocyte migration to inflammatory foci and mediating the activation of endothelial cells, but have also been reported to be associated with fibrosis in other organs [216-222].

Evidence of the involvement of CAMs in MS already exists - genetic variants of CAMs have been associated with MS susceptibility[223] and alpha-4 integrin antagonist is used as a treatment for RRMS[224]. In the context of the evidence in this work implicating glial and fibrotic scar formation in late MS and the expression of CAMs when immune cell infiltration is no longer a prominent lesion feature could support an alternative role for these molecules in late disease. This would make the role of CAMs in the MS trajectory two-fold: firstly, in inflammatory response in early disease promoting damage, and secondly in promoting fibrotic scar formation which increasingly hampers lesion regeneration in late disease. To this end, a study trialling the use of natalizumab in SPMS in a phase 3 clinical trial found that while the treatment did not reduce progression on the primary multicomponent disability endpoint, it reduced the progression on the upper-limb component[225].

NOTCH signalling has been suggested as a valuable target both in the context of chronic inflammatory diseases[226] as well as fibrosis[216]. While its complex relationships with core aspects of MS neuropathology such as remyelination, oligodendrocyte maturation and neurogenesis would need to be taken into consideration[227-231], preliminary animal work with this target has been promising: treating EAE with a NOTCH3 inhibitor has been shown to attenuate EAE[232] and NOTCH1 inhibition has been shown to accelerate remyelination *in vivo* in models of acute demyelination[233]. In this transcriptomic analysis we present evidence of a physiological mechanism in late MS

that could benefit from modulating CAMs in the NOTCH family, potentially as a way of reducing fibrotic scar formation in MS.

4.3.2 Evidence of OG and neuronal dysfunction consistent with fundamental neurodegenerative processes

4.3.2.1 Oxidative and energetic stress in OGs

DGE analysis of OG subclusters identified *MIF* and *SLC22A17* as increased in cases hinted at pathological processes involving oxidoreductase activity and ion transport. This was confirmed with the WGCNA analysis, where cellular compartment GO terms respiratory chain complex, respirasome, mitochondrial respiratory chain complex I and NADPH dehydrogenase complex were highlighted. Molecular function pathways included heat shock protein binding and ATPase regulator activity. Jointly, these terms establish a cellular response consistent with oxidative stress.

The close relationship between metal ion accumulation, oxidative processes and cellular energetics has been explored to some extent in MS. It is well established that oxidative processes are a feature of MS lesions, and it is hypothesised that this at least in part relates to OG death which releases the stored iron, thus inducing oxidative stress[234]. Our analysis reveals transcriptomic changes within surviving OGs in response to this.

Adding to this line of inquiry, the metabolites which have the highest log fold changes and absolute changes when compared to controls, tyrosine, fumerate, acetylCoA and fatty acid play a role in the tricarboxylic acid (TCA) cycle, which has a

central role in cellular energetics and may reflect the energetic aspect of oxidative stress identified in OGs.

4.3.2.2 Changes to protein folding in OGs

WGCNA analysis highlighted that the disease-correlated gene module in oligodendrocytes is also associated with multiple pathways relating to protein folding – protein folding, protein refolding, unfolded protein binding. Multiple diseases associated with this module feature protein misfolding (PrD), protein accumulation (HD) or both protein misfolding and protein accumulation (PD).

Whether these changes to protein folding in OGs relate to the oxidative processes is unclear. Protein folding and oxidoreductase activity have a two-way relationship, whereby oxidative protein folding requires redox processes to establish accurate and stable folding for certain features, but on the other hand protein carbonylation due to oxidation can also affect protein conformation, as discussed in Chapter 3.

Evidence of changes in protein folding were also present in the metabolite analysis, where metabolite accumulation estimation indicates MS has more accumulated (Glc)3(GlcNac)2(Man)9(Asn)1 ($p=0.02$). Glucosylated oligomannose N-lined oligosaccharides ($Glc_xMan_9GlcNAc_2$ where $x=1-3$) are involved in the early stages of glycoprotein biosynthesis and folding[235].

4.3.2.3 Functional evidence of selective vulnerability of CUX2 neurons in response to oxidative stress

Evidence of oxidative stress and energetic changes was also present in neurons. DGE analysis of the total neuronal compartment highlighted genes upregulated in cases when compared to controls that were implicated in oxidoreductase activity, monooxygenase activity, icosanoid receptor binding, amino acid binding, NADP binding, FMN binding and rRNA binding, amongst others.

Analogous analysis of higher resolution clusters revealed that the differential expression of these genes was likely driven by the CUX2 neuronal population. This cell type-specific response provides functional evidence for a previous report that CUX2 is selectively vulnerable neuronal type in MS[130]. Our meta-analysis further pinpointed this vulnerability to two specific CUX2 clusters that were less represented in cases: CUX2 PALMD+ FREM3+ (p=0.0002) and CUX2 RORBhi TXKhi (p=0.0108). *PIK3R3* was also upregulated in this cell subtype; in neurons PI3K signalling can transduce signals from cell surface receptors to specific pathways, is essential for synapse and dendritic spine development and for enduring forms of synaptic plasticity[236]. *CYSLTR2*, was additionally upregulated and is a potent pro-inflammatory and immune modulating lipid mediator which has been observed to be upregulated following an episode of ischemia. Antagonists to this receptor can have a neuroprotective effect[237, 238].

4.3.2.4 Ribosomal effects of MS in neurons

The turquoise and salmon disease-correlated gene expression modules in neurons had several overlaps in pathways associated with their genes as well as a significant overlap in the patients with high correlations with each module. The pathways included ribosomal subunits, nuclear transcribed mRNA catabolic process, co-translation, protein targeting to membrane and rRNA binding. Both also included pathways related to viral infection, such as COVID-19, herpes simplex infection and viral translation.

The increase in translation associated with ribosomal function may be linked to the number of neurite outgrowths/neuronal plasticity. Presence of these processes is evidenced by the number of developmental LR pairs unique to MS, as well as the presence of SMAD signalling and positive regulation of aspartatic-type endopeptidase activity among pathways associated with the orange gene expression module. SMAD has key roles in development and growth factor signalling[239], while aspartate has a role regulating neuronal morphology and plasticity[240, 241]. Robust ribosomal function has been demonstrated to be necessary for growth and maintenance of dendrites *in vitro*[242]. Particularly due to the complex and compartmentalised morphology of neurons, they are capable of modifying their proteomes locally using mRNAs and ribosomes in dendrites and axons distal to the cell soma. This has a central role in synapse formation, axonal growth and synaptic plasticity[243].

Recent evidence has shown that this process is more dynamic than previously thought, and that likelihood of incorporation of novel ribosomal proteins into the ribosome machinery can be influenced by the physiological context of the neuron. Also

relevant to the context of MS with pronounced oxidative stress, ribosomal proteins are susceptible to oxidative stress[244], and following oxidative stress the incorporation of ribosomal proteins can be dynamically enhanced[245].

The viral mechanism most likely connected to ribosomal function in this context is the hijacking process of ribosomal transcriptional machinery which viri are capable of. Ribosomal proteins have been shown to encode dynamic and easily exchangeable ribosomal proteins which may be incorporated into ribosomes which could trigger long lasting effects[246]. Given the well-known link between certain viral infections such as EBV and MS incidence, it is interesting to speculate whether disease-associated changes in ribosomal function may be partly triggered by or linked to viral infection.

4.3.2.5 Transcriptomic changes consistent with presence of neuronal inclusions in MS

Several GO and KEGG terms associated with protein folding/accumulation were identified in the neuronal disease-correlated gene expression module – these included reticulum chaperone complex, protein tag, Parkinson disease and Alzheimer disease, heat shock protein binding and promyelocytic leukaemia (PML) body. Notably, PML body presence is involved in apoptotic death of neurons and has been observed in various contexts of cellular stress, such as interrupted blood supply[247] and familial frontotemporal dementia where it co-localised with intranuclear inclusions[248]. Their most common neuropathological context is polyglutamine conditions, such as

spinocerebellar ataxia type 1 and Huntington's disease[249-253] which are characterised by intranuclear proteins and inclusions.

The neuronal protein inclusions may have other effects such as nuclear pore complex damage, inefficient transcription and aberrant DNA/RNA maintenance [254]. In line with this, this module also shows upregulation of heat shock protein binding which could relate to response to DNA damage such as PML bodies and DNA transcription repressor activity.

The notably high expression of several genes associated with AD pathology such as *APP*, *LRP1*, and *PSEN1* specifically in neurons and most strongly in CUX2 neurons is intriguing. On one hand *APP* and *PSEN1* have been extensively studied in the context of AD mutations, where amyloid-beta peptides are the main component of amyloid plaques found in AD patient brains and upregulation of these pathways in dying neurons could indicate similarities with AD in pathogenesis. Several ligands associated with protein folding are expressed most abundantly in neurons, in particular in excitatory neurons – CUX2, FEZF2, RORB and THEMIS neurons. Of these, CUX2 neurons express the highest levels of these ligands and also exhibit evidence of oxidative stress and a decreasing population. Jointly this poses the possibility that protein misfolding and accumulation of protein aggregates could play some role in MS pathogenesis and could be processes fundamental to neurodegeneration.

On the other hand, the upregulation of these proteins could also be interpreted as being a physiological, protective response in MS (but which malfunctions in AD). *APP*

itself contributes to a number of functions such as neurite outgrowth, neuronal adhesion and axonogenesis. Amyloid beta peptides can also have a role in response to metal ion reducing activity and resulting oxidation: they are lipophilic metal chelators with metal-reducing activity, inhibiting metal-catalysed oxidation of lipoproteins – particularly in the context of an HD mouse model, reduction in APP has an iron accumulation effect[255]. Given the evidence of oxidative response in astrocytes, macrophages and OGs in this data set, this could be evidence of the appropriate neuronal response to this biochemical context.

The upregulation of *LRP1* is more intriguing - it is involved in processes such as intracellular signalling, lipid homeostasis and endocytosis/phagocytosis of apoptotic cells, but is also upregulated in AD brains due to its role in clearance of secreted amyloid precursor protein and beta-amyloid components of amyloid plaques in AD. The high expression of this receptor particularly in CUX2, RORB and VIP neurons, as well as FEZF2, ID2, SST and THEMIS neurons at lower levels is notable.

The joint evidence of oxidative stress, ion imbalance, PML body inclusions, upregulation of AD-associated proteins and the number of gene expression modules with similarities to AD, PD, HD and spinocerebellar ataxia type 1 firmly highlights the similarities of the SPMS neuropathological process in neurons with that of other neurodegenerative conditions. This suggests that a better understanding of cross-disease pathological mechanisms could be of translational value across these conditions.

Chapter 5: Spatial contextualisation and validation

5.1 Introduction

5.1.1 Spatial mapping of CNS omics data

SnRNAseq is enabling a renewed appreciation for cell type heterogeneity and outlining dynamic cell states in the context of health and disease in the human CNS, as well as enabling inference of cell-cell interactions. The study of spatial relationships of these highly dynamic cell types is the next technological and computational challenge currently being tackled in the field, as it aims to dissect how this wealth of information translates spatially.

Such mapping can be achieved to some extent through the use of a limited set of transcript markers through approaches based on fluorescence *in situ* hybridisation (FISH), or of protein markers detected through antibody-dependent IHC. Such methods have been key to outlining the histological organisation of the brain and were used in the identification of several organisational structures in the organ. Currently, the Allen Brain Atlas holds and curates extensive collections of IHC staining of multiple markers. However, despite the scale of such collections there are some limitations inherently associated with these types of data – chiefly the difficulty of quantifying markers and staining patterns – as well as the difficulty of integrating and co-interpreting data derived from different studies given inter-experimental variation.

More novel approaches do not rely on the selection of specific key markers, but allow the generation of spatially-resolved omics data via protocols that do not require tissue dissociation as the initial step, as needed in sc/snRNAseq. For example, spatially resolved transcriptomic methods with increasingly higher resolution have been developed, such as spatial transcriptomics (100 µm diameter/spot)[133], 10x Genomics Visium (50 µm diameter/spot) and slide-seq (10 µm diameter/spot)[13, 15]. These methods do not necessarily provide single-cell level resolution since they are based on a grid-like partitioning of tissue, and therefore the actual resolution achieved depends on the cellular density of the tissue being analysed. Several computational tools are being developed for deconvolution of these data with sc/snRNAseq data, increasing the functional resolution at the time of analysis[153, 256]. Contributions of spatially resolved transcriptomic methods are already highlighting previously unappreciated aspects of the CNS architecture: for instance, three distinct astrocyte types forming layers in the cerebral cortex have been identified using a spatially resolved single-cell approach[257].

Interestingly, new computational tools are being developed which aim to leverage rich archival IHC data, such as those available through the Allen Brain Atlas, for spatial mapping of sc/snRNAseq data, such as the deep learning tools Tangram and DEEPsc which take varied input files relating to spatial organisation and map sc/snRNAseq datasets[258, 259]. More recently an ambitious aim of the field is to map cells spatially exclusively using gene expression information[260].

Spatial protein analysis is also being further developed with mass spectrometry-based spatial proteomics, including in the context of the human CNS[132, 261]. While of lower resolution than spatial transcriptomics methods and conventional IHC approaches, this method can capture hundreds of proteins from small amounts of starting tissue, providing a useful validation method for spatial transcriptomics data since several hundred targets can be validated from a single neighbouring tissue section.

The aims of this Chapter were to apply spatially resolved transcriptomics and proteomics technologies to our SPMS and control CNS tissue, and to collectively consider and interpret the cellular classification and molecules characterisation derived from the different technological modalities applied across the Chapters in this thesis (**Fig 5.1**). These included the datasets reported in Chapter 3 (biochemical information from μFTIR microspectroscopy) and Chapter 4 (single-cell level transcriptomic information from snRNAseq). 10x Visium was used for spatial mapping of transcriptomic data, incorporating the snRNAseq data from Chapter 4 for cell type deconvolution. Jointly, this has generated high-resolution spatial omics information that has not been previously generated for MS. We also performed spatial proteomics, noting that regions of interest for spatial proteomics were carefully chosen to overlap with μFTIR cluster classification where possible. The common spatial features of each section were then used for data alignment and correlation across methods.

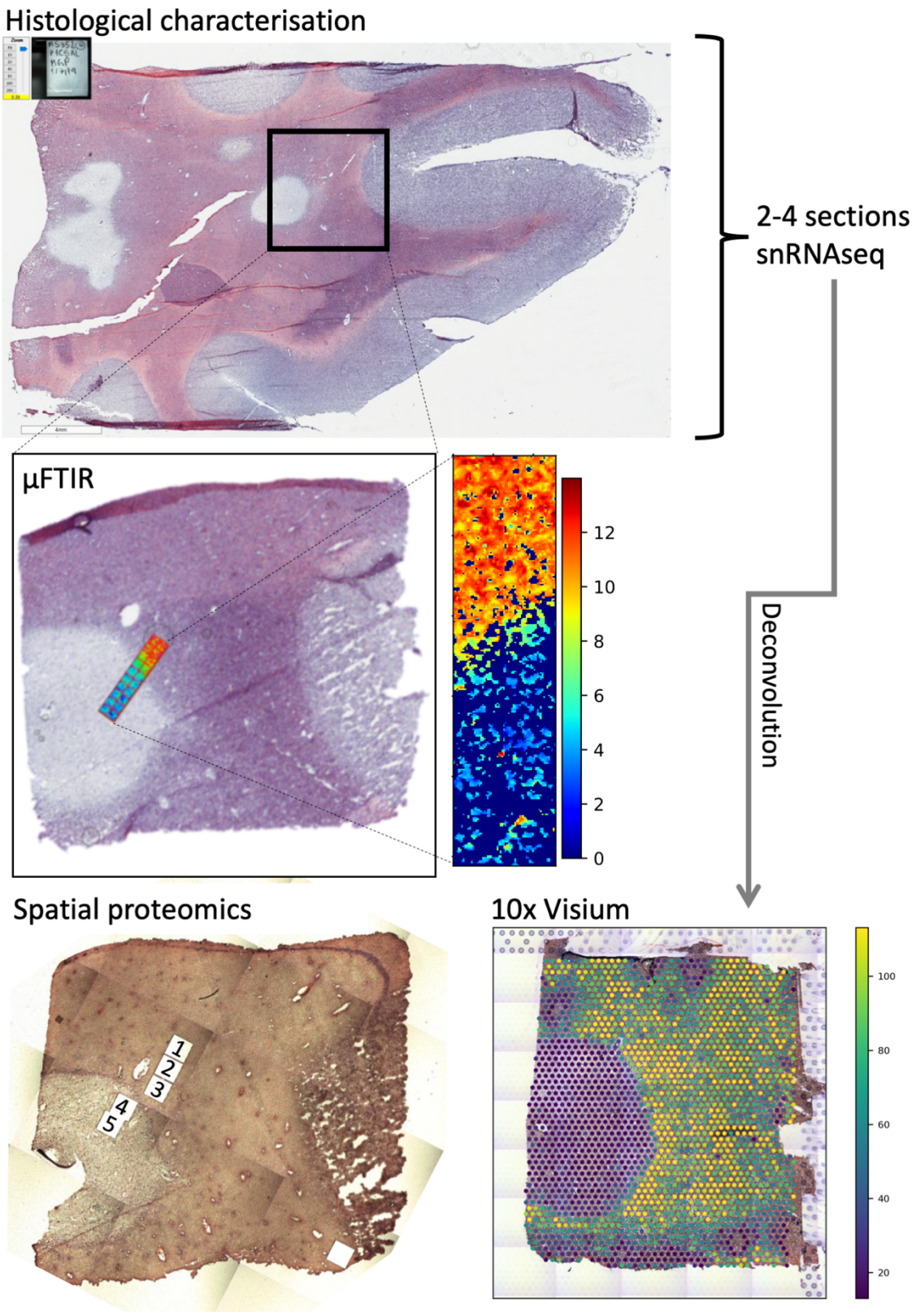


Figure 5.1: Summary of outputs from each technological modality applied in this body of work. In this chapter we correlate the findings across all experimental modalities by comparing clustering classifications and corresponding profiling metrics. Depicted are maps relating to protein content: histological characterisation of ORO staining, µFTIR map of lipid content and 10x Visium map of *MBP* expression.

5.2 Results

5.2.1 Spatial transcriptomics: mapping of differentially abundant cell types

Spatial transcriptomics was performed using the 10x Genomics Visium platform on 5 SPMS and 3 control brain samples and 2 SPMS and 2 control spinal cord samples. Data QC was performed as described previously (Methods section 2.4.4).

In order to maximise the resolution of the 10x Visium data, the annotated snRNAseq data were leveraged and used for deconvolution. This computationally estimated the proportion of each spot likely to be a particular cell type, giving for each value its calculated proportion as a ratio of 1. The top 3 cells with greatest proportion in a given Visium slide spot were recorded as the most statistically likely cells to be represented in that spot. Only cell types with a proportion > 15% were visualised.

Partly due to the rarity of SC post-mortem samples, human snRNAseq data are not publicly available and we did not endeavour to use our SC samples for snRNAseq as this would have required the processing of a substantial proportion of the limited tissue available. In light of this, we opted to deconvolute the SC Visium data with our brain snRNAseq data.

To begin to explore the deconvoluted Visium data, cell types identified as differentially abundant when comparing SPMS cases and controls in the snRNAseq analysis were spatially mapped for the brain samples. Fibroblasts were located within demyelination borders for MS352, MS402 and, to a lesser extent, MS425 which also

had fibroblasts populating the lesion border (Fig 5.2A). MS438 had sparsely distributed fibroblasts within the lesion and a number of them located at the lesion border. MS466 had some fibroblasts within the demyelinated lesion borders, although the higher expression regions were in the perilesional WM. Consistent with the snRNAseq analysis, MicroMacro CPMhi LPLhi populations were identified more frequently in cases than controls (Fig 5.2B). In MS352, MS402 and MS425 these cells were located specifically at the interface between demyelinated lesions and perilesional WM. This cell type was most abundant in MS438 and MS466, where they were localised within demyelinated lesions. In the case of MS438, a clear delineated border was marked without these cell types being detected in the GM, while the border with perilesional WM in MS466 was more diffuse. Fibrous astrocytes were identified in low quantities in most controls, with some detectable in the WM region of C064 (Fig 5.2C). In cases, a high density of fibrous astrocytes was located in demyelinated lesions in MS402, MS425 and, although with diffused borders, in MS352. The distribution was not as localised for MS438 and MS466; MS466 had high levels of this cell type identified in the perilesional WM, while MS438 had lower levels within the lesion and none detected in the GM.

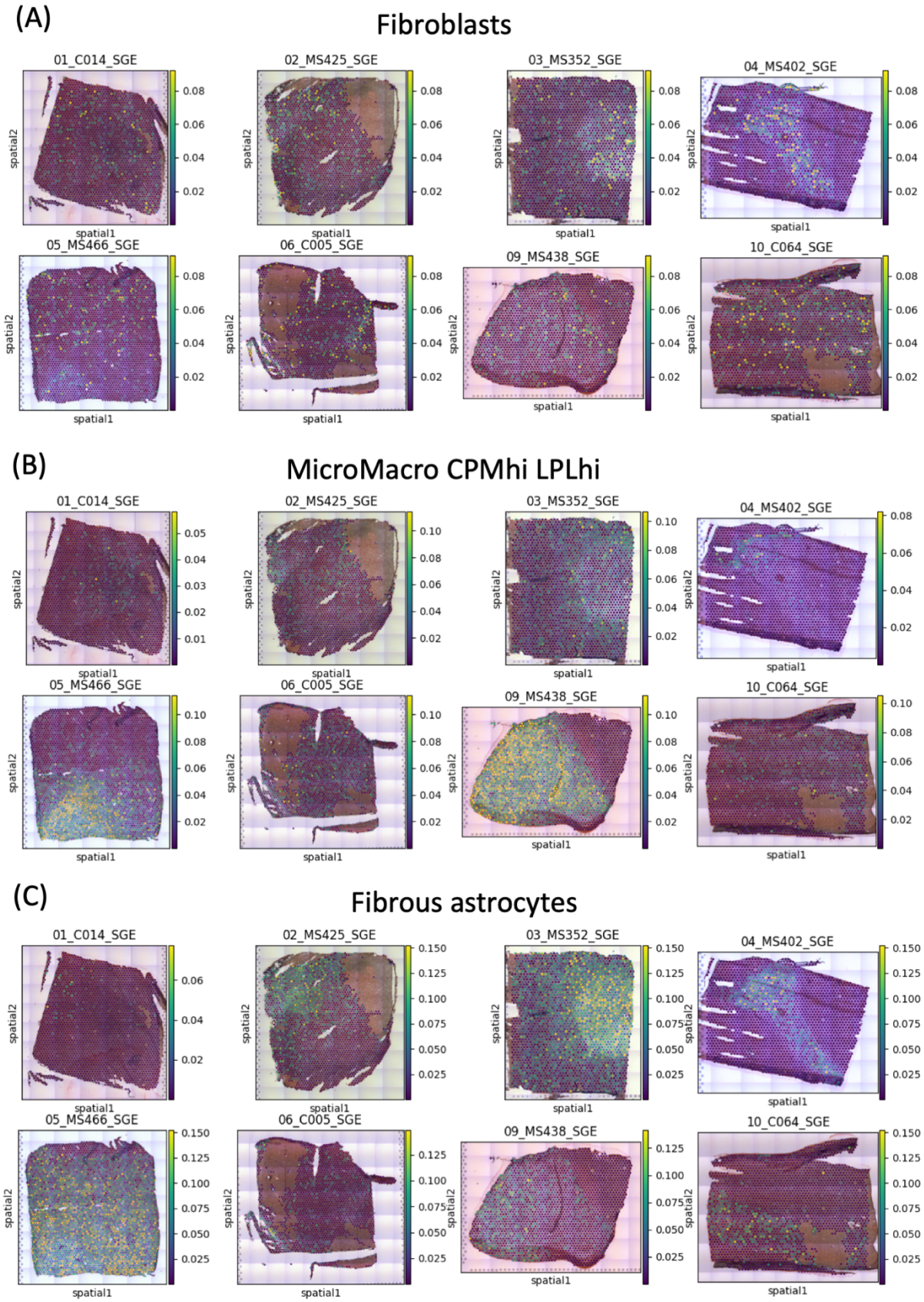


Figure 5.2: Spatial mapping in brain samples of cell types identified to be differentially abundant in snRNAseq analysis.

The same cell types were also mapped onto the SC samples (**Fig 5.3**) MicroMacro CPMhi LPLhi cells tended to localise towards the peripheral regions of the spinal cord, with decreasing presence towards the central GM area. These cells were sparse in the chronically demyelinated MS432. In the active MS376 section, these cells were present and distributed in higher density in the zones compositionally identified to contain an intermediate myelin content on either side of the median fissure. Fibrous astrocytes were identifiable in both controls, particularly toward the periphery of the section but were also present in the GM of C084. In MS432, this cell type was present in high density in the upper left area. In MS376 fibrous astrocytes appeared to be present in areas surrounding the microglial presence with only limited overlapping. Interestingly, fibroblast presence was more abundant in MS432 than in MS376; in the former, it had some overlap with the region populated by fibrous astrocytes. In controls, fibroblasts were mapped in higher density in more meningeal/endothelial tissue and in the GM.

5.2.2 Spatial transcriptomics: cell type co-localisation network analysis

Further to the mapping of cell types of interest, to quantitatively outline the spatial relationship of cell types, we used the deconvolution estimation to generate co-localisation networks. Co-localisation relationships between cell types were recorded in an adjacency matrix and plotted in the form of a network, where the size of each node is proportional to the number of occurrences of that cell type and the edge thickness, colour and length is determined by the number of times the cells represented by the connecting nodes were identified in the same spot.

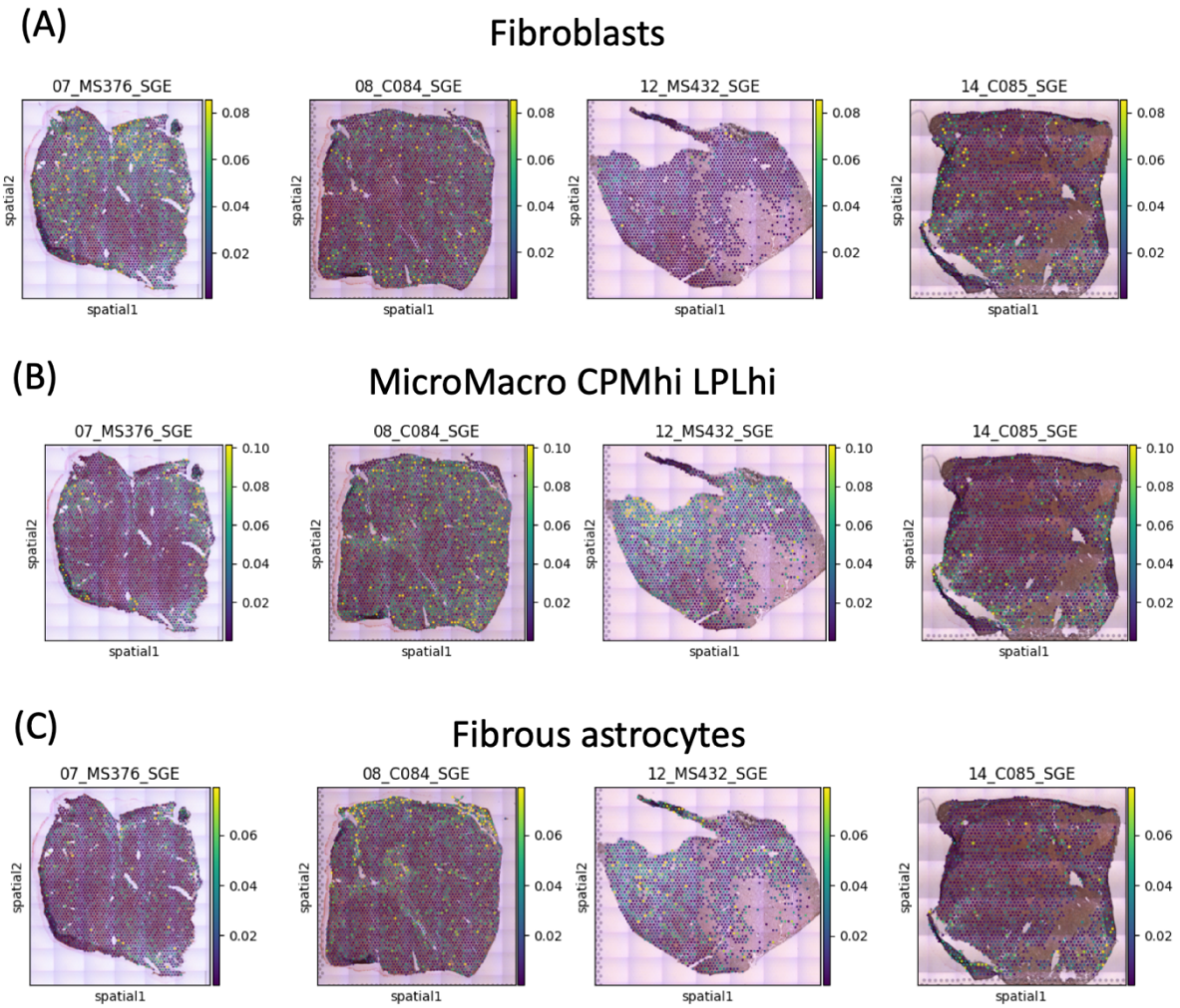


Figure 5.3: Spatial mapping in brain samples of cell types identified to be differentially abundant in snRNAseq analysis.

5.2.2.1 Brain tissue co-localisation networks: an overview

In order to better dissect the spatial relationship between cell types the deconvoluted spatial composition of each sample was individually mapped, the co-localisation network per sample was estimated, architectural composition features were compared to areas scanned with μFTIR. Where regions of homogeneously distributed sample composition were identified, these were classified into sections. The cell types identified populating each section were grouped with a dotted line in the network to compare the co-localisation relationships with the spatial organisation within the section.

An in-depth description of the networks for the different control and SPMS samples is provided in the subsequent sections, but overall the control and SPMS brain co-localisation networks constructed shared several common features that may thus be interpreted as being generic to the brain (Fig 5.4). Both groups included two main subnetworks each with cell co-localisation consistent with WM and GM tissues, respectively. In keeping with the GM architecture, a subnetwork of each group contained closely co-localising neurons, with VIP being the least abundant neuronal cell type, along with OPCs. However, clearly differing co-localisation relationships were also observed when comparing these neuronal compartments: in the control network, CUX2 neurons had the greatest number of strong connections (connecting with 7 other cell types), while in the SPMS network the frequency of CUX2 neurons was lower and these cells only strongly connected with RORB neurons. Moreover, the connectivity of OPCs was also substantially reduced.

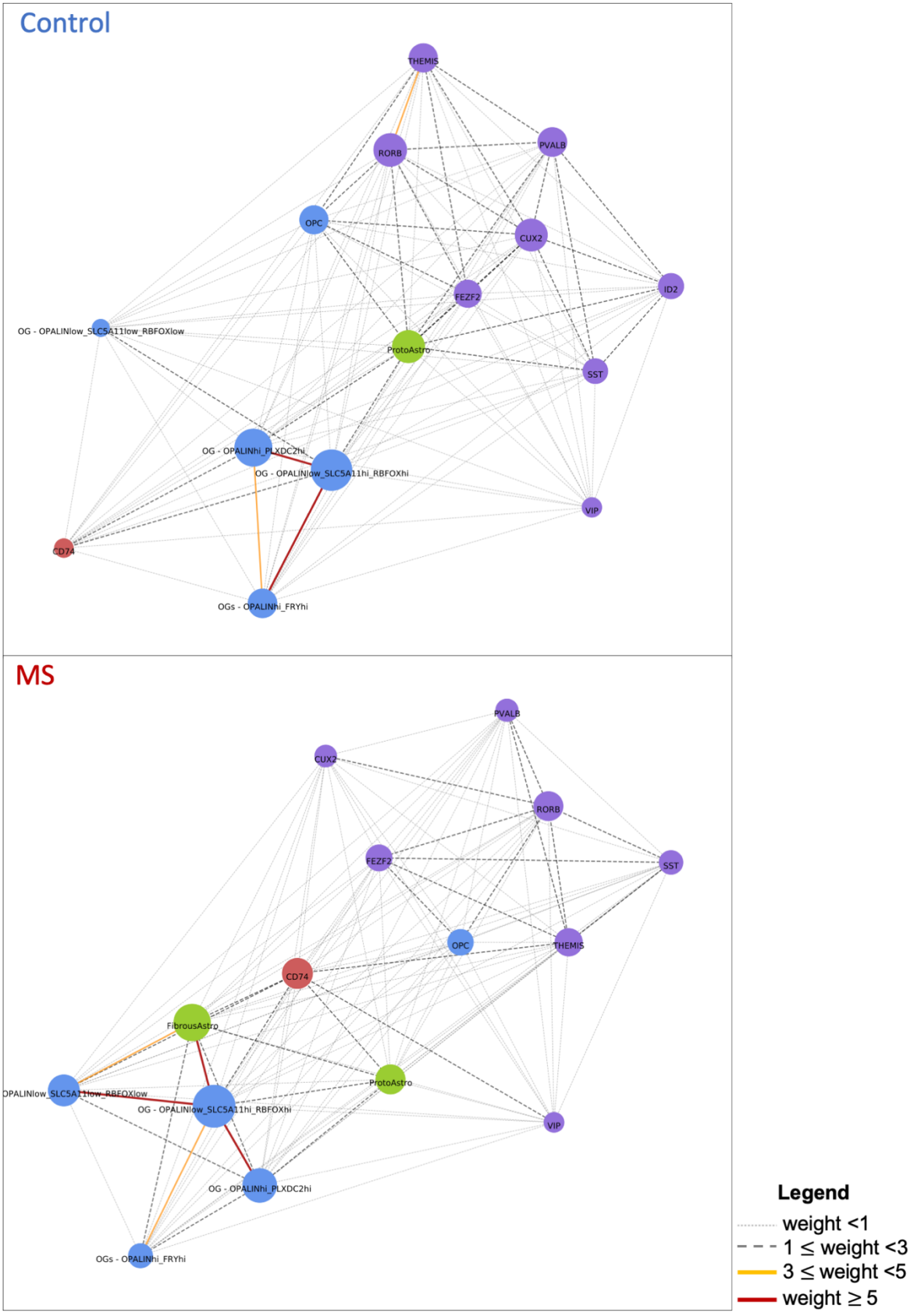


Figure 5.4: Cell type co-localisation analysis for all control and all MS brain samples. Each cell type is represented by a node and each co-occurrence of two cell types contributes to the edge connecting those nodes. For both cases and controls two

subnetworks are identifiable which contain cell types stereotypically present in distinct tissue types: The GM subnetwork includes neuronal cell types and OPCs, while the WM subnetwork includes OGs. Protoplasmic astrocytes were strongly connected to both control WM and GM subnetworks; CD74+ cells were strongly connected to both WM and GM subnetworks in MS. The MS network also had substantially increased representation of fibrous astrocytes.

Both WM subnetwork groups had a high representation of OG subtypes; OG OPALINlow SLC5A11low RBFOXlow was the least represented cluster in controls and more represented in cases, with a very strong co-localisation both with OG OPALINlowSLC5A11hi RBFOXhi and fibrous astrocytes. Fibrous astrocytes, protoplasmic astrocytes and CD74+ cell types were detected more frequently in cases and appeared to be mostly connected to cell types in the WM (consistent with the presence of white matter lesions in our samples).

5.2.2.2 Brain tissue co-localisation networks: controls

Sample C005 was almost entirely constituted of WM and therefore the deconvoluted data identified high contents of OGs and some CD74+ immune cells. Endothelial cells and fibrous astrocytes were present in lower proportions (**Fig 5.5**). The region of this sample with GM was not captured transcriptomically. C014 had both WM and GM within the sampled area, with identifiable neuronal subtypes organised to form cortical layers in the GM and OGs represented in high proportions across the WM (up to 30% of each spot) (**Fig 5.6**). OPCs were present both in WM and GM, as were astrocytes. The overall architecture of the ORO histological image and deconvoluted data were compatible. C064 was mostly GM with a strip of WM-like features, such as OGs. Neuronal cell types could be identified, although for this section the distinct cortical layers were not as discretely partitioned (**Fig 5.7**). This section had a high content of protoplasmic astrocyte and fibrous astrocytes which co-localised with OGs.

