An automated pipeline for extracting quantitative histological metrics for voxelwise MRI-histology comparisons

Daniel Z.L. Kor[*1], Saad Jbabdi[1], Istvan N. Huszar[1], Jeroen Mollink[1], Benjamin C. Tendler[1], Sean Foxley[2], Chaoyue Wang[1], Connor Scott[3], Adele Smart[3], Olaf Ansorge[3], Menuka Pallebage-Gamarallage[3], Karla L. Miller[†1] and Amy F.D. Howard[†1]

[1]Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[2]Department of Radiology, University of Chicago, Chicago, IL, USA.

[3]Neuropathology, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[†]Contributed equally

*Address for Correspondence:

Daniel Zhi Liang Kor

Wellcome Centre for Integrative Neuroimaging,

FMRIB John Radcliffe Hospital, Headington,

Oxford OX3 9DU, UK

Tel: +44 (0) 1865 611450

Email: daniel.kor@ndcn.ox.ac.uk

Abstract

The acquisition of MRI and histology in the same post-mortem tissue sample enables direct correlation between MRI and histologically-derived parameters. However, there still lacks a standardised automated pipeline to process histology data, with most studies relying on heavy manual intervention. Here, we introduce an automated pipeline to extract quantitative histology metrics (stained area fraction) from multiple immunohistochemical (IHC) stains. The pipeline is designed to directly address key IHC artefacts related to tissue staining and slide digitisation. This pipeline was applied to post-mortem human brain data from multiple subjects, relating MRI parameters (FA, MD, R2*, R1) to IHC slides stained for myelin, neurofilaments, microglia and activated microglia. Utilising high-quality MRI-IHC co-registrations, we then performed whole-slide voxelwise comparisons (simple correlation, partial correlation and multiple regression analyses) between multimodal MRI- and IHC-derived parameters. The pipeline was found to be reproducible, robust to some artefacts and generalisable across multiple IHC stains. Our partial correlation results suggest that some simple MRI-IHC correlations should be interpreted with caution, due to the colocalisation of certain tissue features (e.g. myelin and neurofilaments). Further, we find activated microglia to consistently be the strongest predictor of DTI FA, which may suggest sensitivity of diffusion MRI to neuroinflammation. Taken together, these results show the utility of this approach in carefully curating IHC data and performing multimodal analyses to better understand microstructural relationships with MRI.

1 Introduction

Magnetic resonance imaging (MRI) is a powerful tool that can be used to evaluate neurodegenerative disorders in-vivo. MRI techniques have produced quantitative parameters sensitive to macroscopic neuropathological changes [1-3]. However, MRI parameters are generally non-specific and sensitive to multiple factors related to tissue microstructure and neuropathology. Coupled with millimetre resolution, this leads to difficulty in determining the microstructural underpinnings of a given MRI change (e.g., in disease).

Immunohistochemistry (IHC) is a histological staining technique that can be used to address this difficulty. IHC uses primary antibodies to stain target antigens (proteins related to microstructural features of interest) with high specificity. The distributions of antigens are then commonly visualised using the chromogen diaminobenzidine (DAB) to stain the marked proteins brown. To add contextual information and help localise stained tissue features, hematoxylin is often used as a counterstain to mark cell nuclei purple. The acquisition of IHC and MRI in the same post-mortem tissue sample enables direct correlation of MR and histologically-derived metrics. This is a common methodology used to validate MR parameters and improve our microstructural interpretation of MRI. Here, post-mortem MRI functions as a crucial intermediary between the IHC data and in-vivo imaging. Post-mortem MRI shares a common tissue state with IHC, while possessing the same signal forming mechanisms with in-vivo MRI [4].

While many studies relate IHC to MR parameters [5-13], there is still a lack of automated pipelines for extracting quantitative metrics from IHC slides [14]. Most MRI-IHC analyses rely on heavy manual intervention [14-17], with some using semi-quantitative metrics, such as staining intensity scores resembling *low*, *moderate* or *strong* [18,19]. More quantitative metrics can be derived via "colourimetry" to provide a more detailed description of IHC staining. When multiple stains have been applied to the same tissue section, colourimetry utilises the stains' colour information to separate the densities for different stain channels (e.g. separating the DAB-stained densities from background hematoxylin) [20]. Some IHC analyses use the DAB channel's density to approximate the amount of targeted protein within the tissue [18, 21-23]. However, this interpretation is limited as the densities do not scale linearly with protein concentration [17, 22, 24]. To circumvent this, other pipelines extract the stained area fraction (SAF), i.e. the number of pixels stained by DAB within a given area. In most pipelines, SAF is quantified by manually setting a threshold for the DAB channel to segment microstructural tissue compartments from non-specific background staining in regions-of-interest (ROIs) [14-17]. The use of SAF as a quantitative metric has two main issues.

First, the manually-set threshold may not be optimal for any given slide. Manual thresholds are dependent on the operator's expertise. Further, a single threshold is often applied to all slides in a dataset or a batch of slides which have been processed together. Although this avoids intra-observer variability and can be time efficient, it comes at the expense of optimising thresholds for individual slides. This can result in a lack of robustness to key histological artefacts, such as slide-to-slide and within-slide staining intensity variation. Even when a single tissue preparation protocol is followed, slide-to-slide staining intensity variation is introduced due to variations in tissue sample preparation and staining protocol [25-27], which produce artificial (non-biological) differences in stain intensity and colour information. Examples of within-slide staining artefacts include a gradual staining gradient, with stronger staining at one end of the slide progressing to weaker staining at the other, where the axis of the gradient is

related to how the slide is positioned during staining, and striping artefacts from slide digitisation. The latter describes sharp vertical bands of intensity variation or striping across whole slide images, which arise from imperfect illumination and/or the stitching together of multiple strips of the image when scanning the slide [28]. When unaccounted for, these artefacts may impact the extracted IHC metrics.

Second, these manual workflows are time intensive in nature. This restricts research studies to smaller sample sizes (i.e. less slides and/or subjects), and limits IHC analyses to hand-drawn ROIs, as opposed to analysing voxels from the whole slide. Ideally, an automated or semi-automated workflow will allow MRI-IHC correlations to be derived at the MRI voxel level. To achieve this, the workflow will need to be able to 1) to produce whole-slide SAF maps from IHC data, and 2) co-register MRI and SAF maps to enable voxelwise analysis.

Here, we propose an automated IHC processing pipeline, in the context of an end-to-end MRI-IHC workflow, to address these challenges. The automated IHC processing pipeline is able to extract SAF maps from IHC-stained slides for myelin, neurofilaments and microglia. The pipeline was first evaluated on an IHC dataset designed to test the pipeline's reliability. Once evaluated, we applied it on a second dataset containing both IHC and MRI data. We register and correlate these SAF maps on a voxelwise basis with diffusion-MR (fractional anisotropy, FA; mean diffusivity, MD) and relaxometry (R2*, R1) maps. To account for covariance between stains, we use partial correlation to identify the unique variance in MR parameters explained by each targeted protein. Finally, we perform multiple regression with all stains to derive a predictive model of each MR parameter, which may be driven by multiple microstructural features.

2 Data Acquisition

2.1 Immunohistochemistry data

We applied the IHC pipeline to two datasets: one that was acquired specifically to evaluate the reliability of our pipeline and which includes only IHC data, and a second, previously described dataset that includes both IHC and co-registered MRI [9]. Both datasets contain tissue from the same 15 post-mortem brains of patients diagnosed with motor neuron disease (MND) and healthy controls (CTL) (12 x MND, 3 x CTL).

All IHC slides were digitised at x20 objective magnification with the Aperio ScanScope® AT Turbo (Leica Biosystems) high throughput slide scanner. The approximate resolution of each slide is 0.5 µm/pixel. All antibodies were visualised with DAB and the slides were counterstained with hematoxylin.

2.1.1 Evaluation dataset

The "evaluation dataset" was acquired to evaluate our pipeline's performance in terms of reproducibility and robustness to key histological artefacts. Ideally, this data can distinguish true biological variation (e.g. due to differences in pathology, which is assumed to be consistent between adjacent IHC slides) from variance related to staining artefacts and analysis. Consequently, data was collected from adjacent tissue slides. Our analyses using the evaluation dataset make the implicit

assumption that these slides have similar tissue microstructure and that the true biological between-slide variance is low.

For each brain, 12-15 consecutive slides (6-µm thick) were obtained from the primary motor cortex (face region). These adjacent slides were separated into groups of 4-5 slides, where each group was stained for either activated microglia (CD68), myelin (PLP) or phosphorylated neurofilaments (SMI312). The staining protocol is identical to that in [9] and full staining protocols are provided in **Supplementary Materials**. For the evaluation of reproducibility, these slides were first manually quality-checked on a per slide basis. The criterion was based on whether the slides contained artefacts that are unlikely to yield useful SAF estimates from both manual and automated pipeline. These include excessively illuminated slides, significant amounts of inconsistent staining, and slides that have been sectioned poorly.

2.1.2 Multiple-region dataset

We then applied our IHC processing pipeline (Section 3.1) on the main "multiple-region dataset" [9] which will soon be fully available in the Digital Brain Bank [4]. This broader dataset includes IHC and co-registered MRI data from multiple brain regions with varying levels of disease pathology. Here, we consider data from the visual cortex (VC; both hemispheres), anterior cingulate (AC; cingulum bundle, corpus callosum) and hippocampus (HC) which either display less (VC) or some disease pathology (AC, HC) in MND. Sections were stained to visualise activated microglia (CD68), all microglia (Iba1), myelin (PLP) and phosphorylated neurofilaments (SMI312). Prior to any analysis, all slides were quality-checked similarly to the evaluation dataset.

2.2 MRI data

Post-mortem MRI was also acquired for the MND-Control cohort. The data acquisition and preprocessing procedures of multimodal MRI data have been previously described ([9], [11], [29]). In brief, whole brains were imaged in a 7T human scanner (Siemens Healthcare, Erlangen, Germany) using a 1Tx/32Rx head coil. R2* maps were estimated from susceptibility weighted data acquired with a 3D multi-echo GRE sequence (parameters: TEs=2, 8.6, 15.2, 21.8, 28.4, 35 ms with monopolar readout and non-selective RF pulse, TR=38 ms, flip angle=15°, bandwidth=650 Hz/pixel, and in-plane resolution=0.5 × 0.5 mm^2 [11]). R1 maps were estimated from T1-weighted data collected with a multi-TI turbo spin-echo protocol (parameters: TE=14.0 ms, TR=1000 ms, TIs=60, 120, 240, 480, 935 ms, flip angles=90° and 180°, bandwidth=130 Hz/pixel, and in-plane resolution=1.0 × 1.0 mm^2 [9]). Diffusion tensor maps including fractional anisotropy (FA) and mean diffusivity (MD) [30] were estimated from a diffusion-weighted steady-state free precession sequence (parameters: TE=21.0 ms, TR=28.0 ms, flip angles=24° and 94°, bandwidth=393 Hz/pixel, q value=300 cm^{-1}, number of directions/flip angle=120, and in-plane resolution=0.85 × 0.85 mm^2 [29]).

2.3 Data co-registration

MRI and histology data were previously co-registered using FSL's Tensor Image Registration Library (TIRL) as described in detail in [31]. Briefly, FSL-TIRL is a general-purpose image registration framework specifically designed for MRI-microscopy coregistration. We used FSL-TIRL to run 2D-2D co-registration, aligning other stains to PLP from the same subject, as well as the 2D-3D co-registration of PLP with structural MRI in the multi-region dataset. PLP was chosen as the

reference microscopy data due to its strong white/grey matter contrast. All MR parameters maps were first aligned to structural MRI using FSL-FLIRT [38]. The warps generated were then combined with FSL-TIRL to map these MR parameters to the different IHC SAF maps for voxelwise correlations.

3 Methods

This section will first provide an overview of the end-to-end workflow for MRI-histology comparisons (Section 3.1). At the centre of this workflow is our automated IHC processing pipeline, which is described in detail in Section 3.2. The pipeline was designed to be both reproducible and robust to common IHC artefacts. Evaluation of this is described in Section 3.3. The pipeline was then applied to the "multi-region dataset" which includes IHC and co-registered MRI, facilitating voxelwise IHC-MRI comparisons. The final sections describe how we evaluated the quality of the co-registration (Section 3.4), and performed voxelwise MRI-IHC correlation and regression analysis (Sections 3.5 & 3.6) to disentangle the contributions of multiple microstructural features stained with IHC to each MR parameter.

3.1 The MRI-histology workflow

An overview of the end-to-end MRI-IHC analysis workflow is given in Figure 1. The IHC processing pipeline receives high-resolution (0.5 µm/pixel) IHC slides and outputs SAF maps at a given target resolution. Utilising precise MRI-IHC co-registrations generated from the Tensor Image Registration Library (FSL-TIRL) [31], we perform voxelwise MRI-IHC correlation and regression analysis to investigate relationships between multiple MR parameters and multiple proteins at a voxelwise level.

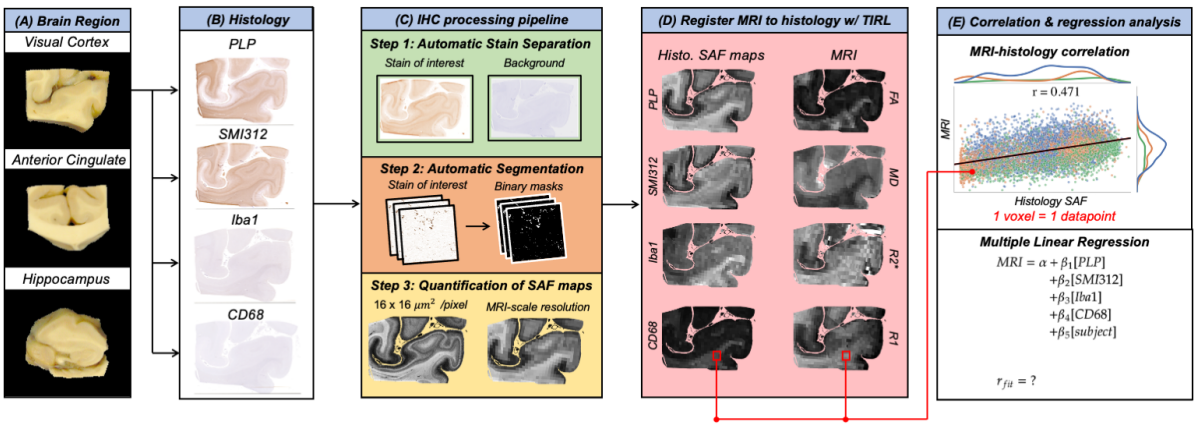

Figure 1: An overview of the MRI-IHC analysis. A: Brain regions are sectioned for immunohistochemistry (IHC). B: Sections are stained for myelin (PLP), neurofilament (SMI312), activated microglia (CD68) & all microglia (Iba1). C: IHC images are processed to output stained area fraction (SAF) maps. D: SAF & MRI maps are then co-registered voxelwise via FSL-TIRL. E: Voxels are pooled to perform pairwise correlations and multiple regression between MRI & SAF. Each datapoint in the scatter plot represents a single voxel; the colours represent data from different brain regions (green for visual cortex, orange for cingulum bundle and corpus callosum, blue for hippocampus).

3.2 Immunohistochemistry Processing Pipeline

The IHC processing pipeline provides a mapping of RGB intensity values of IHC slides to SAF maps. The derivation of SAF maps is comprised of three steps: 1) a data-driven colour matrix derivation to separate the DAB and haematoxylin stains, 2) automatic thresholding to segment the DAB-stained proteins-of-interest from non-specific DAB staining, and 3) calculating the stained area fraction (SAF) map at a specified resolution. In this study, we present SAF maps both at high resolution (16 µm/pixel) and MRI-scale resolution (0.5 mm/pixel).

Ideally, the pipeline should be generalisable to all DAB-stained proteins. In practice, we found visible impact of IHC artefacts in some slides. This motivated the development of two 'configurations' of the pipeline that account for whether the slides showed a strong staining gradient and/or striping artefact ("artefact" configuration) or not ("default" configuration).

We applied the "default" configuration (Figure 2) to microglia-stained slides (Iba1, CD68), where we observed exhibited a less visible staining gradient and/or striping. Conversely, the "artefact" configuration (Figure 3) was used on myelin-stained slides (PLP), where we saw strong staining gradients and/or striping artefacts across all IHC slides in our datasets. The presence or absence of a staining gradient was less consistent in the neurofilament-stained slides (SMI312). Consequently, we manually checked SMI312 slides in both the MND and evaluation datasets before choosing a pipeline configuration.

We now describe each step of the IHC processing pipeline in detail.

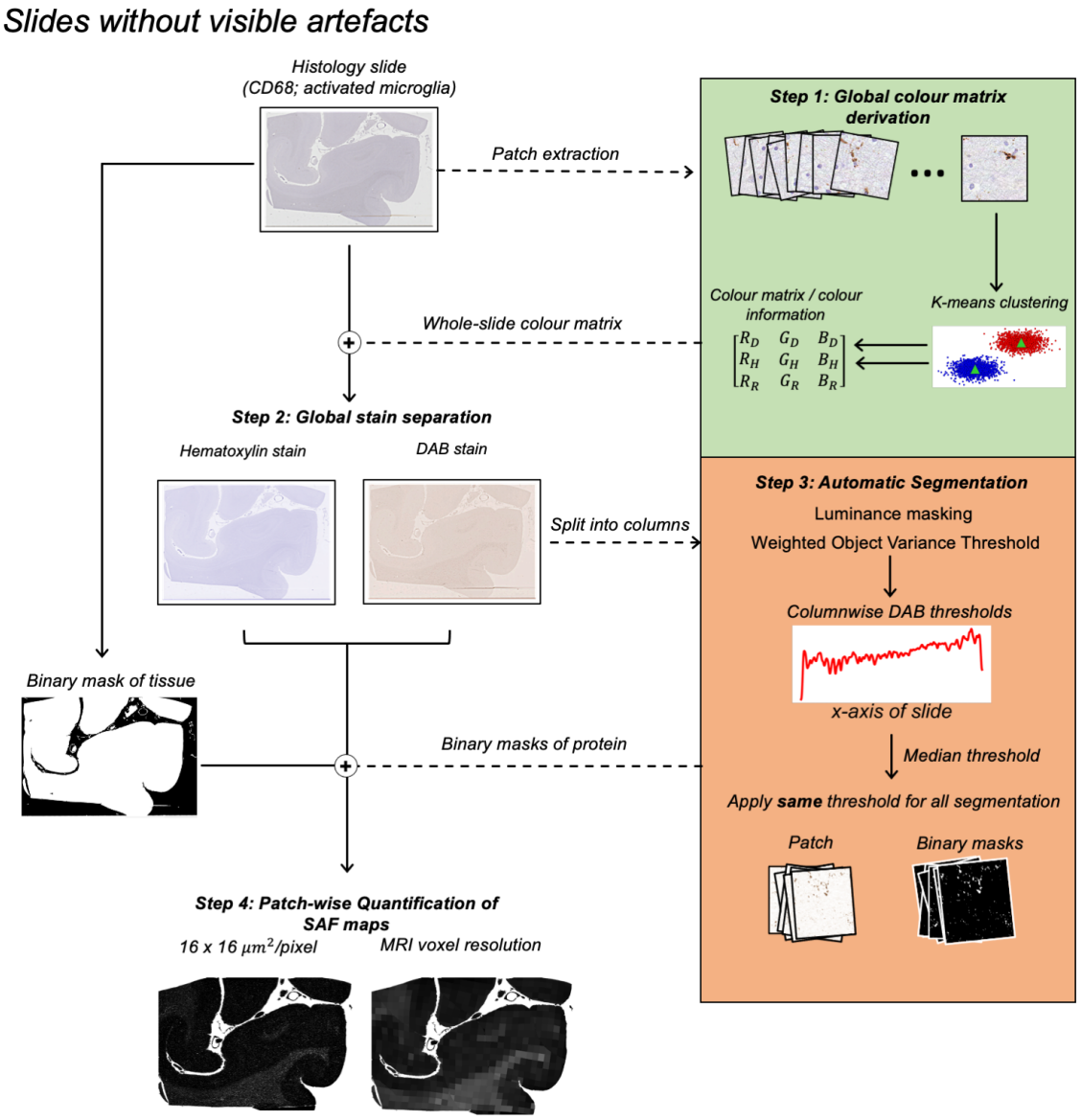

Figure 2: The automated IHC pipeline for stains with little or no intensity gradient and/or stitching artefacts. This usually includes slides of stains specifying structures that sparsely populate the brain tissue, such as microglia (Iba1, CD68) and neurofilament (SMI312) in some cases. For each slide, the pipeline 1) derives a single, global colour matrix to separate DAB from haematoxylin, 2) performs stain separation to isolate the DAB channel, 3) automatically segments the DAB channel with a single median threshold, and 4) calculates the SAF at variable resolution.

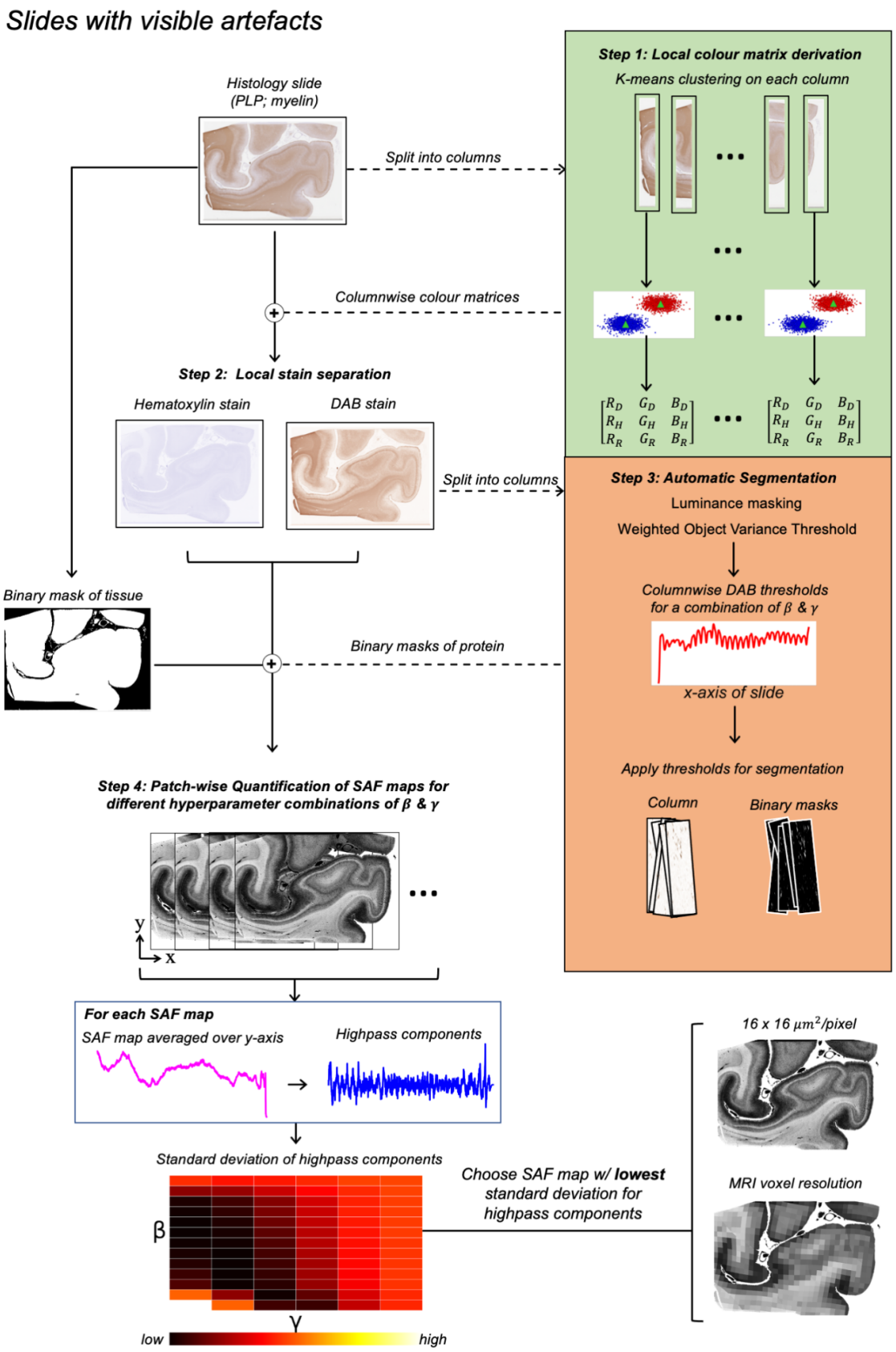

Figure 3: The automated IHC pipeline for stains confounded by staining gradients and/or striping artefacts. For each slide, the pipeline 1) derives multiple, columnar colour matrices from the data, 2) performs stain separation for each column, 3) automatically segments each column's DAB channel

and 4) forms an SAF map. Steps 1-4 are repeated for a range of hyperparameters **β** and **γ**, which are optimised via grid-search to account for within-slide artefacts (Section 3.1.2). For each SAF map, we averaged the map over its height and chose the SAF map with the lowest standard deviation (i.e. least impacted by these artefacts).

3.2.1 Colour Matrix Derivation for Stain Separation

3.2.1.1 Colourimetric Theory

Colourimetric analysis is based on the framework for colour deconvolution [20], a stain separation method based on Beer-Lambert's law. In Beer-Lambert's law, absorbance (A_λ) is used to relate the attenuation of monochromatic incident light ($I_{inc,\lambda}$) passing through the IHC slide to the detected light ($I_{det,\lambda}$):

$$I_{det,\lambda} = I_{inc,\lambda} 10^{-A_\lambda} \tag{1}$$

where A_λ:

$$A_\lambda = \sum_{i=1}^{N} \epsilon_{i,\lambda} C_i \tag{2}$$

Here, N represents the number of stains, $\varepsilon_{i,\lambda}$ denotes the attenuation coefficient of ith stain for light at wavelength λ, and C_i represents the stain density of ith stain. For λ corresponding to red, green and blue colour channels, the stain density of DAB, hematoxylin, and the leftover staining (residual) for a pixel with $I_{det,\lambda}$ on the slide can be computed with a matrix inversion:

$$\begin{bmatrix} C_{DAB} \\ C_{hematoxylin} \\ C_{residual} \end{bmatrix} = \begin{bmatrix} \epsilon_{DAB,red} & \epsilon_{DAB,green} & \epsilon_{DAB,blue} \\ \epsilon_{hema,red} & \epsilon_{hema,green} & \epsilon_{hema,blue} \\ \epsilon_{res,red} & \epsilon_{res,green} & \epsilon_{res,blue} \end{bmatrix}^{-1} \begin{bmatrix} A_{red} \\ A_{green} \\ A_{blue} \end{bmatrix} = \mathbf{M}^{-1} \begin{bmatrix} A_{red} \\ A_{green} \\ A_{blue} \end{bmatrix} \tag{3}$$

This process is called a stain separation. **M**, termed the colour matrix, is a 3 x 3 matrix with rows defining the stains of interest (i), and columns denoting each colour's attenuation coefficient value ($\varepsilon_{i,\lambda}$). The row is the stain's colour vector.