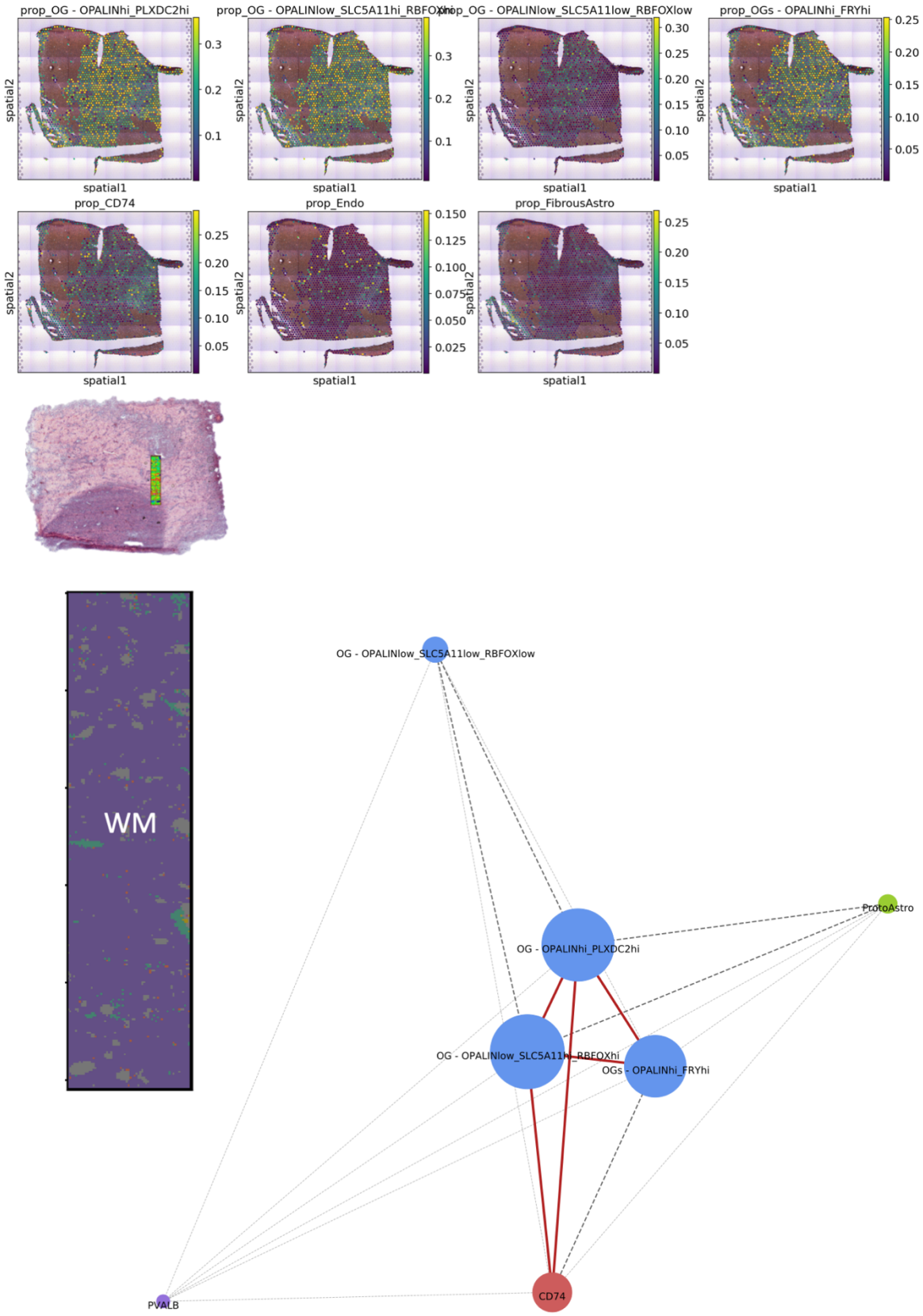


Figure 5.5: Spatial composition and architecture analysis of C005 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data

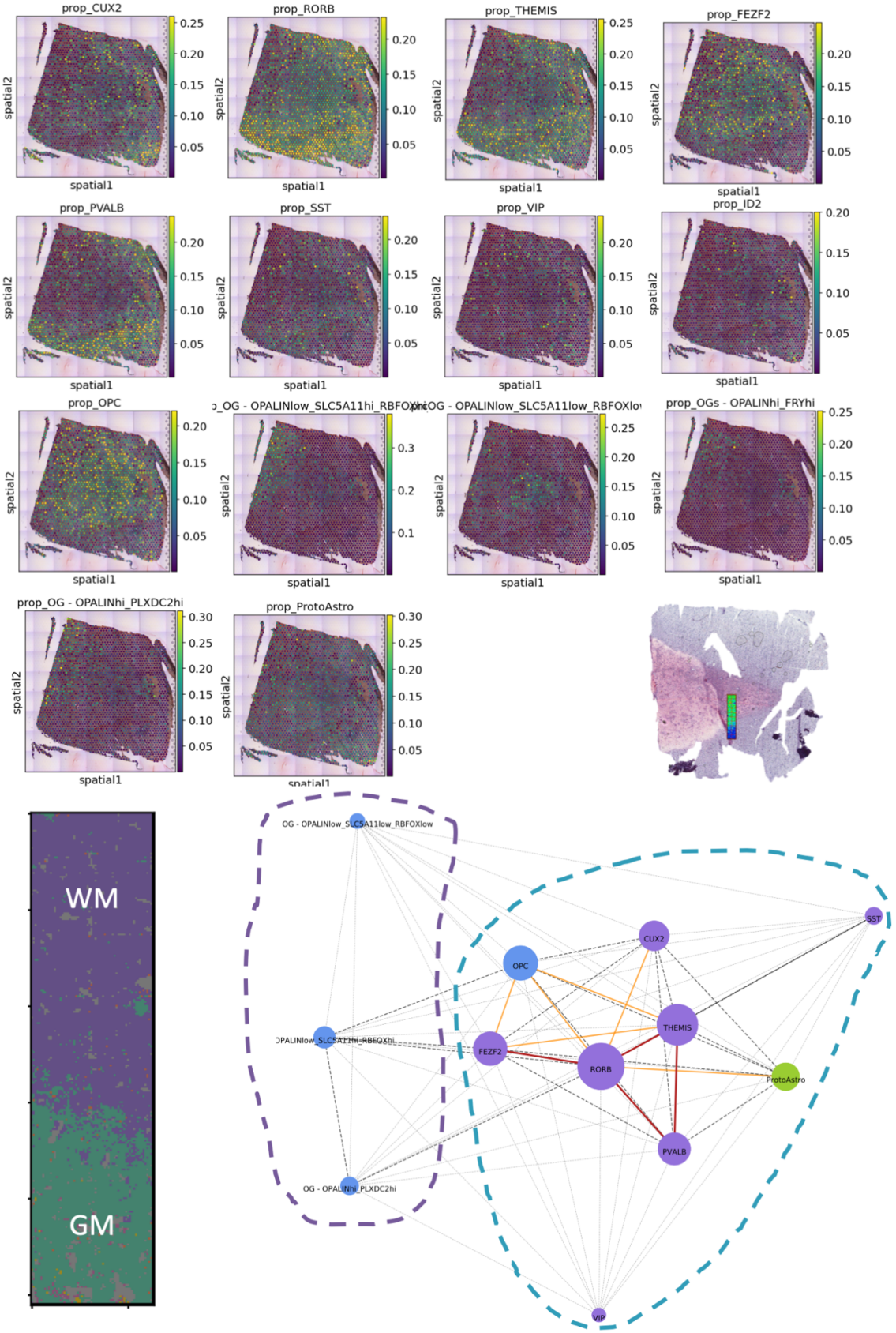


Figure 5.6: Spatial composition and architecture analysis of C014 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data

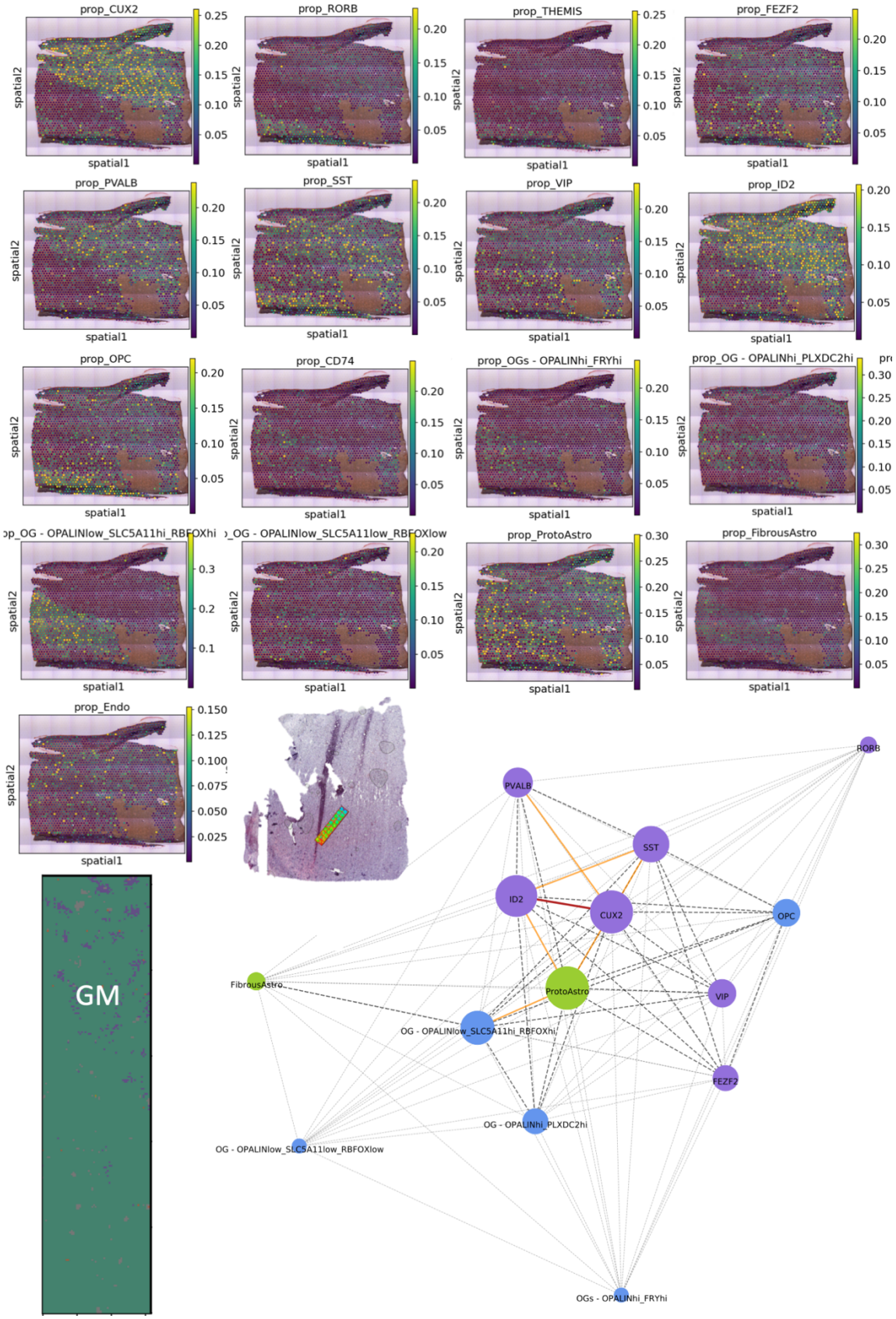


Figure 5.29: Spatial composition and architecture analysis of C064 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data

These features were reflected in the spatial network, where GM cell types were the most frequently represented and had the strongest co-localisation relationships, whilst WM-associated cell types were present in smaller proportions. The 10x Visium section had a greater proportion of WM tissue when compared to the ORO characterisation (noting that ORO staining and Visium were performed in consecutive or near-consecutive 10 µm tissue sections).

5.2.2.3 Brain tissue co-localisation networks: SPMS cases

Patient sample MS352 had a well outlined lesion surrounded by a hypercellular WM; deconvolution of this spatial transcriptomic data identified that the perilesional area was mostly populated by OGs in high proportions and CD74+ immune cells, while the lesion itself was populated by fibrous and protoplasmic astrocytes (**Fig 5.8**). By µFTIR such astrocytic lesions showed the lowest lipid content and highest lipid oxidation. These two main features were reflected in the spatial network and the connections of these cell types.

The MS425 sample included regions of lesion, perilesional WM and GM (**Fig 5.9**). The lesion was well outlined with a clear region of pronounced myelin loss which was reflected in the spatial distribution of cells. The GM region had high proportions of neuronal cell types as well as homogeneously distributed OPCs. The OG OPALINlow SLC5A11low RBFOXlow OG population appeared to line the GM/WM interface, whilst other OG and CD74+ cell types populated the WM area between the lesion and GM.

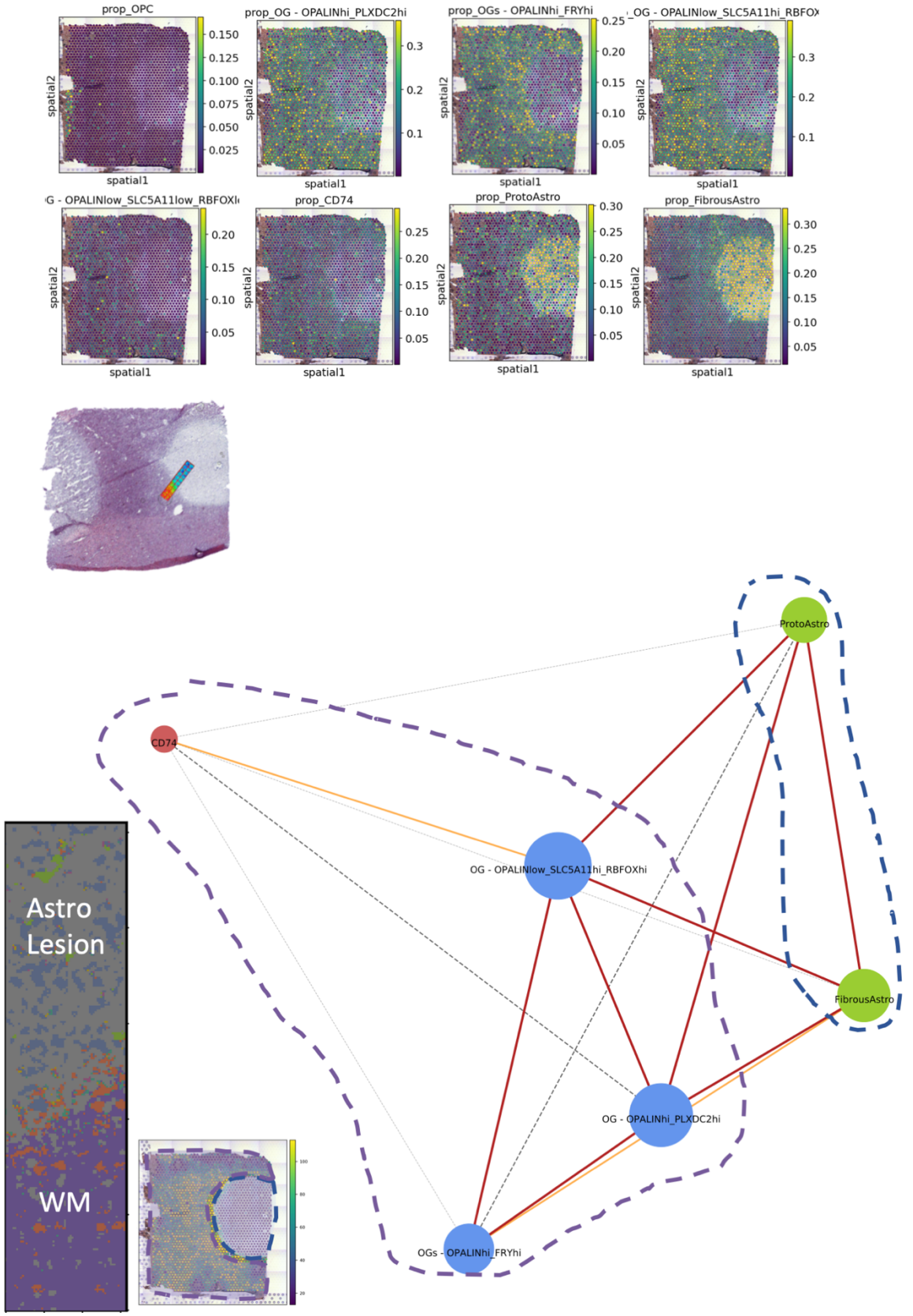


Figure 5.8: Spatial composition and architecture analysis of MS352 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data

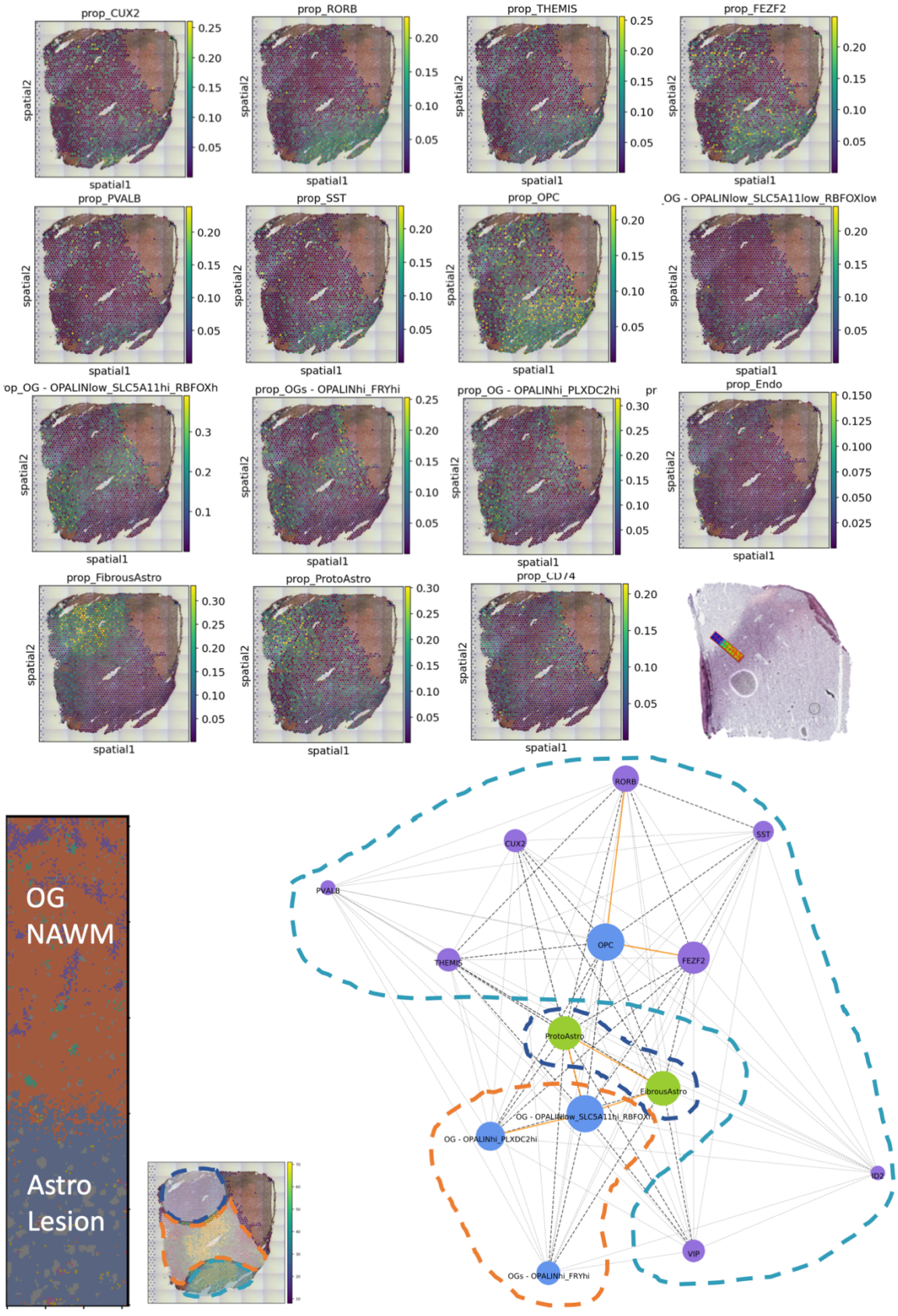


Figure 5.9: Spatial composition and architecture analysis of MS425 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data

Endothelial cell types lined the rim of the lesion and the centre of the lesion was densely populated by fibrous astrocytes. Protoplasmic astrocytes were also present in the lesion, but distributed at a lower density across the WM too. The rich diversity of this sample was represented in the spatial network, with features reflective of GM and WM, as observed in controls, as well as the centrally located astrocytes.

Whilst in some brain samples a distinct tissue architecture could be discerned even with the histological staining alone, this was not the case for MS402. For this sample histological staining revealed a mostly homogeneous sample with low lipid content except for a small region in the corner, with a slightly paler region low in lipid positioned diagonally (Fig 5.10). The µFTIR scanned area from this sample also had low lipid content. However, the spatial transcriptomic deconvolution of MS402 revealed the complex cortical layers and GM features lining the sides of this section with a WM lesion in the centre. As might be expected from the low lipid content observable both in µFTIR and in ORO staining, this sample had <1.5% OGs, with the main cell types populating the lesion being astrocytes (with a particularly dense distribution of fibrous astrocytes and endothelial cells). OPCs were distributed across both the lesion and the GM, with particularly strong co-localisation relationships with neurons.

The overall histology of MS438 indicated substantial damage, suggesting compromised WM integrity with a low MBP content, but also a high lipid content which was densely localised in small rounded areas (Fig 5.11).

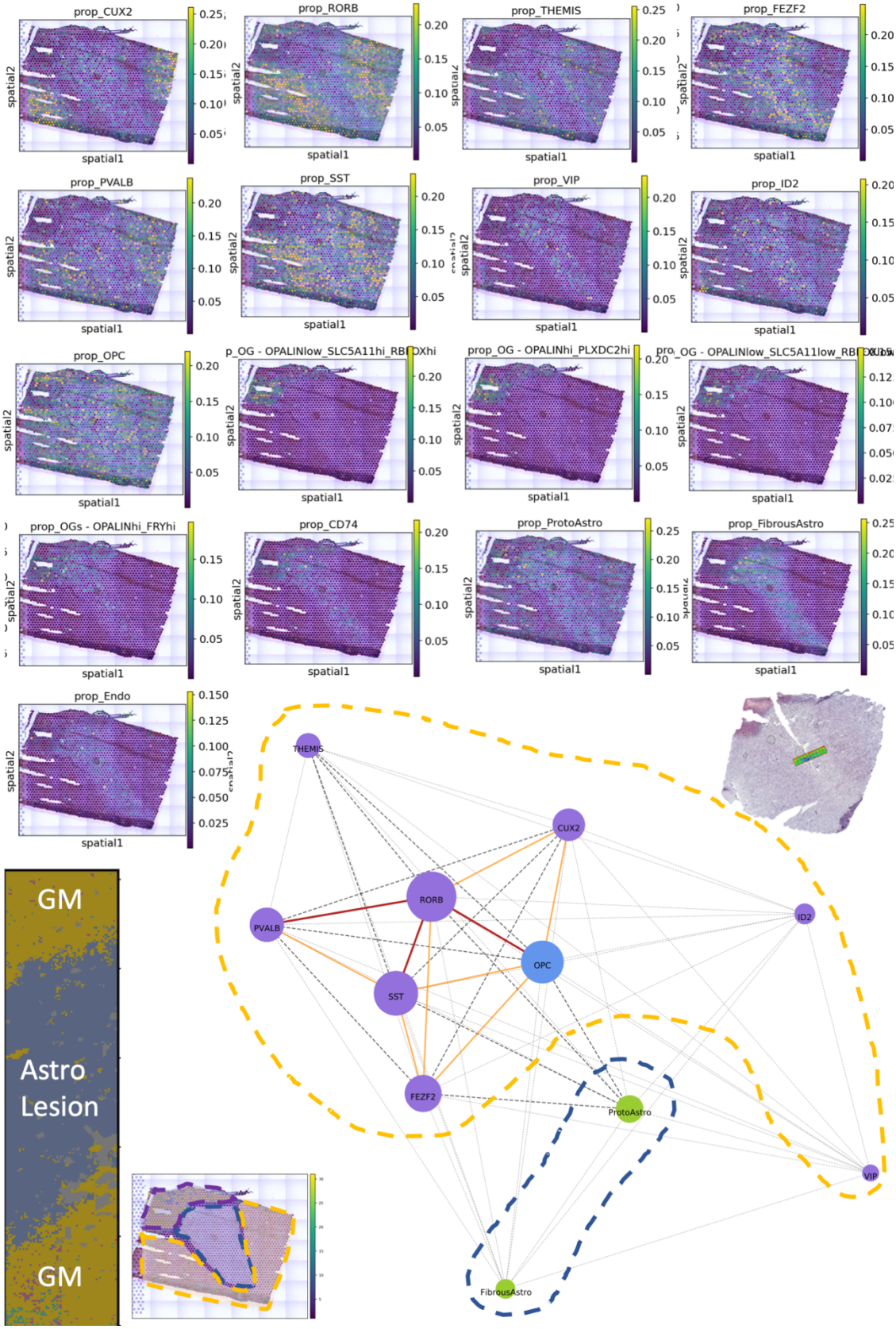


Figure 5.10: Spatial composition and architecture analysis of MS402 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data.

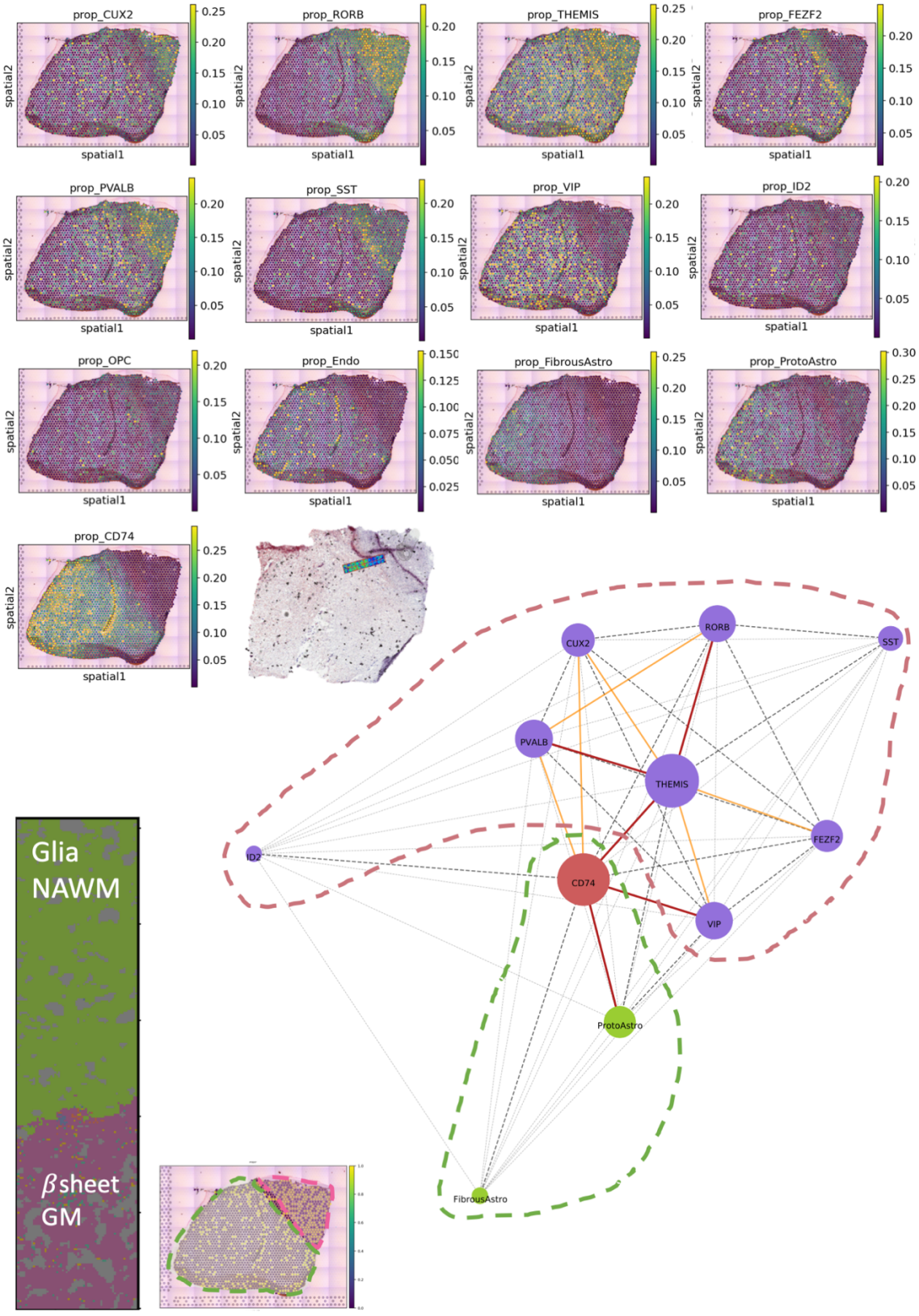


Figure 5.11: Spatial composition and architecture analysis of MS438 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data.

MS438 included a region of GM diagonally outlined with clear cortical layers, although notably with an exceptionally low OPC content. The lesional demyelinated WM had a dense CD74+ immune cell population which co-localised with astrocytes and endothelial cells. These tissue regions were represented in subnetworks of the main co-localisation network, with a cluster of mature OGs analogous to those identified in control WM and a strong relationship between fibrous astrocytes and CD74+ immune cells.

MS466 had a well outlined lesion in the inferior quadrant, and an overall intact lipid content in the surrounding area (Fig 5.12). The lesion was marked by the presence of CD74+ cells and fibrous astrocytes which extended in a gradient beyond the lesion borders with decreasing density of cells further away from the lesion. An opposite gradient of OGs which decreased in density towards the lesion was also present. A small region at the top of the section contained GM-associated neurons and OPCs. The only neuron with enough representation to be in the co-localisation network were RORB neurons. The two major patterns observed on the tissue could be observed in the network, including one subregion with fibrous astrocytes and CD74+ cells with strong relationships as well as strongly connected OGs.

Building on the identifiable substructures associated with known tissue features, as well as cell type localisation observed in cell type maps we summarised major compositional features alongside the corresponding μFTIR data of that corresponding area (Table 5.1).

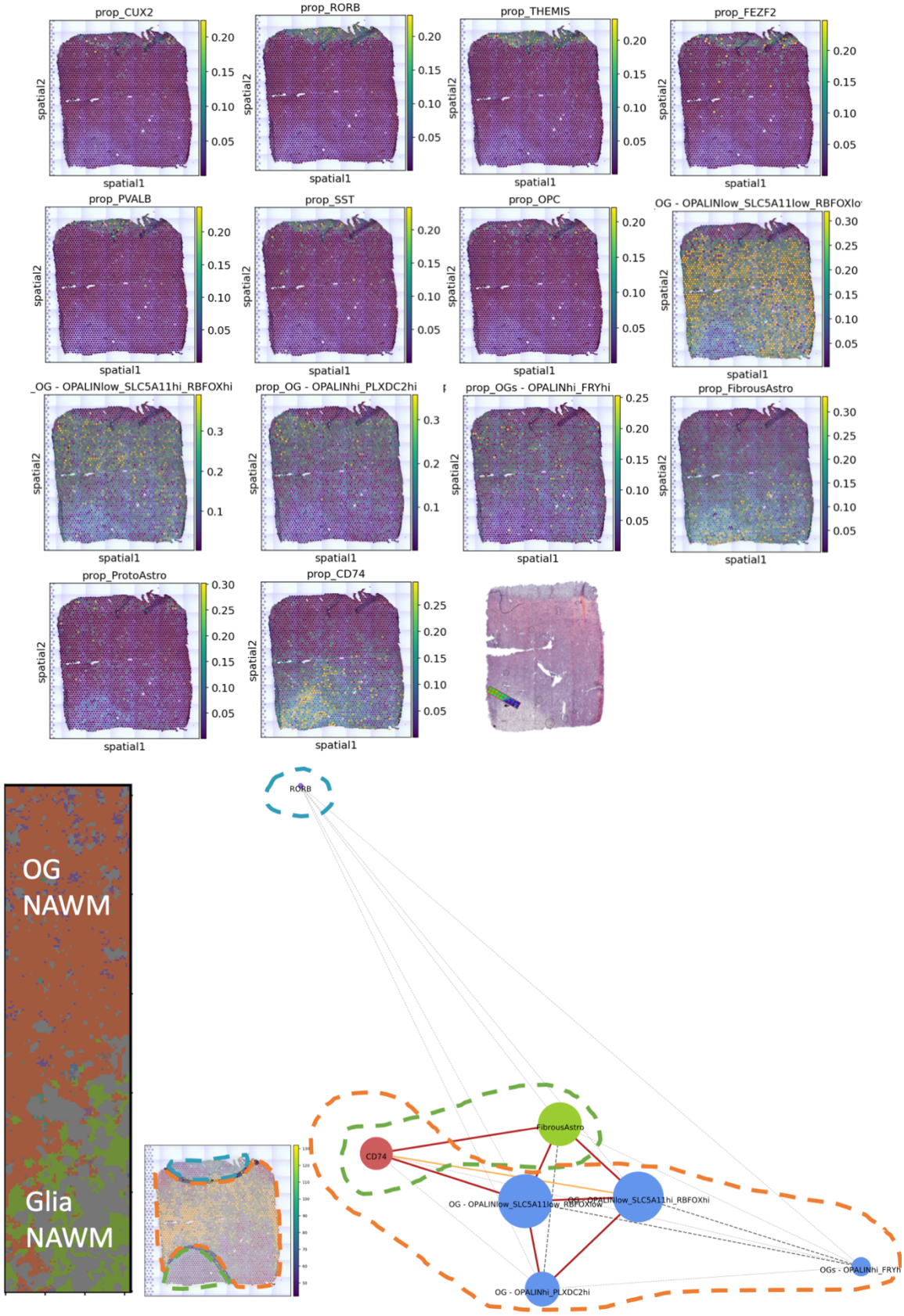


Figure 5.12: Spatial composition and architecture analysis of MS466 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) µFTIR classification map (C) Co-localisation network based on spatial transcriptomic deconvoluted data.

Table 5.1: Spatial cellular profiles associated with each µFTIR cluster

µFTIR cluster	Distinguishing µFTIR features	Spatial composition
WM	Highest mean lipid content Low lipid oxidation Lowest protein content Lowest $\boldsymbol{\beta}$-sheet content	OPALINlow SLC5A11low RBFOXlow OPALINlow SLC5A11hi RBFOXhi OPALINhi PLXDC2hi OPALINhi FRYhi
GM	3rd highest lipid content 3rd lowest lipid oxidation	Neurons OPCs Protoplasmic astrocytes
Les1	Intermediate between WM and GM with 2nd highest lipid content	OPALINlow SLC5A11hi RBFOXhi OPALINhi PLXDC2hi
Les2	Lowest lipid content High lipid oxidation	Astrocytes
Les3	Wide distribution of lipid and protein content	CD74+ immune cells Astrocytes
Les4	Highest lipid oxidation	Neurons OPCs
Les5	Highest $\boldsymbol{\beta}$-sheet content	Neurons

5.2.2.4 Spinal cord co-localisation networks: an overview

The SC co-localisation network of controls reflected the well-organised structure of the SC and the larger representation of WM naturally occurring in the tissue. For control samples OGs were the most frequently observed cell type, followed by astrocytes (**Fig 5.13**). OPALINhi FRYhi and OPALINhi PLXDC2hi OG subtypes were the most likely cell types to co-localise, followed by other mature OGs. Fibrous astrocytes were also likely to co-localise with OPALINhi FRYhi OGs. OGs and astrocytes formed the core of the network with larger representation and strong spatial relationships. CD74+ cells were also represented, although slightly more peripherally. Endothelial cells, when

present, frequently were in the proximity of CD74+ cells. Neurons were identified at a lower frequency.

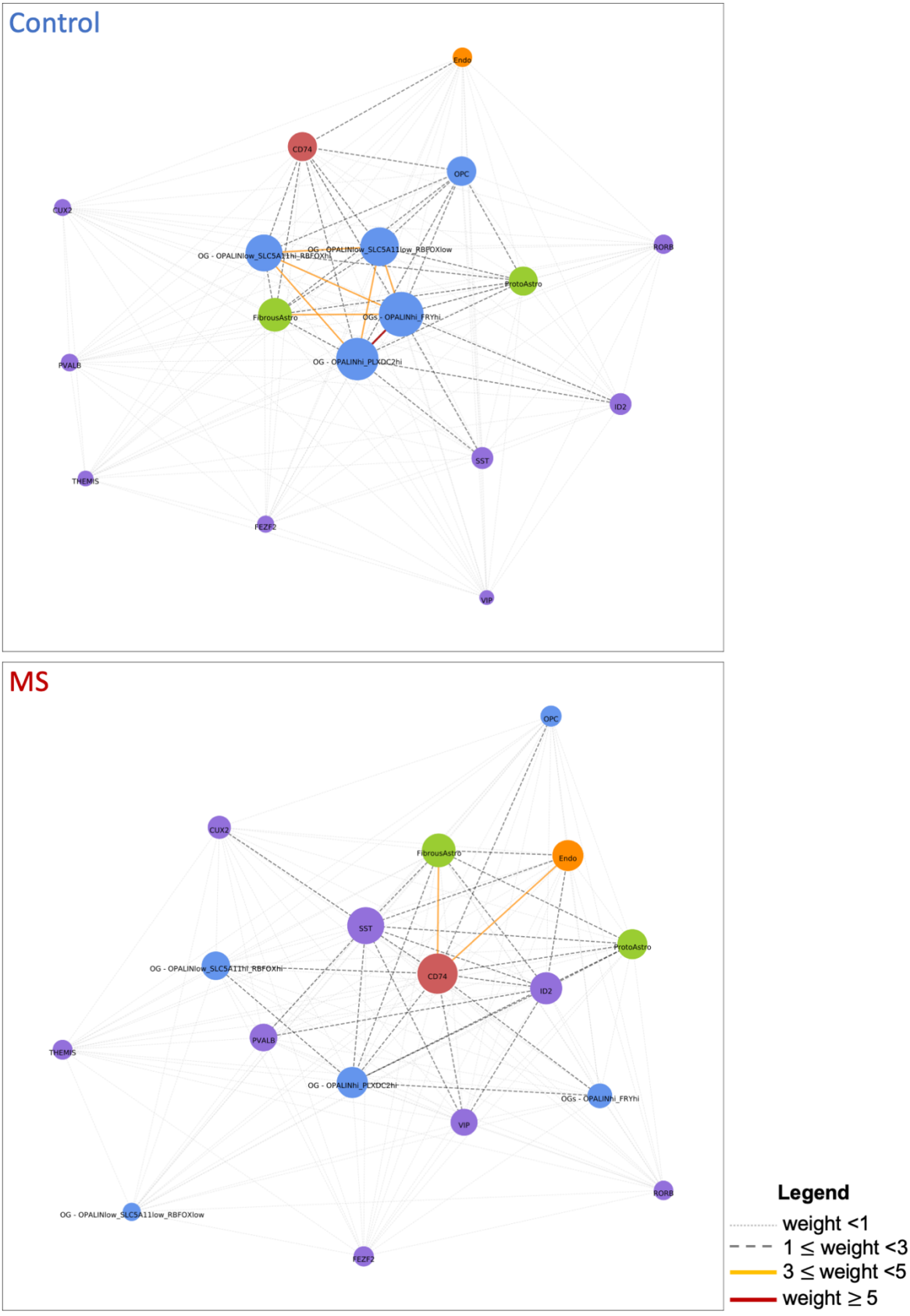


Figure 5.13: Cell type co-localisation analysis for all control SC samples and all MS SC samples. Each cell type is represented by a node and each co-occurrence of two cell types contributes to the edge connecting those nodes.

The co-localisation network of the SPMS samples indicated the MS SC architecture to be substantially altered (Fig 5.13). The most frequently identified cell type was CD74+ immune cells, reflecting increased presence of microglia and macrophages in the vicinity of MS lesions. Endothelial/stromal cells – more frequently detected in MS than controls – continued to have a close spatial relationship with CD74+ cells. Recruitment of these cell types to sites of injury is consistent with SC response to injury, since endothelial cells most likely to migrate in this context are pericytes and fibroblasts. Fibrous astrocytes also tended to be co-localised with CD74+ cells. The lower co-localisation of Fibrous astrocytes and endothelial cells may be indicate the presence of mutually-repelling signals expressed by these cell types, as previously reported in traumatic spinal cord injury response. The central position of CD74+ immune cells reflects both the higher frequency but also the mobility of microglia and macrophages, which may co-localise with many cell types in spite of the distinct architecture of the SC as determined by OGs and neurons. Strikingly, in the disease network OGs have a much less prominent role while SST, ID2, PVALB and VIP neurons are more likely to be identified. OPCs are also peripherally located and have a low likelihood to be co-localised with any cell type other than CD74+ immune cells.

5.2.2.5 Spinal cord co-localisation networks: controls

In the controls deconvolution revealed expected architectural features of the spinal cord, such as high proportions of neuronal cells in the stereotypical GM butterfly-like feature central to the SC horizontal section; the most specifically localized neuronal cell

types were SST and ID2 (Fig 5.14-15). OG and OPC cell types localised to the WM with high accuracy. Fibrous astrocytes had a more diffuse distribution, while protoplasmic astrocytes localised both to the grey matter as well as the outer rim of the SC section. CD74+ immune cells localised in outward-radiating patterns towards the peripheral regions of the spinal cord. Endothelial/stromal cells were specifically localised to tissue features consistent with meningeal layers, such as the anterior median fissure.

Both samples C085 and C084 are thoracic SC sections, although the presence of prominent dorsal horns may be indicative of a SC section at a higher thoracic level with features reminiscent of a cervical cross-section. In C084 the most prominently featured cell types included mature OGs and fibrous astrocytes, with more secondarily co-localised OPCs, protoplasmic astrocytes and CD74+ immune cells (Fig 5.14). For C085 mature OGs were the most frequently detected and most frequently co-localised cell types, followed by CD74+ immune cells, OPCs, endothelial cells, fibrous astrocytes, SST and ID2 neurons (Fig 5.15).

5.2.2.6 Spinal cord co-localisation networks: SPMS cases

Deconvolution of MS376 revealed the varied cellular composition of this sample (Fig 5.16), which may also explain the biochemical complexity of this sample as observed through µFTIR clustering analysis. This sample was marked by a loss of the characteristic GM butterfly feature, with neuronal cell types present in the lower central region of the sample opposite to the anterior median fissure co-localising with a subset of protoplasmic astrocytes.

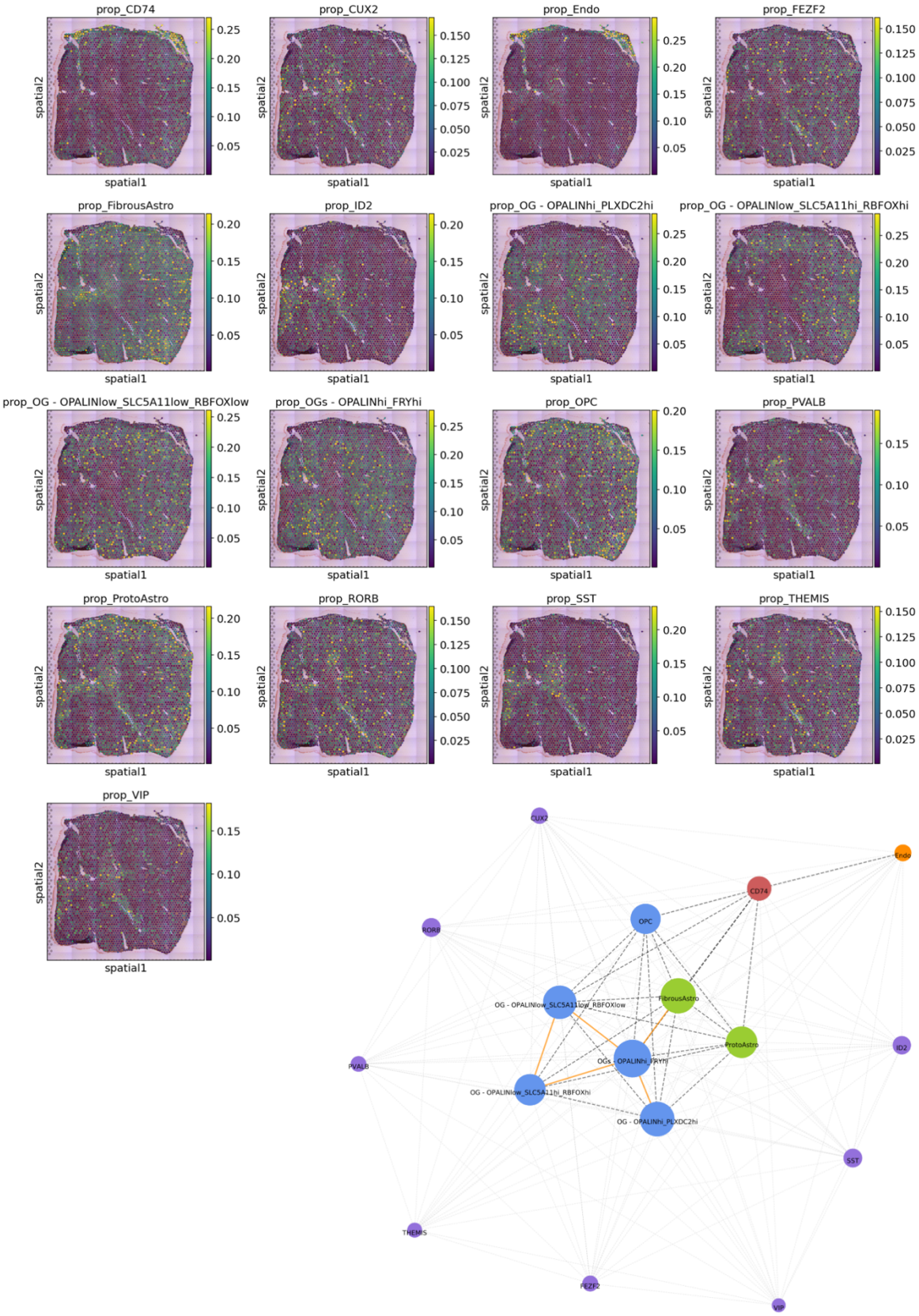


Figure 5.14: Spatial composition and architecture analysis of C084 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) Co-localisation network based on spatial transcriptomic deconvoluted data

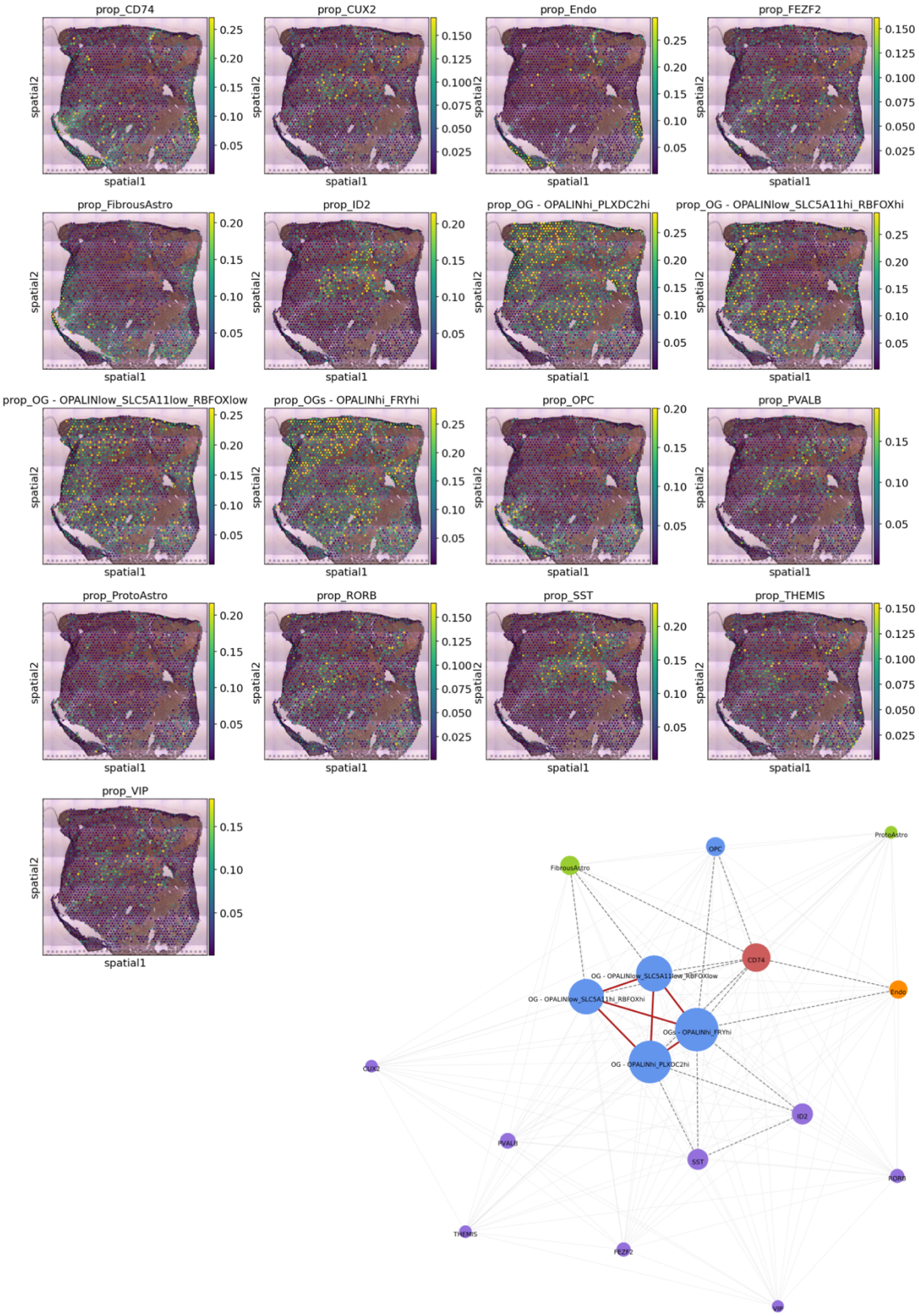


Figure 5.15: Spatial composition and architecture analysis of C085 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) Co-localisation network based on spatial transcriptomic deconvoluted data

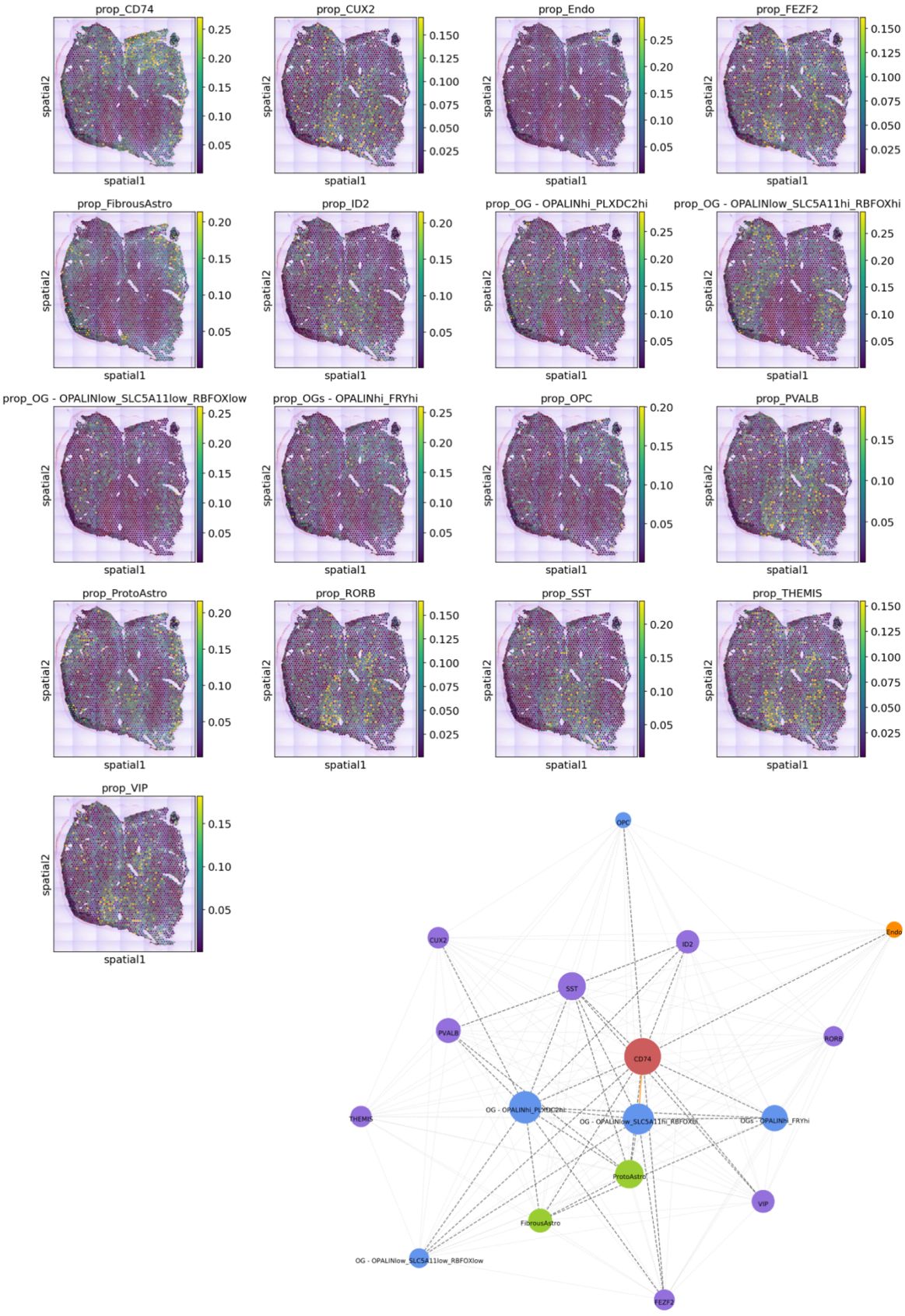


Figure 5.16: Spatial composition and architecture analysis of MS376 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) Co-localisation network based on spatial transcriptomic deconvoluted data

A high influx of CD74+ immune cells was observed immediately to the right of the anterior median fissure which partially overlapped with fibrous astrocytes and OPCs in MS376. To the left of the anterior median fissure another area was observed which contained protoplasmic and fibrous astrocytes. The remaining area in between these three main zones was populated by mature OGs. This complexity was also reflected in the co-localisation network for this sample, which was not centred centralised but rather diffusely distributed reflecting this zonation.

In contrast with the highly complex MS376, MS432 had a notable lack of normal appearing features evident in the low integrity of the sample for handling and sectioning (Fig 5.17). Compositionally, this was most prominently reflected in the lack of mature OGs of this sample and the absence of GM features. A feature in the upper left corner of the sample appeared to have a cellular agglomeration including CD74+ immune cells, endothelial cells, fibrous astrocytes, protoplasmic astrocytes and OPCs. There was a diffuse assignment of SST and ID2 neuronal cell types across the entire sample which is not consistent with sample architecture, but this diffuse assignment may be indicative of the pervasive damaged such that deconvolution cannot be perfectly implemented.

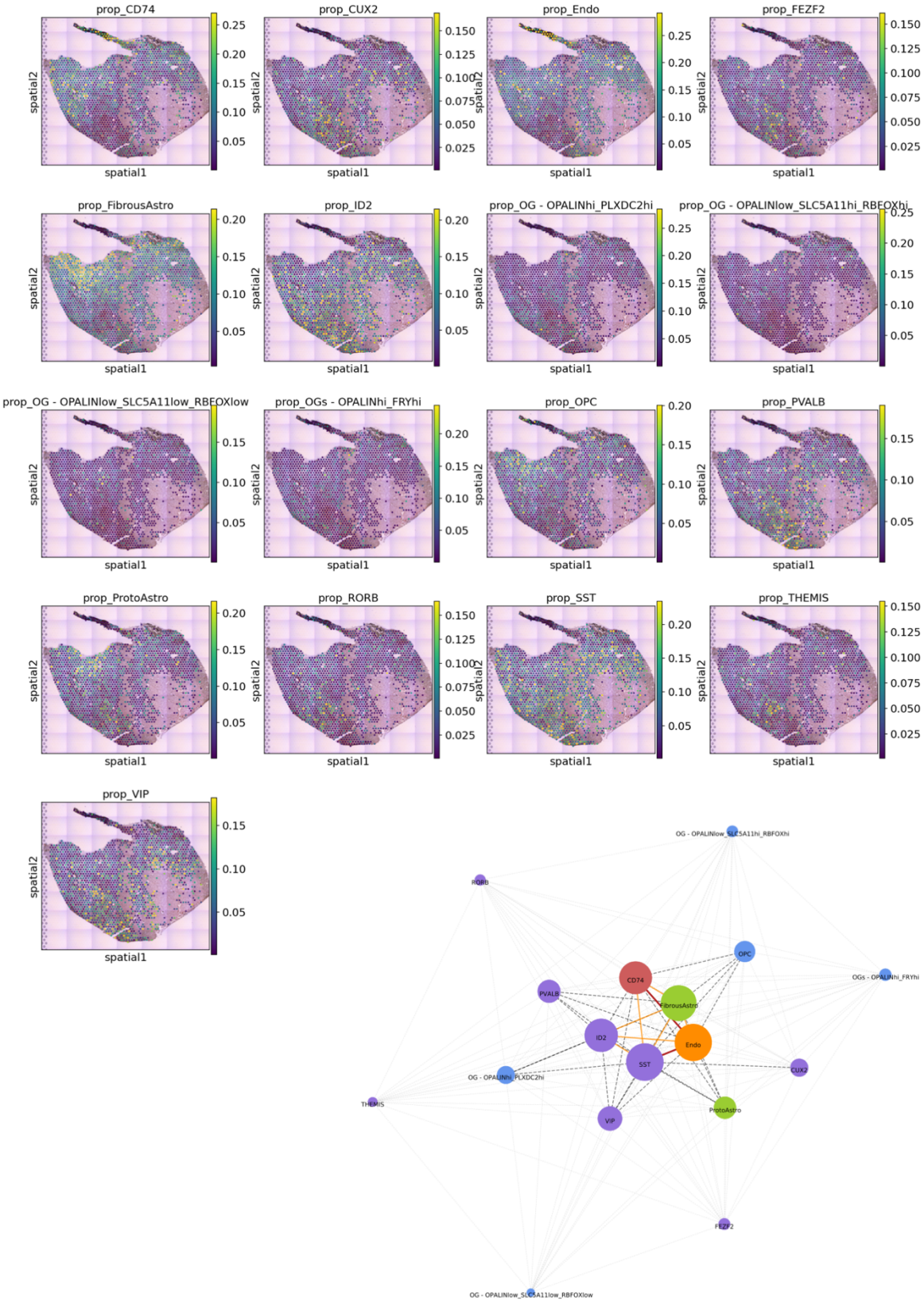


Figure 5.17: Spatial composition and architecture analysis of MS432 (A) Deconvoluted Visium data with plotted proportion of each spot per cell type (B) Co-localisation network based on spatial transcriptomic deconvoluted data

5.2.3 Proteomic characterisation of SPMS and control CNS tissue

5.2.3.1 Proteomic classification of SPMS and control brain samples and interpretation relative to biochemical profiling

In order to proteomically profile the SPMS and control brain samples, regions of interest were selected and laser-capture microdissected. Spatial proteomics was performed for a total of 39 SPMS and 26 control microdissected areas, which were carefully selected to overlap with areas scanned by μFTIR in consecutive or near-consecutive tissue sections. Following QC and pre-processing (Methods section 2.5.4), all samples were grouped using hierarchical clustering (**Fig 5.18**). A total of 5 clusters were identified and compared to group (case vs. control), histology (GM, Lesion, NAWM, WM) and μFTIR cluster.

Brain-Cluster1 was mainly identified in control samples, it corresponded to regions histologically identified as WM and had high protein levels of MBP, PLP1 and CNP. Brain-Cluster2 was only found in MS cases, it corresponded to NAWM regions and had a proteomic profile intermediary between WM and lesion. Brain-Cluster4 was also only found in MS samples, it corresponded to regions with more severely demyelinated lesions and had a greater number of differentially expressed proteins (DEPs). Both of these clusters had high expression of glia-associated GFAP, VIM, and anti-inflammatory ANXA1, for example.

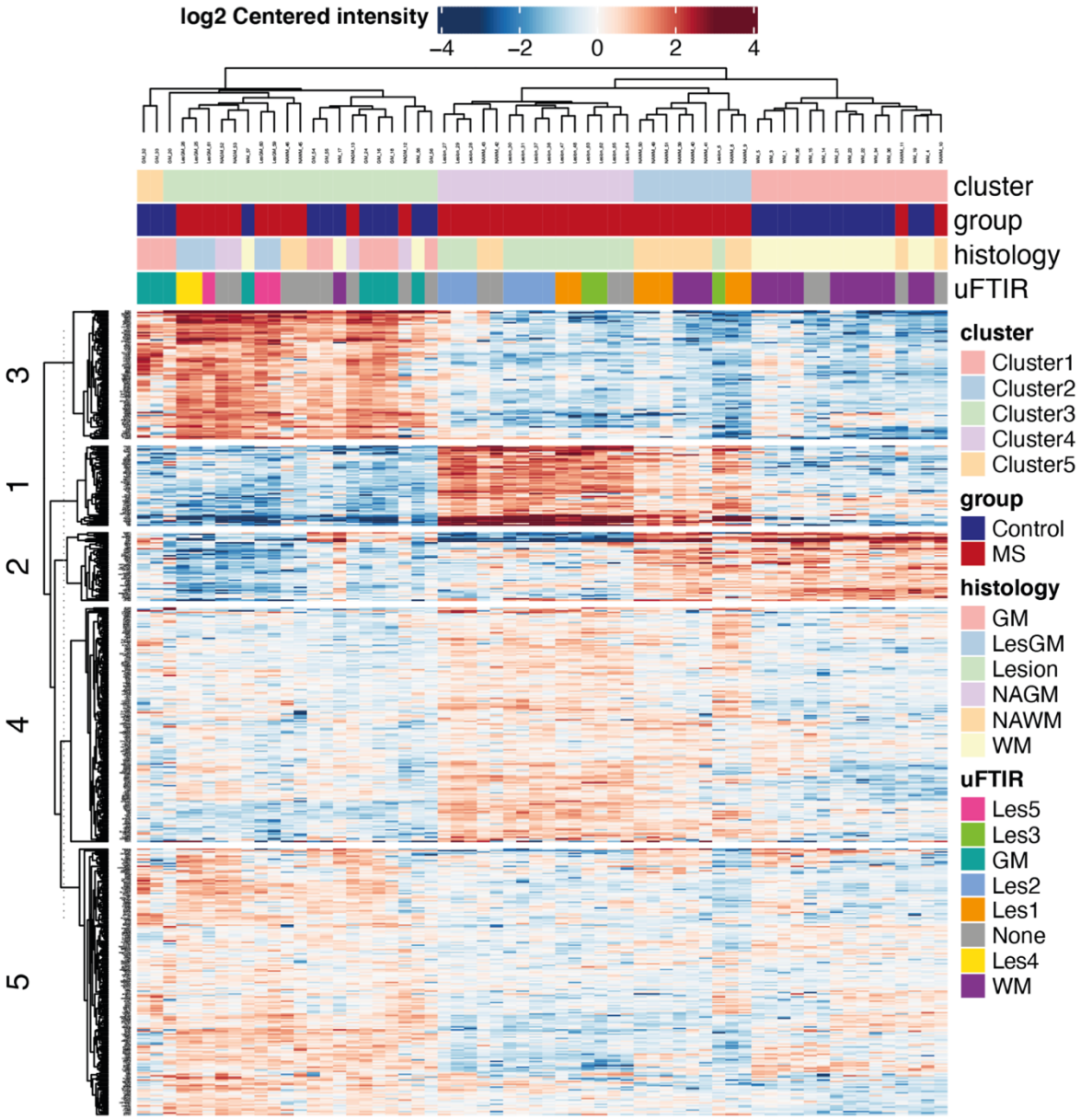


Figure 5.18: Clustering of brain proteomic data identified five main clusters, corresponding to four main tissue types. Cluster 1 corresponded mostly to regions classified as WM from controls; clusters 3 and 5 corresponded to GM from both cases and controls; clusters 2 and 4 were associated with MS patients, cluster 2 with NAWM/early lesion samples and cluster 4 with late/demyelinated samples.

To cross-validate the expression of the genes encoding these proteins with the lesional and peri-lesional architectural features of the tissue, the transcriptional expression of these genes was also mapped using the Visium data (**Fig 5.19**). Indeed, in MS352, MS425 and MS402 *GFAP* and *VIM* tended to co-localise and tended to be present in borders of demyelinated lesions with perilesional WM. In MS466 the borders had a high expression of *GFAP* and *VIM*, although this signal was also present in the demyelinated lesion and extended with high expression levels to the perilesional WM. MS438 had an overall lower expression level of *GFAP* and *VIM*, although the *VIM* expression was not tissue-type specific. *ANXA1* expression was present within demyelinated lesions, and had a particularly high expression in MS466 with highest expression in the lesion and slowly decreasing in expression with increasing distance from the lesion. It was densely localised to lesion borders in MS352 and MS402.

The two remaining proteomic clusters were found in cases and controls and mostly included GM regions. Brain-Cluster3 was the largest of these whilst Brain-cluster5 included only two samples which appeared to have an outlier profile. Brain-Cluster4 was marked by the high expression of proteins associated with neurons such as SYN1, MAP2.

Apart from correlations between the proteomic clustering and histological classification, certain correlation patterns could be observed in the proteomic and µFTIR cluster classification. Regions clustered as WM with µFTIR were more likely to be grouped in the proteomic Brain-Cluster1.

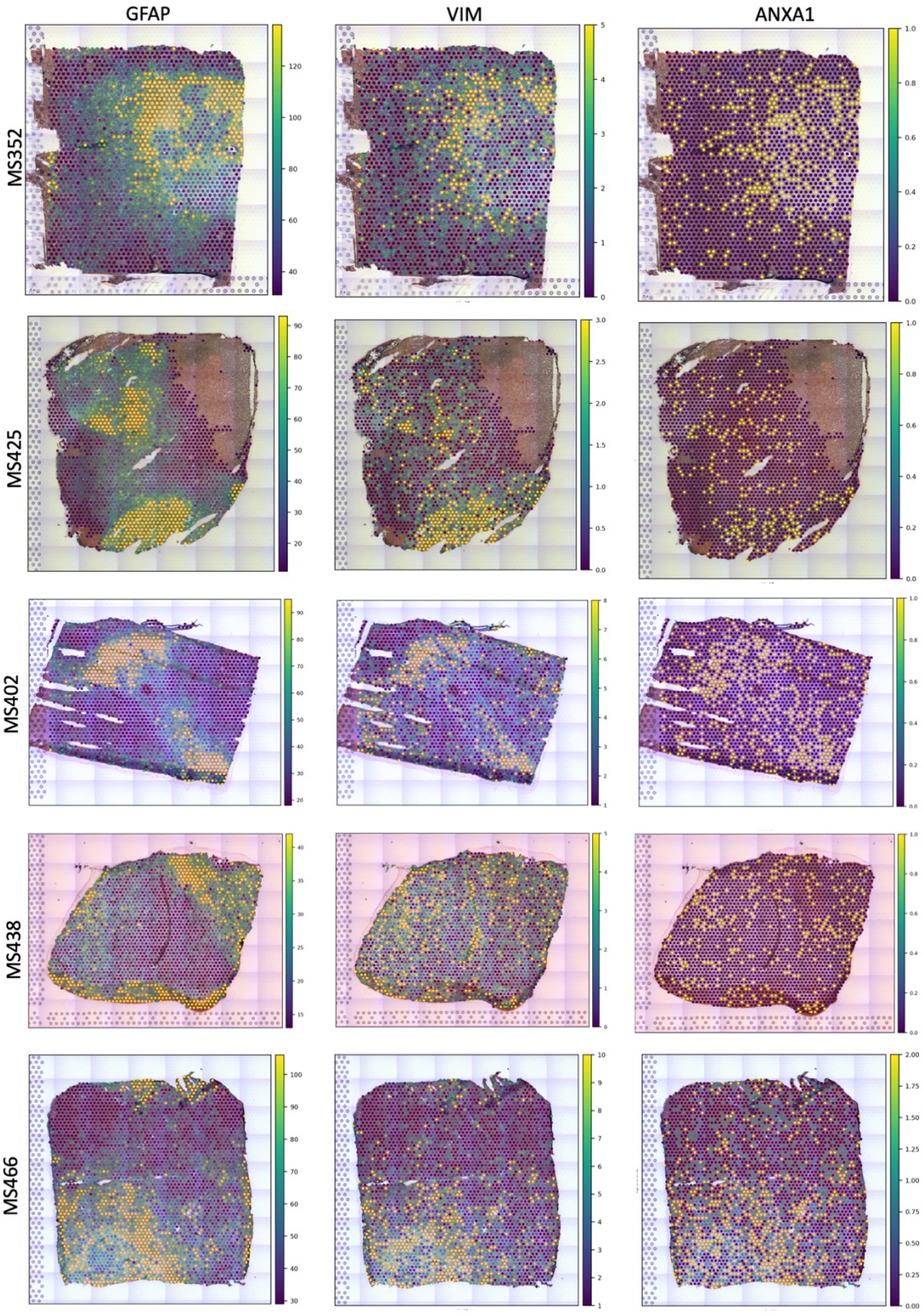


Figure 5.19: Spatial mapping of genes of interest in brain samples

Interestingly, the three µFTIR-WM samples in the Brain-cluster2 are from MS352, a case which biochemically had a region classified as WM although it bordered a demyelinated lesion. This difference in classification indicated that whilst biochemical changes were still minimal and it thus had WM profile, proteomically this region was more similar to pathology-associated Brain-Cluster2. The µFTIR-Les1 cluster was more likely to be grouped under the Brain-Cluster2. Brain-Cluster4 included the µFTIR clusters Les2 and Les3. The large proteomic Brain-Cluster3 included several µFTIR classifications, including µFTIR-GM, µFTIR-Les4 and µFTIR-Les5.

When Brain-Cluster1, associated with WM, was compared to Brain-Cluster2, associated with NAWM/early lesion classification, Brain-Cluster1 had two DEPs, MAP6 (microtubule associated protein 6) and PMP2 (peripheral myelin protein 2), whilst Brain-Cluster2 had ten DEPs including those with notably high expression detected in the heatmap – FLNC and FLNA, VIM, MAOB, HSPB1, ANXA1, ANXA2, TLN1, CAPG and GFAP (Fig 5.18). The top 10 GO enrichment analysis of these DEPs included terms associated with endothelial cell types such as positive regulation of blood vessel endothelial cell migration, cellular response to vascular endothelial growth factor stimulus and platelet aggregation. Other terms included cellular component assembly, positive regulation of organelle organisation, intermediate filament organisation, cell junction assembly, positive regulation of vesicle fusion and homotypic cell-cell adhesion (Fig 5.20).

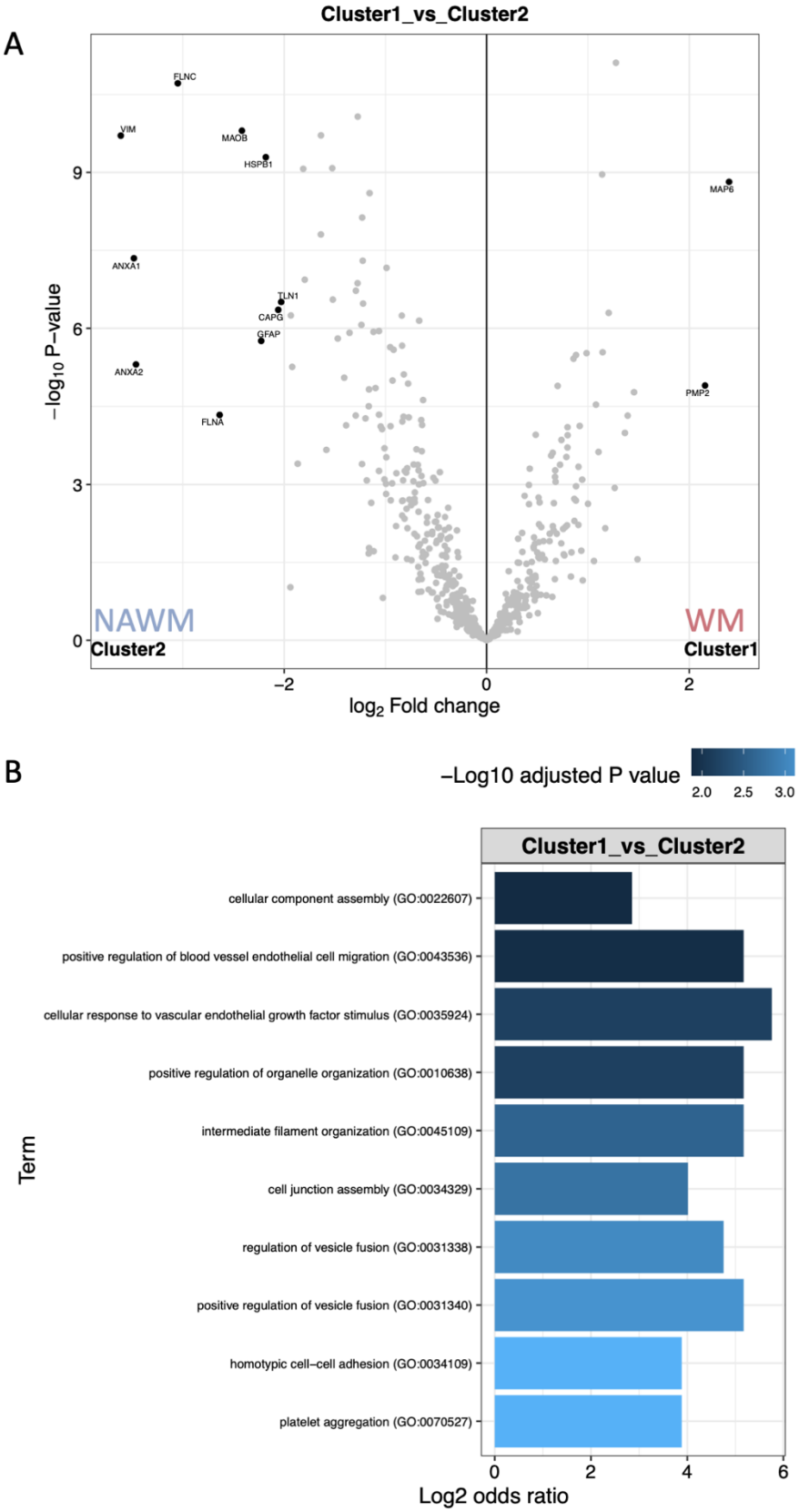


Figure 5.20: Proteomic analysis of brain samples identified 12 DEPs when comparing WM-associated cluster1 and lesion-associated cluster2. The top 10 GO terms resulting from GSEA analysis included terms relating to vesicle function, cell adhesion and cellular organisation.

40 DEPs were identified when comparing Brain-Cluster1 to Brain-Cluster4. The lesion included DEPs associated with glial activation such as GFAP and VIM[262], iron metabolism associated protein FTL, immunoglobulin IGHG1, and ANXA1 and ANXA2. WM DEPs included markers of myelination MBP, PLP1, CNP. The top 10 GO terms of these DEPs were nucleobase catabolic process, positive regulation of potassium ion transport, T-tubule organisation, regulated exocytosis, protein localisation to cell surface, platelet degranulation, myelination, homotypic cell-cell adhesion, platelet aggregation and substantia nigra development (**Fig 5.21**).

5.2.3.2 Proteomic classification of SPMS and control SC samples

In order to proteomically profile SC affected by MS, regions of interest in spinal cord sections were selected and microdissected. Spatial proteomics was performed for a total of 24 SC microdissected areas. Following quality control and pre-processing, all samples were grouped using hierarchical clustering (**Fig 5.22**). Cluster annotation was compared to group (case vs. control), histological (WM, GM, NAWM/Les) and patient classification.

The obtained clusters had a high correlation with histological characterisation, with the main distinction being the identification of two types of pathology-associated clusters. Both pathology-associated clusters were compared to WM. SC-Cluster2 was proteomically more similar to the WM cluster, SC-Cluster1 (**Fig 5.23A**).

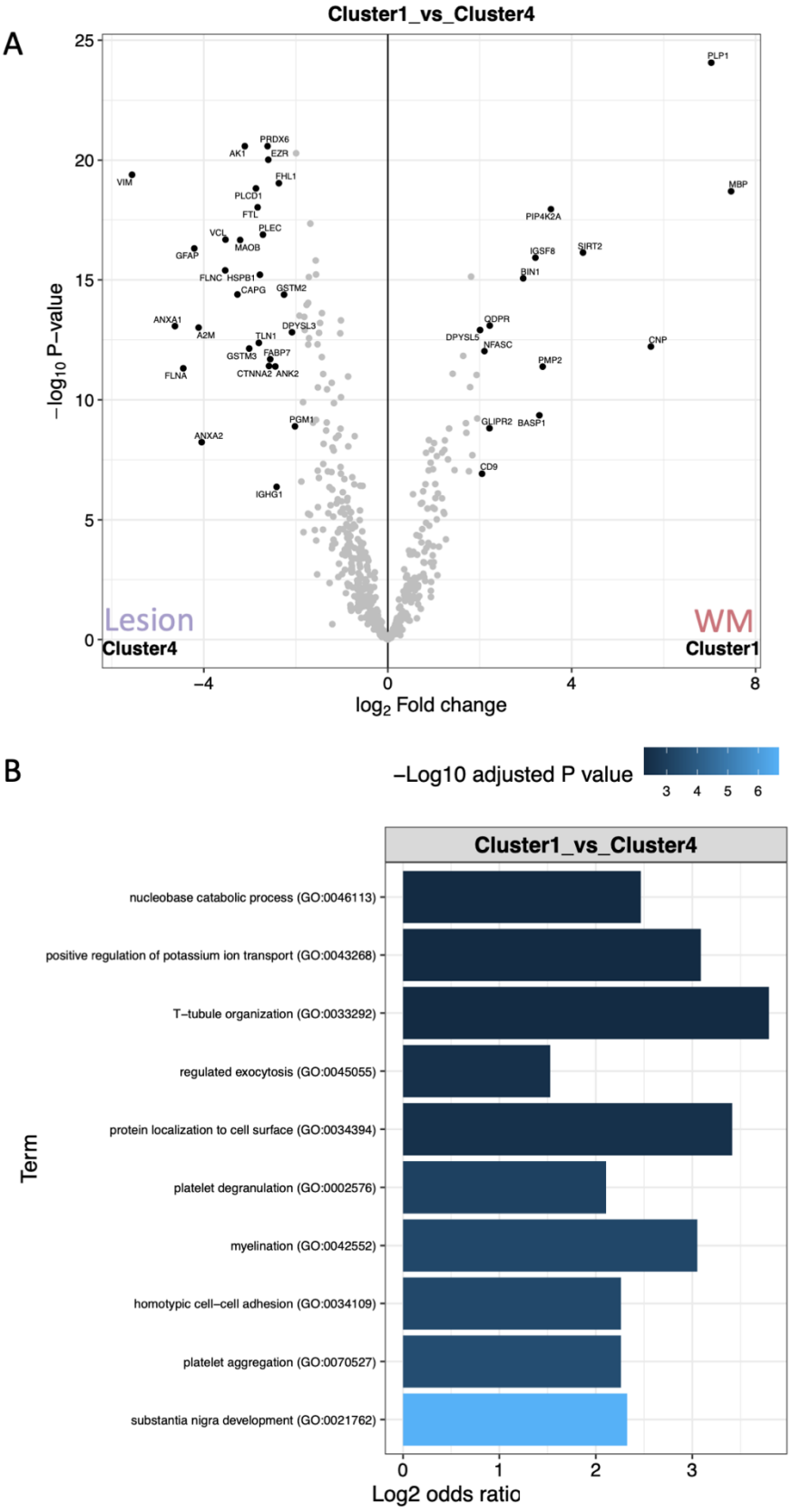


Figure 5.21: Proteomic analysis of brain samples identified 39 DEPs when comparing WM-associated cluster1 and lesion-associated cluster4. The top 10 GO terms resulting from GSEA analysis included terms relating to endothelial cell function, coagulation, myelination, potassium ion transport, cell structure organisation.