M can be defined from literature or empirically measured from single-stained slides [20], with literature values being currently used in the ImageJ Colour Deconvolution plugin [32]. However, using an **M** non-specific to each slide leads to poor stain separation and consequently, inconsistent interpretation during slide-to-slide comparison [25, 27, 33].

3.1.1.2 Data-driven Colour Matrix Derivation

To address the problem of using colour information (**M**) ill-suited to the dataset, we employed a K-means clustering approach to derive colour information directly from the IHC data. This approach is tailored according to whether a slides displays prominent artefacts:

Default: For slides with a less noticeable staining gradient and/or striping (CD68, Iba1), we used the K-means clustering strategy introduced by Geijs et. al. [34] to derive a colour matrix, **M**, specific to the whole slide (Figure 3, Box 2). In brief, we first randomly sampled patches (1024 x 1024 pixels; n=200) from the whole slide image and performed K-means clustering (k=2) to produce 2 cluster centroids corresponding to the colour vectors (i.e. rows in **M**) of hematoxylin and DAB. We then performed a second k-means clustering (k=2), this time on the 200 colour vectors output from the first k-means. This produced two centroids that correspond to the DAB and haematoxylin colour vectors of **M** specific to the slide. Stain separation was then performed using a single M across the whole slide.

Artefact: In slides with more visible staining gradient and/or striping (PLP), we applied K-means locally to account for colour differences along the gradient (Figure 3, Box 2). In our datasets, a staining gradient was present along the horizontal axis of the slide. We addressed this by applying a K-means to columns of data (32-pixels width; height matching the slide) to define $\mathbf{M_{column}}$ on a column-wise basis. Stain separation was then performed locally using $\mathbf{M_{column}}$.

3.2.1.3 Stain Separation

After deriving M, an approach inspired by NNLS was used to estimate the stain density of the DAB channel according to Equation 3. The stain density (D) was then converted to intensity values (I) using $I = log^{-1}{}_{10}(D)$.

3.2.2 Automatic Thresholding for Protein Segmentation

A threshold was applied to the DAB channel to segment DAB related to our structure-of-interest from faint "background" or non-specific DAB. To derive a data-driven threshold, we used a method inspired by weighted object variance (WOV) [35], a histogram-based technique to separate classes of unequal count and variance. This differs from the well-known Otsu's method [36], which assumes equal classes in the histogram.

3.2.2.1 Weighted Object Variance Cost Function

In WOV, the threshold **t** is computed by minimising a cost function comprised of the inter-class variance of both classes:

$$t = argmax_{0 < t < 255} [\sigma_{IC}] \tag{4}$$

where:

$$t_i = argmax_{0 < t < 255} \left[P_0(t)^{1+\delta} * \mu_0(t)^2 + P_1(t) * \mu_1(t)^2 \right] \tag{5}$$

Here, $\boldsymbol{\sigma_{IC}}$ represents the inter-class variance; $\mathbf{P_i(t)}$ and $\boldsymbol{\mu_i(t)}$ are the cumulative probabilities and means of classes i=0 (smaller class) and i=1 (larger class), respectively. Hence, $\boldsymbol{\delta > 0}$ is used to produce a threshold which is closer to the mean of class 0. In [35], Yuan et. al. use $\boldsymbol{\delta}$=1, though in principle, any variable or constant value may be used.

3.2.2.2 Proposed Cost Function

In this study, the two uneven classes were the protein-of-interest (class 0) and the non-specific DAB-stained background (class 1). However, a key distinction is that the protein-of-interest may

either be the larger (PLP in myelin) or smaller (CD68, Iba1 in microglia) class. Consequently, a different δ was used for each stain.

For each stain, δ was decided manually based on the desired segmentation of a specific protein. Further, we weighted $\mathbf{P_1(t)}$ (the background class) with **δ** rather than $\mathbf{P_0(t)}$ as this was found to produce the full range of desired segmentation thresholds. As such, we propose a similar cost function that is again a function of the histogram structure:

$$t_i = argmax_{0 < t < 255} \left[P_0(t) * \mu_0(t)^2 + P_1(t)^{1+\delta} * \mu_1(t)^2 \right] \quad (6)$$

Thresholds $\mathbf{t_i}$ were calculated on a column-wise basis, where **i** represents the column from which the DAB values were extracted (32-pixels width; height matching the slide). A single **δ** was set for each stain, where -1 < **δ** < 1. When **δ** = 0, we obtain the cost function used in Otsu's method [36], which works optimally in cases where both classes have equal count and variance. By adjusting **δ** to be <0, we were able to accurately segment a protein-of-interest (class 0) with a higher count (PLP). Conversely, a **δ** > 0 produced a desired threshold for proteins-of-interest with a lower count (CD68, Iba1).

Before applying this cost function, we first masked the DAB channel with masks generated from the slide's luminance channel (luminance < 0.75). This is to remove the brightest pixels in the column, which were typically related to non-specific DAB-stained background. This masking procedure was found to help balance the two classes in the DAB channel histogram, especially in microglia-stained slides where there was an extreme imbalance between the classes (i.e. activated microglia were very sparse). We then produced thresholds on a column-wise basis to derive a 1D threshold profile, where each datapoint represented a threshold computed from a column.

3.2.2.3 Slides without visible artefacts

Slides with less noticeable staining gradients and/or stitching artefacts are typically found in stains specifying structures that sparsely populate brain tissue (CD68, Iba1 for microglia). Hence, we noted that some columns within the IHC slide may not contain the protein-of-interest. In this case, the local thresholding approach may lead to erroneous segmentation. Given that these IHC slides also did not contain visible sources of within-slide variation, we concluded that a single whole-slide threshold was sufficient for decent segmentation. This whole-slide threshold was estimated directly from the data by using the median of the derived 1D threshold profile.The threshold was applied to the DAB channel to obtain the segmentation mask. This thresholding process was used for the "default" configuration.

Optimisation of hyperparameter δ

As previously described, the hyperparameter δ allows flexibility in deciding what segmentation is desired. To decide a suitable δ for each stain, we extracted a random patch (1024 x 1024 pixels) from at least 8 IHC slides extracted from multiple brain regions and applied the threshold generated from multiple δ values. The patches span structurally distinct tissue types, such as white or grey matter. We asked an expert (MPG) to decide a hyperparameter δ that yields stain segmentation that reflects the underlying tissue biology. We chose one δ per stain: δ=0.05 for Iba1 and CD68. Example segmentations using δ are shown in Figure 4.

3.2.2.4 Slide with visible artefacts

In slides with very visible staining gradients and striping artefacts, columns affected by striping produced DAB histograms with decreased median absolute deviation (MAD) when compared to slides with less striping artefacts. A narrower than usual histogram results in a threshold that oversegements, thereby artificially inflating the SAF in these columns.

To address this, we further weight each column's **δ** to account for how much stitching artefact is present:

$$\delta_i = \alpha\,(1 + \Delta_i)^{\beta} \tag{7}$$

where:

$$\Delta_i = \frac{MAD_i - \sigma(MAD_i)}{\sigma(MAD_i)} \tag{8}$$

Here, **α** is the main weight attached to the cost function. $\mathbf{MAD_i}$ is the median absolute deviation of a DAB channel histogram extracted from the ith column. This value is obtained from all columns in a slide, and represented as a datapoint in a 1D MAD profile. This MAD profile is then smoothed with a 1D Gaussian filter (**σ**=16), with $\boldsymbol{\sigma(MAD_i)}$ denoting the smoothed MAD in the ith column. As such, $\mathbf{\Delta_i}$ represents the ith column's change in structure (MAD) relative to neighbouring columns. A large Δi represents a sudden, large change in the histogram structure, indicative of the ith column being in a striping region. Depending on how much stitching artefact is present, we modulate this weight with a power term **β**. If a stitching artefact is more prominent in an IHC slide, a higher **β** is required.

Using the same cost function in Equation 6 with this new $\mathbf{\delta_i}$, we derived a similar 1D threshold profile for each slide with artefacts. This threshold profile was then smoothed using a 1D Gaussian filter with a kernel size (γ) to prevent any spurious columns from significantly affecting the segmentation. Hence, with a given combination of **α, β** and **γ**, we can generate a resultant 1D threshold profile to be applied to the slide's DAB channel to obtain the segmentation mask. This thresholding process was used for the "artefact" configuration.

Optimisation of hyperparameters α, β, γ

The hyperparameter α was optimised using the same procedure for delta above (default configuration). We chose an α=-0.60 for PLP, and α=-0.30 for SMI312. Example segmentations using these α are shown in Figure 4.

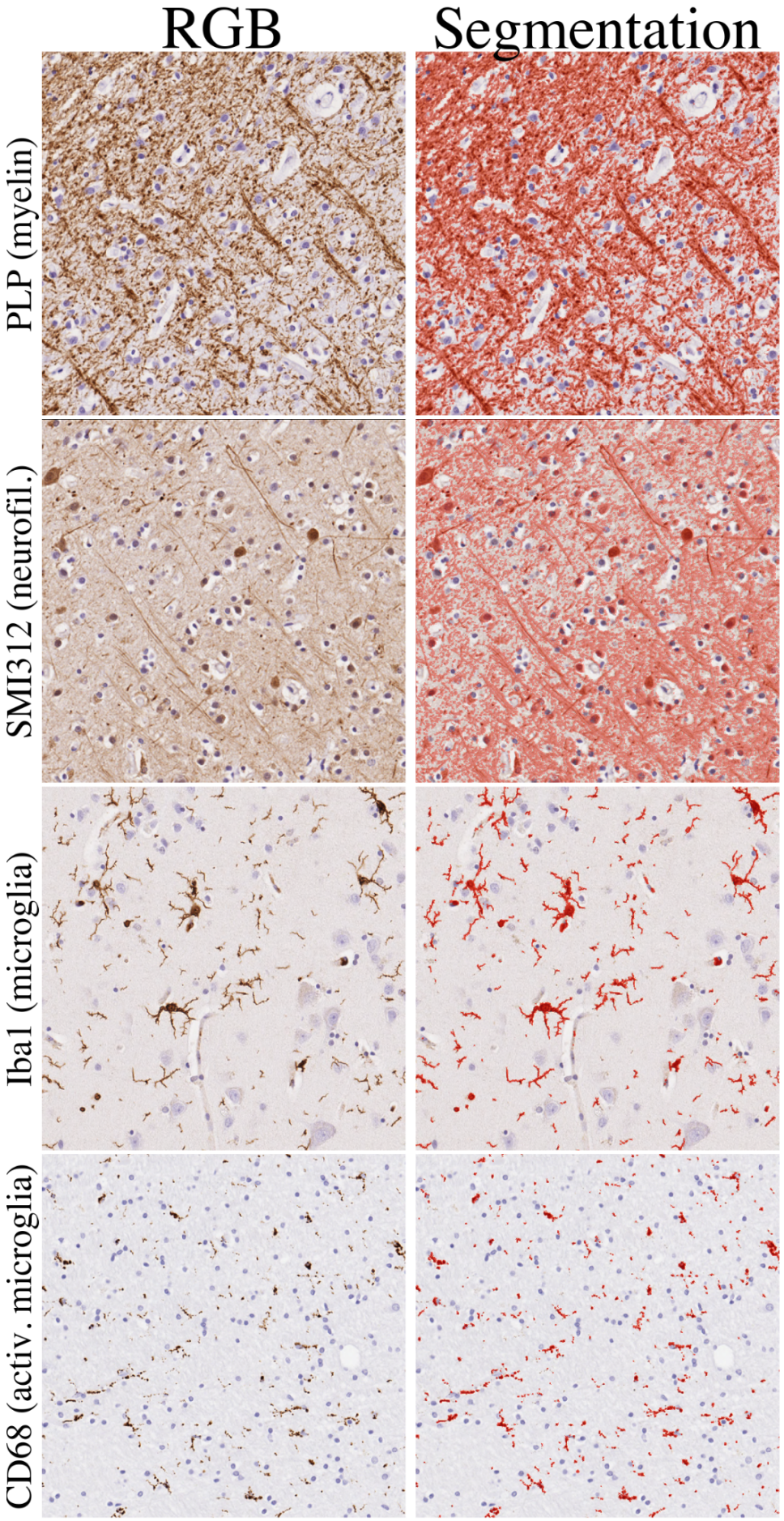
RGB
Segmentation
PLP (myelin)
SMI312 (neurofil.)
Iba1 (microglia)
CD68 (activ. microglia)

Figure 4: Examples of the segmentation (right column; red overlay) produced from the automated IHC pipeline with optimised hyperparameters. The PLP (myelin) and SMI312 (neurofilament) patches were extracted from the WM and GM boundary in the visual cortex, the Iba1 (microglia) patch was sampled from hippocampal GM, and the CD68 (activated microglia) was taken from hippocampal WM.

The hyperparameters **β** and **γ** were optimised using a grid search strategy. We generated multiple SAF maps with different combinations of β and γ and selected the SAF map with the lowest standard deviation of highpass components, i.e. the SAF map least affected by sharp intensity changes. The underlying assumption was that all sharp intensity changes are uncharacteristic of tissue microstructure, and therefore introduced by IHC artefacts. The standard deviation of highpass components is defined properly in "3.1.4.2 Robustness to within-slide artefacts".

3.2.3 Calculating the SAF map

Following segmentation, we computed stained area fraction (SAF) maps at various resolutions. First, we applied Otsu's method [36] on the hematoxylin channel to extract a tissue mask. This tissue mask was then modified with a series of morphological operations (dilation, median filtering, binary hole filling). The SAF was then defined as the ratio of pixels with positive DAB to those in the tissue mask within the local neighbourhood. We chose to generate SAF maps at high resolution (16 x 16 μm^2 per patch) to evaluate its robustness to quickly-varying within-slide artefacts. SAF maps were also generated at MRI resolution (0.5 x 0.5 mm^2 per patch) to test how reproducible the SAF maps are, given that MRI-IHC correlations will only be computed at this resolution.

3.3 Quantitative Evaluation of IHC Pipeline

The IHC pipeline describes how IHC slides are processed to produce SAF maps. The pipeline was evaluated with two criteria: robustness to stitching and staining gradient artefacts, and reproducibility of SAF values produced. The pipeline was applied to the "evaluation dataset" (c.f Section 2.1), where we assume that adjacent slides have similar tissue microstructure and that the true biological between-slide variance is low.

To provide a reference for both criteria, we compared the automatically produced SAF maps with SAF maps derived manually by an expert (MPG), as described in [9]. In brief, a single dataset-specific colour matrix and segmentation threshold was derived for each protein. This was achieved by manually calibrating the threshold in at least 10 randomly selected structurally distinct regions where they define a 'strong', 'moderate' and 'weak' threshold. Here, we segment DAB according to the moderate threshold and compute SAF maps as above.

3.3.1 Robustness to within-slide artefacts

In our IHC slides, the strongest within-slide variation was observed to be along the long axis of the slide, the axis of staining gradients and stitching artefacts (see examples in Figure 4). Consequently, to test the pipeline's robustness to within-slide artefacts, we averaged the high resolution SAF map over its height (y-axis) to get a horizontal SAF plot (Figure 5, box). We then delineated the impact of different within-slide variation sources by splitting the horizontal SAF plot into three via a digital Butterworth filter [37]. The highpass filter isolated sharp changes to the SAF, such as those introduced

by the stitching artefact. The bandpass filter isolated variations introduced by the stitching artefact only. The lowpass filter captured gradual variations introduced by staining gradients.

We computed the relative difference in standard deviation:

$$\%\Delta = \frac{std_{manual} - std_{automated}}{std_{manual}} \tag{9}$$

where a percentage difference (%Δ) of >0 indicates an improved robustness to artefacts in the automatically-derived SAF map relative to its manually-derived counterpart. For each IHC slide, we computed this relative measure for the highpass, bandpass and lowpass plots.

3.3.2 Reproducibility of SAF maps

We evaluated the pipeline's ability to produce reproducible SAF values at MRI-scale resolution (0.5 x 0.5 mm^2/pixel). To enable quantitative evaluation in reproducibility, we co-registered all within-subject IHC slides with FSL-TIRL [31], with the first slice of each subject chosen to be the reference. Absolute difference maps for each pair of co-registered within-subject slides were computed, and normalised pixel-wise by the mean SAF values of both slides. The median of the difference map was calculated for each pairwise comparison of within-subject slides.

3.4 Co-registration of IHC and MRI

After evaluation, the pipeline was applied to data from the wider "multi-region" dataset (c.f. Section 2.2) which includes multi-stain IHC and multi-contrast MRI that was previously co-registered using FSL-TIRL (c.f. Section 2.3) [31]. The quality of co-registration was evaluated according to the alignment of tissue boundaries and white/grey matter (WM/GM) contrast. For each slide, we obtained a WM mask by warping the MRI WM mask (generated using FSL-FAST [39]) to histology space and multiplying by a rough histology WM mask. The latter was calculated using an SAF threshold of 0.5 and a series of morphological operations (closing, opening, eroding). The tissue boundary mask was generated from RGB PLP slides downsampled to the dimensions of the SAF maps. The red channel was thresholded using the median intensity value offset by 25 (i.e. median + 25) to produce a tissue mask. The contours of the tissue masks were then overlaid on both the high resolution SAF maps (16 x 16 μm^2 patches), SAF maps matching MRI data resolution and resampled MR parameter maps.

3.5 Voxelwise correlation and regression analysis

We performed correlations across multiple brain regions (2 x VC, 1 x AC, 1 x HC) between MR parameters (FA, MD, R2*, R1) and IHC SAF maps from one control brain (CTL 1) and one MND brain (MND 1) taken from the multiple-region dataset. This serves as a proof-of-concept of how our IHC processing pipeline can be used to investigate relationships between multiple MR parameters and multiple proteins at a voxelwise level.

3.5.1 Outlier detection and removal

Prior to MRI-IHC comparison, non-tissue pixels were first removed from SAF maps by discarding pixels lower than the 5th percentile. Outliers were then defined using an approach that accounts for the covariance of the data. Each datapoint was given a weight (range 0-1) after optimising a Huber

objective function (tuning constant=2.5) with a iteratively reweighted least squares approach [40]. Datapoints further from the mean had lower weights, and points with weight values less than 0.75 were excluded.

3.5.2 Statistical analysis

For MRI-IHC analyses, the MR parameter maps were resampled into PLP space and the histology SAF was calculated for each MR voxel. We then pooled voxelwise data across regions and brains and assumed that our chosen MR parameters (FA, MD, R1, R2*) could be represented as a linear combination of our DAB-stained proteins. First, univariate linear regression was used to determine the correlation between each pairing of MR parameter vs. IHC SAF. Second, we used partial correlation to disentangle the unique variance of each MR parameter that is explained by a given IHC stain. Each IHC stain was used separately as an individual covariate. We also covaried for subject (i.e . CTL 1: 0, MND 1: 1) to account for between-subject confounds, such as post-mortem interval, fixative and subject age. We then incorporated all stains and subject covariates into a single analysis to produce the final partial correlation coefficient.

Last, we performed multiple regression using all proteins to derive a predictive model for each MR parameter, which could be driven by multiple microstructural features. The subject parameter was again included. Here, we computed the relative importance measure [41] of each stain i.e. the averaged relative contribution of each predictor (stain) in explaining the variance of the outcome variable (MR parameter) after the predictor is added to the model.

3.6 Application to multiple subjects for MRI-PLP

One of the principal benefits of this workflow (Sections 3.1-3.3) is that it can be applied to larger datasets, with multiple subjects and brain regions, with minimal manual intervention. To demonstrate this, we applied the workflow (Sections 3.1-3.3) to an additional 41 brain samples (14 x VC, 1 x AC, 17 x HC) taken from 10 separate subjects (CTL 2 & 3; MND 2-9) in the multiple-region dataset. For these nine subjects, co-registration between additional IHC stains (SMI312, Iba1, CD68) and MR data is still a work-in-progress. Consequently, we correlated MRI with PLP and compared the output from 10 subjects with those from the 2 subjects previously analysed (CTL 1 & MND 1). Again, the different subjects were modelled as a fixed effect to account for between-subject confounds, such as post-mortem intervals.

4 Results

4.1 Quantitative Evaluation of Pipeline

To evaluate the overall performance of our automated pipeline, we analysed data from the “evaluation dataset” (Section 2.1), which includes IHC data acquired from adjacent tissue sections. The PLP slides were processed using the automated pipeline with the artefact configuration (Figure 3), whilst SMI312, CD68 and Iba1 were processed using the default pipeline (Figure 2). Figure 5 compares SAF maps from both our automated pipeline and a more typical “manual” processing pipeline. We see representative examples of how the stitching artefact (blue arrows) and/or a staining gradient artefact (green arrows) can affect SAF maps when using the manual processing method (first row), and how these artefacts are reduced with the proposed “automated” pipeline. Notably, both configurations of

the automated pipeline are shown to reduce the impact of stitching artefacts. In the pipeline's default configuration, this can be attributed to the use of a slide-specific colour matrix, which results in better stain separation. This is compared to the manual method that uses a single colour matrix for the whole dataset. The automated pipeline's "artefact" configuration was designed to also reduce staining gradient artefacts, as observed. In slides with less obvious artefacts (SMI312 and Iba1), the manual and automated pipelines produce similar results.

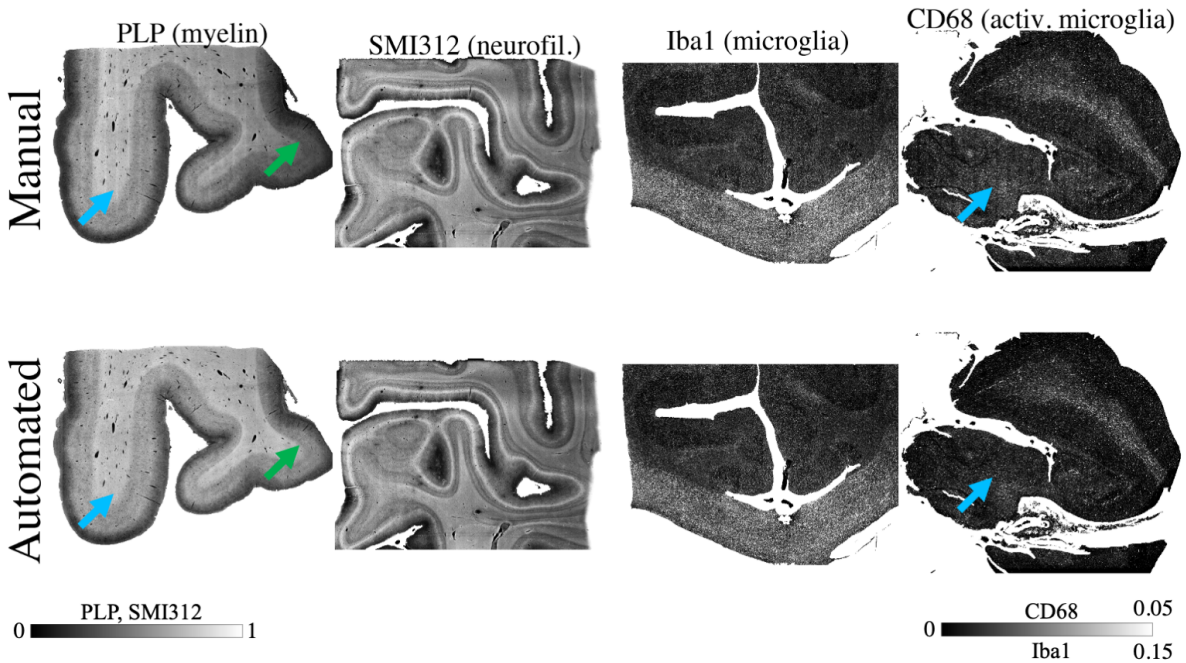

Figure 5: How within-slide artefacts can be reduced using the automated pipeline. Manually-derived SAF maps (top row) show some noticeable stitching artefacts (cyan arrows) and staining gradients (green arrows), especially in PLP IHC slides. These artefacts are reduced in the proposed automatic processing pipeline (bottom row). For cases where the artefacts are less visible (SMI312 and Iba1 in the example here), the manual and automated pipelines produce similar results. Examples here show PLP slides processed with the pipeline configuration that corrects for staining gradient artefacts ("artefacts" configuration), while SMI312, Iba1 and CD68 slides were processed with the pipeline's "default" configuration.

To quantify the relative reduction in artefacts, we then collapse the SAF along the vertical axis (take the mean) and calculate the standard deviation of the horizontal SAF. The relative percent change in standard deviation (Equation 9) for each slide is plotted in Figure 6. A value >0 suggests that the automated method produces SAF maps with less variation relative to the manually-derived SAF maps, which are thus less likely to have been impacted by artefacts. The SAF is filtered using lowpass, bandpass and highpass filters to isolate the effects of different artefacts (staining gradients, striping and high frequency noise respectively). For all filters, we observed the relative change in standard deviations to be >0 for almost all IHC slides in both PLP and SMI312 (Figure 6), suggesting a reduction in image artefacts for the automated pipeline. The results for CD68 are less conclusive. The automated pipeline typically reduces the impact of the stitching artefact (bandpass typically > 0), but performs slightly worse than the manual pipeline with respect to the highpass and lowpass filters.

We then investigated the reproducibility of SAF maps from the manual and automated pipelines (Figure 7) i.e.the consistency of SAF values across adjacent slides where true biological variation is

low. Reproducibility was quantified by taking the median of the SAF difference map between slides. Each difference map was calculated from the pairwise comparison of SAF maps for all within-subject slides. Both methods were found to have similar reproducibility with a median percentage change of around 5%, 9% and 20% in PLP, SMI312 and CD68, respectively. The variance was also similar between methods. Note that fewer slides were used when evaluating CD68 (9/15 subjects; 37/73 total slides) than in SMI312 (13/15 subjects; 55/72 total slides) and PLP (15/15 subjects; 73/73 total slides). The CD68 slides in the evaluation dataset were often observed to have inconsistent staining and/or microstructure and were excluded during manual quality control.