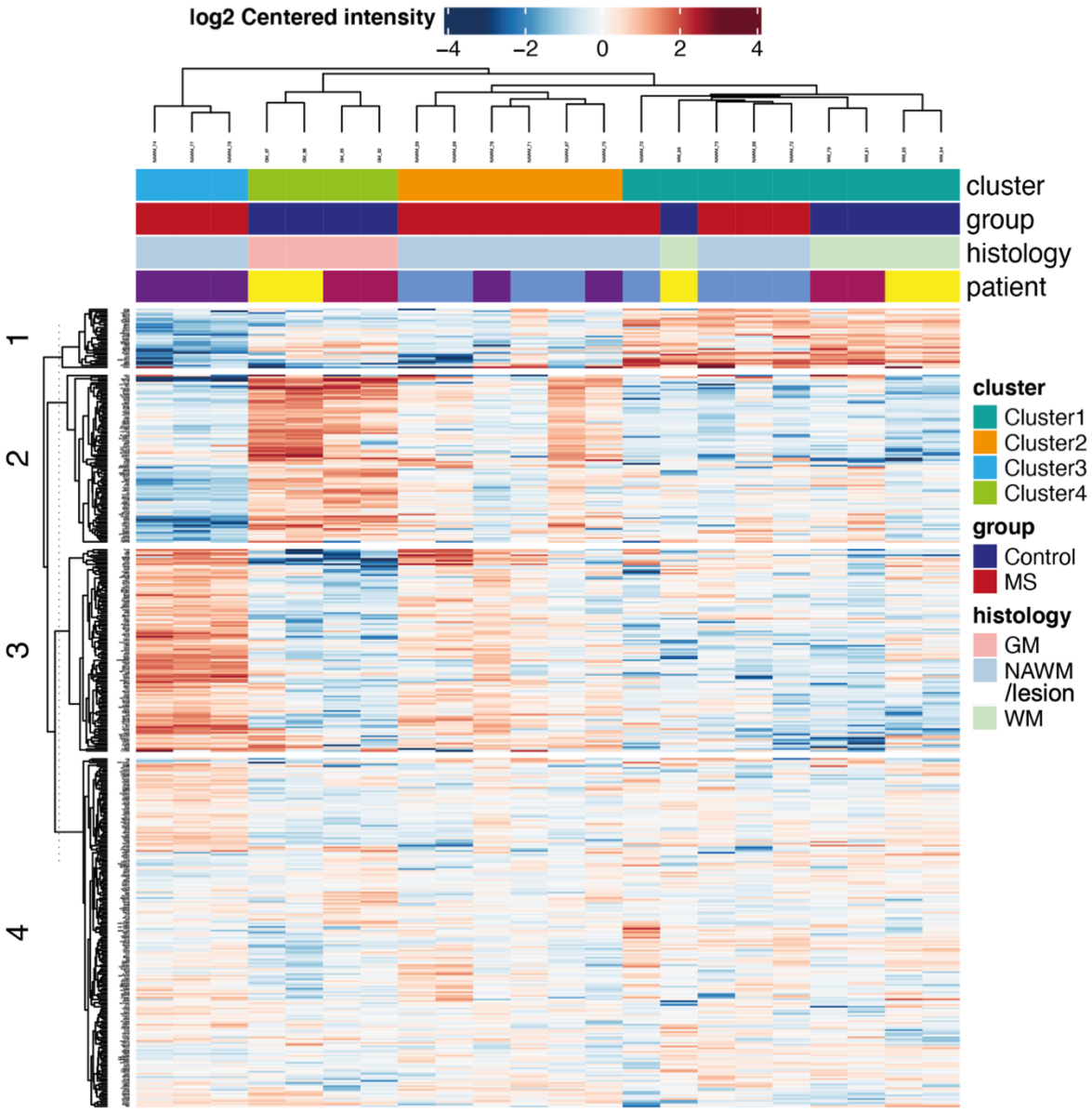


Figure 5.22: Clustering of spinal cord proteomic data identified four main clusters, corresponding to four main tissue types. Clusters 2 and 3 were associated with pathological NAWM/lesion from cases, whilst Cluster 1 was associated with healthy WM and Cluster 4 was associated with GM regions.

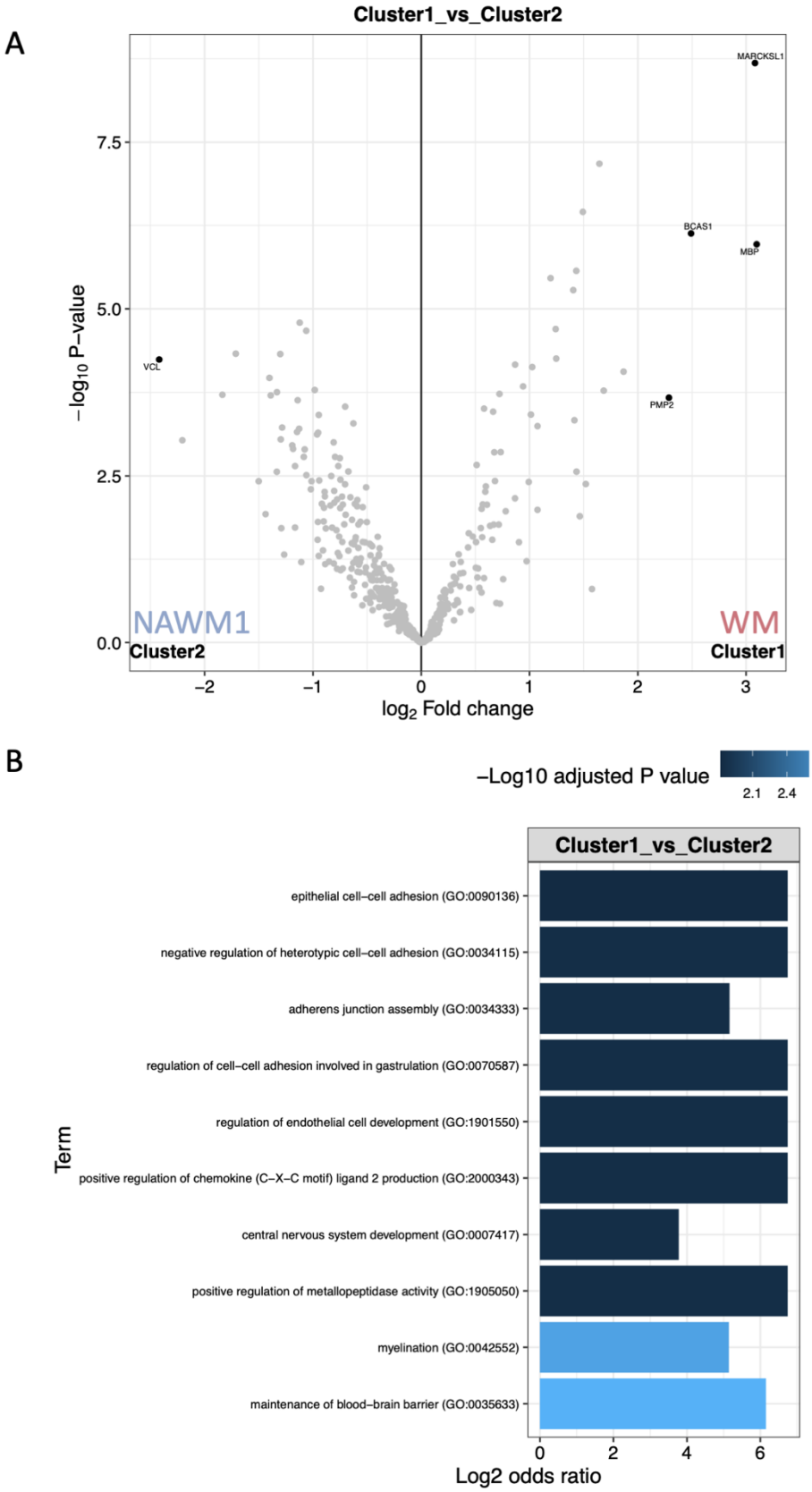


Figure 5.23: Proteomic analysis of SC samples identified only five DEPs when comparing WM-associated cluster1 and NAWM/lesion-associated cluster2. The top 10 GO terms resulting from GSEA analysis included terms relating to epithelial cell function, blood-brain barrier function and CNS development.

SC-Cluster2 had higher VCL, a cytoskeletal protein associated with cell-cell and cell-matrix junctions, while SC-Cluster1 had higher levels of proteins associated with myelination such as MBP, PMP2 and BCAS1; MARCKSL1 was also upregulated in SC-Cluster1, and is involved with cytoskeletal homeostasis.

GO analysis of biological processes associated with the differentially expressed proteins (DEPs) included terms associated with cell-cell adhesion, endothelial cell development, CNS development, metallopeptidase activity, myelination and blood brain barrier maintenance (Fig 5.23B). SC-Cluster 3 had a substantially more distinct proteomic expression when compared to the WM. There were 35 DEPs when comparing these two clusters (Fig 5.24A). GO terms associated with DEPs include calcium ion homeostasis, neurotransmitter transport, chemical synaptic transmission, positive regulation of vacuole organisation, negative regulation of protein metabolic process, myelination, anterograde trans-synaptic signalling and substantia nigra development (Fig 5.24B). Based on this analysis and on myelination as estimated by histological characterisation of these samples, SC-Cluster2 was considered to likely be an earlier lesion cluster, while SC-Cluster3 was considered to be a later lesion cluster.

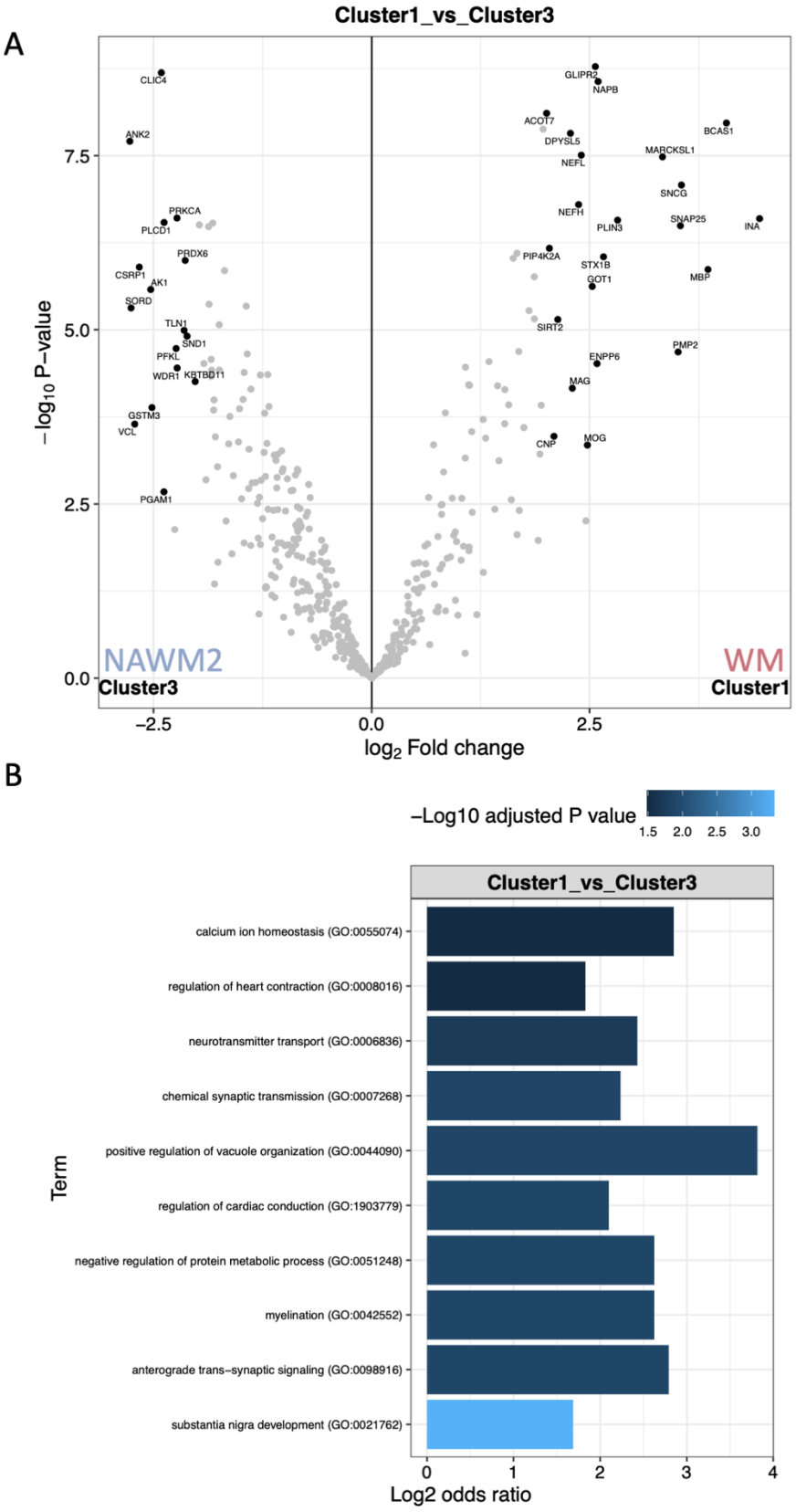


Figure 5.24: Proteomic analysis of SC samples identified over 35 DEPs were identified when comparing WM-associated cluster1 and NAWM/lesion-associated cluster3. The top 10 GO terms resulting from GSEA analysis included terms relating to myelination, calcium ion homeostasis, synaptic communication and CNS development.

5.2.4 Comparison between biochemical metrics, and spatial proteomics and transcriptomics

The μFTIR biochemical profiling findings were also more directly correlated with spatial proteomic metrics. For brain samples, this was accomplished by capturing areas classified by μFTIR separately, performing proteomic analysis and taking the mean intensity for the regions with that cluster classification. This enabled correlation of the mean biochemical component per cluster and the mean normalised LFQ intensity of the proteins of interest for that cluster. For the SC samples, due to the less well-defined μFTIR cluster borders (**Fig 3.30**), this targeted proteomic capture based on the μFTIR cluster mapping was not possible. Therefore, the mean values for the μFTIR clusters were compared to the mean values for the proteomics clusters.

As detailed in the prior sections, both approaches generated a cluster which corresponded to WM (μFTIR: WM; Proteomics: Cluster1), GM (μFTIR: GM; Proteomics: Cluster4), and a cluster which corresponded to severely demyelinated areas in MS432 (μFTIR: Les2; Proteomics: Cluster3). The main difference in the clustering accomplished by each method was that μFTIR captured higher resolution clusters for MS376, so μFTIR Les1, Les3 and Les4 correspond to subtly distinct areas of what under the proteomic classification was classified as a single cluster, Cluster2. For ease of comparison, the proteomic SC clusters will be referred to as WMCluster1, GMCluster4, MS432Cluster3 and LesCluster2.

5.2.4.1 Correlation between lipid content metrics in µFTIR and myelin metrics in spatial proteomics and transcriptomics

In order to validate the assessment that changes to lipid content observed with µFTIR are associated with myelination, the correlation between the lipid estimate with µFTIR and the normalised LFQ intensity of proteins associated with myelin content (MBP, PLP1) and myelination/remyelination (PADI2) were estimated.

In the brain samples, MBP levels had a strong correlation with lipid levels estimated by µFTIR (Pearson's R = 0.928, SEM = 2.573 x 10^{-3}) (**Fig 5.25**). PMP2 also had a strong correlation with lipid levels (Pearson's R = 0.850, SEM = 1.551 x 10^{-2}). PLP1 (Pearson's R = 5.582, SEM = 1.928 x 10^{-1}) and PADI2 (Pearson's R = 0.199, SEM = 6.682 x 10^{-1}) did not have strong correlations with lipid content. Interestingly, specifically in Les1, Les2 and Les3 there appeared to be an anti-correlation between lipid content and PADI2 expression, where the samples with lowest lipid expression had the highest PADI2 expression.

These metrics were also assessed spatially, whereby lipid assessed by ORO, lipid content mapped by µFTIR, and spatial gene expression of *MBP* and *PADI2* were compared (**Fig 5.26**). A high fidelity between these spatial metrics was observable across all brain samples, except for MS438. This sample had a distinct lipid distribution which could be observed with ORO staining and the lipid map, which had a high maximum value comparable to other samples. Here the lipid content was less evenly distributed and was instead concentrated in small punctate areas across the tissue.

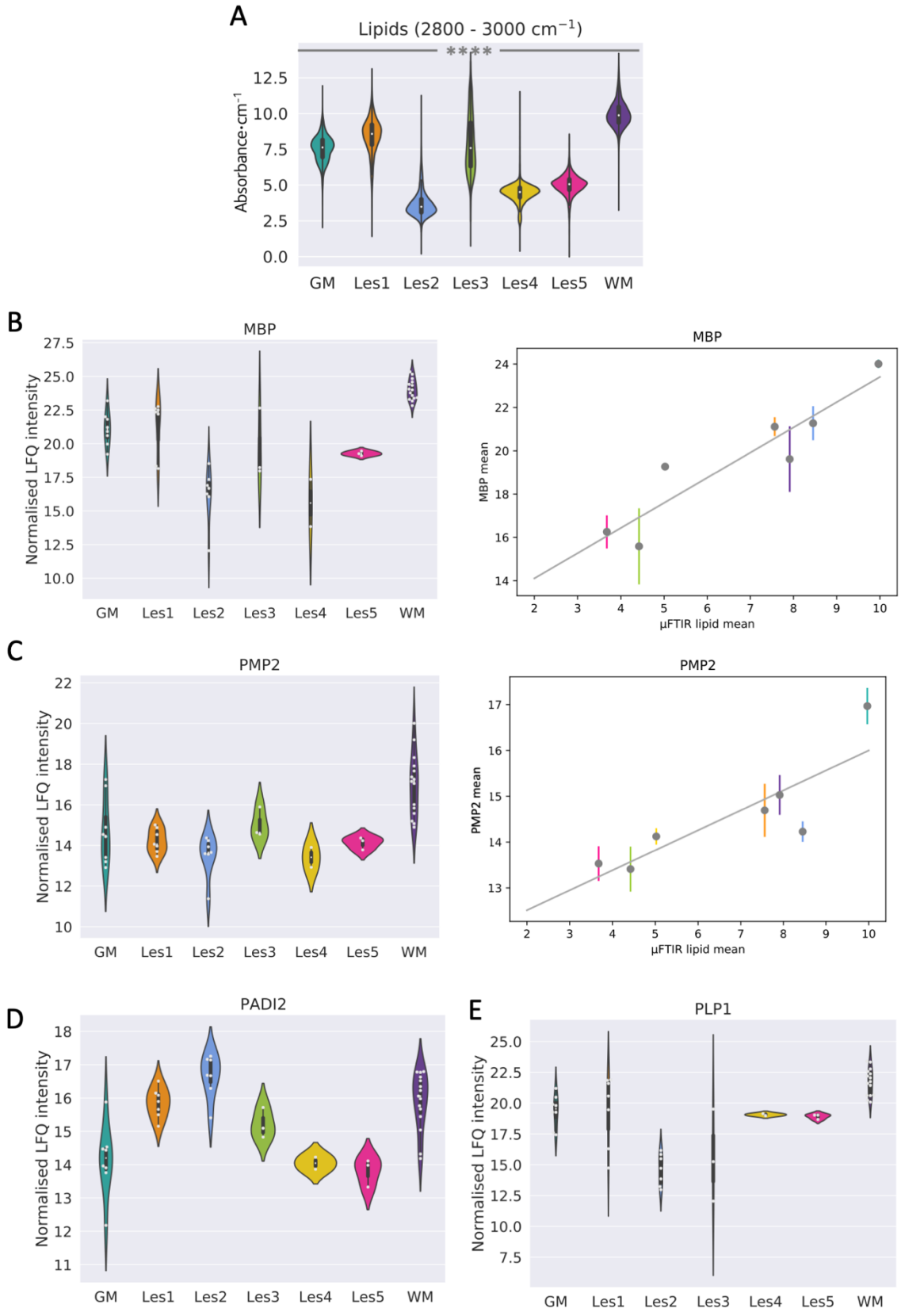


Figure 5.25: Comparison of biochemical and proteomic myelination metrics in brain samples. (A) Lipid content estimated by integration of spectral window 2,800-3,000 cm^{-1} for each µFTIR cluster (B) Normalised LFQ intensity of MBP for regions corresponding to

the μFTIR classification and correlation between MBP and μFTIR content (Pearson's R = 0.928, SEM = 2.573 x 10^{-3}) (C) Normalised LFQ intensity of PMP2 for regions corresponding to the μFTIR classification and correlation between PMP2 and μFTIR content (Pearson's R = 0.850, SEM = 1.551 x 10^{-2}) (D) Normalised LFQ intensity of PADI2 for regions corresponding to the μFTIR classification, this protein had low correlation with lipid content (Pearson's R = 0.199, SEM = 6.682 x 10^{-1}) (E) Normalised LFQ intensity of PLP1 for regions corresponding to the μFTIR classification, this protein had low correlation with lipid content (Pearson's R = 5.582, SEM = 1.928 x 10^{-1}).

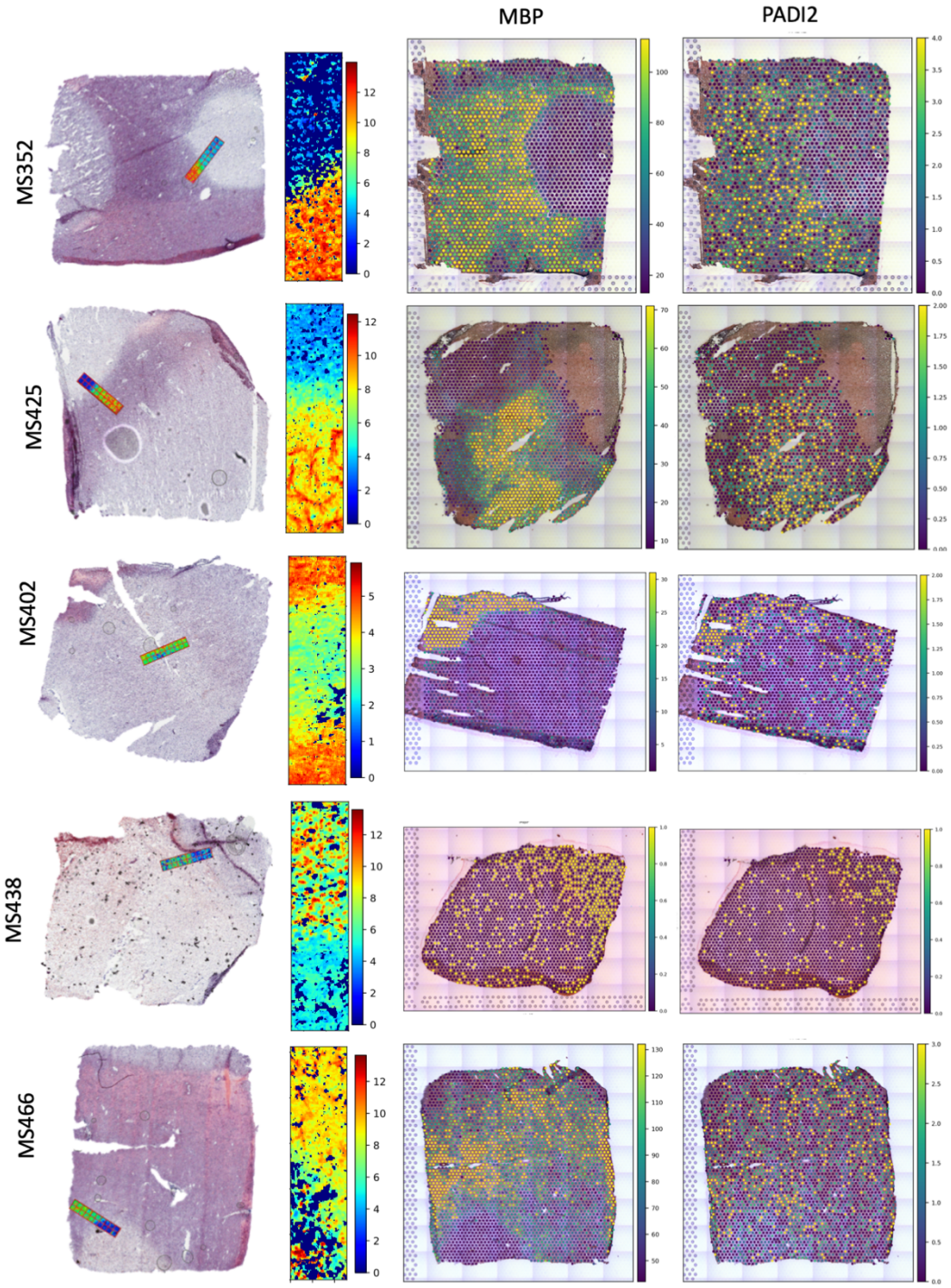


Figure 5.26: Visualisation of lipid estimation in brain samples by histological staining with ORO, μFTIR and myelin content with MBP and PADI2 as a metric for remyelination.

The maximum *MBP* expression in this sample, however, was up to 130 times lower than other samples (MS466), indicating that this lipid content did not to correspond to normal myelin expression. Interestingly, in spite of the low correlation between lipid levels and *PADI2* expression, the spatial distribution of *PADI2* had similar patterns with regions of high *MBP* expression, albeit at a substantially lower density than *MBP* expression.

Many of the findings in the brain appeared to be analogous in SC, with WMCluster1 and the µFTIR WM cluster being the ones with highest MBP content and lipid content, respectively (Fig 5.27). Of the two lesional clusters, the mean MBP expression of MS432 Cluster 3, similarly to the lipid content of µFTIR Les2, was the lowest. This correlation was not present in PLP1 expression, where GMCluster4 has the lowest mean value of all clusters. Similar to the observation in brain samples, the PADI2 expression was highest in MS432Cluster3 that had the most lipid loss (indicated by µFTIRLes2), followed by LesCluster2. Spatially the *MBP* gene expression content was more diffuse than the lipid content detected with ORO and µFTIR (Fig 5.28). Moreover, *PADI2* expression was less localised to regions of high MBP content and was diffuse across the entire SC section.

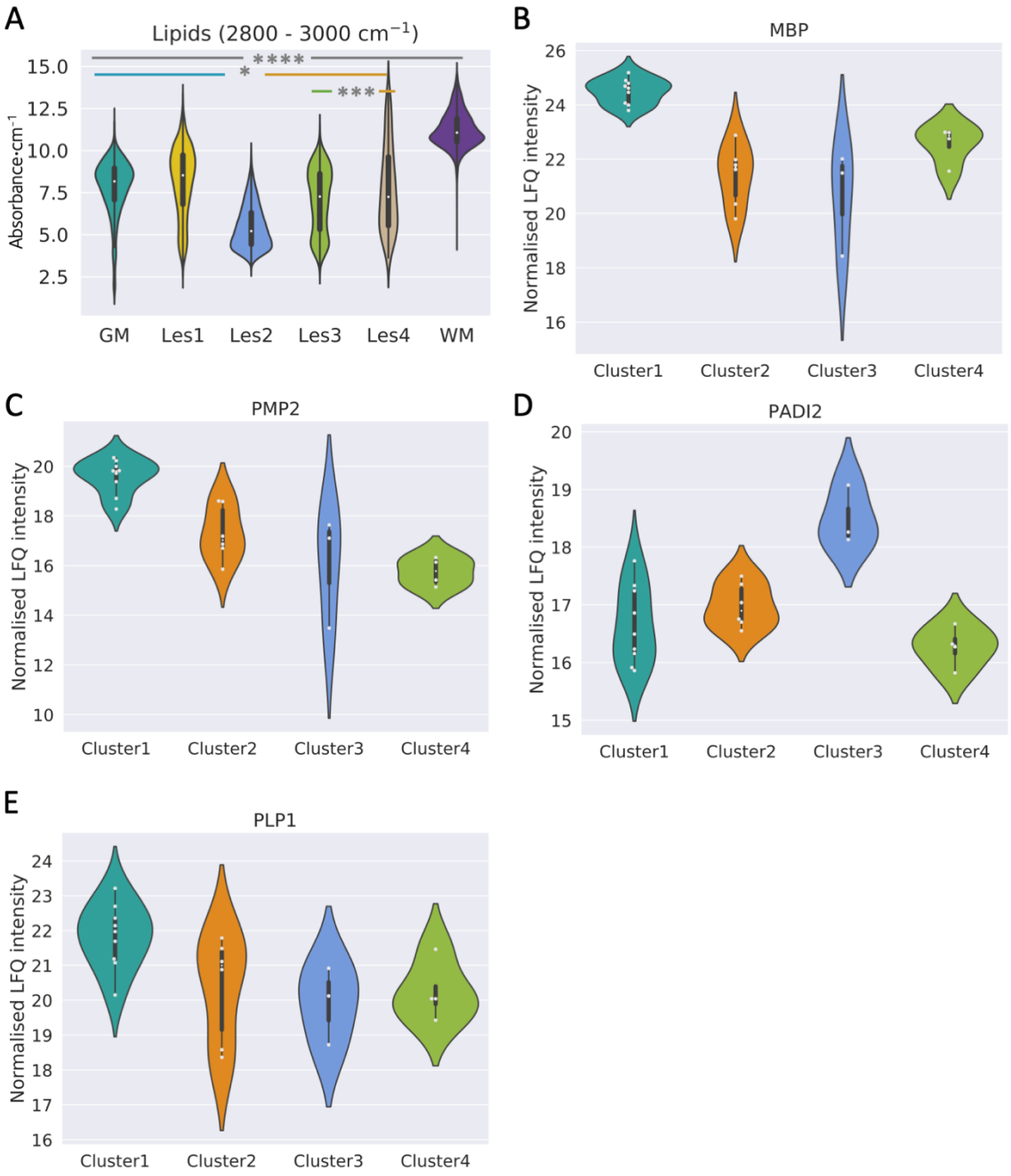


Figure 5.27: Comparison of SC lipid content estimated by µFTIR and proteomic metrics of myelination. (A) Lipid content estimated by integration of spectral window 2,800-3,000 cm^{-1} for each µFTIR cluster (B) Normalised LFQ intensity of MBP for proteomic clusters (C) Normalised LFQ intensity of PMP2 for proteomic clusters (D) Normalised LFQ intensity of PADI2 for proteomic clusters (E) Normalised LFQ intensity of PLP1 for proteomic clusters

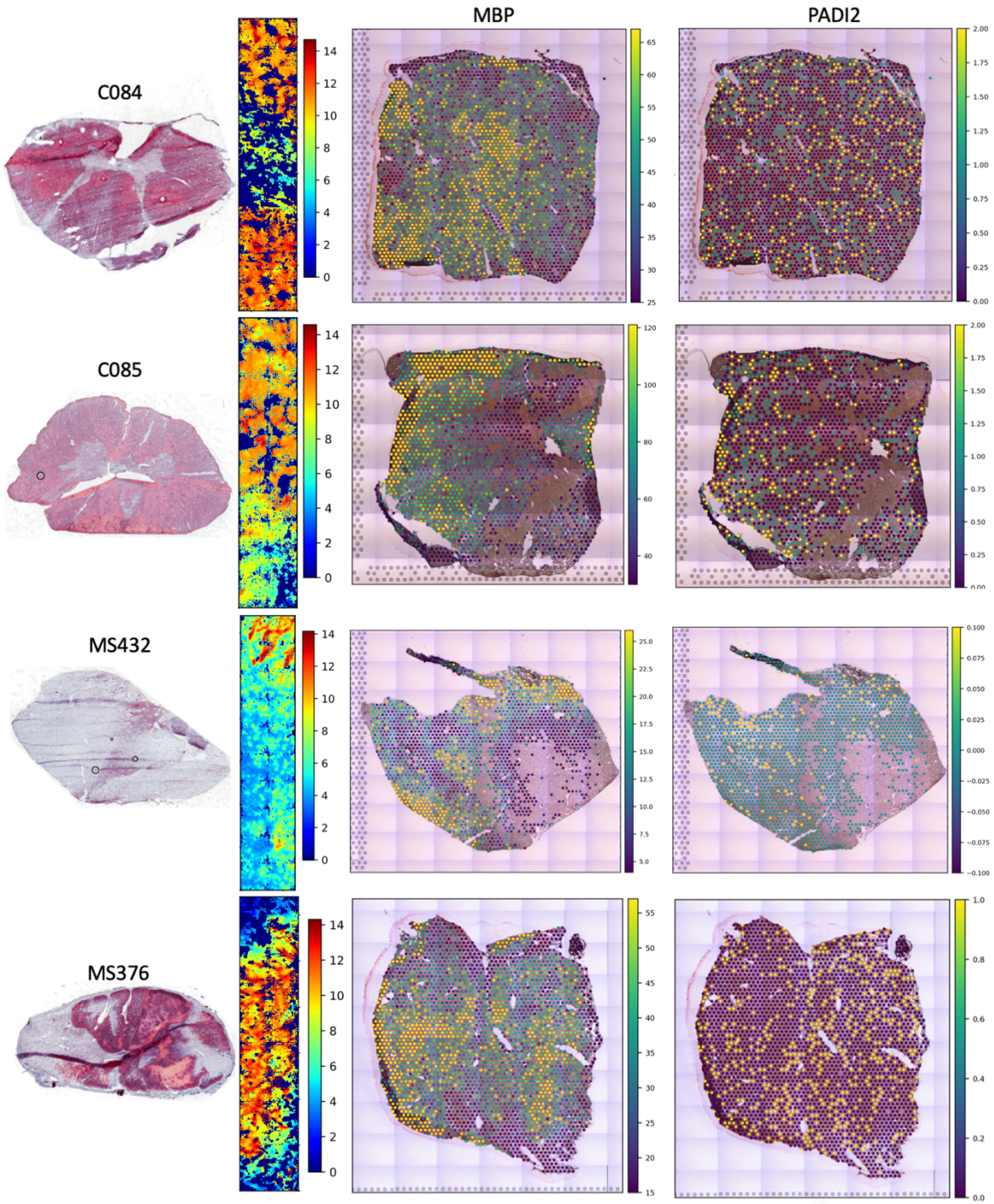


Figure 5.28: Visualisation of lipid estimation in SC samples by histological staining with ORO, µFTIR and myelin content with MBP and PADI2, as a metric for remyelination.

5.2.4.2 Comparison of lipid oxidation metrics in µFTIR and spatial proteomics and transcriptomics

The lipid oxidation was estimated using µFTIR for each cluster and compared to the normalised LFQ intensity of proteins involved in response to oxidative events, PARK7, PRDX1, PRDX2, PRDX5 and PRDX6. Due to the involvement of iron ion storage in oxidation, FTL expression correlation to lipid oxidation was also estimated (**Fig 5.29**).

In the brain, PRDX5 had a very strong correlation to lipid oxidation (Pearson's R = 0.9582, SEM = 6.709×10^{-4}). PARK7 (Pearson's R = 0.073, SEM = 8.759×10^{-1}), PRDX1 (Pearson's R = -0.327, SEM = 4.734×10^{-1}), PRDX2 (Pearson's R = 0.284, SEM = 5.366×10^{-1}), PRDX6 (Pearson's R = -0.176, SEM = 9.702×10^{-1}) and FTL (Pearson's R = -0.091, SEM = 8.468×10^{-1}) did not have strong correlations with lipid oxidation, however, with the exception of PRDX2, the expression levels of these proteins were consistently distributed across clusters: Les2 and Les3 had the highest expression levels, whilst GM had the lowest along with WM.

Evidence of expression of these proteins at the mRNA level was not present in the transcriptomic data, and therefore gene expression of other genes that have been reported to be involved in oxidation, *MIF* and *MAOB*, were mapped (**Fig 5.30**). Both genes were highly detected in GM regions, and in the case of MS352, MS402 and MS466 in a diffuse pattern at the interface between lesion and perilesional WM. The overall maximum gene expression in cases was higher than in controls for both genes, with the exception of MS352 (*MIF*) and MS425 (*MAOB*).

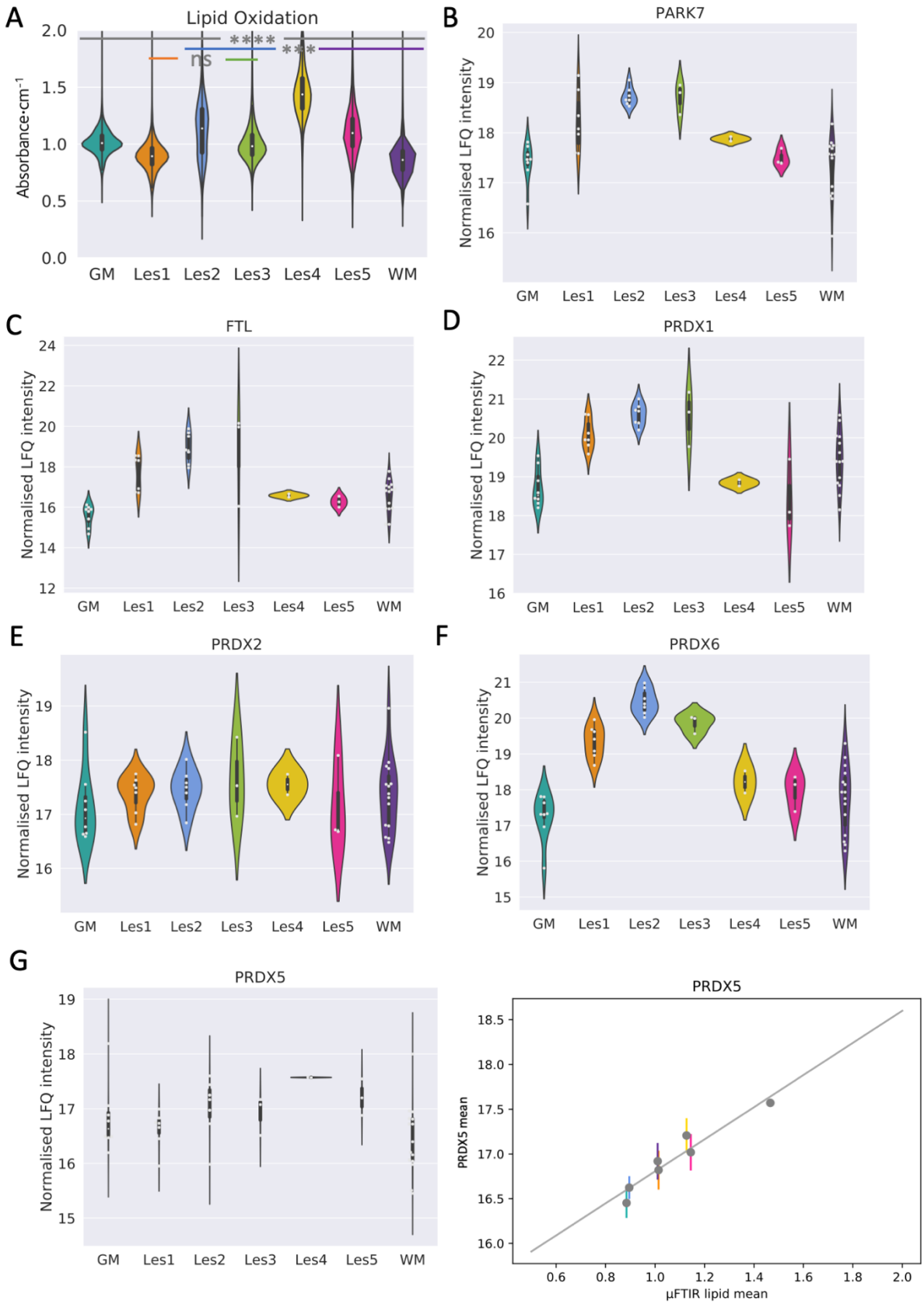


Figure 5.29: Comparison of biochemical and proteomic oxidation metrics in brain samples. (A) Lipid oxidation content estimated integration of carbonyl ester spectral window 1,700-1,740cm^{-1} normalised by total lipid content for each µFTIR cluster (B)

Normalised LFQ intensity of PARK7 for regions corresponding to the µFTIR classification, this protein had no evidence of correlation with lipid oxidation (Pearson's R = 0.073, SEM = 8.759×10^{-1}) (C) Normalised LFQ intensity of FTL for regions corresponding to the µFTIR classification, this protein had no evidence of correlation with lipid oxidation (Pearson's R = -0.091, SEM = 8.468×10^{-1}) (D) Normalised LFQ intensity of PRDX1 for regions corresponding to the µFTIR classification, this protein had no evidence of correlation with lipid oxidation (Pearson's R = -0.327, SEM = 4.734×10^{-1}) (E) Normalised LFQ intensity of PRDX2 for regions corresponding to the µFTIR classification, this protein had no evidence of correlation with lipid oxidation (Pearson's R = 0.284, SEM = 5.366×10^{-1}) (F) Normalised LFQ intensity of PRDX6 for regions corresponding to the µFTIR classification, this protein had no evidence of correlation with lipid oxidation (Pearson's R = -0.176, SEM = 9.702×10^{-1}) (G) Normalised LFQ intensity of PRDX5 for regions corresponding to the µFTIR classification, and correlation of mean PRDX5 values in relation to mean lipid oxidation content per cluster (Pearson's R = 0.9582, SEM = 6.709×10^{-4}).