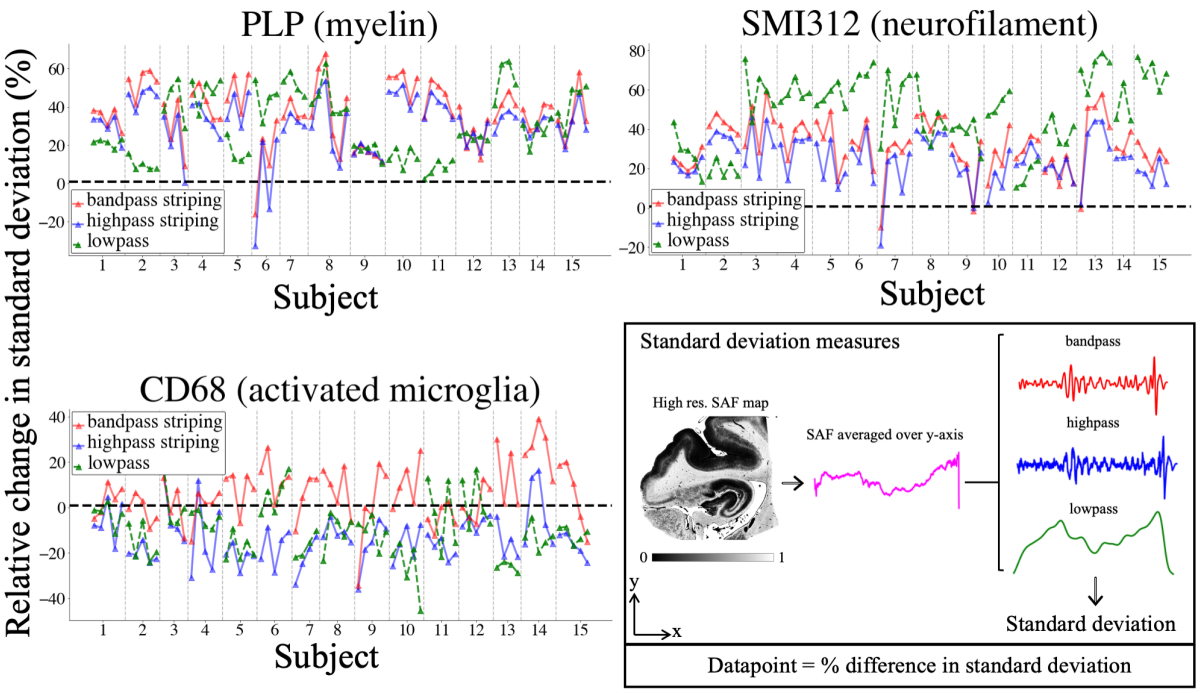

Figure 6: Quantitatively comparing the impact of within-slide artefacts in the manual and automated processing pipelines, as measured using the relative change in standard deviation ((std_{manual} - $std_{automatic}$) / std_{manual}). Standard deviation measures (box) are derived by 1) averaging the high-resolution (16 x 16 μm^2/pixel) SAF map along the y-axis, 2) filtering it to produce 3 components (bandpass, highpass and lowpass), and 3) computing the components' standard deviation. The lowpass filter isolates the staining gradient artefact, whilst the stitching artefact and other high frequency noise are captured with the bandpass and highpass filters respectively. Each IHC slide is represented as a single datapoint, and slides from the same subject are connected by lines. A value >0 indicates that the automatically-derived SAF map is less affected by the associated artefact.

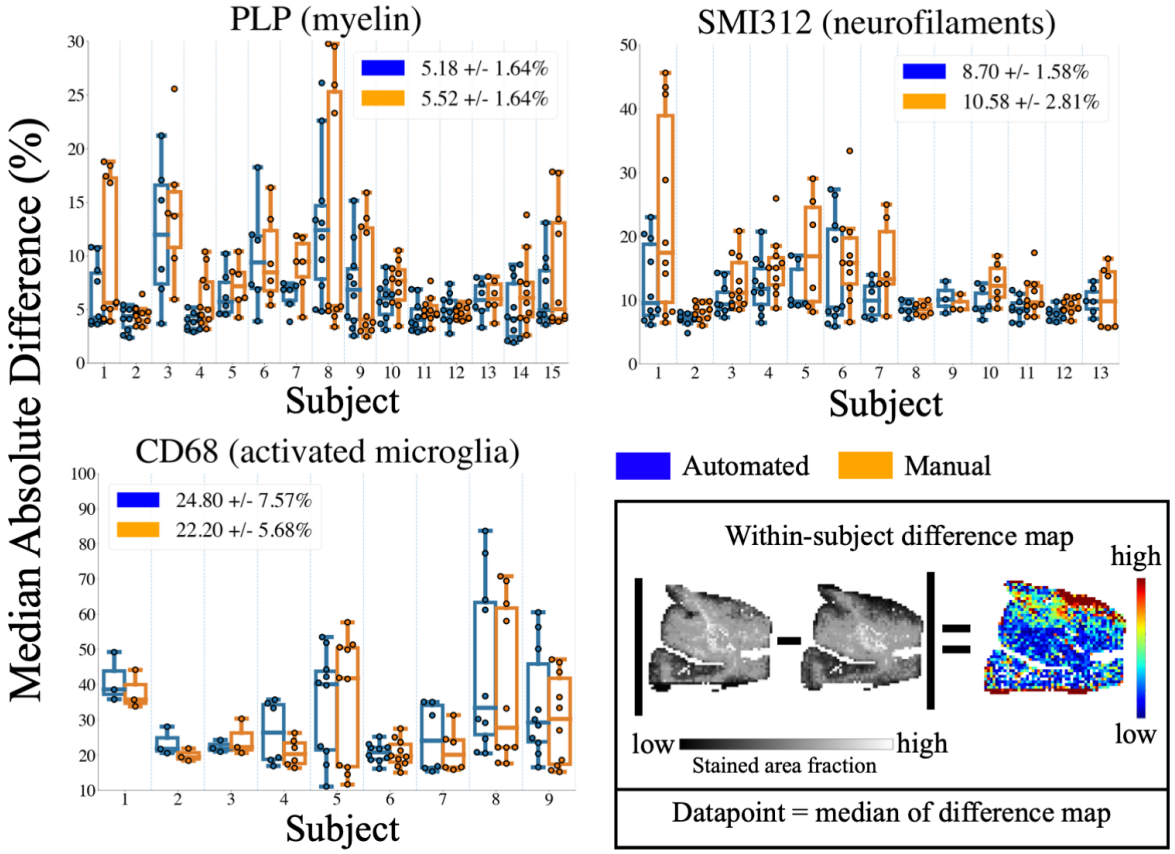

Figure 7: The reproducibility of SAF values across adjacent tissue sections, as measured by the median absolute difference. The median absolute difference (box) was computed by 1) taking the absolute difference of a pair of SAF maps produced from IHC slides extracted from the same region and subject, 2) normalising this difference map by the mean of both SAF maps, and 3) taking the median to represent one datapoint. We used this measure to compare reproducibility of both manually- (orange) and automatically-derived (blue) SAF maps for PLP, SMI312 and CD68. The median and median absolute variance of data points are also shown for each method (legends). Prior to reproducibility analysis, all slides were quality-checked. Slides deemed unsuitable were excluded. Note the change in scale-bar for CD68, where the difference values were generally larger when compared to PLP and SMI312.

4.2 Qualitative Evaluation of Co-registration

The automated pipeline was found to notably reduce artefacts and produce SAF maps of similar reproducibility to those manually derived by an expert. As such, we applied our automated pipeline on the multiple-region dataset. The multiple-region dataset contains co-registered IHC and post-mortem MRI data acquired for the same brain regions, facilitating voxelwise MRI-IHC analysis.

Our ability to perform meaningful voxelwise MRI-IHC analysis relies heavily on the quality of the data co-registration. Figures 8 and 9 evaluate the quality of the co-registration for CTL 1 and MND 1 respectively. The figures show visually good alignment of both the edges of the tissue mask (green contours) and the WM/GM interface (red contour). Similar results were seen in the co-registration of MR parameter and PLP SAF maps of all other subjects in the wider dataset (CTL 2-3; MND 2-9, Figure S2 **(Supplementary Materials)**.

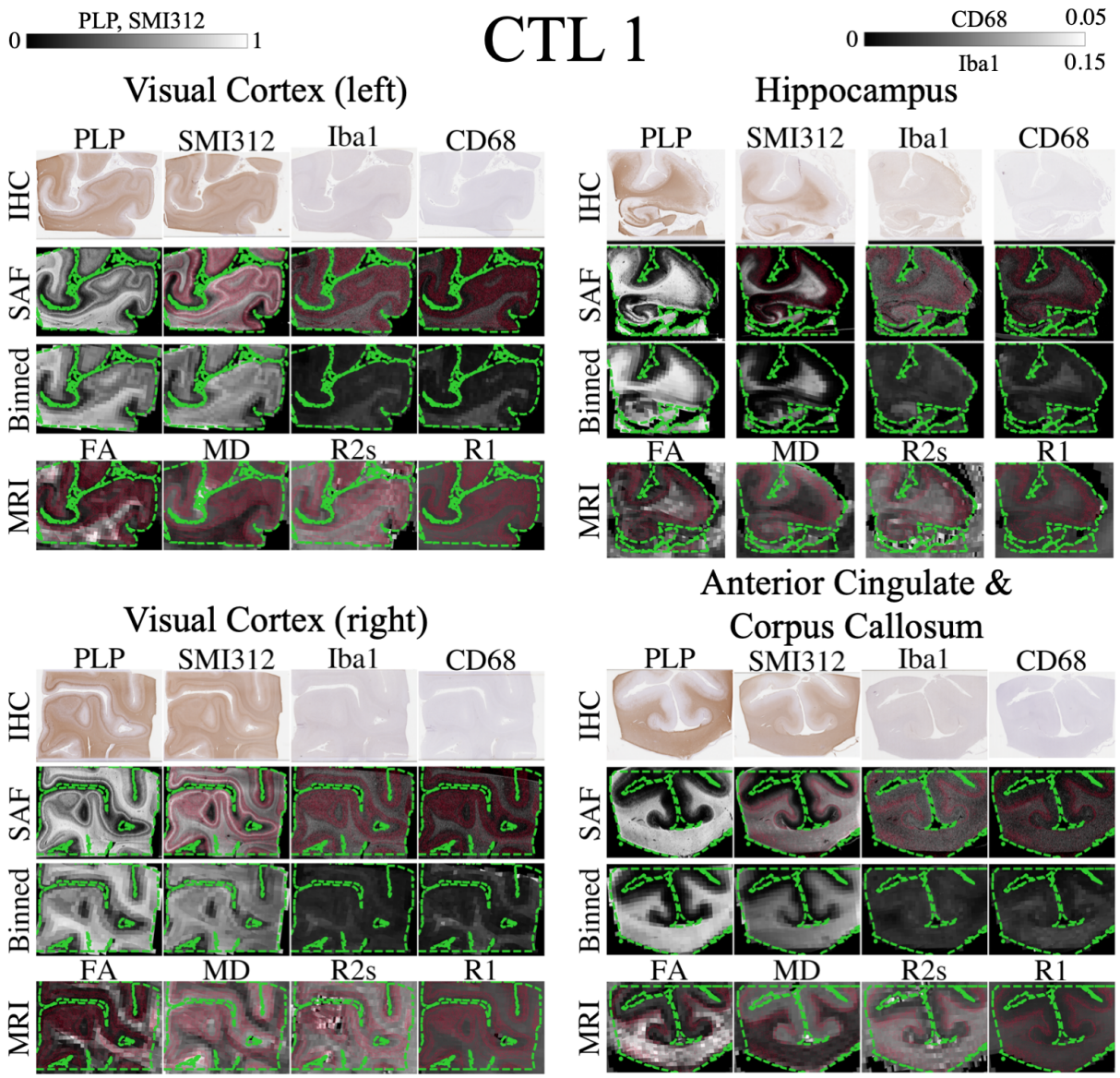

Figure 8: Registration evaluation for all brain regions in CTL 1. In each brain region, contours of the tissue mask (green dashed) are overlaid on the co-registered SAF maps at high resolution (first row, 1 pixel here represents SAF calculated in a 16 x 16 μm^2 patch), SAF maps matching MRI resolution (second row) and MR parameter maps (third row). The white and grey matter (WM/GM) interface is shown in red. The tissue boundaries are closely aligned and the high registration accuracy enables us to perform meaningful voxelwise MR-histology correlations.

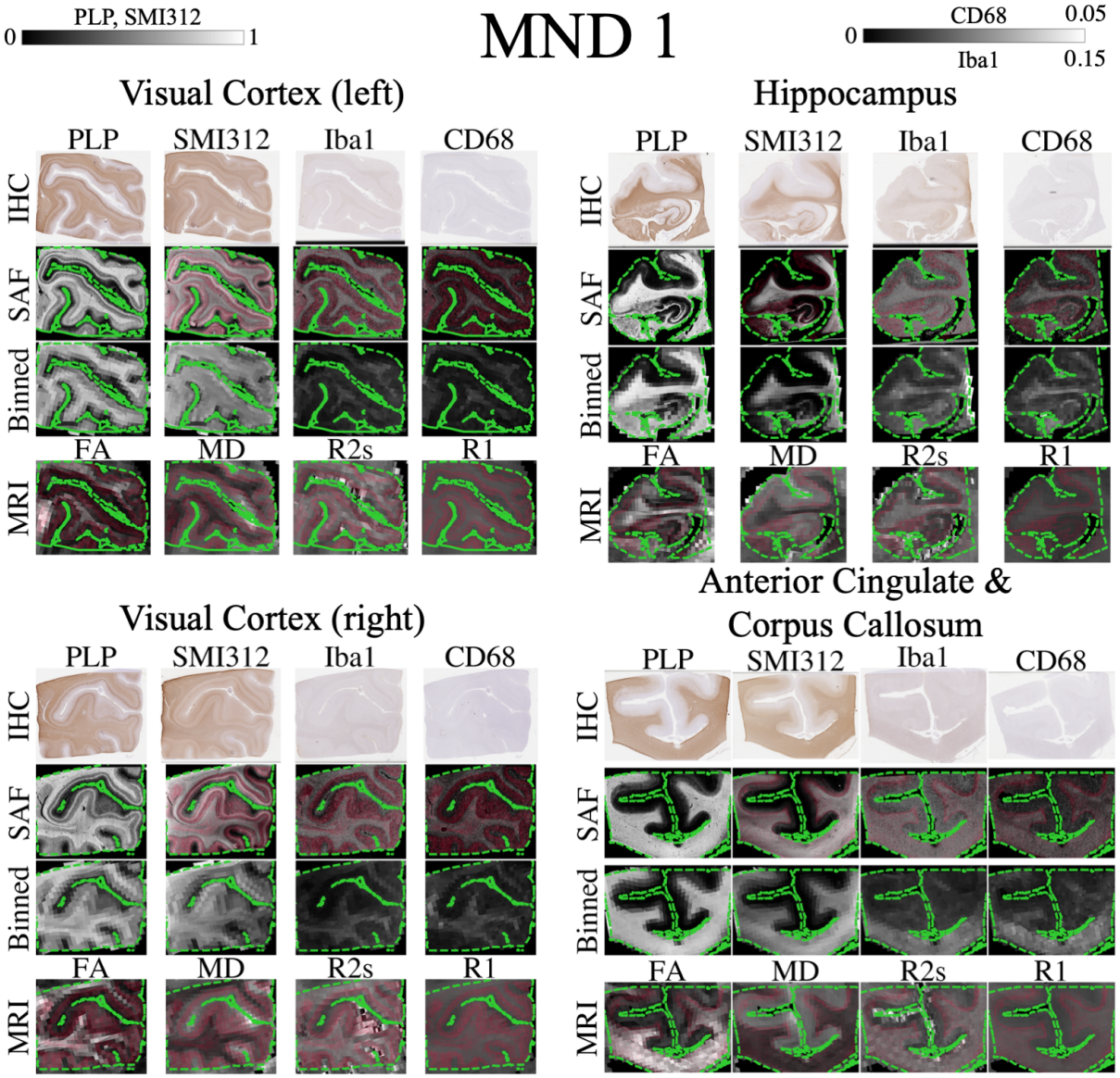

Figure 9: Registration evaluation for all brain regions in MND 1. Results are displayed as previously described in Figure 8. Again, we see close alignment of tissue boundaries.

4.3 Correlation and regression Analysis

Voxelwise correlation and regression analysis was performed using data from all slides in the multiple-stain study with 1 CTL and 1 MND subject. A total of 8 IHC sides were mapped onto MRI data. For each MR parameter, up to 3% of voxels were classified at outliers and removed, resulting in an average sample size of 6506 voxels (FA=5846, MD=5852, R2*=8956, R1=5373). Of these voxels, an average of 2246 voxels were classified as WM (FA=2039, MD=2029, R2*=3041, R1=1877).

4.3.1 Univariate linear regression of MRI and IHC

Univariate linear regression was used to relate each pairing of MR parameter vs. IHC SAF. For this analysis, voxels were pooled across brain regions and subjects. Figure 9 shows scatter plots relating MR parameters and SAF from multiple stains, using voxels from both WM and GM. Univariate linear

regression indicates high correlations between MR parameters (FA, MD, R2*, R1) and both PLP (myelin) and SMI312 (neurofilaments) (|r|=0.35-0.66). While correlations with Iba1 appear relatively low (|r|=0.035-0.280), we note some strikingly high correlations between CD68 (activated microglia) and FA (r=0.56) or MD (r=-0.39). Finally, we observe apparent nonlinear trends on some plots, particularly between FA and PLP or SMI312. This is likely due to FA being only indirectly related to myelin or neurofilaments, where various other tissue features such as dispersion (which is not accounted for in the current study) would also affect the FA values. A similar scatterplot using voxels from WM only is shown in **Supplementary Materials** (Figure S1).

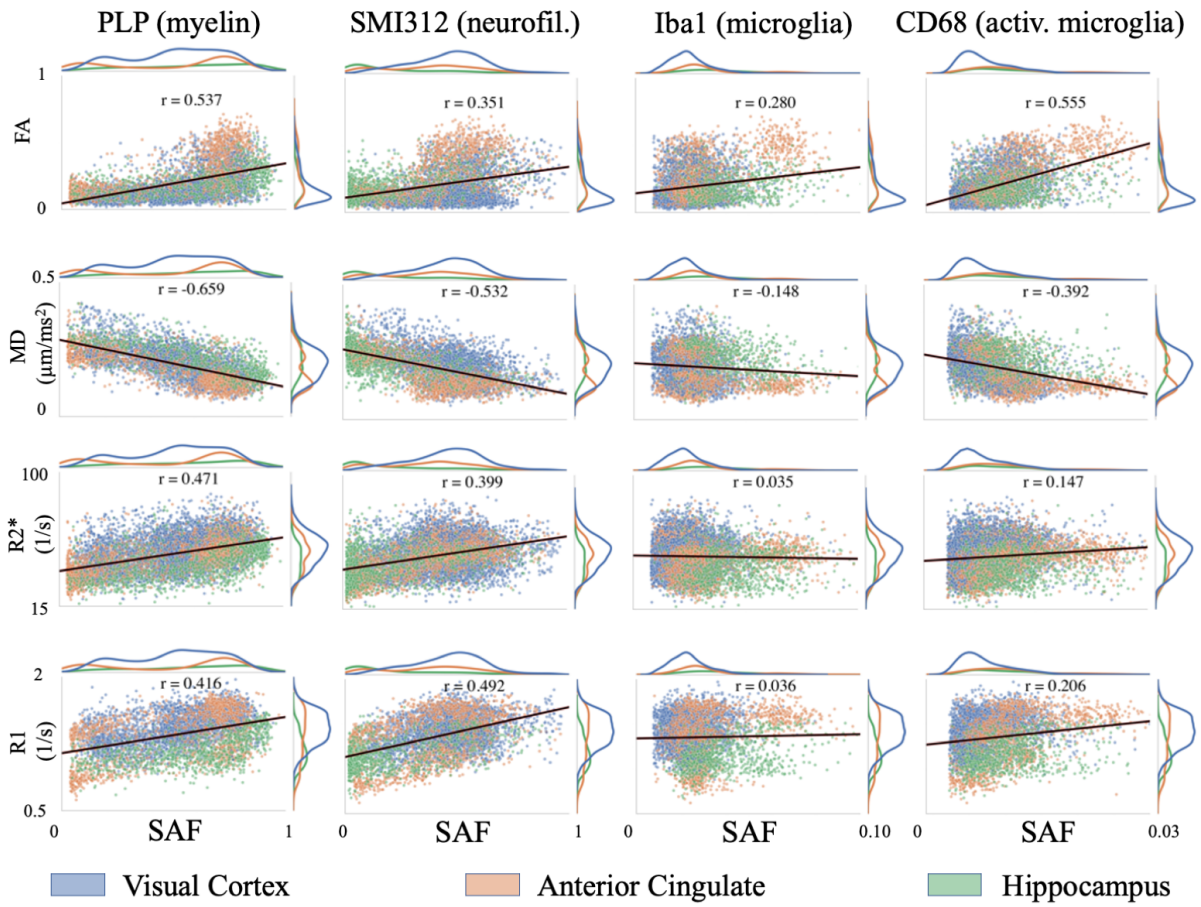

Figure 10: Correlating MR parameters (DTI FA, MD, R2* and R1) with IHC SAF. The line of best fit (black line) and corresponding Pearson correlation coefficients, r, are overlaid. The visual cortex (blue), anterior cingulate (orange) and hippocampus (green) provide good dynamic range for the MR parameters and SAFs.

4.3.2 Partial correlations of MRI with IHC

Partial correlations were then used to delineate the unique variance of each MR parameter that is explained by an IHC stain. Note that subject was included as a covariate in all partial correlation calculations to account for between-subject confounds. Figure 11 shows the correlation coefficients when including voxels from both WM and GM, and Figure 12 shows the coefficients for WM only. The top row (Figure 11, "None") indicates the correlation coefficients between MR parameters and

IHC SAF, where we included the subject as a covariate. The partial correlation coefficient was then calculated whilst controlling for one of the other stains (second to fifth rows, confounding variable indicated in the centre). When accounting for all covariates simultaneously ("All"), we get the unique variance of an MR parameter explained by that specific IHC SAF.

4.3.3.1 WM & GM

Figure 11 "All" reveals which stain explains the most unique variance in each MR parameter. When accounting for all other stains, FA is best explained by CD68 (r=0.38), MD by PLP (r=-0.42), R2* by PLP (r=0.30) and R1 by SMI312 (r=0.36).

When looking at the partial correlation coefficients (second to fifth row), we observe interesting interplay between PLP (myelin) and SMI312 (neurofilaments). Notably, the correlation coefficient between FA and SMI312 is greatly reduced after accounting for PLP (r=0.35 to 0.026), whilst the correlation coefficient between FA and PLP largely remains after accounting for SMI312 (r=0.54 to 0.44). Similar effects are observed between MD and PLP/SMI312, and between R2* and PLP/SMI312. Together, this suggests that there is unique variance in FA, MD and R2* explained by myelin in particular, but that associations with neurofilaments may primarily reflect a spatial covariance with myelin rather than a direct relationship. The opposite effect is seen in R1. The correlation coefficient between R1 and PLP decreases significantly when covarying for PLP (r=0.46 to 0.19), but the correlation coefficient between R1 and SMI312 is not as reduced when accounting for PLP (r=0.54 to 0.36). This implies that R1 is sensitive to neurites in general, rather than myelin itself.

Figure 11 also shows interplay between Iba1 (all microglia) and CD68 (activated microglia), especially with diffusion-MR parameters. The high correlation of FA with CD68 only marginally decreases after accounting for Iba1 (r=0.56 to 0.51). Conversely, the apparent correlation with Iba1 vanishes after accounting for CD68 (r=0.28 to 0.061). We note similar trends when considering correlations of MD with CD68/Iba1.

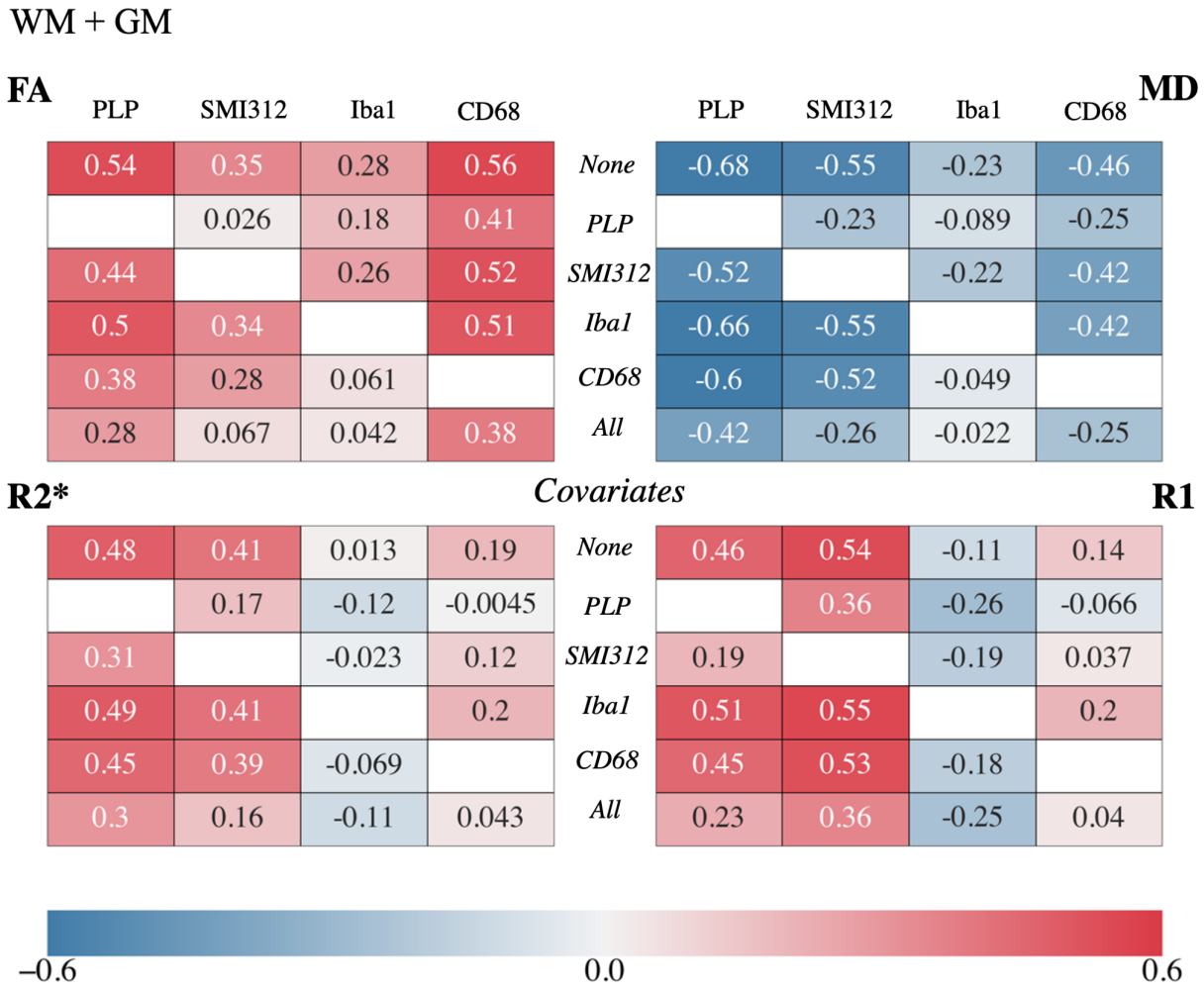

Figure 11: Partial correlation analysis between MR parameters (FA, MD, R2* and R1) and IHC stains. In each quadrant, the top row ("None") gives the correlation coefficients when accounting for a subject confound only. The partial correlation coefficient was then calculated whilst controlling for one of the other stains (italicised labels in the second to fifth rows). The last row ("All") describes all covariates being partialled out. *PLP (myelin); SMI312 (neurofilaments); Iba1 (microglia); CD68 (activated microglia).*

4.3.3.2 WM only

Figure 12 shows the same analysis for voxels found in the WM only. When accounting for all other stains, FA was best explained by CD68 (r=0.26), MD by CD68 (r=-0.18), R2* by Iba1 (r=-0.11) and R1 by Iba1 (r=-0.23). When compared to our previous results for WM and GM voxels, we see consistent results for FA but different for all other MR parameters. The correlation coefficient between the four MR parameters and PLP/SMI312 are greatly reduced and are often close to zero.This suggests that the correlations of MR parameters with PLP and SMI312 being primarily driven by WM and GM contrast. The same is not seen for CD68 and Iba1. When considering WM voxels only, CD68 explains the most unique variance with FA. Notably, this correlation remains relatively unchanged when accounting for other stains (r= 0.33 to 0.26). Similar behaviour is observed when relating CD68 with MD and Iba1 with relaxometry (R1/R2*).