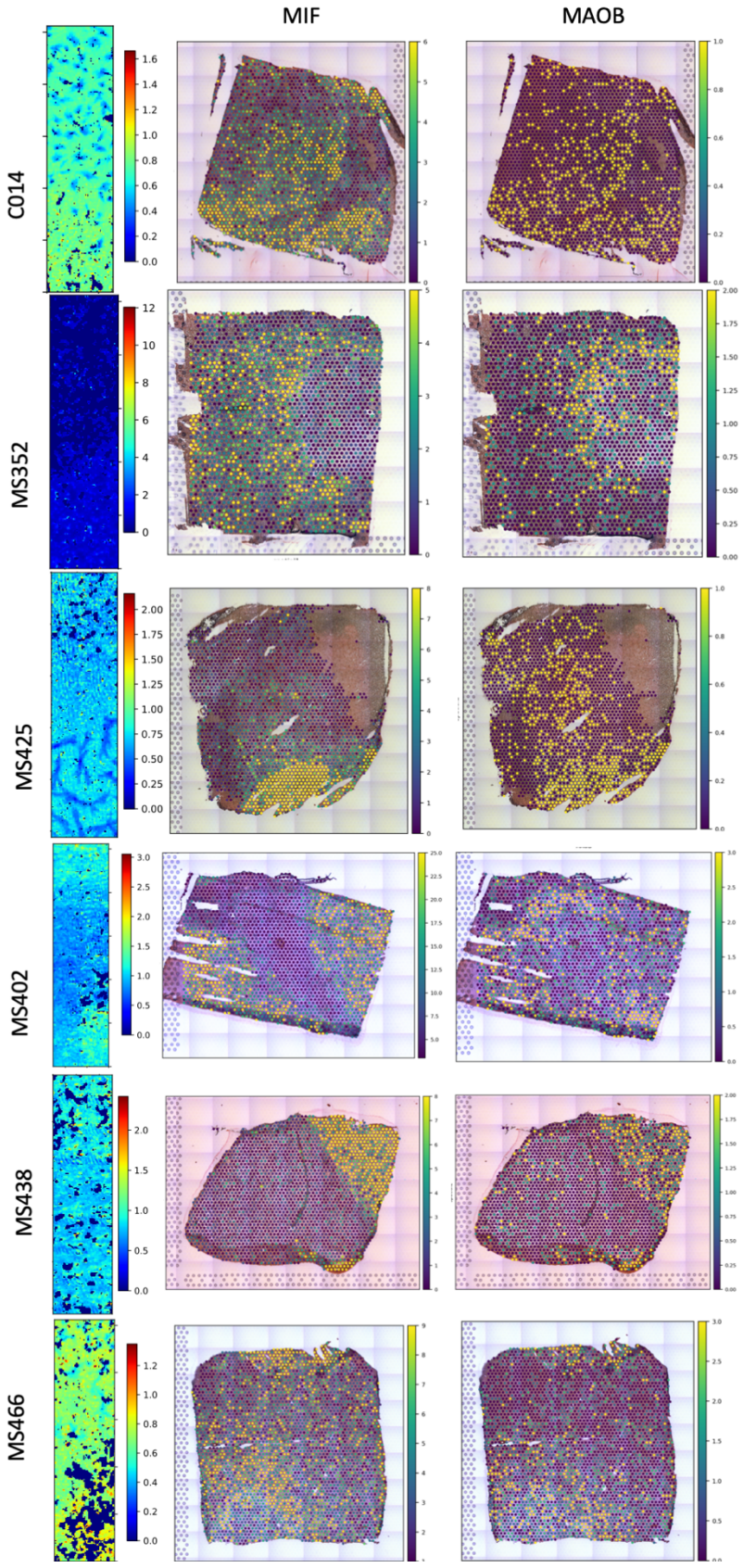


Figure 5.30: Spatial visualisation brain sample of µFTIR maps of lipid oxidation and gene expression of MIF and MAOB, genes associated with oxidative processes.

The SC samples had smaller variation in lipid oxidation levels, although consistently with the brain data lesional samples had higher expression of PARK7, PRDX1 and PRDX6 when compared to control GM and WM (Fig 5.31). Expression levels of these proteins were more pronounced in MS432Cluster3, which had signs of more advanced demyelination and damage. Reflective of the categorical data, spatial maps of SC lipid oxidation in SC had small variation (Fig 5.32). *MIF* and *MAOB* gene expression in cases, unlike in the brain, was equal to or lower than controls. There were no identifiable patterns correlating to SC architectural features.

5.2.4.3 Comparison of protein folding modifications in µFTIR, spatial proteomics and transcriptomics

µFTIR analysis of the brain tissue identified alterations in protein secondary structure composition across µFTIR clusters. In these changes Les1 and Les2 had a small increase in α-helix content, whilst Les4 and Les5 had a decreased content when compared to control WM and GM. To assess whether there were protein expression changes in proteins involved in protein folding, expression of PDIA3, HSPA5, HSP90B1, HSPA1B and HSPB1 were correlated with mean α-helix content per cluster (Fig 5.33).

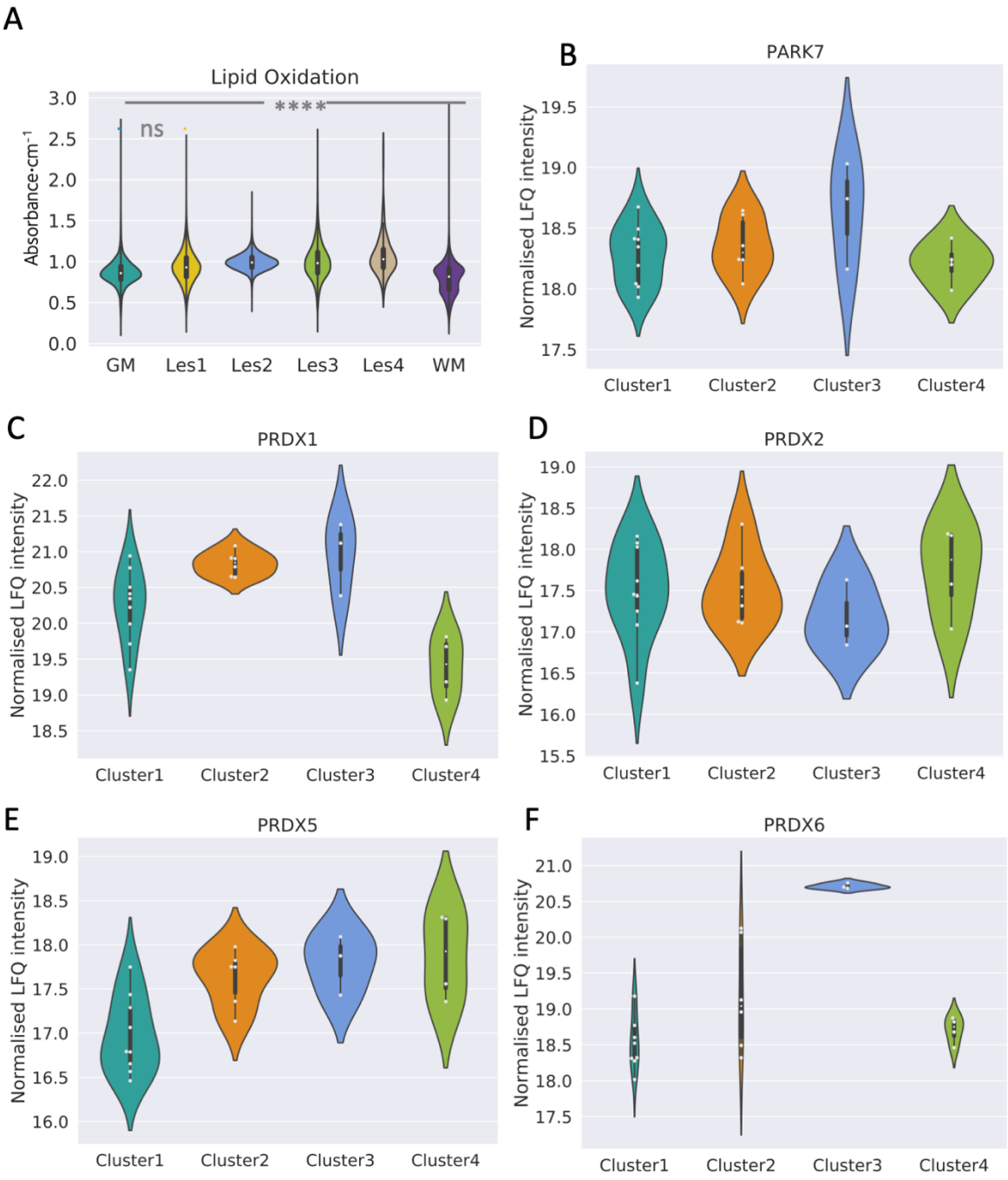


Figure 5.31: Comparison of SC lipid oxidation estimated by µFTIR and proteomic metrics of oxidation. (A) Lipid oxidation content estimated integration of carbonyl ester spectral window 1,700-1,740cm^{-1} normalised by total lipid content for each µFTIR cluster (B) Normalised LFQ intensity of PARK7 for regions corresponding to the proteomic classification (C) Normalised LFQ intensity of FTL for regions corresponding to the proteomic classification (D) Normalised LFQ intensity of PRDX1 for regions corresponding to the proteomic classification (E) Normalised LFQ intensity of PRDX2 for regions corresponding to the proteomic classification (F) Normalised LFQ intensity of PRDX6 for regions corresponding to the proteomic classification (G) Normalised LFQ intensity of PRDX5 for regions corresponding to the proteomic classification.

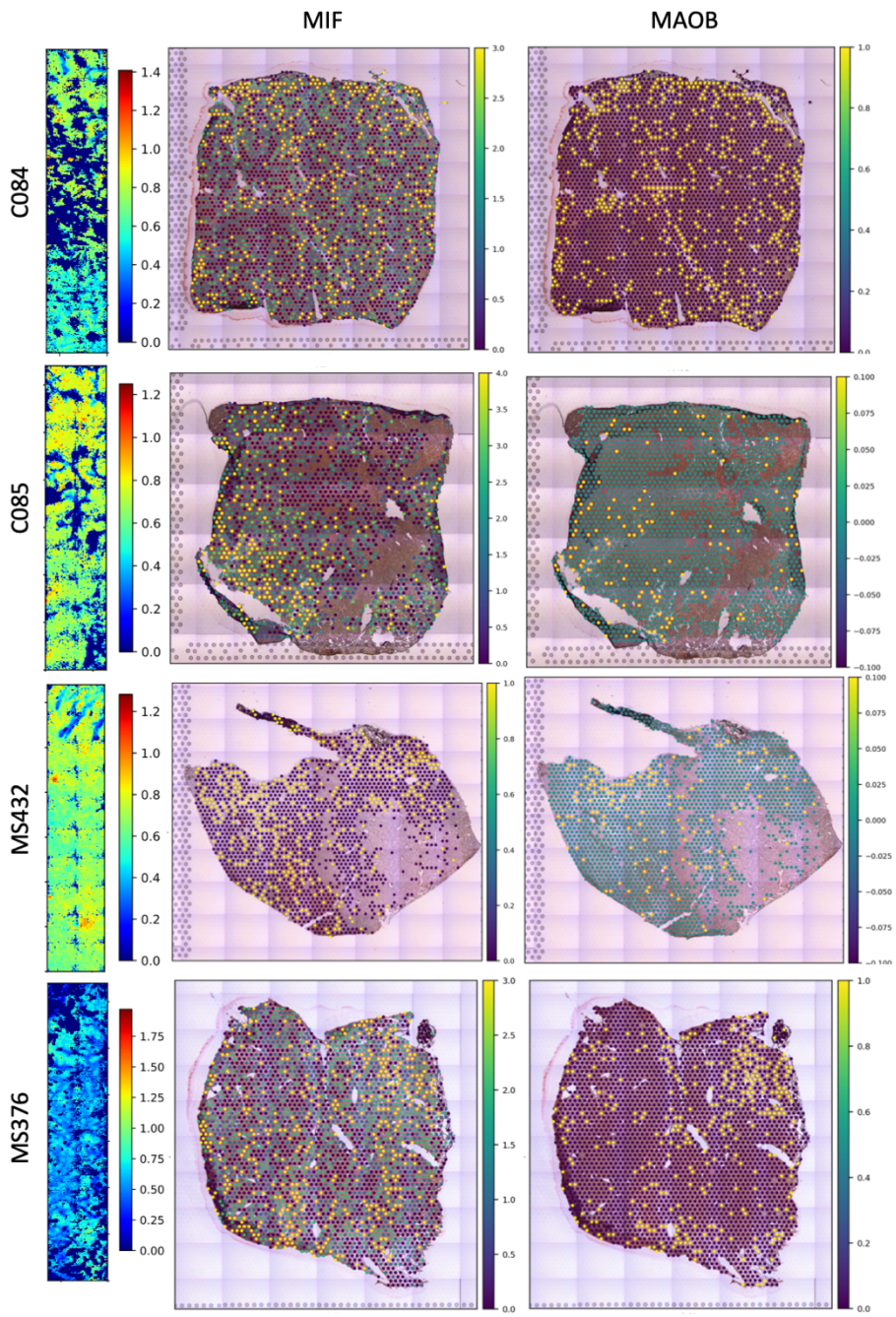


Figure 5.32: Spatial visualisation of SC sample lipid oxidation maps by μFTIR and gene expression of MIF and MAOB, genes associated with oxidative processes.

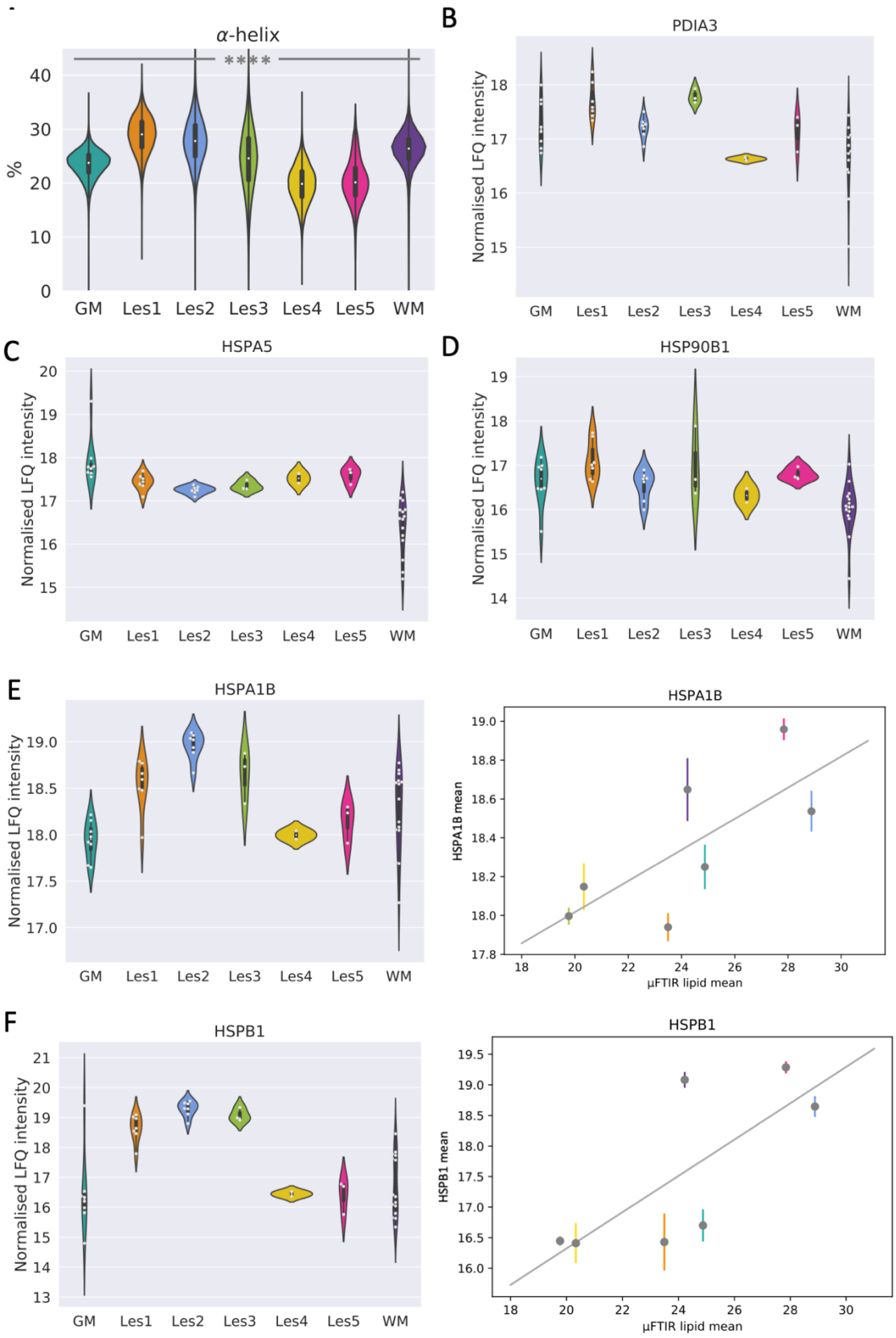


Figure 5.33: Comparison of brain protein secondary structure content estimated by µFTIR and proteomic metrics of protein folding processes (A) α-helix protein content calculated using the 1,500-1,700 cm^{-1} amide window for each µFTIR cluster (B)

Normalised LFQ intensity of PDIA3 for regions corresponding to the µFTIR classification, this protein did not have a strong correlation with $\boldsymbol{\alpha}$-helix protein content (Pearson's R = 0.494, SEM = 2.602 x 10^{-1}) (C) Normalised LFQ intensity of HSPA5 for regions corresponding to the µFTIR classification, this protein did not have a strong correlation with $\boldsymbol{\alpha}$-helix protein content (Pearson's R = -0.278, SEM = 5.460 x 10^{-1}) (D) Normalised LFQ intensity of HSP90B1 for regions corresponding to the µFTIR classification, this protein did not have a strong correlation with $\boldsymbol{\alpha}$-helix protein content (Pearson's R = 0.292, SEM = 5.256 x 10^{-1}) (E) Normalised LFQ intensity of HSPA1B for regions corresponding to the µFTIR classification, and correlation of mean HSPA1B values in relation to mean lipid content per cluster (Pearsons R = 0.736, SEM = 5.920 x 10^{-2}) (F) Normalised LFQ intensity of HSPB1 for regions corresponding to the µFTIR classification, and correlation of mean HSPB1 values in relation to mean lipid content per cluster (Pearsons R = 0.751, SEM = 5.157 x 10^{-2}).

The two strongest correlations were with HSPA1B (Pearson's R = 0.736, SEM = 5.920×10^{-2}) and HSPB1 (Pearson's R = 0.751, SEM = 5.157×10^{-2}). PDIA3 (Pearson's R = 0.494, SEM = 2.602×10^{-1}), HSPA5 (Pearson's R = -0.278, SEM = 5.460×10^{-1}) and HSP90B1 (Pearson's R = 0.292, SEM = 5.256×10^{-1}) did not have strong correlations, although PDIA3 and HSP90B1 expression appeared to have similar patterns of expression across the clusters – mainly high expression levels in Les3 and Les1 followed by Les5, Les2 and GM with lowest expression in Les4 and WM.

Spatial transcriptomic expression of *HSPB1* and *HSPA1B* were assessed (**Fig 5.34**). GM regions had the highest expression levels of both *HSPB1* and *HSPA1B*; GM levels in cases exceeded that of GM in controls two-fold for *HSPA1B* and 1.2-fold for *HSPB* for most samples, although MS466 had notably higher levels of expression of both genes than other cases (both were almost three times higher than other cases). *HSPB1* expression was particularly high in the lesion/perilesion WM interface. This pattern could also be observed in *HSPA1B*, although the signal was more diffuse and reached greater distances from the lesion (particularly visible feature in MS466).

Certain trends observed in the brain samples could also be observed in SC samples, particularly in HSPB1 and HSPA1B protein expression (**Fig 5.35**). MS432 Cluster3 which was the case with more advanced demyelination, analogous to µFTIRLes2, had both the highest α-helix content and the highest HSPA1B and HSPB1 protein levels; the second highest expression level was that of LesCluster2, further supporting a role for these proteins in WM pathology.

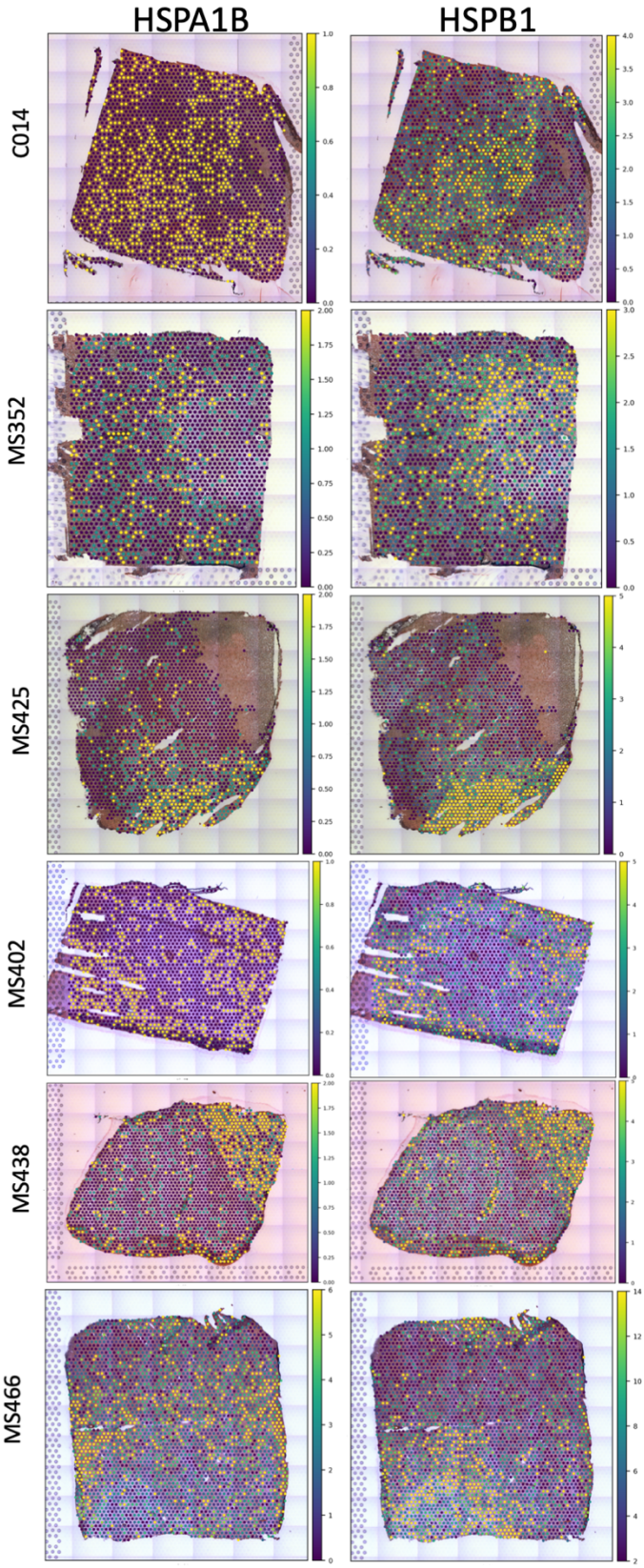


Figure 5.34: Spatial visualisation in brain samples of HSPA1B and HSPB1 gene expression; HSPA1B and HSPB1 are associated with protein folding processes.

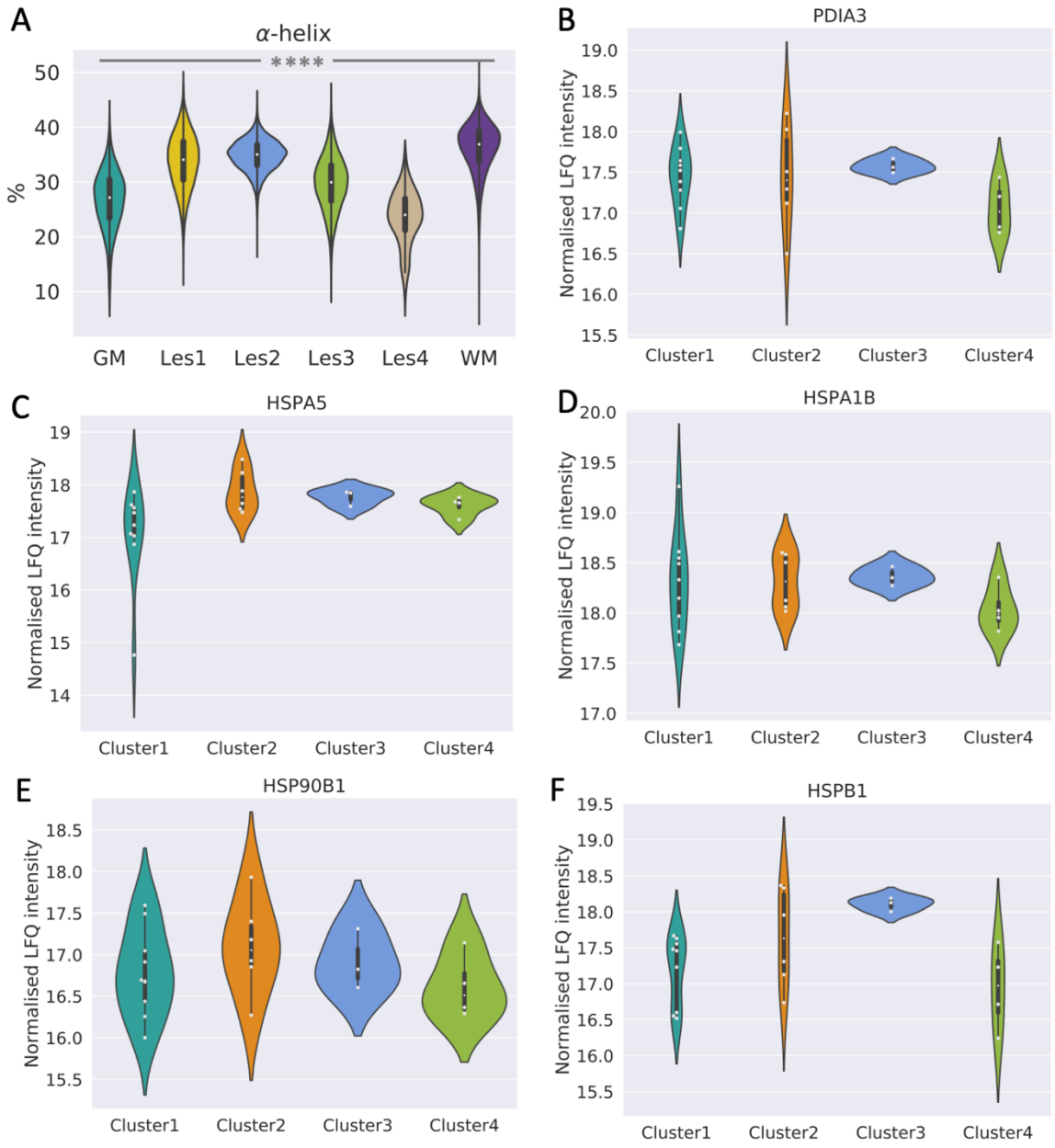


Figure 5.35: Comparison of SC protein secondary structure content estimated by µFTIR and proteomic metrics of protein folding processes (A) $\boldsymbol{\alpha}$-helix protein content calculated using the 1,500-1,700 cm^{-1} amide window for each µFTIR cluster (B) Normalised LFQ intensity of PDIA3 for regions corresponding to the proteomic classification (C) Normalised LFQ intensity of HSPA5 for regions corresponding to the proteomic classification (D) Normalised LFQ intensity of HSPA1B for regions corresponding to the proteomic classification (E) Normalised LFQ intensity of HSP90B1 for regions corresponding to the proteomic classification (F) Normalised LFQ intensity of HSPB1 for regions corresponding to the proteomic classification.

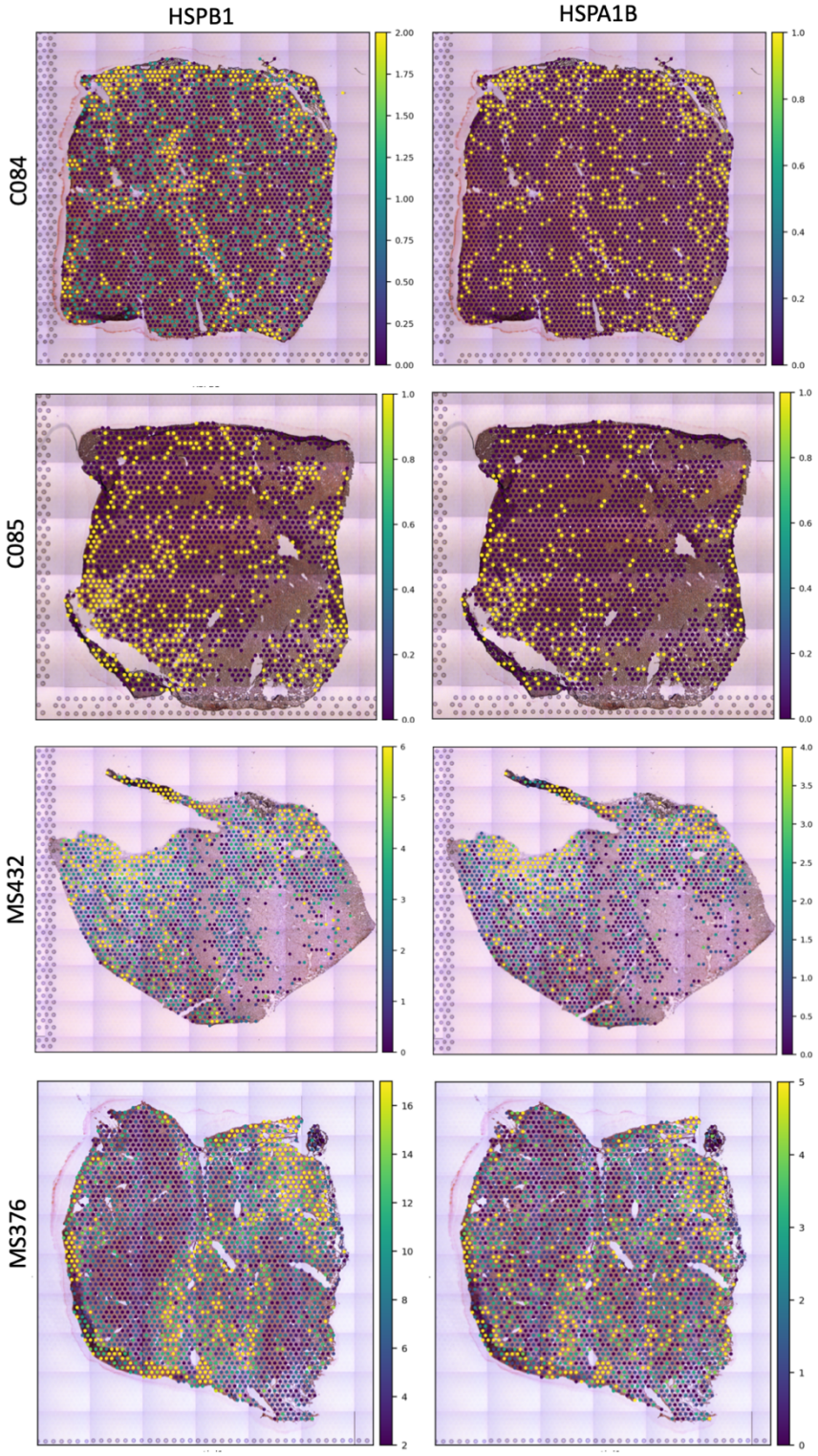


Figure 5.36: Spatial visualisation in SC samples of HSPA1B and HSPB1 gene expression; HSPA1B and HSPB1 are associated with protein folding processes.

Spatially, *HSPB1* transcript expression coincided with GM features in C084, the control sample with best resolved architecture (Fig 5.36). In cases, expression of these genes was localised particularly to areas with evidence of demyelination/damage.

5.2.5 Integrated sample annotation across modalities

Finally, sample region classifications from all technological modalities in this work were compiled (Table 5.2). This included histological, µFTIR biochemical, and spatial proteomic and transcriptomic classification, which were incorporated by comparing and correlating the composition of different tissue regions relative to the µFTIR clustering. With this, we produced an integrated annotation which revealed subtle captured differences across technologies. In this comparison, the lowest resolution classification was that performed with histological staining, followed by proteomics and µFTIR clustering had the highest resolution.

The WM and GM clusters were mostly consistent across modalities, with the exception of C064 – a sample with a feature with higher lipid content, identified as WM using histology and cell types identified by Visium, although classified as GM in both µFTIR and proteomics. The WM/GM border was unconventional for this sample.

Due to the evidence both in stereotypical demyelinated lesions as well as surrounding areas, we did not use 'lesion' or 'NAWM' terms. We also opted to use timeline-agnostic names, composed of two parts first a classification based on demyelination (mild, intermediate, severe) and specific secondly compositional features (Table 5.3; Fig 5.43). The most WM-like lesion was identified in MS352, which was

biochemically and transcriptionally classified as WM/NAWM, and yet proteomically was classified as being lesional. Two other distinct perilesional clusters were identified, 'Intermediate with leukocytes' – a lesion characterised by presence of leukocytes and phagocytic cells in the centre of the lesion, and 'Intermediate with glia' - characterised by the presence of astrocytes as well as CD74+ immune cells.

Finally, a severely demyelinated lesion, 'severe with astrogliosis', characterised by clearly distinct borders with the surrounding tissue and was robustly classified across technological modalities and was represented in MS352, MS425 and MS402 marked by substantial astrocyte as well as fibroblast presence.

Two classifications of GM with evidence of pathology were also identified. In MS425 and MS402, 'GM no astrocytes' areas were characterised by the presence of neurons and OPCs, but not astrocytes. MS402 µFTIR characterisation also indicated that the GM region of MS402 had the highest lipid oxidation across all clusters – since the GM of MS425 was not scanned, it is unknown whether this region had the same feature. Similar to MS402, the lesion in MS438 bordered directly with the GM although this GM, labelled 'GM with $\boldsymbol{\beta}$-sheet+' was compositionally and biochemically distinct: it had both lack of OPCs and astrocytes, and was almost exclusively composed of neurons. This cluster had the highest β-sheet content, although the other µFTIR-scanned GM area from MS402 also had high β-sheet content when compared to other clusters according to µFTIR analysis. Jointly, these analyses indicate that GM in the vicinity of

lesions also underwent disease-associated alterations, mainly in the OPC and astrocyte populations but also in relation to lipid oxidation and an increase in β-sheet content.

Table 5.2: Tissue region classification per sample across modalities and integrated annotation

	Histology (H&ORO)	µFTIR	Proteomic	10x Visium	Integrated annotation
C005	WM	WM	Cluster1	All OGs	WM
C014	WM	WM	Cluster1	All OGs, except OPALINhi FRYhi	WM
	GM	GM	Cluster3	Neurons OPCs ProtoAstro	GM
C064	WM	GM	Cluster3	All OGs FibrousAstro	WM
	GM			Neurons OPCs ProtoAstro	GM
MS352	Lesion	Les2	Cluster4 (lesion)	FibrousAstro ProtoAstro	Severe with astrogliosis
	Perilesion/ NAWM	WM	Cluster2 (lesion)	All OGs, except OPALINlow SLC5A11low RBFOXlow	WM with soluble factors
MS425 (top)	Lesion	Les2	Cluster4 (lesion)	FibrousAstro ProtoAstro	Severe with astrogliosis
	Perilesion/ NAWM	Les1	Cluster2 (lesion)	All OGs, except OPALINlow SLC5A11low RBFOXlow ProtoAstro CD74+ immune cells	Mild with glia
MS425 (bottom)	Lesion		Cluster3	Neurons OPCs	GM no astrocytes
MS402	Lesion	Les2	Cluster4 (lesion)	FibrousAstro ProtoAstro	Severe with astrogliosis
	Perilesion/ NAWM	Les4	Cluster3	Neurons OPCs	GM no astrocytes
MS438	Lesion	Les3	Cluster4 (lesion)	CD74+ immune cells Leukocytes ProtoAstro FibrousAstro	Intermediate with leukocytes
	Perilesion/ NAWM	Les5	Cluster3	Neurons	GM with $\boldsymbol{\beta}$-sheet+
MS466	Lesion	Les3	Cluster2 (lesion)	CD74+ immune cells Leukocytes FibrousAstro	Intermediate with leukocytes
	Perilesion/ NAWM	Les1	Cluster1	All OGs CD74+ immune cells FibrousAstro	Mild WM proteomics with glia

Table 5.3: Summary of WM lesion types based on lipid quantification from µFTIR

	WM control	WM with soluble factors	Mild WM proteomics with glia	Mild with glia	Intermediate with leukocytes	Intermediate with leukocytes with lipid-laden phagocytes	Severe with astrogliosis
Samples present in	C005 C014 C028 C048	MS352	MS466	MS425	MS466	MS438	MS352 MS402 MS425
Lipid content	9.966± 4.09 $x10^{-3}$	9.966± 4.09 $x10^{-3}$	8.453± 8.71 $x10^{-3}$	8.453± 8.71 $x10^{-3}$	7.913± 1.92 $x10^{-2}$	7.913± 1.92 $x10^{-2}$	3.677± 7.16 $x10^{-3}$
Proteomics	Cluster1	Cluster2	Cluster1	Cluster2	Cluster2	Cluster4	Cluster4
µFTIR	WM	WM	Les 1	Les 1	Les 3	Les 3	Les 2
Histology	WM	NAWM	NAWM	NAWM	Lesion	Lesion	Lesion
10xVisium	OGs	OGs	CD74+ cells FibrousAstro	CD74+ cells Astro	CD74+ cells leukocytes Astro	CD74+ cells leukocytes No OGs	Astro Fibroblasts No OGs

5.2.6 Exemplary targeted integrated analysis

5.2.6.1 Targeted investigation of CSPG4^{+} OPC populations

CSPG4+ expression has been shown to be upregulated in OPCs in the context of injury as a result of macrophage activity[263], and given the significant role of CSPG4+ OPC populations in remyelination response to injury, in particular in CNS scar formation, we investigated *CSPG4* expression spatially and in OPC subpopulations (**Fig 5.37**). Spatial mapping indicated that, unlike CNS scar architecture in the context of traumatic injury, *CSPG4* expression was localised in tissue areas surrounding injuries, but rarely localised within the lesional area. There was also no striking pattern of *CSPG4* expression at lesion rims. We also compared gene expression within the two OPC populations for cases and controls to ascertain whether this phenomena was also present in MS. In the small PDGFRA+ PLP1+ BCAShi OPC population believed to be involved in remyelination, there was a trend towards reduced mean *CSPG4* expression (cases: 0.7724± 0.8216;

controls: 2.3281±0.2436), although this difference was not statistically significant (Welch'st-test, pval=0.0822). In the PDGFRA+ PLP1- OPC population there was no substantial difference in means (cases: 1.623 ± 0.0525; controls: 1.537 ±0.0704), and the differences were not statistically significant (t-test, pval = 0.3160).