WM only

FA	PLP	SMI312	Iba1	CD68		MD	PLP	SMI312	Iba1	CD68
	-0.032	0.11	0.2	0.33	*None*		0.098	-0.086	-0.11	-0.22
		0.12	0.21	0.33	*PLP*			-0.11	-0.14	-0.23
	-0.057		0.2	0.32	*SMI312*		0.12		-0.11	-0.21
	-0.08	0.11		0.27	*Iba1*		0.13	-0.085		-0.19
	-0.074	0.091	0.011		*CD68*		0.13	-0.072	0.014	
	-0.1	0.11	0.035	0.26	*All*		0.15	-0.1	-0.019	-0.18

R2*					*Covariates*	R1			
	-0.056	-0.023	-0.13	-0.063	*None*		-0.15	0.036	-0.25	-0.05
		-0.01	-0.12	-0.058	*PLP*			0.077	-0.22	-0.032
	-0.052		-0.13	-0.062	*SMI312*		-0.16		-0.25	-0.051
	-0.032	-0.021		0.013	*Iba1*		-0.1	0.039		0.075
	-0.049	-0.019	-0.12		*CD68*		-0.15	0.039	-0.25	
	-0.028	-0.015	-0.11	0.014	*All*		-0.11	0.063	-0.23	0.075

Figure 12: Partial correlation analysis between MR parameters (FA, MD, R2* and R1) and IHC stains for voxels from the WM only. The interpretation is similar to Figure 11. Note that correlations between FA and MD with CD68 (activated microglia) remain relatively high, even after accounting for all other stains. High correlations also remain between R1 and R2* with Iba1 (microglia).

4.3.3 Multiple linear regression of MRI with IHC

Multiple linear regression was performed to derive a predictive model for each MR parameter that is driven by multiple stains. Each MR parameter was modelled as a linear combination of all stains (explanatory variables) and the subject predictor (confounding variable). Figure 13A (WM & GM) shows the regression coefficients and the total fitted correlation coefficients (i.e. square root of the coefficient of determination). We note how using multiple stains together better explains the variance in MR parameters. This implies that there exists a complex relationship between microstructure and MRI. Figure 13B shows the relative importance of each predictor, and Figure 14B shows similar analysis for white matter only. When comparing Figures 13B and 14B with Figures 11 and 12, we observe that for each MR parameter, the stains with the highest relative importance are the same as the stains with highest partial correlations. Finally, the "subject" predictor was found to have high relative importance in predicting R1 in all voxels, and MD/R2*/R1 in WM voxels only. This implies that between-subject confounds substantially influence MR parameters.

WM + GM

(A) Multiple regression offset and slopes

	Offset	PLP	SMI312	Iba1	CD68	Subject	r_{fit}
FA	-0.00142*	0.173	0.0429	0.252	10.5	-0.0112	0.645
MD [$\mu m^2/ms$]	0.286	-0.112	-0.070	-0.0629*	-2.55	0.0366	0.745
R2* [1/s]	39.2	16.5	9.04	-61.2	80.7	-3.73	0.536
R1 [1/s]	1.04	0.220	0.382	-2.48	1.34	0.234	0.710

(B) Relative Importance [%]

	PLP	SMI312	Iba1	CD68	Subject
FA	36.0	11.9	7.04	44.5	0.55
MD	45.0	25.7	1.69	14.5	13.1
R2*	51.3	30.1	2.90	3.86	11.8
R1	18.7	30.6	3.15	2.93	44.6

Figure 13: Multiple linear regression predicting MR parameters using multiple IHC stains. A "subject" variable was also included to account for confounds such as post-mortem interval and age effects. As the predictors are unitless, all offsets and regression slopes are given in units of MR parameters. A: The regression coefficients and correlation coefficients r_{fit}. B: The relative importance of each stain describes the amount of variance it can explain in an MR parameter, averaged over all permutations of multiple regressions that include the specific stain. Values are normalised across stains to get a unit percentage.

WM

(A) Multiple regression offset and slopes

	Offset	PLP	SMI312	Iba1	CD68	Subject	r_{fit}
FA	0.255	-0.173	0.144	0.410*	8.81	-0.0247	0.353
MD [μm^2/ms]	0.118	0.0610	-0.0323	-0.0531*	-1.40	0.0310	0.429
R2* [1/s]	61.2	-2.22*	-0.889*	-55.2	19.6*	-5.54	0.455
R1 [1/s]	1.52	-0.217	0.0878	-2.71	1.89	0.268	0.637

(B) Relative Importance [%]

	PLP	SMI312	Iba1	CD68	Subject
FA	4.32	8.75	15.1	69.2	2.66
MD	6.43	2.40	4.99	16.2	70.0
R2*	0.47	1.97	22.6	3.62	71.3
R1	4.54	2.39	5.88	0.443	86.8

Figure 14: Multiple linear regression now computed from voxels located in the tissue's WM. These tables are interpreted similar to Figure 13.

4.4 Application to multiple subjects for MR-PLP

The primary benefit of the automated pipeline is that it can be rapidly applied across many subjects with multiple stains. We demonstrate this by applying this pipeline to a larger dataset of 10 subjects to perform a similar univariate linear regression analysis between MRI and PLP only. For this analysis of 2 healthy controls (CTL 2 & 3) and 8 MND patients (MND 2-9), we used 41 PLP IHC slides. Following outlier removal, we had an average of 31,047 voxels per MR parameter (FA=28,849, MD=28,598, R2*=44,134, R1=22,608). In WM, we had an average of 10,724 voxels per MR parameter (FA=10,040, MD=9954, R2*=14,832, R1=8072).

Figure 15 compares the analysis result of MRI-IHC correlations across many subjects (CTL 2,3 & MND 2-9; right column) to our previous result with only 2 subjects (CTL 1 & MND 1; left column). Similar trends and regression coefficients were seen for all parameters, with the results for FA and MD being particularly consistent.

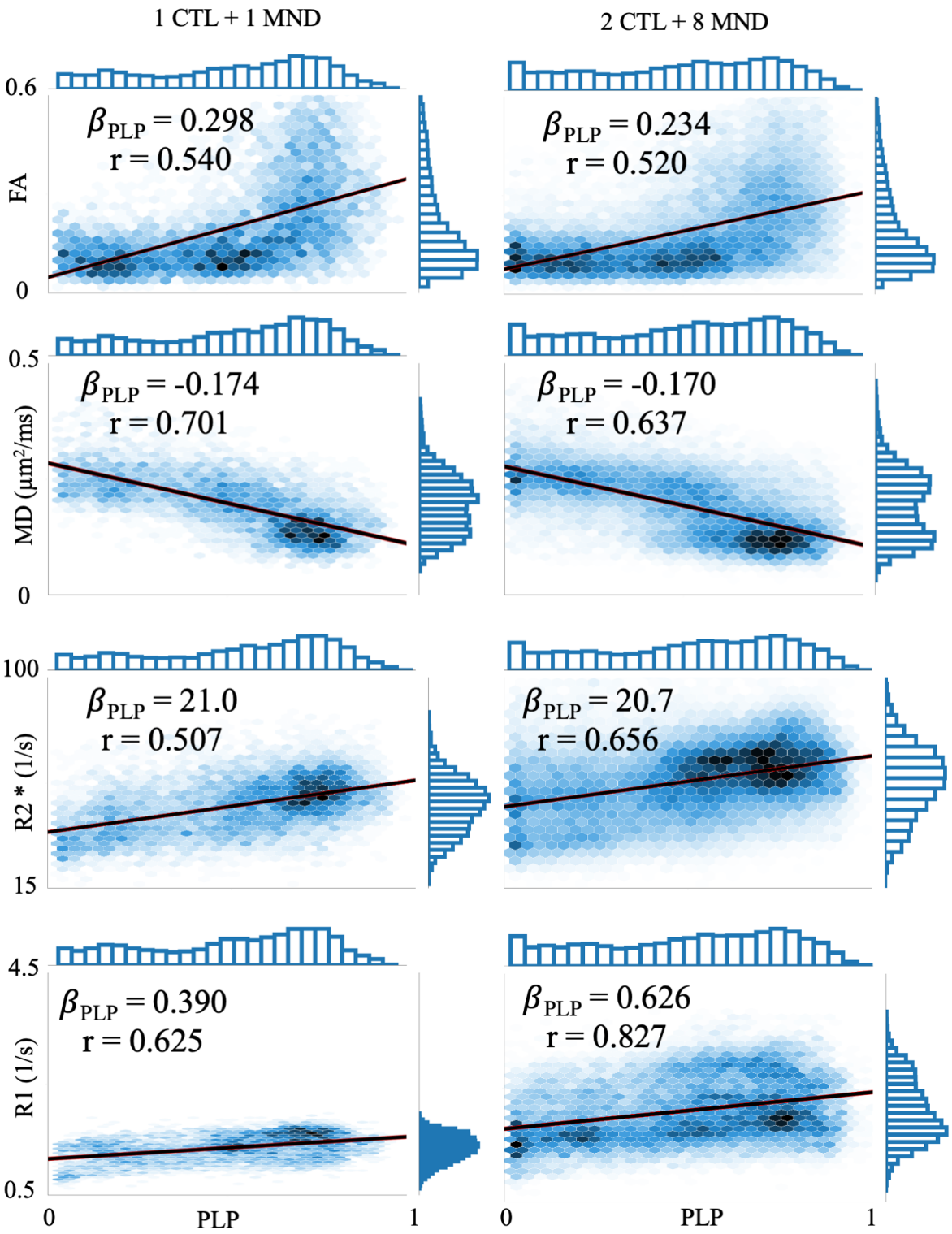

Figure 15: Density plots showing the relationship of different MR parameters (rows) with PLP (myelin). We compare the MR-PLP regression coefficient (β_{PLP}) derived from analysis across many subjects (CTL 2,3 & MND 2-9; right column) with analysis across 2 subjects (CTL 1 & MND 1; left column). We do this to understand if our automated workflow is applicable to larger studies. Note that we modelled the different subjects as a fixed effect to account for between-subject confounds, such as post-mortem intervals. Evenso, we see consistent β_{PLP} in FA, MD and R2*.

5 Discussion

To facilitate voxelwise analyses comparing MRI with histology, we introduce an automated pipeline to extract quantitative histology metrics (stain area fraction) from multiple immunohistochemical (IHC) stains. Using high-quality co-registration, where 2D histology slides have been mapped to 3D MRI in the same tissue, we then perform whole-slide voxelwise comparisons between multimodal MRI- and our IHC-derived parameters. Here, the pipeline was applied to post-mortem human brain data from multiple subjects, comparing IHC slides stained for myelin (PLP), neurofilaments (SMI312), microglia (Iba1) and activated microglia (CD68) to MR parameters of FA, MD R2* and R1.

When compared to previous literature, our approach has two key advantages. First and foremost, we develop an automated, data-driven IHC processing pipeline that can be applied to multiple IHC stains and which explicitly aims to minimise the impact of artefacts related to tissue staining or slide digitisation. While most MRI-IHC studies use manual IHC processing methods, some more automated approaches have been employed. This includes K-means clustering [42], Otsu's method [43], and hysteresis thresholding [44] for segmentation of different microstructural sources [13]. However, none of these methods explicitly address IHC artefacts which we found could bias results. Our pipeline shows a more rigorous treatment of processing IHC which aims to minimise key non-biological sources of variation. Further, the automated nature of our IHC pipeline and its generalisability to multiple IHC stains, allows for rapid scaling up of this analysis to larger research studies. Combined with recent advances in MRI-microscopy co-registration [31], the pipeline can help pave the way for a standardised IHC processing protocol for future MRI-IHC studies [14].

Second, we perform voxelwise MRI-IHC analysis, comparing multiple IHC stains with MR parameters from the same tissue sample. Previous studies are often restricted to ROI-based, rather than voxelwise, analysis [10, 18, 23, 42]. A voxelwise analysis uses data from all tissue voxels in the brain sample, thereby eliminating bias that may be introduced based on which ROI is selected. Hence, this inspires more confidence in the derived MR-histology correlations. Further, most studies correlate MRI with only a single, or a few histological markers [10, 18, 23, 42, 45]. In this work, we relate parameters derived from multimodal MRI with multiple IHC stains in the human brain. This facilitates more complex analyses, such as the partial correlation and multiple regression, which help determine how each microstructural source might either directly or indirectly contribute to MRI signal changes while accounting for other co-localised microstructure.

5.1 IHC Processing Pipeline

We evaluated the proposed IHC pipeline to determine 1) its robustness to IHC artefacts, and 2) its ability to produce reproducible SAF values. Here, we analysed a separate "evaluation dataset" which included 4-5 adjacent slides (per stain) sampled from each subject. We use this dataset to separate sources of biological variability, which is assumed to be minimised in adjacent slides, from SAF variations associated with tissue processing or IHC analysis.