5.2.6.2 Targeted investigation of S100β gene expression

Following the identification of the *GFAP*-rich regions surrounding in particular the "severe with astrogliosis" lesions, we aimed to further characterise this tissue area. As astrocytes are recruited to sites of CNS traumatic injury to form a glia-limitans-like area, markers for astrocyte end-feet *GFAP*, *AQP4* and *S100β* were mapped in to assess whether these co-localised in an injury-associated glia-limitans, recapitulating features of the BBB glia-limitans [264, 265] (**Fig 5.38**). Surprisingly, we found that these markers not only did not co-localise clearly with a single astrocytic population (**Fig 5.39**).

Apart from astrocytes, in the CNS *S100β* is also known to be expressed in OGs with a prominent role particularly in remyelination, therefore, we compared *S100β* expression map to OG distribution in tissue as well (**Fig 5.39**). We were yet again surprised to find that, while this protein appeared to have niches of highly concentrated distribution in MS, particularly surrounding areas with an "severe with astrogliosis" classification (MS352, MS402 and MS425), its expression pattern did not obviously overlap with that of any particular cell population identified. The OG subtypes it had most similarity in distribution to were OG OPALINlow SLC5A11hi RBFOXhi, including the OG OPALINhi_FRYhi population identified as more abundant in MS.

In order to better elucidate the source of the *S100β* expression, we compared the mean expression across cases and controls for astrocytes and oligodendrocytes. We compared the mean expression in astrocytes which was found to not be significant (t-test, pval = 0.2225). Notably, however, the astrocyte population marked by *S100β* expression – fibrous astrocytes – is more abundant in cases, as discussed in chapter 4. In a similar comparison, we observed that the mean expression in oligodendrocytes was significantly lower (Welch's t-test, $p = 6.3402 \times 10^{-64}$) suggesting that unlike traumatic CNS injury its role in remyelination is overall less pronounced in MS.

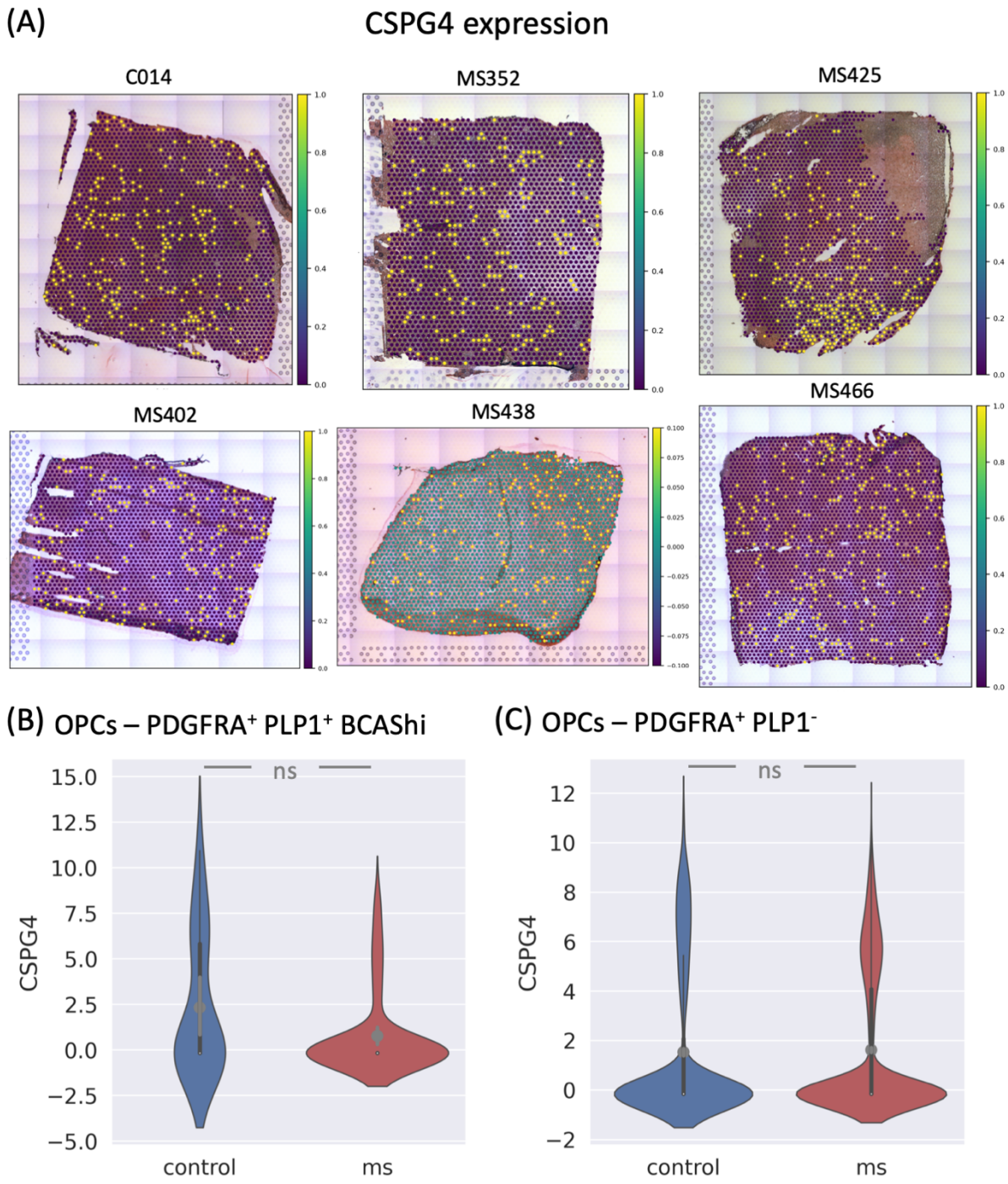


Figure 5.37: Assessment of CSPG4 gene expression spatially and in OPCs (A) Spatial mapping of CSPG4 expression in example control and MS. (B) MS vs. control CSPG4 snRNAseq expression distribution in PDGFRA^{+} PLP1^{+} BCAShi OPCs, with mean indicated with grey dot. The mean expression between both groups was not statistically significant (Welch'st-test, pvalue=0.0822) (C) MS vs. control snRNAseq expression distribution in PDGFRA^{+} PLP1^{-} OPCs, with mean indicated with grey dot. The mean expression between both groups was statistically significant (t-test, pval = 0.3160).

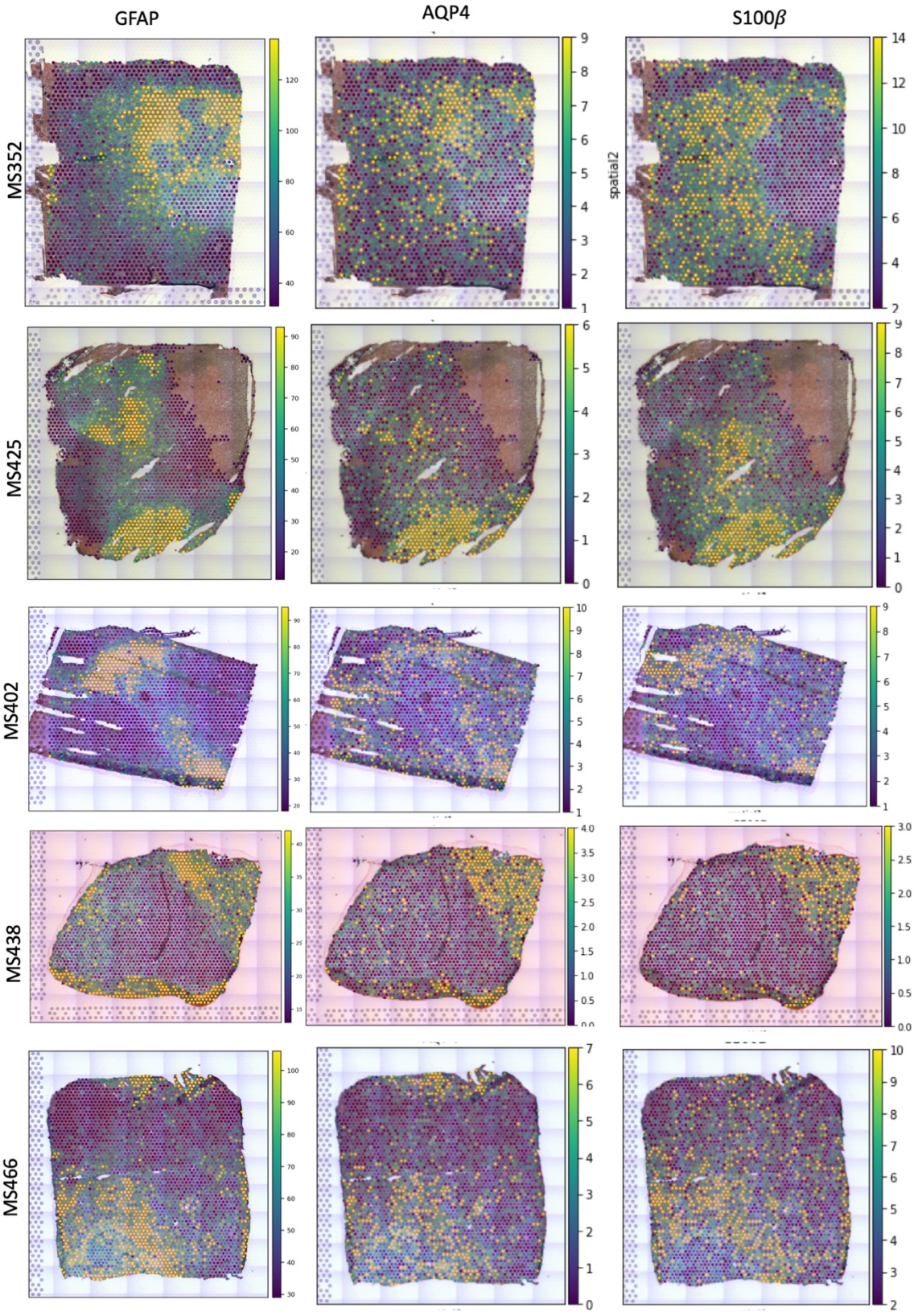


Figure 5.38: Visualisation of spatial gene expression of glia limitans astrocyte markers GFAP, AQP4 and S100β.

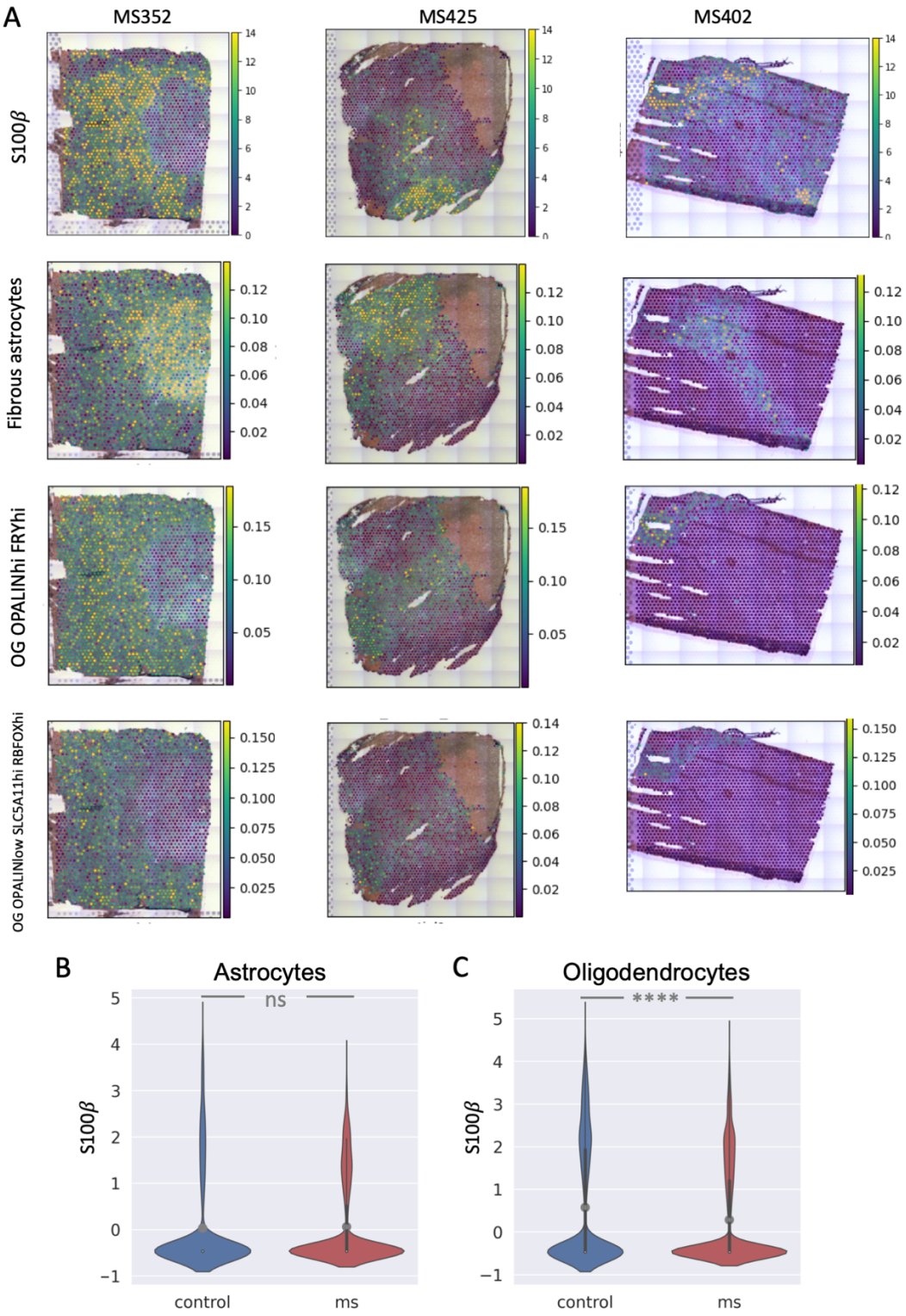


Figure 5.39: Identifying S100β-expressing populations. (A) Comparison of S100β spatial gene expression to astrocyte and OG subpopulations. (B) MS vs. control snRNAseq expression distribution in astrocytes, with mean indicated with grey dot. The mean expression between both groups is not statistically significant (t-test, pval = 0.2225) (C) MS vs. control snRNAseq expression distribution in oligodendrocytes, with mean indicated with grey dot. The mean expression difference between the groups is statistically significant (Welch'st-test, pvalue=6.3402 x10^{-64}).

5.3 Discussion

5.3.1 Spatial transcriptomics reveals differential cellular co-localisation networks in the SPMS CNS relative to controls that correlates with high-resolution biochemical profiles

By performing spatial transcriptomics with tandem snRNAseq-dependent deconvolution, we were able to identify key high-resolution features associated with CNS architecture and pathology. As expected, the GM/WM division was prominently observed in control and cases. Network analyses showed that the GM cellular co-localisation cluster was mainly composed of neuronal cell types, OPCs and protoplasmic astrocytes, whilst the WM co-localisation cluster was enriched for mature OGs. Interestingly, when considering the OG distribution across the tissue sections, a gradient of *SLC5A11* and *RBFOX* expression in OPALINlow OGs represented a spatial gradient whereby OPALINlow SLC5A11low RBFOXlow types tended to be at the border between WM and GM, and expression of these genes increased further away from this border. This may be indicative of a layered spatial gene expression gradient among OGs within the WM which has not been previously described, but which may in some ways be analogous to the spatially-resolved cortical layers of neurons or astrocytes that have been previously described[257]. Such a spatial gradient of OGs may represent OGs in the process of maturation migrating from the GM – which is rich in OPCs – to the WM.

Despite the GM/WM division being largely maintained in the SPMS brain tissue, pathology-associated differences relative to the controls were nevertheless observable. In the GM subnetwork CUX2 neurons were less abundant and less connected to other cell types. In the WM subnetwork CD74 cells and fibrous astrocytes were more abundant; the strongest co-localisation relationships remained between mature OGs, although here fibrous astrocytes were also prominently co-localised with OPALINlow populations. This relationship may relate to the role of astrocytes in OPC-to-OG differentiation[266].

Having identified the network-co-localisation patterns which characterised the WM and GM regions in the brain as classified by μFTIR clustering, we hypothesised that other μFTIR classifications in cases might have compositional relevance. The mapped distribution of cell types, co-localisation network and μFTIR classification were jointly analysed for each MS sample. Indeed, certain patterns emerged in the co-localisation networks which corresponded well with the main identifiable compositional features: the Les2 μFTIR cluster corresponded to the presence of fibrous and protoplasmic astrocytes; Les1 corresponded to CD74+ immune cells and mature OGs; and Les3 corresponded to strong co-localisation relationships between CD74+ and protoplasmic astrocytes.

The co-localisation networks for the SC case and control samples also revealed several key pathology-associated differences. In the control networks, neurons have a peripheral role in the network. The most frequently present and co-localised cell types

are OGs and fibrous astrocytes, followed by protoplasmic astrocytes, CD74+ immune cells and the neuronal subtypes SST and ID2.

In MS, CD74+ immune cells are more frequently present and tend to co-localise with fibrous astrocytes and endothelial cell types. This hints at increased proliferation and/or migration of both CD74+ immune cells and endothelial cells when compared to controls. The segregation between cell types characteristic of WM and GM was no longer detectable in the network, with neurons, immune cells, astrocytes and certain OGs frequently co-localising in the same spot. Interestingly, in MS the only cell type OPCs frequently co-localised with were CD74+ immune cells, whilst in controls they co-localised frequently with other cell types as well, such as mature OGs and astrocytes. Overall, MS-associated pathology causes substantial architectural changes in the SC, with a decrease in OG representation and a concomitant more prominent role of CD74+ immune and endothelial cell types.

5.3.2 Proteomic lesion classification in brain and SC

Hierarchical clustering of proteomic data collected from brain and SC samples indicated that in both CNS tissue types there are two main pathology-associated clusters could be classified - one associated with less severe demyelination and the second associated with more severe demyelinated. The heatmaps of differentially expressed proteins reveal a proteomic spectrum between WM and severely demyelinated lesions, where intermediary clusters had proteomic features from both WM and lesions.

For the brain samples, the overall segregation of tissue areas histologically denoted to be WM and NAWM by hierarchical clustering indicated that proteomics can reveal that NAWM is not identical to WM, with the NAWM showing an increase in proteins such as GFAP (indicative of astrocytes), VIM, HSPB1, ANXA1 and ANXA2 (the latter can have an anti-inflammatory effect). GSEA analysis also revealed that NAWM proteins were associated with pathways relating to endothelial cell migration, cellular response to vascular endothelial growth factor and platelet aggregation, for instance. A comparison of proteomic clusters enriched for WM versus lesional tissue areas revealed additional differences including IGHG1 and FTL expression in lesions, and higher PLP1 and MBP expression in the WM (which were not higher relative to the NAWM). GSEA analysis of this comparison indicated differences in pathways associated with nucleobase catabolic process, regulation of potassium ion transport, regulated exocytosis, protein localisation to cell surface, platelet aggregation and degranulation, myelination, cell-cell adhesion and substantia nigra development, for example.

For the SC samples, comparisons of proteomic clusters enriched for WM and for NAWM/lesional tissue areas revealed milder and more pronounced differences in the NAWM/lesion clusters relative to the WM, with lower expression of MBP which is constituent of myelin, BCAS1 which has been shown to be expressed in myelinating OGs, and PMP2 which has a role stabilising myelin. GSEA analysis of the SC samples included terms related to cell adhesion, endothelial cell development, CNS development, maintenance of blood-brain barrier, myelination, metallopeptidase activity, chemokine

ligand production, calcium homeostasis and regulation of protein metabolic processes. The latter is intriguing in light of the μFTIR findings of increased estimated protein content in advanced lesions. A reduced rate of protein metabolic process could delay protein turnover, and result in an accumulation of proteins detectable biochemically. Many of the identified pathways are in line with the pathway analysis of WGCNA transcriptomic analysis (see chapter 4), confirming the presence of demyelination of lesions, changes to calcium homeostasis and presence of CNS development signals particularly in late lesions.

The proteomic analysis suggests certain similarities in SPMS lesions in the brain and SC, for examples in relation to the role of endothelial cells, although differences were also noted, for example increased ANXA1 and ANXA2 expression in NAWM/lesional tissue areas was observed in the brain but not in the SC samples.

5.3.3 Correlating biochemical and proteomic changes

5.3.3.1 MBP expression levels correlate with lipid content measured by μFTIR in the brain

Several studies have performed μFTIR analysis of MS and have attributed the changes in lipid content in the 2800-3000 cm^{-1} spectral window to the process of demyelination[158, 159]. While intuitively plausible and in accordance with known MS pathological features, this has not, to our knowledge, been explicitly validated. Therefore, we compared the estimated changes in lipid content to myelin content based on proteins associated with myelin.

The proteins chosen for this comparison were MBP, PLP1 and PADI2 due to their various roles associated with myelination. Of these, PADI2 and PLP1 had a low correlation with total lipid content; this could relate to the role these proteins have in remyelination [267, 268]. In particular, this is supported by PADI2 being most highly expressed in the lesional WM clusters correlating with most severe demyelination – in the brain with Les2 and in the SC with MS432Cluster3.

MBP is the second most abundant protein in myelin and constitutes 8% of total myelin protein, making this a fair indicator of myelin[269], and was the protein with the strongest correlation with the µFTIR estimated lipid levels ($R = 0.928 \pm 2.573 \times 10^{-3}$). A comparison of the µFTIR lipid map, ORO lipid staining and the transcriptomic map of MBP also revealed a consistent architecture across most MS brain cases.

Interestingly, the ratio of the maximum gene expression in relation to the maximum lipid content in each sample ranged between 5.8 and 10 across all samples, except for MS438 which had a ratio of 0.077. In this sample the region with a higher lipid content had a low MBP content, and the region with lowest lipid content had the highest MBP content - suggesting that there was no correlation between lipid and MBP content in this sample. Notably, this sample had a distinct histological profile with low ORO staining and low sample integrity. The overall distribution of lipid as estimated by µFTIR was also distinct – lipid was concentrated in small round clusters. Given the composition of this sample, with a high content of foam cell-like phagocytes, these data indicate that the

high lipid content and low MBP content co-occur due to the presence of phagocytosing microglia/macrophages.

Thus, while the interpretation of lipid content as an indicator of myelination in the brain is justified in most cases, MS438 showcases how this may at times give an incomplete understanding of lesion pathology.

The spatial correlation between myelination and MBP levels was not as reliably featured in the SC samples. While in C084 and MS376 the tissue architecture was revealed to be as expected, this was not the case with MS432 and C085. This may partly be due to the more dense composition of the SC due to particularly high myelin content – thus making permeabilization (and hence RNA and protein capture) a process sensitive to changes even within a given SC classification. MS432 specifically already had lower lipid levels as indicated by µFTIR and ORO analysis, and therefore the lack of a stereotypical SC architecture transcriptomically is likely to reflect the advanced level of damage in this sample.

5.3.3.2 PRDX5 correlates with µFTIR lipid oxidation quantification

In order to correlate µFTIR-estimated lipid oxidation with proteomic evidence, protein expression of a panel of proteins which have roles related to oxidation were selected from the detected proteins. Among these were peroxiredoxins which are an antioxidant mechanism which neutralise ROS and NOS, and are known to have a role in neurodegeneration. FTL was also quantified, given the role of this protein in iron storage

which may affect oxidative processes and PARK7, given its suggested role as a sensor for oxidative stress.

PRDX5 had a very strong correlation with the µFTIR-estimated oxidation ($R = 0.928 \pm 2.573 \times 10^{-3}$), but there was no strong correlation with the other proteins. This may in part relate to the fact that, although PRDX1, PRDX2 and PRDX3 expression has been detected in MS previously and appeared upregulated in MS lesions compared to surrounding NAWM, this expression seems to correlate more strongly with early/active lesions [270-272] which are less likely to be represented in this sample set of SPMS lesions. While PRDX5 is present in astrocytes in acute lesions displaying inflammation, it is also present in post-reactive astrocytes in chronic lesions with abated inflammation[273]. Thus, the strong correlation between oxidation levels and PRDX5 protein expression may relate to the enduring PRDX5 signal in chronic/late lesions that are more characteristic of SPMS.

It is worth noting that PARK7, FTL, PRDX1 and PRDX6 had similar expression patterns across clusters. In particular, Les1, Les2 and Les3, identified through 10x composition analysis to be the WM lesion clusters, had high co-expression of these proteins. In particular, Les2 which had the most evidence of demyelination had the highest expression of these proteins. Thus, the lack of correlation of these proteins may relate to a WM-specific mechanisms; the inclusion of GM in the correlation analysis may confound the correlation estimate.

5.3.3.3 No strong correlation between proteins involved in protein-folding and changes to protein secondary structure folding estimated by µFTIR

µFTIR analysis predicted a small increase in α-helix content in Les1 and decrease in Les3, Les4 and Les5, and in these a corresponding increase in β-sheet content in lesion clusters. Here we compared the α-helix content to the levels of proteins involved in protein folding: PDIA3, HSPB1, HSPA5, HSP90B1 and HSPA1B. HSPB1 and HSPA1B had the strongest correlation with α-helix content measured by µFTIR, although these correlations are still inconclusive. There was evidence of involvement of these proteins with pathological processes, regardless of their involvement specifically in the changes to protein secondary structure content. They were most highly expressed in WM lesions, with increased presence in severely demyelinated lesions both in the brain (Les2) and SC (MS432Cluster3).

Spatial mapping of these genes revealed that HSPB1 tended to localise at the interface of the MS352, MS402 and MS466 lesions with the WM perilesion. In MS425 and MS438 it tended to localise in the GM. HSPA1B, on the other hand, was localised mostly in the GM of MS 425, MS402, MS438. The highest expression, however, was in the perilesion of MS466. In the SC both genes tended to co-localise in regions of active damage. Given the concomitant decrease in α-helix content and increase in β-sheet content specifically in the GM, the lack of specificity to the GM puts into question the role of these proteins in this alteration in protein secondary structure content.

Finally, we highlight that there are well known biological reasons why the correlation between the transcriptome and the proteome is known to be imperfect. Among these are post-translational modifications to the transcriptome which have an impact on protein synthesis (e.g.: splicing events) as well as disparity between the time of translation and protein degradation (e.g.: cases of protein remaining deposited in tissue long after gene expression for that protein has ceased or stabilised). This highlights the drawbacks in treating protein data as a validation, such as proteomic data or IHC staining, and highlights the value of other RNA-targetting technologies for validation, such as FISH.

5.3.4 Integrated annotation and degree of pathological severity classification

By classifying the tissue types based on myelination (histology) and composition (10xVisium), we have identified two regions consistent with previous lesion classification from the Bö/Trapp and Vienna consensus system: inflammatory and demyelinating (MS438 and MS466) and inflammatory rim, hypocellular centre and demyelinating (MS352, MS402 and MS425).

However, in many regards our analysis reveals a complexity in SPMS lesion characterisation that goes beyond existing classification systems which are largely based on RRMS. For instance, many definitions of inflammatory include the presence of leukocytes[67, 105], which are not always prominent in SPMS lesions, with glial cell types such as activated astrocytes having more prominent pro-inflammatory roles in chronic lesions. Secondly, in RRMS there is less evidence of diffuse damage in areas surrounding

lesions, while here we identify several pathology-associated features in these areas which have quantitative differences. µFTIR characterization of the perilesion indicated the reduction of lipid content is not limited to demyelinated lesions, but also extended to a lesser degree to the surrounding tissue. Combined, spatial proteomic and 10xVisium analysis also revealed a number of signals in the perilesion associated with damage, such as presence of antibodies and complement proteins, which may be associated with the pathological mechanisms causing diffuse injury and expansion of SPMS lesions.

We have used our integrated dataset to create an unbiased and quantitative approach to classify several degrees of pathology which are timeline-agnostic and which do not strictly rely on nomenclature such as 'lesion' or 'NAWM', operating under the assumption that whether histologically evident or not, in SPMS areas from a same sample have varying degrees of damage. While many classification systems attempt to create a timeline based on the degree of demyelination, there is sufficient evidence of remyelination in our analysis – including an *PADI2* gene and protein expression and *S100β* gene expression in OGs, that we opted to not use lipid content as a vector for timeline prediction.

In these degrees of damage that we observed, the first was biochemically consistent with the WM, whilst having the presence of soluble factors associated with pathology such as antibodies and complement proteins. The second had some lipid loss, a small increase in amide and type IV collagen with slightly increased presence of astrocytes and microglia. The third had diffuse demyelination without clear borders with

surrounding areas and variable overall lipid content, and a particularly high microglial population with evidence that these were myelin laden phagocytes. Finally, the last had large regions of focal demyelination with clear borders and a large population of astrocytes, with some microglia and fibroblasts contained within the demyelinated borders; these areas were marked by high protein and collagen content. The rim of these areas also was lined by GFAP-expressing astrocytes, which may be a glia-limitans-like feature.

Certain features can be generalized based on these classifications – for instance, the more lipid loss in a region, the higher its amide and collagen content, as well as its astrocyte population. The presence of GFAP-expressing astrocytes surrounding areas of severe demyelination is associated with clear borders of areas with demyelination, and these areas also are populated by fibroblasts and have high type IV collagen content, associated with CNS scar formation.

5.3.5 Evidence of non-conventional CNS scarring in SPMS

In differential abundance analysis of snRNAseq data we identified that fibrous astrocytes, activated phagocytes and fibroblasts were more abundant in cases than in controls. Due to the roles of these cell types in scar formation in the CNS, we hypothesised the presence of scar formation in MS lesions. To investigate this further, the localisation of these cell types was mapped to compare to known CNS scarring architectural features observed in the context of traumatic injury[173, 204, 262], including fibrotic scar components as well as glial scar components.

5.3.5.1 Evidence of a fibrotic component in MS CNS scarring

Abundant evidence of fibrotic scar formation specifically in active MS exists, both in the context of human samples and murine EAE models[209, 274].

Here we present compositional, ligand-receptor interaction and architectural evidence of fibrotic scar formation in chronically demyelinated SPMS lesions. This was first evidenced by snRNAseq analysis, which identified differential abundance of fibroblasts when comparing cases to controls, with cases having the larger abundance. Interestingly, DGE analysis indicates that the endothelial compartment of controls had several differentially expressed genes associated with cell cycling/mitotic processes suggesting decreased proliferation in the endothelial compartment of cases. A known mechanism which likely enables the presence of fibrotic processes in the absence of cell cycling has been previously described, where myelin engulfment by endothelial cells confers to these cells the ability to stimulate the production of fibrotic components[275].

Several disease specific ligand-receptor interactions between extracellular matrix (ECM) proteins synthesised by fibroblasts and pericytes and cell adhesion molecule (CAM) receptors expressed by other cell types supported the presence of fibrotic scarring as well: CNS fibrotic scarring has ECM deposits expressed by fibroblasts and pericytes in cavity of injury as a way of maintaining the physical integrity of an injured area, and CAM expression is key for fibrotic scar development[276].

Spatially this hypothesis is supported by the localisation of fibroblasts in the lesion core of samples with evidence of glial scar formation – in particular, in areas classified as

'severe with astrogliosis' in MS352, MS402 and MS425. The areas of highest fibroblast content in these samples were also the cluster with highest COL4A2 levels, a protein which has been shown to be necessary for the formation of the fibrotic component of CNS scarring[204].

The presence of ECM deposits in these areas may at least in part enlighten spectral interpretation of μFTIR data. Consistently μFTIR studies of MS have observed an increase in absorbance in the amide region in the context of MS, however the biological interpretation of this feature has remained elusive (see chapter 3). The spectral fingerprint is notably complex, but as one of its main features the absorbance of the amide region for collagen is notably high. Moreover, collagen deposition has been linked with a greater β-sheet content within a given sample. Both of these features are consistent with our findings in μFTIR.

5.3.5.2 Glial CNS scarring component is present in MS chronic lesions but lack CSPG4+ OPCs

Given existing reports of the presence of glial components of CNS scars in chronically demyelinated MS lesions[106], this was the first component to be investigated. Among brain samples MS352, MS402 and MS425, the regions classified as "severe with astrogliosis" in integrated annotation had architectural evidence of glial scar formation: these lesions had a strong astrocytic presence with a GFAP-expressing astrocyte barrier surrounding the demyelinated area. Notably, however, this barrier was diffused, not containing areas of damage; diffused astrocyte barrier formation has been

previously observed in other contexts, and has been linked to proteomic alterations such as altered integrin $\beta 1$ subunit[277].

Since this barrier formation has been described to result from interlocking neuronal end-foot processes, we mapped spatially the expression of end-foot process markers observed in the BBB glia-limitans – *GFAP*, *AQP4* and *S100β* - to see if these gene expression patterns were recapitulated surrounding these areas of interest. Surprisingly, we found that not only did these markers not co-localise as a barrier surrounding areas of severe demyelination, the S100β expression did not always co-localise with astrocytes. The cell type it appeared to co-localise most frequently with were OGs. This suggests that while there is evidence of astrocyte behaviour in SPMS consistent with CNS gliotic scar formation, we did not observe gene expression patterns likening the glia-limitans astrocyte barrier at borders of severe damage to those observed in the BBB glia-limitans. This could indicate an inherent difference in these structures, or could indicate an SPMS-associated altered astrocyte behaviour. Impaired containment of damaged areas would be consistent with SPMS pathology, which includes increase in lesion size and our observations of diffused damage surrounding severely demyelinated areas.

In accordance with classical glial scar architecture, the lesion borders of these brain samples also contained a small number of myelin-phagocytosing phagocytes. A notable absence of OPCs as well as CSPG4$^+$ OPCs was a major distinguishing factor for MS lesions when compared to glial scars in the context of traumatic injury. This was

observable both by OPC mapping and *CSPG4* spatial gene expression. Given previous evidence of macrophage populations upregulating *CSPG4* gene expression, it was surprising that there not only not was a statistically significant increase in *CSPG4* expression in MS, but also that there was an observable trend of lower *CSPG4* expression in the $PDGFRA^{+}$ $PLP1^{+}$ BCAShi OPC population believed to have a role in remyelination.

This finding has a substantial impact on our understanding CNS scarring in the context of MS. CSPG4+ OPCs have a role in inhibiting developmental signals from neurons severed by traumatic injury; dystrophic bulbs of neurons are attracted to these OPCs and form synaptoids – synaptic-like structures which prevent further neurite development within the vicinity of the damaged area. Thus, an absence of these cell types would enable unhindered neurite outgrowth. This may be reflected in the large number of ligand-receptor interactions associated with neurite outgrowth and developmental processes observed in the snRNAseq data, as well as certain GSEA identified GO terms in the proteomic data.

5.3.5.3 Incomplete glia-limitans may be linked to GM pathology

We also present a CNS scarring process with potential particular impact on neuronal populations. The lack of *CSPG4* expression would remove the chemical barrier preventing the neurite outgrowths from extending to areas of extensive damage. Moreover, here we observed evidence that the *GFAP*+ astrocytic barrier did not isolate GM areas in the cases observed in this work, isolating only regions of surrounding WM.

The combined absence of CSPG4+ OPCs, low CSPG4 expression and lack of a physical barrier isolating the GM from the lesional area remove both chemical and physical barriers to neurite outgrowth into areas laden with pro-inflammatory and oxidative processes. This outlines ominous microenvironmental implications for neuronal cell bodies populating the GM, particularly in light of their limited neurogenesis potential.

μFTIR classification identified two biochemically atypical GM classifications – one most strongly associated with lipid oxidation with decreased astrocytic population compared to control GM **(Fig. 3.32)** and the second with the highest β-sheet content across all clusters and reduced OPC population when compared to control GM **(Fig. 3.33)**. Transcriptomically we identified evidence neuropathology associated with both these features. Oxidoreductase activity was identified to be particularly affecting CUX2 neurons **(Fig. 4.15-16)**. WGCNA analysis also identified numerous pathways which have links to protein folding, such as ribosomal components, as well as pathways linked to PML bodies which are strongly linked to familial frontotemporal dementia and polyglutamine conditions characterised by intranuclear proteins and inclusions[249-253] (see section 4.3.2.5).

While we hypothesise that collagen deposition in lesions may be associated with the increase in amide absorption and increased β-sheet content in WM areas, the proteomic evidence does not support this within the GM**,** raising the question as to whether, similarly to other neurodegenerative disorders, SPMS has a protein aggregation element in the neuronal cells.

5.3.6 S100β expression interfaces with 'severe with astrogliosis' classification

The observation of highly concentrated S100β gene expression spatially in MS is intriguing, particularly since the effects of S100β are highly dependent on its concentration.

In the nanomolar range, S100β levels modulate neurite outgrowth, synaptogenesis and long-term plasticity, acting as a neurotrophic factor[278]. Historically S100 expression was believed to be specific to astrocytes, although it has been shown that it is also expressed in OG lineage cells, in particular in CSPG4+ OPCs and in myelinating OGs in the brain and SC[279, 280].

In higher concentrations at a micromolecular order of magnitude, however, S100β can have a deleterious role. It has been reported to induce oxidative damage by causing overexpression of inducible nitric oxide synthase and subsequent release of nitric oxide[281]. It also has a pro-inflammatory role, stimulating the activation or astrocytes and microglia through a cytokine cycle as well as activating nuclear factor-κB, which has a prominent role in inflammatory responses[282]. Interestingly, elevated S100β also impairs oligodendrogenesis and myelination[283, 284]. It has also been reported to exacerbate brain damage and periinfarct reactive astrogliosis and microgliosis in focal ischemia[285]. In the context of neurodegeneration, S100β overexpression has been associated with increase amyloid-beta accumulation in animal models of AD[286]. Inhibition in an AD mouse model with the small molecule pentamidine attenuated reactive gliosis and reduced neuronal loss[287].

S100β expression is increased in the context of CNS injury. Its detection above a certain threshold in patients and, in particular, for extended lengths of time can reliably predict brain injury outcome, including brain death or mortality[288]. It may assessed in conjunction with GFAP expression, with median serum levels increased by 33.4-fold for GFAP and 2.1 fold for S100β when comparing levels of patients who died to those who survived[289]. Characteristically, its expression is enhanced in the context of astrogliosis where astrocytes proliferate and become activated, swarming regions of insult or chronic inflammatory states[290]. In other systems, it has been linked to pathways promoting fibrosis due to signalling through the fibroblast growth factor pathway[291], and, in the context of the CNS, may functionally aid in orchestrating the formation of CNS scarring.

S100β has also been associated with MS and been suggested as a biomarker, having been used to measure treatment efficacy[292], and drug target[293]. Its expression is most elevated in RRMS when compared to PMS[294], with an increase in S100β astrocyte-derived expression in actively demyelinating lesions[293]. Its inhibition in the EAE mouse model reduced nitric oxide synthase activity and inflammation, ameliorating MS symptoms demonstrably by lowering disability scores[295].

Here we have observed that the mean S100β expression in oligodendrocytes is higher in controls than cases, indicating that overall OG expression is lower. While the mean astrocyte S100β expression is unchanged, it is among the markers used for the

fibrous astrocyte population which are significantly more abundant in cases than controls.

The highly localised S100β expression adjacent to "severe with astrogliosis" areas, which have the most evidence of damage, including most severe lipid loss, astrogliosis and oxidation, may indicate that these are areas of active demyelination – similarly to how in AD an increase in S100β precedes β-amyloid plaques[286], here we may observe an S100β expression from the larger astrocyte population concentrated in an area preceding MS-associated damage. This would be consistent with the indication that these areas show early signs of damage.

Alternatively, since spatially its expression tends to co-localise with OGs, it may in fact be indicative of active myelination processes, only in a highly localised manner – thus, although overall the mean OG expression is lower, this expression may be concentrated in a localised effort to remyelinate. This theory would be supported by the fact that the regions with high S100β expression also have high OG incidence and PADI2 expression.

An inspection of the S100β gene expression patterns in comparison to deconvoluted cell type maps shows that there is a small overlap between astrocytes and S100β, with a larger overlap with OGs, which suggests this expression is likely associated with more than one cell type. Even if the source of expression is not dominated by a single cell type, given the close relationship between its concentration

and its effects, the highly localised signal is deleterious regardless of the source/cause of expression.

The clinical implications of these findings suggest that inhibition of S100β in SPMS may require careful S100β regulation, so as to attenuate its pro-inflammatory and pro-oxidative effect without hindering the already low OG expression, necessary for remyelination and oligodendrogenesis.

Chapter 6: Conclusions

The two main objectives of this project were to (1) spatially capture the changes occurring across the multi-level functional range of the cells – the transcriptome, proteome and biochemical profiles - in the SPMS and control CNS and (2) use this approach to elucidate spatially-resolved pathological processes in SPMS lesions. In accomplishing these objectives, we have gained insight both into this type of methodological approach as well as key biological insights into SPMS pathology.

6.1 Integrated spatially resolved multiomic profiling: a proof-of-concept

Technologically, this project acts as a proof-of-concept of integrated spatially resolved tissue profiling. Each technology individually provided molecular information which could be contextualised and interpreted within the existing body of literature, building on currently available findings to generate new insights.

For instance, µFTIR analysis of MS has generally used lipid content as an estimate for demyelination, although without explicit validation of this assumption[158, 159]. By correlating µFTIR and proteomic data we showed that lipid content estimated by µFTIR correlates strongly with MBP protein expression. However, we showed that gross assumptions based on lipid content alone can be confounded: demyelinating lesions can still contain myelin laden phagocytes, such that while µFTIR captures a high lipid content, the MBP gene expression is severely depleted – indicating the relevance of combining multiple tissue-profiling approaches.

We have also found correlative evidence that the biochemical classification of a sample area using clustering driven by the amide region (1,500-1,700 cm^{-1}) from µFTIR spectra has corresponding cell-composition patterns, which were observable in deconvoluted 10x Visium maps and summarised by the use of co-localisation networks.

6.1.1 Unsupervised machine learning for classification of pathology

A key methodological approach for the analysis of these datasets was the use of unsupervised machine learning for data partitioning. This revealed the heterogeneous spectrum of lesion states across SPMS in an unbiased and systematic approach. We identified that particularly the µFTIR classification simultaneously had a high clustering resolution – with the largest total number of clusters – but also had high spatial resolution. Particularly in the WM/GM transitions and in the borders of demyelinated lesions, the µFTIR clustering outlined borders between distinct tissue types, with a clearer and higher resolution segregation and higher spatial precision than traditional ORO and H&E imaging. Current lesion classification methods rely on targeted staining assessing leukocyte infiltration, myelination and inflammation markers and can require multiple sections for lesion classification[67, 68, 104, 105, 107, 108].

We envisage that the clustering approach here presented could be leveraged in future for streamlined lesion classification, bypassing the need for staining and use of multiple sections per lesion for classification (important in light of CNS sample rarity). Systematic collection of µFTIR data of SPMS lesions in a representative dataset could be used for algorithmic training, enabling an unbiased, label-free and standardised histological lesion classification approach which avoids the subjective and RRMS-centric criteria of current classification methods. In light of the observed correlations between co-localising cell types and certain biochemical profiles, this could even go beyond sample partitioning and be used for predictive analysis. For instance, using multi-modal datasets such as this one for training a semi-supervised machine-learning algorithm, a

computational method to predict likely cell types present in a sample based on biochemical signatures could be used to contain experimental costs and aid in targeted inquiries.

6.1.2 Anchoring integrated analysis on spatial elements

Apart from comparisons and correlations across specific inquiries in each data set, we also integrated all analyses by summarising classification findings for each area of interest across modalities. Based on lesion severity indicators, the classifications were notably consistent, albeit with variable resolution. A key aspect both in identifying patterns of intra-sample heterogeneity as well as in the process of integration was the preservation of the spatial component across modalities. Preserving the tissue architecture enabled us to enquire multi-level data from neighbouring histological areas, while mapping in each sample these subareas of interest spatially in relation to one another.

6.1.3 Limitations and future directions of spatially resolved multiomic approaches

This methodological approach did not lack challenges and highlighted areas which require further development for future work.

Other more practical insights were also valuable. Partly due to the high sensitivity of the methods here applied, batch effects were a limiting factor for data analysis. For instance, batch effects complicated the joint clustering of the 10x visium data across sections, and were observable in the μFTIR data when samples were profiled in different

experimental sessions. We attributed the latter to changes in light source, instrumentation and alignment between sessions. The utility and potential of these methods would significantly benefit from improved approaches to control for batch effects.

This could be performed experimentally – in the case of 10x Visium, potentially by coupling this technology with the development of a compatible multiplexing-like approach. In the case of µFTIR, this could be addressed by creating a synthetic standard sample scanned at every session in order to identify and mathematically subtract spectral features that are batch-associated. Computational approaches such as ComBat[17] or Harmony[144] compatible with these data set types could be a valuable alternative. A factor which interferes with direct repurposing of these already existing resources and poses an analytical/statistical challenge is that these methods are often designed with underlying assumptions of independence suitable for sc/snRNAseq data. Unlike single-cell datasets, 10x Visium, spatial proteomics and µFTIR do not have an assumption of independence on a pixel-to-pixel basis due to their grid-like acquisition and, in the case of µFTIR and 10x Visium, also multicellular, approach. The µFTIR dataset, represented by a spectrum, is made up of datapoints collected to describe a single continuous spectrum, making the statistically and mathematically further removed from sc/snRNAseq methods.

Coupled with the limitations regarding batch effects of µFTIR data is the limited sample area that can be scanned at a time with current instrumentation. The three main

factors which limit the scanning area in µFTIR data collection are (1) the fact that the detector in the µFTIR system requires cooling intermittently with liquid nitrogen, setting a cap on scanning time since the acquisition is compromised if this time is exceeded or liquid nitrogen cooling is performed during acquisition; (2) the need to acquire all samples within a single session, due to lack of effective methods to control for batch effect; and (3) limited software capacity to acquire and process large scanned areas.

Theoretically, µFTIR has a remarkable and as of yet untapped potential to be used as a powerful, unlabelled histological characterisation tool, however, for this to be feasible in practice it would be paramount to develop the instrumentation and software to effectively and competitively collect µFTIR data on a larger spatial scale.

While here we demonstrate both the technical feasibility as well as biological relevance for a multiomic integrative approach, we acknowledge that for this to become a standardised approach accomplishable on large-scale studies, sample alignment would need to be streamlined or a suitable alternative multiomic approach would need to be considered.

Experimentally, this could take the form of collecting multi-modal data from the same sample/cells, in approaches such as those being developed for single-cell suspension samples with technologies such as CITE-seq[22] or commercially available technologies such as the 10x chromium single cell multiome ATAC and gene expression. Same-cell multiome approaches in spatial transcriptomics are still in their infancy – one of the more recent developments in the field has been the coupling of 10x Visium with

IHC – which, while limited in target selection, provides a valuable alternative to traditional H&E. In the case of analysis of sequential sections, such as for the application of technologies of a destructive nature, computational tools could be a powerful ally in sample alignment. Tools used for sequential imaging data such as MRI imaging and/or geospatial imaging could be repurposed. An entirely distinct approach is also plausible, and that is of bypassing image alignment entirely and computationally predict tissue architecture with tools such as Tangram[258] and DEEPsc[259], which aim to leverage protein data from different samples to map spatially single-cell transcriptomic data.

Finally, in this work we opted for an approach where we cluster each modality separately and compared the classifications and features of each area of interest across modalities. We foresee that multiomic approaches could benefit from joint clustering approaches, with integration at an earlier stage of analysis. This could take the form of iterative clustering, or a unified clustering of all data elements jointly. Due to the different nature, orders of magnitude and formats of data, this may prove to be mathematically and statistically challenging, but would allow high resolution segregation of cell types and states across multiple features simultaneously, providing a holistic understanding of co-occurring molecular features across different functional levels.

6.1.4 Sample availability limitations and contingency approaches

One limitation of this study was the relatively small cohort, although in fact this dataset is comparable to that used in other snRNAseq studies of this type of study[130, 131, 296-299] and relates both to sample rarity and sample quality, which is particularly

relevant for spatially resolved transcriptomic analysis. In particular, the limited number and sample quality of the spinal cord prevented the inclusion of SC samples in the snRNAseq assays. This resulted in the cell type annotation being focused on brain cell types, rather than SC cell types – having consequences for the deconvolution of the spatial SC data. The small SC representation also made comparison between brain and SC pathology statistically vulnerable, but was included in this project as an initial community resource given the paucity of high-resolution omic data for the SC in MS.

Ultimately, we aimed to address the sample availability limitation in three main ways, firstly by multiplexing our snRNAseq data – this will have aided in distinguishing patient-specific features from batch effects – secondly, by including previously published snRNAseq data sets of MS and finally by controlling for batch effects computationally. By combining these measures with cross-validation and correlations across modalities, we expect that our findings are robustly validated and should be reproducible in other sample cohorts.

Two long-term and interdependent solutions to the reoccurring challenges with sample availability are the combined increasing availability of molecular profiling datasets as well as increasing availability of experimental and computational batch correction methods. The combination of these factors will be valuable to overcome difficulties with sample rarity and enable more statistically robust and overall representative analyses.

6.1.5 Practical considerations for future multi-omic approaches

Apart from several practical considerations in data collection and data analysis for multi-omic approaches, perhaps one of the more widely applicable insights generated by this project has been the importance of a strategic and clear goal for this type of project. In light of the increasing popularisation of this type of strategy, it has rapidly become apparent that a technological goal (multi-omic integration of multiple methods) no longer contributes significantly to the scientific community unless the dataset is large enough to be classified as an atlas and can be reused with statistical robustness for other inquiries.

Thus, for smaller sample sets such as the one we report here, two strategies remain which are enriching for the scientific field: (1) targeted strategy: having a clear goal, hypothesis or biological inquiry based on previous work/available literature or (2) exploratory strategy: generating and pursuing a hypothesis based on the analysis of the generated data set(s).

The first is the more time and financially conservative option, since by limiting the scope of the inquiry one may be able to reduce the costs and volume of data to generate, store, process, annotate and analyse.

In this project, commencing prior to the establishment of multi-omic approaches, we used strategy an exploratory strategy. On one hand, this will facilitate reusability of this data by other users for specific inquiries since data from all analyses can be made available without the need for users to repeat the entire pre-processing and analysis process. However, this came at a cost in time, particularly in light of the continually

growing library of potential analysis methods and tools increasing the required analysis for each method, with several potential lines of inquiry available to pursue in each given analysis step. While this highlights the immense utility of the data for hypothesis generation and for reuse, it also makes this type of project impractical for current models of scientific communication and publication which benefit from clearly outlined narratives.

Ultimately the main restraining boundary of the exploratory strategy is the dataset with the smallest number of variables – in the case of this project, we targeted our inquiries by discretising µFTIR data into features of interest (oxidation, lipid content, protein folding content); like this, we constrained the scope of inquiry to pursue with the other methods which capture thousands of variables each. For other multi-omic modalities, however, this is not always an option, thus highlighting the importance of a clear goal/hypothesis to be achieved with the multiomic approach.

6.2 A comprehensive view of SPMS molecular landscape

This broad technological approach generated a multilevel holistic molecular profile of SPMS. Here we present for the first time in the literature macromolecular characterisation with µFTIR specifically of SPMS in post-mortem human brains, including characterisation of WM and GM tissue areas surrounding severe lesions. Through this analysis, we have both been able to recapitulate in a more quantitative fashion several key pathology-associated features previously observed in MS lesions such as decreased

lipid content, increased amide content and lipid oxidation, but have also shown evidence of the presence of such alterations in both WM and GM areas surrounding severe lesions.

6.2.1 Biological contextualisation of biochemical alterations in SPMS

Through integrative analysis with proteomic and transcriptomics data, we have biologically contextualised these biochemical findings, for instance with the aforementioned validation of lipid content correlating with indicators of myelin content.

We also presented evidence of ongoing lipid oxidation in SPMS, where increased lipid oxidation levels were observed in WM lesions and correlated with the level of demyelination. Proteomically, oxidation levels correlated with PRDX5 protein expression. Through transcriptomic analysis we demonstrated that the two cell types with most evidence of oxidative response were OGs and neurons. In OGs, we report upregulation of oxidation-associated *MIF* and *SLC22A17* gene expression in MS with DGE analaysis, as well as WGCNA analysis which associated this cell type with GO terms associated with oxidation. This strengthens the observed association between lipid oxidation and severity of demyelination, since both OGs and neurons which have most evidence of oxidative stress are also the cell types most vulnerable to processes of demyelination. Interestingly, among neuronal cell types showing hallmarks of increased oxidation the CUX2 neuronal population had the most significant signs of oxidative response. This

population has notably been previously reported to be selectively vulnerable in the context of MS[130].

We also observed substantial proteomic changes in SPMS. Consistently with previous biochemical analyses of MS, we have also observed an increase in amide content in areas of damage in SPMS. We observed that amide levels were higher in regions with more severe lipid loss, although while in controls we observed that the amide content had a linear anti-correlation relationship, we have observed that this relationship was no longer linear in cases and was more akin to a second-degree polynomial – highlighting a more nuanced and complex relationship between these macromolecules in a pathological context.

Beyond total amide content, we reported evidence of alterations to protein folding content in SPMS, with a statistically significant increase in β-sheet content with concomitant decrease in α-helix content. This type of alteration has previously been observed in the context of neurodegenerative disorders such as AD, PD and HD, but had never, to our knowledge, been reported in the context of MS. In the context of other neurodegenerative disorders, this is generally associated to protein misfolding and protein aggregation in the form of plaques. Protein deposition of certain extracellular matrix proteins such as FN1 in MS have been previously reported[300], and here we present evidence of type IV collagen protein deposition in areas of damage. The spectral profile of type IV collagen has a demonstrably higher β-sheet and β-turn content[301], and may contribute to this spectral feature. We also observed cell type-specific

evidence of alterations to protein processing – this was most notable in OGs, with WGCNA correlation pathways such as protein folding, protein refolding and unfolded protein binding as well as associating pathways linked to protein misfolding such as PrD, HD and PD. Notably, however, the increased β-sheet content is most pronounced in the pathological GM which has little proteomic evidence supporting collagen deposition. Given the susceptibility of neurons to protein aggregation under certain circumstances[302] as observed in neurodegenerative disorders such as AD and PD, the neurodegenerative-like features of SPMS and the pathway hits in WGCNA analysis linking the neuronal compartment to neurodegenerative disorders with protein aggregation elements and evidence of ribosomal alterations, this work presents a case that protein aggregation and/or misfolding – whether translationally or post-translationally – may be an as of yet undescribed element of MS.

Alternatively, the µFTIR findings indicating β-sheet accumulation which have been reoccurringly observed in proteinopathies and assumed to be exclusively evidence of protein aggregates in β-sheet structures[137, 303], may reflect a more fundamental aspect of neurodegeneration common across conditions including ones believed to not have protein folding involvement. This could be in the form of an overrepresentation of proteins with this type of feature, which may be overexpressed in neurodegeneration due to, for instance, their neuroprotective or inflammatory role.

6.2.2 CNS scar formation in SPMS

We present evidence that cell types associated with CNS scar formation are more abundant in SPMS than controls – this includes fibrous astrocytes, fibroblasts and a phagocytosing macrophage/microglial population. Notably, we have not observed an increase in OPC populations and, in particular, we have not observe CSPG4+ OPCs expected to be present in CNS scars[173]. These cells have a crucial role in hindering developmental processes of neurites through CSPG expression as well as by attracting dystrophic bulbs to form stabilising synaptoids. This removes the chemical barriers in place preventing neuronal processes from reaching lesional areas rich in pro-inflammatory signals.

Our spatial analysis supported the hypothesis of CNS scarring, particularly in areas with the most severe demyelination and astrogliosis classified as "severe with astrogliosis". These areas had high GFAP expressing regions in the rim of these areas of damage associated with the perilesional glia-limitans barrier, with a centre populated by microglia/macrophages and fibroblasts. We also identified that areas with these features also had high type IV collagen deposition, which is consistent with the role of fibroblasts in CNS scar formation. However, astrocytic barriers did not have evidence of perfect interlocking -having more diffused signals – and lacked features of the BBB-associated glia limitans, such as absence of co-localising astrocyte end-feet markers. Moreover, there were no astrocytic barriers isolating GM areas. This poses the possibility of faulty CNS scar formation mechanisms in SPMS. Ineffective lesion isolation in SPMS would be consistent with pathological features associated with late disease, such as diffused injury

and lesion expansion. Our integrated analysis demonstrates that, while lesion-surrounding areas may not always have biochemical evidence of injury, they may already present pro-inflammatory soluble factors such as complement proteins.

Altered cellular behaviour in SPMS could, for instance, be linked to cellular exhaustion due to chronic activation or cellular senescence. MS is characterised both by a high cellular turnover during the relapsing remitting period as a response to damage as well as the presence of cellular stressors such as oxidative/energetic stress and DNA damage, both being events which may trigger cellular senescence[304]. Consistently with this view, we observe evidence of cell-cycle arrest in the MS endothelial compartment, with control-specific differential gene expression of several genes associated with cell cycling. Further investigation of CNS scarring in MS could help elucidate these mechanisms and further our understanding of SPMS. One methodological approach to probe this inquiry further would be to leverage the well-established protocols for CNS scar formation in the context of injury in animal models, comparing scar formation across controls, EAE and senescence models[305] – or, perhaps even EAE induction in conjunction with senescence[306].

Should these assays further support cellular senescence is a contributing factor for faulty lesion isolation and therefore lesion expansion and diffused injury in SPMS, several approved drugs target senescent cell anti-apoptotic pathways which could be repurposed, including, among others, Desatinib, Quercetin, Fisetin and Geldanamycin[307].

6.2.3 Highly localised S100β gene expression

We also present an exemplary case of a context in which spatially resolved transcriptomic analysis can be key to outlining potentially pathological gene expression, which would not have been captured in a non-spatially resolved approach. In areas surrounding highly chronic lesions with evidence of CNS scar formation we have observed regions highly expressing S100β. Expressed both in astrocytes and OGs, highly localised expression of this gene is intriguing due to its unique role which is highly susceptible to its concentrations: in low concentrations it has a role in CNS maintenance and long term plasticity[278] and myelination[279, 280], while in high concentrations it can induce oxidation and have a pro-inflammatory role[279, 280]. EAE assays suggest that S100β inhibition had a beneficial impact, reducing nitric oxide synthase activity and inflammation and ultimately ameliorating MS symptoms – this would be consistent with highly localised expression of S100β creating pockets of damage and demyelination. However, the evidence that at least part of the gene expression here observed is associated with OGs suggests the scenario is more complex and that SPMS S100β would require careful regulation, so as to not impede remyelinating processes.

In vivo data suggests S100β inhibitor Pentamidine may strike this balance suitable, having been shown to ameliorate paralysis, reduce astrocyte reactivity whilst still facilitating myelination and recovery[308].

6.2.4 SPMS-specific sample classification and heterogeneity

Many existing lesion classification systems for MS focus on features associated with active lesions. In the absence of immune cell infiltration criteria, the remaining

established binary metrics for lesion such as classification, myelination and inflammation, limit the scope for lesion classification in SPMS[67, 68, 104, 105, 107, 108], where dues to the presence of expanding lesions there is no space for nuanced interpretation of regions with more or less damage.

In this work we have observed substantial lesion heterogeneity – this was particularly pronounced in biochemical profiling, where three WM classifications associated with pathology were identified, as well as two classifications associated with GM pathology. Through analysis of deconvoluted spatial transcriptomics data, patterns linking certain cluster classifications with cellular composition emerged – this was primarily identified in controls, where sample-specific spatial co-localisation networks contained subnetworks associated with characteristically WM and GM cell types. This correlation between compositional and biochemical classification was further explored in lesions, with correlations which could be reproducibly recapitulated across samples. Proteomically, we observed two pathology-associated WM clusters which captured a health-to-disease continuum in proteomic profile.

While it would be consistent with previous classification systems in the MS field to use the lipid content level to generate a predicted health-to-disease continuum[67, 68, 104, 105, 107, 108], we also observed evidence supporting the presence of remyelination, such as PADI2 expression and S100β expression in OGs. Therefore, we integrated these clustering approaches and outlined a timeline-agnostic classification approach which enables classification of pathological severity using quantitative metrics, thus outlining

the subtle degrees of pathology in SPMS without the assumption of continual demyelination. In future work in this field, we hypothesise this approach could be harnessed in conjunction with longitudinal MRI and *in vivo* NMR spectroscopy data to disentangle the timeline of SPMS lesion progression.

6.3 Final remarks

We have endeavoured to harness high-resolution and high throughput methods to maximise the utility of rare samples, in the process encountering a wide range of findings in technical as well as biological fronts. Remarkably, within four data sets not only have we been able to observe a range of features associated with MS pathology, we have also been able to pin-point these processes to specific cell types and tissue-specific events linked to these features. This includes the evidence of oxidation and alterations to protein processing in OGs and in neurons, as well as an outlining of events in SPMS analogous to injury-associated CNS scarring and deviations from typical responses. We have also demonstrated an example of highly localised gene expression of a gene with a physiological role highly susceptible to changes in its concentration, outlining the degree of complexity in gene expression lost in absence of a spatially resolved element. We believe jointly these findings provide substantial evidence of this technological approach not only for disease characterisation and hypothesis generation, but also in potential enabling clinical insight for conditions such as rare diseases, where availability of biological material may be a limiting factor.

Notably, we highlight that this approach can be employed to profile neurological tissue obtained post mortem, thus providing a framework to profile neurological disease as well as the central nervous system in the absence of neurological disease, thus providing molecular insight into the CNS whilst bypassing ethical concerns with obtainment of samples via biopsy. Not only does this provide an experimental method for future interrogation into disease but also maintenance of neurological health. We also outline analytical strategies to understand the cellular architecture, biochemical profile and cell activation states of the homeostatic central nervous system.

Secondly, we demonstrate this method provides multiple avenues of inquiry into drug targets for drugs that are already available and approved, thus providing a cost effective approach for maximising available resources. With increasingly publicly available datasets – all of which apart from neurological disease, also include controls without neurological disease-, the growing body of work profiling the molecular landscape of the human CNS could be key to enabling healthy ageing of the central nervous system in the future.

Bibliographic references

1. Wyss-Coray, T., *Ageing, neurodegeneration and brain rejuvenation.* Nature, 2016. **539**(7628): p. 180-186.
2. Sender, R. and R. Milo, *The distribution of cellular turnover in the human body.* Nature Medicine, 2021. **27**(1): p. 45-48.
3. Nyberg, L., et al., *Memory aging and brain maintenance.* Trends in Cognitive Sciences, 2012. **16**(5): p. 292-305.
4. Liu, Z., L. Ding, and B. He, *Integration of EEG/MEG with MRI and fMRI.* IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society, 2006. **25**(4): p. 46-53.
5. Condon, B., *Magnetic resonance imaging and spectroscopy: how useful is it for prediction and prognosis?* The EPMA journal, 2011. **2**(4): p. 403-410.
6. The Lancet, N., *Brain banking: more effective strategies needed.* Lancet Neurol, 2013. **12**(11): p. 1035.
7. De Jager, P.L., et al., *A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research.* Scientific Data, 2018. **5**(1): p. 180142.
8. Nativio, R., et al., *An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease.* Nature Genetics, 2020. **52**(10): p. 1024-1035.
9. Morello, G., et al., *Integrative multi-omic analysis identifies new drivers and pathways in molecularly distinct subtypes of ALS.* Scientific Reports, 2019. **9**(1): p. 9968.
10. Conesa, A. and S. Beck, *Making multi-omics data accessible to researchers.* Scientific Data, 2019. **6**(1): p. 251.
11. Diaz-Ortiz, M.E. and A.S. Chen-Plotkin, *Omics in Neurodegenerative Disease: Hope or Hype?* Trends Genet, 2020. **36**(3): p. 152-159.
12. Yamawaki, T.M., et al., *Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling.* BMC Genomics, 2021. **22**(1): p. 66.
13. Stickels, R.R., et al., *Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2.* Nature Biotechnology, 2021. **39**(3): p. 313-319.

14. Aldridge, S. and S.A. Teichmann, *Single cell transcriptomics comes of age.* Nature Communications, 2020. **11**(1): p. 4307.
15. Rodriques, S.G., et al., *Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.* Science (New York, N.Y.), 2019. **363**(6434): p. 1463-1467.
16. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony.* Nature Methods, 2019. **16**(12): p. 1289-1296.
17. Zhang, Y., G. Parmigiani, and W.E. Johnson, *ComBat-seq: batch effect adjustment for RNA-seq count data.* NAR Genomics and Bioinformatics, 2020. **2**(3).
18. Hie, B., B. Bryson, and B. Berger, *Efficient integration of heterogeneous single-cell transcriptomes using Scanorama.* Nature Biotechnology, 2019. **37**(6): p. 685-691.
19. Gaublomme, J.T., et al., *Nuclei multiplexing with barcoded antibodies for single-nucleus genomics.* Nature Communications, 2019. **10**(1): p. 2907.
20. Miao, Z., et al., *Multi-omics integration in the age of million single-cell data.* Nature Reviews Nephrology, 2021. **17**(11): p. 710-724.
21. Kaufmann, M., et al., *Identifying CNS-colonizing T cells as potential therapeutic targets to prevent progression of multiple sclerosis.* Med, 2021. **2**(3): p. 296-312.e8.
22. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature Methods, 2017. **14**(9): p. 865-868.
23. Absinta, M., et al., *A lymphocyte-microglia-astrocyte axis in chronic active multiple sclerosis.* Nature, 2021. **597**(7878): p. 709-714.
24. Jonkman, L.E., et al., *Normal Aging Brain Collection Amsterdam (NABCA): A comprehensive collection of postmortem high-field imaging, neuropathological and morphometric datasets of non-neurological controls.* Neuroimage Clin, 2019. **22**: p. 101698.
25. Dendrou, C.A., G. McVean, and L. Fugger, *Neuroinflammation — using big data to inform clinical practice.* Nature Reviews Neurology, 2016. **12**(12): p. 685-698.
26. Walton, C., et al., *Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition.* Multiple sclerosis (Houndmills, Basingstoke, England), 2020. **26**(14): p. 1816-1821.
27. Filippi, M., et al., *Multiple sclerosis.* Nature Reviews Disease Primers, 2018. **4**(1): p. 43.
28. Di Filippo, M., et al., *Multiple sclerosis and cognition: synaptic failure and network dysfunction.* Nature Reviews Neuroscience, 2018. **19**(10): p. 599-609.
29. Lassmann, H., *Pathogenic Mechanisms Associated With Different Clinical Courses of Multiple Sclerosis.* Frontiers in immunology, 2019. **9**: p. 3116-3116.
30. Fisniku, L.K., et al., *Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis.* Brain, 2008. **131**(3): p. 808-817.

31. Lassmann, H., J. van Horssen, and D. Mahad, *Progressive multiple sclerosis: pathology and pathogenesis.* Nat Rev Neurol, 2012. **8**(11): p. 647-56.
32. *Clinical demographics of multiplex families with multiple sclerosis. Multiple Sclerosis Genetics Group.* Ann Neurol, 1998. **43**(4): p. 530-4.
33. Compston, A. and A. Coles, *Multiple sclerosis.* Lancet, 2008. **372**(9648): p. 1502-17.
34. Harirchian, M.H., et al., *Worldwide prevalence of familial multiple sclerosis: A systematic review and meta-analysis.* Multiple Sclerosis and Related Disorders, 2018. **20**: p. 43-47.
35. Robertson, N.P., et al., *Age-adjusted recurrence risks for relatives of patients with multiple sclerosis.* Brain, 1996. **119 (Pt 2)**: p. 449-55.
36. O'Gorman, C., et al., *Modelling Genetic Susceptibility to Multiple Sclerosis with Family Data.* Neuroepidemiology, 2013. **40**(1): p. 1-12.
37. Baranzini, S.E. and J.R. Oksenberg, *The Genetics of Multiple Sclerosis: From 0 to 200 in 50 Years.* Trends in Genetics, 2017. **33**(12): p. 960-970.
38. Patsopoulos Nikolaos, A., et al., *Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility.* Science, 2019. **365**(6460): p. eaav7188.
39. Baranzini, S.E., *The genetics of autoimmune diseases: a networked perspective.* Current Opinion in Immunology, 2009. **21**(6): p. 596-605.
40. Dendrou, C.A., L. Fugger, and M.A. Friese, *Immunopathology of multiple sclerosis.* Nat Rev Immunol, 2015. **15**(9): p. 545-58.
41. Rosati, G., *The prevalence of multiple sclerosis in the world: an update.* Neurological Sciences, 2001. **22**(2): p. 117-139.
42. Langer-Gould, A., et al., *Incidence of multiple sclerosis in multiple racial and ethnic groups.* Neurology, 2013. **80**(19): p. 1734-9.
43. Simpson, S., Jr., et al., *Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis.* J Neurol Neurosurg Psychiatry, 2011. **82**(10): p. 1132-41.
44. Olsson, T., L.F. Barcellos, and L. Alfredsson, *Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis.* Nature Reviews Neurology, 2017. **13**(1): p. 25-36.
45. Ahlgren, C., A. Odén, and J. Lycke, *A nationwide survey of the prevalence of multiple sclerosis in immigrant populations of Sweden.* Multiple Sclerosis Journal, 2011. **18**(8): p. 1099-1107.
46. Berg-Hansen, P., et al., *Prevalence of multiple sclerosis among immigrants in Norway.* Mult Scler, 2015. **21**(6): p. 695-702.
47. Aranow, C., *Vitamin D and the immune system.* J Investig Med, 2011. **59**(6): p. 881-6.
48. Handel, A.E., et al., *An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis.* PLoS One, 2010. **5**(9).

49. Zhou, Y., et al., *Genetic loci for Epstein-Barr virus nuclear antigen-1 are associated with risk of multiple sclerosis.* Mult Scler, 2016. **22**(13): p. 1655-1664.
50. Levin, L.I., et al., *Primary infection with the Epstein-Barr virus and risk of multiple sclerosis.* Ann Neurol, 2010. **67**(6): p. 824-30.
51. Sundqvist, E., et al., *Epstein-Barr virus and multiple sclerosis: interaction with HLA.* Genes Immun, 2012. **13**(1): p. 14-20.
52. Nielsen, T.R., et al., *Effects of infectious mononucleosis and HLA-DRB1*15 in multiple sclerosis.* Mult Scler, 2009. **15**(4): p. 431-6.
53. Sundqvist, E., et al., *Cytomegalovirus seropositivity is negatively associated with multiple sclerosis.* Mult Scler, 2014. **20**(2): p. 165-73.
54. Rivers, T.M., D.H. Sprunt, and G.P. Berry, *OBSERVATIONS ON ATTEMPTS TO PRODUCE ACUTE DISSEMINATED ENCEPHALOMYELITIS IN MONKEYS.* J Exp Med, 1933. **58**(1): p. 39-53.
55. Rivers, T.M. and F.F. Schwentker, *ENCEPHALOMYELITIS ACCOMPANIED BY MYELIN DESTRUCTION EXPERIMENTALLY PRODUCED IN MONKEYS.* J Exp Med, 1935. **61**(5): p. 689-702.
56. Fritz, R.B., C.H. Chou, and D.E. McFarlin, *Induction of experimental allergic encephalomyelitis in PL/J and (SJL/J x PL/J)F1 mice by myelin basic protein and its peptides: localization of a second encephalitogenic determinant.* J Immunol, 1983. **130**(1): p. 191-4.
57. Soos, J.M., et al., *Accelerated induction of experimental allergic encephalomyelitis in PL/J mice by a non-V beta 8-specific superantigen.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(13): p. 6082-6086.
58. Bettelli, E., et al., *Myelin oligodendrocyte glycoprotein-specific T and B cells cooperate to induce a Devic-like disease in mice.* J Clin Invest, 2006. **116**(9): p. 2393-402.
59. Friese, M.A., et al., *Opposing effects of HLA class I molecules in tuning autoreactive CD8+ T cells in multiple sclerosis.* Nat Med, 2008. **14**(11): p. 1227-35.
60. Stys, P.K., et al., *Will the real multiple sclerosis please stand up?* Nature Reviews Neuroscience, 2012. **13**(7): p. 507-514.
61. Aboul-Enein, F., et al., *Preferential loss of myelin-associated glycoprotein reflects hypoxia-like white matter damage in stroke and inflammatory brain diseases.* J Neuropathol Exp Neurol, 2003. **62**(1): p. 25-33.
62. Rodriguez, M. and B. Scheithauer, *Ultrastructure of multiple sclerosis.* Ultrastruct Pathol, 1994. **18**(1-2): p. 3-13.
63. Henderson, A.P., et al., *Multiple sclerosis: distribution of inflammatory cells in newly forming lesions.* Ann Neurol, 2009. **66**(6): p. 739-53.

64. Seewann, A., et al., *Diffusely Abnormal White Matter in Chronic Multiple Sclerosis: Imaging and Histopathologic Analysis.* Archives of Neurology, 2009. **66**(5): p. 601-609.
65. Wingerchuk, D.M., et al., *International consensus diagnostic criteria for neuromyelitis optica spectrum disorders.* Neurology, 2015. **85**(2): p. 177.
66. Lucchinetti, C., et al., *Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination.* Ann Neurol, 2000. **47**(6): p. 707-17.
67. Ferguson, B., et al., *Axonal damage in acute multiple sclerosis lesions.* Brain, 1997. **120**(3): p. 393-399.
68. van der Valk, P. and C.J. De Groot, *Staging of multiple sclerosis (MS) lesions: pathology of the time frame of MS.* Neuropathol Appl Neurobiol, 2000. **26**(1): p. 2-10.
69. Gan, L., et al., *Converging pathways in neurodegeneration, from genetics to mechanisms.* Nature Neuroscience, 2018. **21**(10): p. 1300-1309.
70. Guzman-Martinez, L., et al., *Neuroinflammation as a Common Feature of Neurodegenerative Disorders.* Frontiers in Pharmacology, 2019. **10**: p. 1008.
71. Ponath, G., C. Park, and D. Pitt, *The Role of Astrocytes in Multiple Sclerosis.* Frontiers in immunology, 2018. **9**: p. 217-217.
72. Trapp, B.D., et al., *Axonal transection in the lesions of multiple sclerosis.* N Engl J Med, 1998. **338**(5): p. 278-85.
73. Kirby, L., et al., *Oligodendrocyte precursor cells present antigen and are cytotoxic targets in inflammatory demyelination.* Nature Communications, 2019. **10**(1): p. 3887.
74. Neely, S.A., et al., *New oligodendrocytes exhibit more abundant and accurate myelin regeneration than those that survive demyelination.* bioRxiv, 2020: p. 2020.05.22.110551.
75. Peterson, J.W., et al., *Transected neurites, apoptotic neurons, and reduced inflammation in cortical multiple sclerosis lesions.* Ann Neurol, 2001. **50**(3): p. 389-400.
76. Wherry, E.J., *T cell exhaustion.* Nat Immunol, 2011. **12**(6): p. 492-9.
77. Howell, O.W., et al., *Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis.* Brain, 2011. **134**(Pt 9): p. 2755-71.
78. Choi, S.R., et al., *Meningeal inflammation plays a role in the pathology of primary progressive multiple sclerosis.* Brain, 2012. **135**(Pt 10): p. 2925-37.
79. Neher, J.J., et al., *Inhibition of Microglial Phagocytosis Is Sufficient To Prevent Inflammatory Neuronal Death.* The Journal of Immunology, 2011. **186**(8): p. 4973.
80. Mayo, L., et al., *Regulation of astrocyte activation by glycolipids drives chronic CNS inflammation.* Nat Med, 2014. **20**(10): p. 1147-56.

81. Zeis, T., et al., *Molecular changes in white matter adjacent to an active demyelinating lesion in early multiple sclerosis.* Brain Pathol, 2009. **19**(3): p. 459-66.
82. Fischer, M.T., et al., *Disease-specific molecular events in cortical multiple sclerosis lesions.* Brain, 2013. **136**(Pt 6): p. 1799-815.
83. Mahad, D., et al., *Mitochondrial defects in acute multiple sclerosis lesions.* Brain, 2008. **131**(Pt 7): p. 1722-35.
84. Mahad, D.J., et al., *Mitochondrial changes within axons in multiple sclerosis.* Brain, 2009. **132**(Pt 5): p. 1161-74.
85. Campbell, G.R., et al., *Mitochondrial DNA deletions and neurodegeneration in multiple sclerosis.* Ann Neurol, 2011. **69**(3): p. 481-92.
86. Smith, K.J., R. Kapoor, and P.A. Felts, *Demyelination: the role of reactive oxygen and nitrogen species.* Brain Pathol, 1999. **9**(1): p. 69-92.
87. Haider, L., et al., *Oxidative damage in multiple sclerosis lesions.* Brain, 2011. **134**(Pt 7): p. 1914-24.
88. Vladimirova, O., et al., *Oxidative damage to DNA in plaques of MS brains.* Mult Scler, 1998. **4**(5): p. 413-8.
89. Qin, J., et al., *Oxidized phosphatidylcholine is a marker for neuroinflammation in multiple sclerosis brain.* J Neurosci Res, 2007. **85**(5): p. 977-84.
90. Bizzozero, O.A., et al., *Elevated protein carbonylation in the brain white matter and gray matter of patients with multiple sclerosis.* J Neurosci Res, 2005. **81**(5): p. 687-95.
91. Wang, X. and E.K. Michaelis, *Selective neuronal vulnerability to oxidative stress in the brain.* Front Aging Neurosci, 2010. **2**: p. 12.
92. French, H.M., et al., *Oxidative stress disrupts oligodendrocyte maturation.* Journal of Neuroscience Research, 2009. **87**(14): p. 3076-3087.
93. Thorburne, S.K. and B.H.J. Juurlink, *Low Glutathione and High Iron Govern the Susceptibility of Oligodendroglial Precursors to Oxidative Stress.* Journal of Neurochemistry, 1996. **67**(3): p. 1014-1022.
94. Bagnato, F., et al., *Tracking iron in multiple sclerosis: a combined imaging and histopathological study at 7 Tesla.* Brain, 2011. **134**(Pt 12): p. 3602-15.
95. Friese, M.A., et al., *Acid-sensing ion channel-1 contributes to axonal degeneration in autoimmune inflammation of the central nervous system.* Nat Med, 2007. **13**(12): p. 1483-9.
96. Ouardouz, M., et al., *Glutamate receptors on myelinated spinal cord axons: I. GluR6 kainate receptors.* Ann Neurol, 2009. **65**(2): p. 151-9.
97. Kornek, B., et al., *Distribution of a calcium channel subunit in dystrophic axons in multiple sclerosis and experimental autoimmune encephalomyelitis.* Brain, 2001. **124**(6): p. 1114-1124.

98. Boscia, F., et al., *Altered Expression of Ion Channels in White Matter Lesions of Progressive Multiple Sclerosis: What Do We Know About Their Function?* Frontiers in Cellular Neuroscience, 2021. **15**(214).
99. Correale, J., M. Marrodan, and E.E. Benarroch, *What is the role of axonal ion channels in multiple sclerosis?* Neurology, 2020. **95**(3): p. 120.
100. Kapoor, R., et al., *Blockers of sodium and calcium entry protect axons from nitric oxide-mediated degeneration.* Annals of Neurology, 2003. **53**(2): p. 174-180.
101. Friese, M.A., et al., *Acid-sensing ion channel-1 contributes to axonal degeneration in autoimmune inflammation of the central nervous system.* Nature Medicine, 2007. **13**(12): p. 1483-1489.
102. Schattling, B., et al., *TRPM4 cation channel mediates axonal and neuronal degeneration in experimental autoimmune encephalomyelitis and multiple sclerosis.* Nature Medicine, 2012. **18**(12): p. 1805-1811.
103. Charcot, J.M., *Histologie de la sclerose en plaques.* Gazette des hospitaux, 1868(41): p. 554-55.
104. Bö, L., et al., *Detection of MHC class II-antigens on macrophages and microglia, but not on astrocytes and endothelia in active multiple sclerosis lesions.* J Neuroimmunol, 1994. **51**(2): p. 135-46.
105. Lucchinetti, C.F., et al., *Distinct Patterns of Multiple Sclerosis Pathology Indicates Heterogeneity in Pathogenesis.* Brain Pathology, 1996. **6**(3): p. 259-274.
106. Kuhlmann, T., et al., *An updated histological classification system for multiple sclerosis lesions.* Acta Neuropathol, 2017. **133**(1): p. 13-24.
107. Brück, W., et al., *Monocyte/macrophage differentiation in early multiple sclerosis lesions.* Ann Neurol, 1995. **38**(5): p. 788-96.
108. Brück, W., et al., *Oligodendrocytes in the early course of multiple sclerosis.* Ann Neurol, 1994. **35**(1): p. 65-73.
109. Sorensen, P.S. and M. Blinkenberg, *The potential role for ocrelizumab in the treatment of multiple sclerosis: current evidence and future prospects.* Therapeutic advances in neurological disorders, 2016. **9**(1): p. 44-52.
110. Hauser, S.L., et al., *Ofatumumab versus teriflunomide in multiple sclerosis.* New England Journal of Medicine, 2020. **383**(6): p. 546-557.
111. Rudick, R.A., et al., *Natalizumab plus interferon beta-1a for relapsing multiple sclerosis.* New England Journal of Medicine, 2006. **354**(9): p. 911-923.
112. Polman, C.H., et al., *A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.* New England Journal of Medicine, 2006. **354**(9): p. 899-910.
113. Fox, R.J., et al., *Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis.* New England Journal of Medicine, 2012. **367**(12): p. 1087-1097.
114. O'Connor, P., et al., *Randomized trial of oral teriflunomide for relapsing multiple sclerosis.* New England Journal of Medicine, 2011. **365**(14): p. 1293-1303.

115. Group, I.M.S.S., *Interferon beta-1b is effective in relapsing-remitting multiple sclerosis: I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial.* Neurology, 1993. **43**(4): p. 655-655.
116. Johnson, K., et al., *Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind, placebo-controlled trial.* Neurology, 1995. **45**(7): p. 1268-1276.
117. Hauser, S.L. and B.A.C. Cree, *Treatment of Multiple Sclerosis: A Review.* The American Journal of Medicine, 2020. **133**(12): p. 1380-1390.e2.
118. Montalban, X., et al., *Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis.* N Engl J Med, 2017. **376**(3): p. 209-220.
119. Kappos, L., et al., *Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study.* The Lancet, 2018. **391**(10127): p. 1263-1273.
120. Giovannoni, G., et al., *A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis.* New England Journal of Medicine, 2010. **362**(5): p. 416-426.
121. Fox, R.J., et al., *Phase 2 Trial of Ibudilast in Progressive Multiple Sclerosis.* New England Journal of Medicine, 2018. **379**(9): p. 846-855.
122. Faissner, S., et al., *Progressive multiple sclerosis: from pathophysiology to therapeutic strategies.* Nature Reviews Drug Discovery, 2019. **18**(12): p. 905-922.
123. Kalkers, N.F., et al., *The effect of the neuroprotective agent riluzole on MRI parameters in primary progressive multiple sclerosis: a pilot study.* Mult Scler, 2002. **8**(6): p. 532-3.
124. Chataway, J., et al., *Efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis (MS-SMART): a phase 2b, multiarm, double-blind, randomised placebo-controlled trial.* Lancet Neurol, 2020. **19**(3): p. 214-225.
125. Caggiula, M., et al., *Neurotrophic factors in relapsing remitting and secondary progressive multiple sclerosis patients during interferon beta therapy.* Clinical Immunology, 2006. **118**(1): p. 77-82.
126. Gravesteijn, A.S., et al., *Neuroprotective effects of exercise in people with progressive multiple sclerosis (Exercise PRO-MS): study protocol of a phase II trial.* BMC Neurology, 2020. **20**(1): p. 177.
127. Tourbah, A., et al., *MD1003 (high-dose biotin) for the treatment of progressive multiple sclerosis: A randomised, double-blind, placebo-controlled study.* Mult Scler, 2016. **22**(13): p. 1719-1731.
128. Birnbaum, G. and J. Stulc, *High dose biotin as treatment for progressive multiple sclerosis.* Mult Scler Relat Disord, 2017. **18**: p. 141-143.
129. Baker, M.J., et al., *Using Fourier transform IR spectroscopy to analyze biological materials.* Nat Protoc, 2014. **9**(8): p. 1771-91.
130. Schirmer, L., et al., *Neuronal vulnerability and multilineage diversity in multiple sclerosis.* 2019(1476-4687 (Electronic)).

131. Jäkel, S., et al., *Altered human oligodendrocyte heterogeneity in multiple sclerosis.* Nature, 2019. **566**(7745): p. 543-547.
132. Davis, S., et al., *Development of a Sensitive, Scalable Method for Spatial, Cell-Type-Resolved Proteomics of the Human Brain.* Journal of proteome research, 2019. **18**(4): p. 1787-1795.
133. Ståhl, P.L., et al., *Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.* Science, 2016. **353**: p. 78-82.
134. Bellisola, G. and C. Sorio, *Infrared spectroscopy and microscopy in cancer research and diagnosis.* Am J Cancer Res, 2012. **2**(1): p. 1-21.
135. Goormaghtigh, E., et al., *Protein secondary structure content in solution, films and tissues: redundancy and complementarity of the information content in circular dichroism, transmission and ATR FTIR spectra.* Biochim Biophys Acta, 2009. **1794**(9): p. 1332-43.
136. Petibois, C., et al., *Biochemical Aspects of Overtraining in Endurance Sports.* 2003.
137. Benseny-Cases, N., et al., *Microspectroscopy (μ FTIR) reveals co-localization of lipid oxidation and amyloid plaques in human Alzheimer disease brains.* Anal Chem, 2014. **86**(24): p. 12047-54.
138. Traag, V.A., L. Waltman, and N.J. van Eck, *From Louvain to Leiden: guaranteeing well-connected communities.*
139. Huang, Y., D.J. McCarthy, and O. Stegle, *Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference.* Genome Biology, 2019. **20**(1): p. 273.
140. *Popscle GitHub*. [cited 2021; Available from: https://github.com/statgen/popscle/
141. Huang, X. and Y. Huang, *Cellsnp-lite: an efficient tool for genotyping single cells.* Bioinformatics, 2021. **37**(23): p. 4569-4571.
142. Wolock, S.L., R. Lopez, and A.M. Klein, *Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data.* Cell Syst, 2019. **8**(4): p. 281-291.e9.
143. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1): p. 15.
144. Tran, H.T.N., et al., *A benchmark of batch-effect correction methods for single-cell RNA sequencing data.* Genome Biology, 2020. **21**(1): p. 12.
145. Becht, E., et al., *Dimensionality reduction for visualizing single-cell data using UMAP.* Nature Biotechnology, 2019. **37**(1): p. 38-44.
146. Consortium, C.O.-M.-o.B.A., et al., *A blood atlas of COVID-19 defines hallmarks of disease severity and specificity.* medRxiv, 2021: p. 2021.05.11.21256877.
147. Crowell, H.L., et al., *muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data.* Nature Communications, 2020. **11**(1): p. 6077.

148. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics, 2012. **16**(5): p. 284-7.
149. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics (Oxford, England), 2010. **26**(1): p. 139-140.
150. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**(1): p. 559.
151. Hou, R., et al., *Predicting cell-to-cell communication networks using NATMI.* Nature Communications, 2020. **11**(1): p. 5011.
152. Alghamdi, N., et al., *A graph neural network model to estimate cell-wise metabolic flux using single cell RNA-seq data.* bioRxiv, 2021: p. 2020.09.23.310656.
153. Lopez, R., et al., *Multi-resolution deconvolution of spatial transcriptomics data reveals continuous patterns of inflammation.* bioRxiv, 2021: p. 2021.05.10.443517.
154. Griffiths, P.R. and J.A. de Haseth, *Introduction to Vibrational Spectroscopy.* Fourier Transform Infrared Spectrometry, 2007: p. 1-18.
155. Caine, S., et al., *The application of Fourier transform infrared microspectroscopy for the study of diseased central nervous system tissue.* Neuroimage, 2012. **59**(4): p. 3624-40.
156. Wehbe, K., et al., *Discrimination between two different grades of human glioma based on blood vessel infrared spectral imaging.* Analytical and bioanalytical chemistry, 2015. **407**(24): p. 7295-7305.
157. Martel, C., et al., *Diagnosis of idiopathic amyotrophic lateral sclerosis using Fourier-transform infrared spectroscopic analysis of patient-derived skin.* Analyst, 2020. **145**(10): p. 3678-3685.
158. Choo, L.P., et al., *Infrared spectroscopic characterisation of multiple sclerosis plaques in the human central nervous system.* Biochim Biophys Acta, 1993. **1182**(3): p. 333-7.
159. LeVine, S.M. and D.L. Wetzel, *Chemical analysis of multiple sclerosis lesions by FT-IR microspectroscopy.* Free Radic Biol Med, 1998. **25**(1): p. 33-41.
160. Lamba, O.P., D. Borchman, and W.H. Garner, *Spectral characterization of lipid peroxidation in rabbit lens membranes induced by hydrogen peroxide in the presence of Fe2+/Fe3+ cations: a site-specific catalyzed oxidation.* Free Radic Biol Med, 1994. **16**(5): p. 591-601.
161. Lamba, O.P., et al., *Spectroscopic detection of lipid peroxidation products and structural changes in a sphingomyelin model system.* Biochim Biophys Acta, 1991. **1081**(2): p. 181-7.
162. Chan, W.-K., et al., *Oxidative Amide Synthesis and N-Terminal α-Amino Group Ligation of Peptides in Aqueous Medium.* Journal of the American Chemical Society, 2006. **128**(46): p. 14796-14797.

163. Curtis, J.M., et al., *Protein carbonylation and metabolic control systems.* Trends in endocrinology and metabolism: TEM, 2012. **23**(8): p. 399-406.
164. Nie, B., J. Stutzman, and A. Xie, *A Vibrational Spectral Maker for Probing the Hydrogen-Bonding Status of Protonated Asp and Glu Residues.* Biophysical Journal, 2005. **88**(4): p. 2833-2847.
165. Hienerwadel, R., et al., *Protonation of Glu L212 following QB- formation in the photosynthetic reaction center of Rhodobacter sphaeroides: evidence from time-resolved infrared spectroscopy.* Biochemistry, 1995. **34**(9): p. 2832-43.
166. Hellwig, P., et al., *Carboxyl group protonation upon reduction of the Paracoccus denitrificans cytochrome c oxidase: direct evidence by FTIR spectroscopy.* FEBS Lett, 1996. **385**(1-2): p. 53-7.
167. Zarkovic, K., *4-hydroxynonenal and neurodegenerative diseases.* Mol Aspects Med, 2003. **24**(4-5): p. 293-303.
168. Kabuta, T., et al., *Aberrant molecular properties shared by familial Parkinson's disease-associated mutant UCH-L1 and carbonyl-modified UCH-L1.* Hum Mol Genet, 2008. **17**(10): p. 1482-96.
169. Brotherton, T.E., et al., *Localization of a toxic form of superoxide dismutase 1 protein to pathologically affected tissues in familial ALS.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5505-10.
170. Choo, L.P., et al., *In situ characterization of beta-amyloid in Alzheimer's diseased tissue by synchrotron Fourier transform infrared microspectroscopy.* Biophys J, 1996. **71**(4): p. 1672-9.
171. Szczerbowska-Boruchowska, M., et al., *Biomolecular investigation of human substantia nigra in Parkinson's disease by synchrotron radiation Fourier transform infrared microspectroscopy.* Arch Biochem Biophys, 2007. **459**(2): p. 241-8.
172. Belbachir, K., et al., *Collagen types analysis and differentiation by FTIR spectroscopy.* Analytical and Bioanalytical Chemistry, 2009. **395**(3): p. 829-837.
173. Cregg, J.M., et al., *Functional regeneration beyond the glial scar.* Experimental Neurology, 2014. **253**: p. 197-207.
174. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
175. Dalle-Donne, I., et al., *Protein carbonylation, cellular dysfunction, and disease progression.* J Cell Mol Med, 2006. **10**(2): p. 389-406.
176. Bizzozero, O.A., et al., *Elevated protein carbonylation in the brain white matter and gray matter of patients with multiple sclerosis.* (0360-4012 (Print)).
177. Rabek, J.P., W.H. Boylston, 3rd, and J. Papaconstantinou, *Carbonylation of ER chaperone proteins in aged mouse liver.* Biochem Biophys Res Commun, 2003. **305**(3): p. 566-72.

178. England, K. and T. Cotter, *Identification of carbonylated proteins by MALDI-TOF mass spectroscopy reveals susceptibility of ER.* Biochem Biophys Res Commun, 2004. **320**(1): p. 123-30.
179. Lovergne, L., et al., *An infrared spectral biomarker accurately predicts neurodegenerative disease class in the absence of overt symptoms.* Scientific Reports, 2021. **11**(1): p. 15598.
180. Horgan, C.C., et al., *Image-guided Raman spectroscopy probe-tracking for tumor margin delineation.* J Biomed Opt, 2021. **26**(3).
181. Kroupa, K.R., et al., *Raman needle arthroscopy for in vivo molecular assessment of cartilage.* Journal of Orthopaedic Research, 2021. **n/a**(n/a).
182. Grindberg, R.V., et al., *RNA-sequencing from single nuclei.* Proceedings of the National Academy of Sciences, 2013. **110**(49): p. 19802.
183. Slyper, M., et al., *A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors.* Nature Medicine, 2020. **26**(5): p. 792-802.
184. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nature Biotechnology, 2018. **36**(5): p. 411-420.
185. Absinta, M., et al., *A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis.* Nature, 2021. **597**(7878): p. 709-714.
186. Senanayake, V.K., et al., *Metabolic dysfunctions in multiple sclerosis: implications as to causation, early detection, and treatment, a case control study.* BMC Neurology, 2015. **15**(1): p. 154.
187. Pineda-Torra, I., et al., *Disrupted Lipid Metabolism in Multiple Sclerosis: A Role for Liver X Receptors?* Frontiers in Endocrinology, 2021. **12**(345).
188. Fard, M.K., et al., *BCAS1 expression defines a population of early myelinating oligodendrocytes in multiple sclerosis lesions.* Science translational medicine, 2017. **9**(419): p. eaam7816.
189. Fingerle-Rowson, G., et al., *The p53-dependent effects of macrophage migration inhibitory factor revealed by gene targeting.* Proceedings of the National Academy of Sciences, 2003. **100**(16): p. 9354.
190. Nguyen, M.T., et al., *The cytokine macrophage migration inhibitory factor reduces pro-oxidative stress-induced apoptosis.* J Immunol, 2003. **170**(6): p. 3337-47.
191. Wei, J., et al., *An Iron Metabolism-Related SLC22A17 for the Prognostic Value of Gastric Cancer.* OncoTargets and therapy, 2020. **13**: p. 12763-12775.
192. Konishi, H., et al., *Astrocytic phagocytosis is a compensatory mechanism for microglial dysfunction.* Embo j, 2020. **39**(22): p. e104464.
193. Iram, T., et al., *Megf10 Is a Receptor for C1Q That Mediates Clearance of Apoptotic Cells by Astrocytes.* The Journal of Neuroscience, 2016. **36**(19): p. 5185.

194. Pfisterer, U., et al., *Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis.* Nature Communications, 2020. **11**(1): p. 5038.
195. Manfredini, F., et al., *Rehabilitation Improves Mitochondrial Energetics in Progressive Multiple Sclerosis: The Significant Role of Robot-Assisted Gait Training and of the Personalized Intensity.* Diagnostics (Basel), 2020. **10**(10).
196. Bonora, M., et al., *ATP synthesis and storage.* Purinergic Signal, 2012. **8**(3): p. 343-57.
197. Miljković, D., et al., *A comparative analysis of multiple sclerosis-relevant anti-inflammatory properties of ethyl pyruvate and dimethyl fumarate.* J Immunol, 2015. **194**(6): p. 2493-503.
198. Kuračka, Ľ., et al., *Multiple Sclerosis: Evaluation of Purine Nucleotide Metabolism in Central Nervous System in Association with Serum Levels of Selected Fat-Soluble Antioxidants.*
199. Negrotto, L. and J. Correale, *Amino Acid Catabolism in Multiple Sclerosis Affects Immune Homeostasis.* The Journal of Immunology, 2017. **198**(5): p. 1900.
200. Vanlandewijck, M., et al., *A molecular atlas of cell types and zonation in the brain vasculature.* Nature, 2018. **554**(7693): p. 475-480.
201. Markowitz, C.E., et al., *The treatment of multiple sclerosis with inosine.* Journal of alternative and complementary medicine (New York, N.Y.), 2009. **15**(6): p. 619-625.
202. Tsakiris, I., et al., *Carboxypeptidase-M is regulated by lipids and CSFs in macrophages and dendritic cells and expressed selectively in tissue granulomas and foam cells.* Laboratory Investigation, 2012. **92**(3): p. 345-361.
203. Boven, L.A., et al., *Myelin-laden macrophages are anti-inflammatory, consistent with foam cells in multiple sclerosis.* Brain, 2006. **129**(2): p. 517-526.
204. Kawano, H., et al., *Role of the lesion scar in the response to damage and repair of the central nervous system.* Cell and Tissue Research, 2012. **349**(1): p. 169-180.
205. van Horssen, J., et al., *Extensive extracellular matrix depositions in active multiple sclerosis lesions.* Neurobiol Dis, 2006. **24**(3): p. 484-91.
206. van Horssen, J., C.D. Dijkstra, and H.E. de Vries, *The extracellular matrix in multiple sclerosis pathology.* J Neurochem, 2007. **103**(4): p. 1293-301.
207. Mohan, H., et al., *Extracellular matrix in multiple sclerosis lesions: Fibrillar collagens, biglycan and decorin are upregulated and associated with infiltrating immune cells.* Brain Pathol, 2010. **20**(5): p. 966-75.
208. Yahn, S.L., et al., *Fibrotic scar after experimental autoimmune encephalomyelitis inhibits oligodendrocyte differentiation.* Neurobiol Dis, 2020. **134**: p. 104674.
209. Dorrier, C.E., et al., *CNS fibroblasts form a fibrotic scar in response to immune cell infiltration.* Nature Neuroscience, 2021. **24**(2): p. 234-244.

210. Mercer, P.F. and R.C. Chambers, *Coagulation and coagulation signalling in fibrosis.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2013. **1832**(7): p. 1018-1027.
211. Liu, D., Z. Ke, and J. Luo, *Thiamine Deficiency and Neurodegeneration: the Interplay Among Oxidative Stress, Endoplasmic Reticulum Stress, and Autophagy.* Molecular neurobiology, 2017. **54**(7): p. 5440-5448.
212. Fumagalli, M., et al., *Pathophysiological Role of Purines and Pyrimidines in Neurodevelopment: Unveiling New Pharmacological Approaches to Congenital Brain Diseases.* Frontiers in Pharmacology, 2017. **8**: p. 941.
213. Maiuolo, J., et al., *Regulation of uric acid metabolism and excretion.* International Journal of Cardiology, 2016. **213**: p. 8-14.
214. Gupte, V.V., et al., *Overexpression of fibroblast growth factor-10 during both inflammatory and fibrotic phases attenuates bleomycin-induced pulmonary fibrosis in mice.* Am J Respir Crit Care Med, 2009. **180**(5): p. 424-36.
215. Wang, P., et al., *Basic fibroblast growth factor reduces scar by inhibiting the differentiation of epidermal stem cells to myofibroblasts via the Notch1/Jagged1 pathway.* Stem Cell Research & Therapy, 2017. **8**(1): p. 114.
216. Hu, B. and S.H. Phan, *Notch in fibrosis and as a target of anti-fibrotic therapy.* Pharmacol Res, 2016. **108**: p. 57-64.
217. Henderson, N.C., et al., *Targeting of α v integrin identifies a core molecular pathway that regulates fibrosis in several organs.* Nature medicine, 2013. **19**(12): p. 1617-1624.
218. Xu, X., et al., *Changes of ECM and CAM gene expression profile in the cirrhotic liver after HCV infection: analysis by cDNA expression array.* World journal of gastroenterology, 2005. **11**(14): p. 2184-2187.
219. Henderson, N.C. and D. Sheppard, *Integrin-mediated regulation of TGFβ in fibrosis.* Biochimica et biophysica acta, 2013. **1832**(7): p. 891-896.
220. Hallahan, D.E., L. Geng, and Y. Shyr, *Effects of Intercellular Adhesion Molecule 1 (ICAM-1) Null Mutation on Radiation-Induced Pulmonary Fibrosis and Respiratory Insufficiency in Mice.* JNCI: Journal of the National Cancer Institute, 2002. **94**(10): p. 733-741.
221. Shijubo, N., et al., *Circulating intercellular adhesion molecule-1 (ICAM-1) antigen in sera of patients with idiopathic pulmonary fibrosis.* Clinical & Experimental Immunology, 1992. **89**(1): p. 58-62.
222. Schenkel, A.R., et al., *Different susceptibilities of PECAM-deficient mouse strains to spontaneous idiopathic pneumonitis.* Experimental and molecular pathology, 2006. **81**(1): p. 23-30.
223. Dardiotis, E., et al., *Gene variants of adhesion molecules predispose to MS: A case-control study.* Neurology. Genetics, 2019. **5**(1): p. e304-e304.

224. Rudick, R.A. and A. Sandrock, *Natalizumab: alpha 4-integrin antagonist selective adhesion molecule inhibitors for MS.* Expert Rev Neurother, 2004. **4**(4): p. 571-80.
225. Kapoor, R., et al., *Effect of natalizumab on disease progression in secondary progressive multiple sclerosis (ASCEND): a phase 3, randomised, double-blind, placebo-controlled trial with an open-label extension.* Lancet Neurol, 2018. **17**(5): p. 405-415.
226. Christopoulos, P.F., et al., *Targeting the Notch Signaling Pathway in Chronic Inflammatory Diseases.* Frontiers in Immunology, 2021. **12**: p. 1194.
227. John, G.R., et al., *Multiple sclerosis: re-expression of a developmental pathway that restricts oligodendrocyte maturation.* Nature medicine, 2002. **8**(10): p. 1115-1121.
228. Genoud, S., et al., *Notch1 control of oligodendrocyte differentiation in the spinal cord.* The Journal of cell biology, 2002. **158**(4): p. 709-718.
229. Stidworthy, M.F., et al., *Notch1 and Jagged1 are expressed after CNS demyelination, but are not a major rate-determining factor during remyelination.* Brain, 2004. **127**(9): p. 1928-1941.
230. Seifert, T., et al., *Notch1 and its ligand Jagged1 are present in remyelination in a T-cell-and antibody-mediated model of inflammatory demyelination.* Acta neuropathologica, 2007. **113**(2): p. 195-203.
231. Aparicio, E., et al., *The notch signaling pathway: its role in focal CNS demyelination and apotransferrin-induced remyelination.* Journal of neurochemistry, 2013. **127**(6): p. 819-836.
232. Jurynczyk, M., et al., *Notch3 Inhibition in Myelin-Reactive T Cells Down-Regulates Protein Kinase C θ and Attenuates Experimental Autoimmune Encephalomyelitis.* The Journal of Immunology, 2008. **180**(4): p. 2634.
233. Fan, H., et al., *Effect of Notch1 gene on remyelination in multiple sclerosis in mouse models of acute demyelination.* J Cell Biochem, 2018. **119**(11): p. 9284-9294.
234. Sheykhansari, S., et al., *Redox metals homeostasis in multiple sclerosis and amyotrophic lateral sclerosis: a review.* Cell Death & Disease, 2018. **9**(3): p. 348.
235. Petrescu, A.J., et al., *The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding.* Embo j, 1997. **16**(14): p. 4302-10.
236. Gross, C. and G. Bassell, *Neuron-specific regulation of class I PI3K catalytic subunits and their dysfunction in brain disorders.* Frontiers in Molecular Neuroscience, 2014. **7**: p. 12.
237. Zhang, X.-Y., et al., *HAMI 3379, a CysLT2 receptor antagonist, attenuates ischemia-like neuronal injury by inhibiting microglial activation.* Journal of Pharmacology and Experimental Therapeutics, 2013. **346**(2): p. 328-341.

238. Saad, M., et al., *Montelukast, a cysteinyl leukotriene receptor-1 antagonist protects against hippocampal injury induced by transient global cerebral ischemia and reperfusion in rats.* Neurochemical research, 2015. **40**(1): p. 139-150.
239. Jovanovic, V.M., et al., *BMP/SMAD Pathway Promotes Neurogenesis of Midbrain Dopaminergic Neurons In Vivo and in Human Induced Pluripotent and Neural Stem Cells.* J Neurosci, 2018. **38**(7): p. 1662-1676.
240. Errico, F., et al., *Free D-aspartate regulates neuronal dendritic morphology, synaptic plasticity, gray matter volume and brain activity in mammals.* Translational Psychiatry, 2014. **4**(7): p. e417-e417.
241. Nicoletti, C.G., et al., *Oral D-Aspartate enhances synaptic plasticity reserve in progressive multiple sclerosis.* Mult Scler, 2020. **26**(3): p. 304-311.
242. Slomnicki, L.P., et al., *Requirement of Neuronal Ribosome Synthesis for Growth and Maintenance of the Dendritic Tree.* J Biol Chem, 2016. **291**(11): p. 5721-5739.
243. Holt, C.E., K.C. Martin, and E.M. Schuman, *Local translation in neurons: visualization and function.* Nature Structural & Molecular Biology, 2019. **26**(7): p. 557-566.
244. Mirzaei, H. and F. Regnier, *Identification of yeast oxidized proteins: Chromatographic top-down approach for identification of carbonylated, fragmented and cross-linked proteins in yeast.* Journal of Chromatography A, 2007. **1141**(1): p. 22-31.
245. Fusco, C.M., et al., *Neuronal ribosomes exhibit dynamic and context-dependent exchange of ribosomal proteins.* Nature Communications, 2021. **12**(1): p. 6127.
246. Mizuno, C.M., et al., *Numerous cultivated and uncultivated viruses encode ribosomal proteins.* Nature Communications, 2019. **10**(1): p. 752.
247. Hayashi, T., et al., *Induction of PML immunoreactivity in rat brain neurons after transient middle cerebral artery occlusion.* Neurol Res, 2001. **23**(7): p. 772-6.
248. Mackenzie, I.R., et al., *A family with tau-negative frontotemporal dementia and neuronal intranuclear inclusions linked to chromosome 17.* Brain, 2006. **129**(Pt 4): p. 853-67.
249. Takahashi, J., et al., *PML nuclear bodies and neuronal intranuclear inclusion in polyglutamine diseases.* Neurobiol Dis, 2003. **13**(3): p. 230-7.
250. Chort, A., et al., *Interferon β induces clearance of mutant ataxin 7 and improves locomotion in SCA7 knock-in mice.* Brain, 2013. **136**(Pt 6): p. 1732-45.
251. Yamada, M., et al., *Interaction between neuronal intranuclear inclusions and promyelocytic leukemia protein nuclear and coiled bodies in CAG repeat diseases.* Am J Pathol, 2001. **159**(5): p. 1785-95.
252. Lee, J., et al., *ATRX induction by mutant huntingtin via Cdx2 modulates heterochromatin condensation and pathology in Huntington's disease.* Cell Death Differ, 2012. **19**(7): p. 1109-16.

253. Skinner, P.J., et al., *Ataxin-1 with an expanded glutamine tract alters nuclear matrix-associated structures.* Nature, 1997. **389**(6654): p. 971-4.
254. Bitetto, G. and A. Di Fonzo, *Nucleo–cytoplasmic transport defects and protein aggregates in neurodegeneration.* Translational Neurodegeneration, 2020. **9**(1): p. 25.
255. Berggren, K., et al., *Amyloid Precursor Protein Haploinsufficiency Preferentially Mediates Brain Iron Accumulation in Mice Transgenic for The Huntington's Disease Mutation.* J Huntingtons Dis, 2017. **6**(2): p. 115-125.
256. Kleshchevnikov, V., et al., *Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics.* bioRxiv, 2020: p. 2020.11.15.378125.
257. Bayraktar, O.A., et al., *Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map.* Nature Neuroscience, 2020. **23**(4): p. 500-509.
258. Biancalani, T., et al., *Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram.* Nature Methods, 2021. **18**(11): p. 1352-1362.
259. Maseda, F., Z. Cang, and Q. Nie, *DEEPsc: A Deep Learning-Based Map Connecting Single-Cell Transcriptomics and Spatial Imaging Data.* Frontiers in Genetics, 2021. **12**: p. 348.
260. Loher, P. and N. Karathanasis, *Machine Learning Approaches Identify Genes Containing Spatial Information From Single-Cell Transcriptomics Data.* Frontiers in Genetics, 2021. **11**: p. 1743.
261. Dyring-Andersen, B., et al., *Spatially and cell-type resolved quantitative proteomic atlas of healthy human skin.* Nature Communications, 2020. **11**(1): p. 5587.
262. Moeendarbary, E., et al., *The soft mechanical signature of glial scars in the central nervous system.* Nature Communications, 2017. **8**(1): p. 14787.
263. Liu, Y., et al., *Myelin Debris Stimulates NG2/CSPG4 Expression in Bone Marrow-Derived Macrophages in the Injured Spinal Cord.* Frontiers in Cellular Neuroscience, 2021. **15**.
264. El-Khoury, N., et al., *Astrocyte End-Feet in Germinal Matrix, Cerebral Cortex, and White Matter in Developing Infants.* Pediatric Research, 2006. **59**(5): p. 673-679.
265. Bradbury, E.J. and E.R. Burnside, *Moving beyond the glial scar for spinal cord repair.* Nature Communications, 2019. **10**(1): p. 3879.
266. Domingues, H.S., et al., *Oligodendrocyte, Astrocyte, and Microglia Crosstalk in Myelin Development, Damage, and Repair.* Frontiers in Cell and Developmental Biology, 2016. **4**: p. 71.
267. Pereira, G.B., et al., *Targeted deletion of the antisilencer/enhancer (ASE) element from intron 1 of the myelin proteolipid protein gene (Plp1) in mouse reveals that*

the element is dispensable for Plp1 expression in brain during development and remyelination. J Neurochem, 2013. **124**(4): p. 454-65.
268. Falcão, A.M., et al., *PAD2-Mediated Citrullination Contributes to Efficient Oligodendrocyte Differentiation and Myelination.* Cell Reports, 2019. **27**(4): p. 1090-1102.e10.
269. Jahn, O., S. Tenzer, and H.B. Werner, *Myelin Proteomics: Molecular Anatomy of an Insulating Sheath.* Molecular Neurobiology, 2009. **40**(1): p. 55-72.
270. Nijland, P.G., et al., *Astroglial PGC-1alpha increases mitochondrial antioxidant capacity and suppresses inflammation: implications for multiple sclerosis.* Acta Neuropathologica Communications, 2014. **2**(1): p. 170.
271. Schreibelt, G., et al., *Protective effects of peroxiredoxin-1 at the injured blood–brain barrier.* Free Radical Biology and Medicine, 2008. **45**(3): p. 256-264.
272. Voigt, D., et al., *Expression of the Antioxidative Enzyme Peroxiredoxin 2 in Multiple Sclerosis Lesions in Relation to Inflammation.* International Journal of Molecular Sciences, 2017. **18**(4).
273. Holley, J.E., et al., *Peroxiredoxin V in multiple sclerosis lesions: predominant expression by astrocytes.* Mult Scler, 2007. **13**(8): p. 955-61.
274. Dias, D.O., et al., *Pericyte-derived fibrotic scarring is conserved across diverse central nervous system lesions.* Nature Communications, 2021. **12**(1): p. 5501.
275. Zhou, T., et al., *Microvascular endothelial cells engulf myelin debris and promote macrophage recruitment and fibrosis after neural injury.* Nature Neuroscience, 2019. **22**(3): p. 421-435.
276. Zhu, Y., et al., *Fibronectin Matrix Assembly after Spinal Cord Injury.* Journal of neurotrauma, 2015. **32**(15): p. 1158-1167.
277. Peng, H., et al., *Integrins and dystroglycan regulate astrocyte wound healing: the integrin beta1 subunit is necessary for process extension and orienting the microtubular network.* Dev Neurobiol, 2008. **68**(5): p. 559-74.
278. Sorci, G., et al., *The many faces of S100B protein: when an extracellular factor inactivates its own receptor and activates another one.* Ital J Anat Embryol, 2010. **115**(1-2): p. 147-51.
279. Hachem, S., et al., *Spatial and temporal expression of S100B in cells of oligodendrocyte lineage.* Glia, 2005. **51**(2): p. 81-97.
280. Deloulme, J.C., et al., *Nuclear expression of S100B in oligodendrocyte progenitor cells correlates with differentiation toward the oligodendroglial lineage and modulates oligodendrocytes maturation.* Mol Cell Neurosci, 2004. **27**(4): p. 453-65.
281. Hu, J., et al., *S100 beta stimulates inducible nitric oxide synthase activity and mRNA levels in rat cortical astrocytes.* J Biol Chem, 1996. **271**(5): p. 2543-7.
282. Lam, A.G., et al., *Mechanism of glial activation by S100B: involvement of the transcription factor NFkappaB.* Neurobiol Aging, 2001. **22**(5): p. 765-72.

283. Santos, G., et al., *Impaired oligodendrogenesis and myelination by elevated S100B levels during neurodevelopment.* Neuropharmacology, 2018. **129**: p. 69-83.
284. Santos, G., et al., *S100B Impairs Oligodendrogenesis and Myelin Repair Following Demyelination Through RAGE Engagement.* Frontiers in Cellular Neuroscience, 2020. **14**.
285. Mori, T., et al., *Overexpression of human S100B exacerbates brain damage and periinfarct gliosis after permanent focal ischemia.* Stroke, 2008. **39**(7): p. 2114-21.
286. Sheng, J.G., et al., *Overexpression of the neuritotrophic cytokine S100beta precedes the appearance of neuritic beta-amyloid plaques in APPV717F mice.* J Neurochem, 2000. **74**(1): p. 295-301.
287. Cirillo, C., et al., *S100B Inhibitor Pentamidine Attenuates Reactive Gliosis and Reduces Neuronal Loss in a Mouse Model of Alzheimer's Disease.* Biomed Res Int, 2015. **2015**: p. 508342.
288. Bloomfield, S.M., et al., *Reliability of S100B in predicting severity of central nervous system injury.* Neurocrit Care, 2007. **6**(2): p. 121-38.
289. Vos, P.E., et al., *GFAP and S100B are biomarkers of traumatic brain injury.* Neurology, 2010. **75**(20): p. 1786.
290. Sofroniew, M.V., *Molecular dissection of reactive astrogliosis and glial scar formation.* Trends in Neurosciences, 2009. **32**(12): p. 638-647.
291. Chuang, C.T., et al., *S100B is required for high glucose-induced pro-fibrotic gene expression and hypertrophy in mesangial cells.* Int J Mol Med, 2015. **35**(2): p. 546-52.
292. O'Connell, K.E., et al., *The use of cytokine signature patterns: separating drug naïve, interferon and natalizumab-treated multiple sclerosis patients.* Autoimmunity, 2014. **47**(8): p. 505-11.
293. Barateiro, A., et al., *S100B as a Potential Biomarker and Therapeutic Target in Multiple Sclerosis.* Mol Neurobiol, 2016. **53**(6): p. 3976-3991.
294. Bartosik-Psujek, H., et al., *Total tau and S100b proteins in different types of multiple sclerosis and during immunosuppressive treatment with mitoxantrone.* Acta Neurol Scand, 2011. **123**(4): p. 252-6.
295. Di Sante, G., et al., *The S100B Inhibitor Pentamidine Ameliorates Clinical Score and Neuropathology of Relapsing-Remitting Multiple Sclerosis Mouse Model.* Cells, 2020. **9**(3).
296. Wheeler, M.A., et al., *MAFG-driven astrocytes promote CNS inflammation.* Nature, 2020. **578**(7796): p. 593-599.
297. Al-Dalahmah, O., et al., *Single-nucleus RNA-seq identifies Huntington disease astrocyte states.* Acta Neuropathologica Communications, 2020. **8**(1): p. 19.

298. Agarwal, D., et al., *A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders.* Nature Communications, 2020. **11**(1): p. 4183.
299. Maniatis, S., et al., *Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis.* Science, 2019. **364**(6435): p. 89-93.
300. Stoffels, J.M.J., et al., *Fibronectin aggregation in multiple sclerosis lesions impairs remyelination.* Brain, 2013. **136**(1): p. 116-131.
301. Noreen, R., et al., *Detection of collagens in brain tumors based on FTIR imaging and chemometrics.* Analytical and Bioanalytical Chemistry, 2011. **401**(3): p. 845-852.
302. Lim, J. and Z. Yue, *Neuronal aggregates: formation, clearance, and spreading.* Developmental cell, 2015. **32**(4): p. 491-501.
303. Miller, L.M., M.W. Bourassa, and R.J. Smith, *FTIR spectroscopic imaging of protein aggregation in living cells.* Biochim Biophys Acta, 2013. **1828**(10): p. 2339-46.
304. Gorgoulis, V., et al., *Cellular Senescence: Defining a Path Forward.* Cell, 2019. **179**(4): p. 813-827.
305. Kõks, S., et al., *Mouse models of ageing and their relevance to disease.* Mechanisms of Ageing and Development, 2016. **160**: p. 41-53.
306. Peferoen, L.A., et al., *Ageing and recurrent episodes of neuroinflammation promote progressive experimental autoimmune encephalomyelitis in Biozzi ABH mice.* Immunology, 2016. **149**(2): p. 146-156.
307. Oost, W., et al., *Targeting senescence to delay progression of multiple sclerosis.* Journal of molecular medicine (Berlin, Germany), 2018. **96**(11): p. 1153-1166.
308. Barros, C., et al., *S100B inhibition protects from chronic experimental autoimmune encephalomyelitis.* Brain Communications, 2022. **4**(3): p. fcac076.