Our results demonstrate that the proposed pipeline shows robustness to common IHC artefacts (Figures 5 and 6), while maintaining similar reproducibility to manually-derived SAF values (Figure 7). These artefacts were particularly prominent in more densely-stained slides such as PLP (myelin),

which almost always had some form of observable within-slide artefact. The artefacts were reduced in our proposed IHC processing pipeline, as seen in the positive relative change in the standard deviation of highpass (i.e. probing all sharp intensity changes), bandpass (i.e. stitching artefacts) and lowpass (i.e. staining gradient) components for most IHC slides.

In comparison, CD68 (activated microglia) appears less reproducible and there is a less clear improvement (<0 in standard deviation measures) to within-slide artefacts in most slides. However, this is not entirely unexpected for IHC slides specifying sparser and smaller structures (i.e. microglia soma sizes are approximately 10 μm [46]), where we expect a higher likelihood of biological variability between adjacent slides. Consequently, this may make the SAF maps appear less reproducible. Further, the staining gradient artefact was not visibly present in CD68-stained slides, though some striping was visible on the manually-derived SAF maps. The automated pipeline was able to reduce this striping, as seen in the improved bandpass standard deviation metric. Finally, we note that many CD68 slides were excluded during manual quality control due to poor illumination, inconsistent staining and/or significant tissue tearing. These artefacts are related to non-idealities of tissue processing and image digitisation that typically affect batches of slides which are processed together. Notably, no CD68 slides were excluded during QC of the separately acquired multiple-region dataset.

Our results (Figure 7) indicate similar reproducibility between automatically- and manually-derived SAF maps. However, increased reproducibility does not necessarily imply a more accurate segmentation, where SAF values reflect the true tissue microstructure. For example, if adjacent slides are impacted by similar staining gradients, the SAF maps may be biassed in the same way and thus appear highly reproducible. By correcting for these artefacts, the SAF's reproducibility may appear reduced. Consequently, reproducibility results should be interpreted alongside how much a SAF map has been corrected for artifactual within-slide variations.

5.2 MRI-IHC analysis

The multiple-region dataset includes MRI and IHC data that has been carefully co-registered using FSL-TIRL, facilitating meaningful voxelwise analysis. Our correlation results agree with previous studies for myelin [14, 16, 18, 23] and neurofilaments [42] with FA, MD, R2* and R1. In comparison, microglia's contribution to MR is relatively poorly understood with few studies reporting MR correlations with microglia in the human brain. Our simple correlations suggest Iba1 (all microglia) correlates positively with FA and negatively with MD, similar to that reported by Grussu et. al. in the human spinal cord [42].

5.2.1 Myelin and neurofilaments contributions to MRI

Our partial correlation results for PLP (myelin) and SMI312 (neurofilaments) demonstrate pitfalls in performing simple correlations between MRI and a single histology stain without controlling for other tissue features. An example is the correlation of FA, MD and R2* with SMI312 when all tissue (WM and GM) voxels are used. We observe how the correlation coefficients are greatly diminished after accounting for PLP. Conversely, the correlations with PLP do not decrease as much after accounting for SMI312. This suggests that there is unique variance in FA, MD and R2* explained by myelinated axons in particular, and that univariate associations with neurofilaments may primarily reflect a spatial covariance with myelin rather than a direct relationship. The opposite pattern is seen for R1, suggesting that R1 is sensitive to neurites in general, rather than myelin itself. This is noteworthy,

since some studies report myelin correlating with R1 [21, 47-49] while not controlling for neurofilaments.

In white matter only, PLP and SMI312 have considerably reduced associations with all MR parameters (Figure 12). This may suggest that our PLP's and SMI312's associations with MRI are being primarily driven by white and grey matter contrast.

Together, our results indicate considerable overlap between myelin and neurofilament in explaining variation in MR, and that much of this relationship may be driven by microstructural differences in white and grey matter. This could help explain the large range of MR-myelin correlation coefficients reported in two recent reviews (Figure 4 and 5 in [14], Figure 4 in [16]).

5.2.2 MRI's sensitivity to microglia

Our work suggests several MR parameters' potential sensitivity to microglia. CD68 correlates positively with DTI's FA and negatively with MD. The correlations remain after accounting for Iba1, though the correlations of Iba1 with FA and MD vanish after accounting for CD68. This points to unique variance in both MR parameters being substantially explained by activated microglia, rather than microglia in general. Notably, these results hold in both analyses considering all voxels (WM & GM), and white matter only.

As activated microglia is a known biomarker of inflammation [50, 51], our results may also suggest DTIs sensitivity to neuroinflammation, where FA would increase, and MD decrease, with heightened neuroinflammation. Intriguingly, this is the opposite trend to that reported in previous in-vivo studies: FA was shown to be lower with increased neuroinflammation found in Parkinson's disease [52] and obesity [53]. However, these studies do not correlate FA with activated microglia, making direct comparison with our results difficult. Future work will seek to elucidate how microglia directly or indirectly affects diffusion-MR signals by using more sophisticated biophysical models [54, 55].

Another example is Iba1's correlations with relaxometry in WM (Figure 12). Iba1's correlations with R2* and R1 remain unchanged after accounting for CD68, while CD68's correlations are diminished when accounting for Iba1. This suggests that all microglia, as opposed to just activated microglia, may relate to relaxometry measures. Surprisingly, microglia correlates negatively with R1 and R2*. whilst iron is known to correlate positively with R1 and R2* and colocalize with activated microglia [56, 57]. Here, we do not see considerable correlation between CD68 and relaxometry parameters. Future analyses will incorporate iron-stained IHC to further elucidate the unique contributions of iron and glia to R1 and R2*.

5.3 Application of Workflow to Multiple Subjects for MRI-PLP

Finally, the pipeline was applied to data from a further 10 subjects to demonstrate its generalisability to larger-scale MRI-IHC studies. This analysis would be extremely time-consuming if conducted manually, highlighting the value of the automated pipeline. As expected, our results were generally consistent (2 subjects, 8 brain samples vs. 10 subjects, 41 brain samples) with R1 showing the most discrepancy. Preliminary investigations suggest the R1 variation across subjects may be influenced by differing post-mortem intervals and fixatives (not shown) which are artefacts of the MRI processing protocol, rather than the proposed workflow. Nonetheless, the similarity of MR-PLP trends and correlation coefficients gives confidence that the pipeline is consistent in deriving IHC SAF maps and MRI-IHC relationships from larger multi-subject multi-region datasets.

5.4 Limitations

There are several limitations to this study. First, our analyses do not include IHC slides stained for iron, which is known to particularly affect R2* and R1 [22, 23, 47, 59]. We are currently working to incorporate the analysis of iron-stained slides (ferritin) into our future MRI-IHC studies.

Second, we quantify the IHC stain using a stained area fraction, which may not scale linearly with protein concentration. Similarly, MR metrics may be nonlinearly or indirectly related to each aspect of the stained tissue microstructure, though we only include linear analyses. While our results may be useful for showing key trends between different MR parameters and different proteins, the MRI-IHC slopes may not be directly comparable to other studies. Proton induced X-ray emission and mass spectrometry have shown promise in deriving MR-histology slopes that can predict accurate MR parameter values given different concentrations of myelin and/or iron [47, 59]. While the former is non-destructive, the latter may damage the tissue during preparation. Regardless, fully quantitative methods are still rarely used in MRI-histology studies [14].

Third, our results may be affected by subject-specific confounds which are known to affect MR parameters but are not related to the biology of interest, such as post-mortem interval [60] and fixation time [61]. Our results suggest that these confounds may be driving significant variation in MD, R2* and R1, as reflected in the subject predictor's high relative importance (Figures 13B & 14B) and the variable MRI-IHC trends in R1 (Figure 15, 2 versus 10 subjects). This highlights the need to consider non-biological confounds or sources of variance in MRI-IHC studies. To address this, we added subject as a covariate in both our partial correlations and multiple linear regression.

Finally, to demonstrate the utility of the pipeline we analysed brain regions which are not known to display substantial disease pathology. In end-stage MND, substantial disease pathology is usually in the primary motor cortex [62]. Future work will focus on analysing brain regions with variable pathology across many subjects, to better understand microstructural signature of MND pathology.

6 Conclusion

We introduce an automated, data-driven histology processing pipeline that extracts stained area fraction maps from whole immunohistochemically (IHC) stained slides. By design, the pipeline is both generalisable to multiple IHC stains and robust to artefacts from tissue staining and/or slide digitisation. The pipeline was applied to co-registered MRI and IHC data from post-mortem human brains where we perform multiple linear analyses (simple correlation, partial correlations and multiple regression) to investigate voxelwise relationships between MRI- and IHC-derived parameters (FA, MD, R1 and R2* against SAF for myelin, neurofilaments and microglia). Our results emphasise the need to simultaneously analyse multiple stains when validating MR parameters, as to avoid misleading inference due to the spatial covariance of multiple microstructural features. Interestingly, we found several MR parameters' sensitivity to activated microglia--a key biomarker of neuroinflammation. Together, our results demonstrate the pipeline as a valuable tool for IHC analyses and future investigations relating MR signals to disease pathology.

Authors' Contributions

DZLK conceived the study design, developed the histology processing pipeline, implemented the MRI-histology workflow and conducted data analyses and drafted the manuscript. **SJ** conceived the study design, contributed to the design of the MRI-histology workflow and data analyses. **INH** developed the TIRL registration platform, performed the MRI-histology co-registration and advised on data analyses. **JM** contributed to the histology processing pipeline. **BCT** analysed the diffusion MRI data. **SF** developed the MRI protocols and acquired the MRI data. **CW** analysed the R2* data. **CS** contributed to the acquisition of the multiple-region dataset. **AS** performed the histological processing, curated histological data, and provided expert feedback on histology analysis. **OA** designed the brain sampling strategy and provided expert feedback on histological analysis. **MPG** optimised histological protocols, designed the brain sampling strategy, performed histology analysis, evaluated histological segmentations and provided expert feedback on histological analysis. **KLM** conceived the study design, designed the post-mortem MRI protocol, contributed to MRI-histology workflow and all data analysis. **AFDH** conceived the study design, contributed to the MRI-histology workflow and all data analysis, and helped draft the manuscript.

Declaration of Competing Interests

None

Acknowledgements

The authors are grateful to the donors and benefactors of the Oxford Brain Bank that kindly provided all human tissues for this study. AFDH and KLM contributed equally to this work. Data acquisition was funded by the grant MR/K02213X/1 from the Medical Research Council (MRC). DZLK is supported by the Hrothgar Singaporean Clarendon Scholarship and the Nuffield Department of Clinical Neurosciences studentship. AFDH and INH are supported by the EPSRC and MRC grants (grants EP/L016052/1 and MR/L009013/1). KLM, AFDH, INH and SJ are supported by the Wellcome Trust (grants WT202788/Z/16/A and WT215573/Z/19/Z). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z).

References

[1] Johnson KA, Fox NC, Sperling RA, Klunk WE. Brain Imaging in Alzheimer Disease. *Cold Spring Harb Perspect Med*. 2012;2(4):a006213. doi:10.1101/cshperspect.a006213

[2] Geraldes R, Ciccarelli O, Barkhof F, et al. The current role of MRI in differentiating multiple sclerosis from its imaging mimics. *Nat Rev Neurol*. 2018;14(4):199-213. doi:10.1038/nrneurol.2018.14

[3] Grolez G, Moreau C, Danel-Brunaud V, et al. The value of magnetic resonance imaging as a biomarker for amyotrophic lateral sclerosis: a systematic review. *BMC Neurol*. 2016;16(1):155. doi:10.1186/s12883-016-0672-6

[4] Tendler BC, Hanayik T, Ansorge O, et al. *The Digital Brain Bank, an Open Access Platform for Post-Mortem Datasets*. Neuroscience; 2021. doi:10.1101/2021.06.21.449154

[5] Goubran M, Hammond RR, de Ribaupierre S, et al. Magnetic resonance imaging and histology correlation in the neocortex in temporal lobe epilepsy. *Ann Neurol*. 2015;77(2):237-250. doi:10.1002/ana.24318

[6] Fjær S, Bø L, Lundervold A, et al. Deep Gray Matter Demyelination Detected by Magnetization Transfer Ratio in the Cuprizone Model. *PLOS ONE*. 2013;8(12):e84162. doi:10.1371/journal.pone.0084162

[7] Fjær S, Bø L, Myhr KM, Torkildsen Ø, Wergeland S. Magnetization transfer ratio does not correlate to myelin content in the brain in the MOG-EAE mouse model. *Neurochemistry International*. 2015;83-84:28-40. doi:10.1016/j.neuint.2015.02.006

[8] Dusek P, Bahn E, Litwin T, et al. Brain iron accumulation in Wilson disease: a post mortem 7 Tesla MRI - histopathological study. *Neuropathol Appl Neurobiol*. 2017;43(6):514-532. doi:10.1111/nan.12341

[9] Pallebage-Gamarallage M, Foxley S, Menke RAL, et al. Dissecting the pathobiology of altered MRI signal in amyotrophic lateral sclerosis: A post mortem whole brain sampling strategy for the integration of ultra-high-field MRI and quantitative neuropathology. *BMC Neurosci*. 2018;19(1):11. doi:10.1186/s12868-018-0416-1

[10] Yano R, Hata J, Abe Y, et al. Quantitative temporal changes in DTI values coupled with histological properties in cuprizone-induced demyelination and remyelination. *Neurochem Int*. 2018;119:151-158. doi:10.1016/j.neuint.2017.10.004

[11] Wang C, Foxley S, Ansorge O, et al. Methods for quantitative susceptibility and R2* mapping in whole post-mortem brains at 7T applied to amyotrophic lateral sclerosis. *Neuroimage*. 2020;222:117216. doi:10.1016/j.neuroimage.2020.117216

[12] Abe Y, Komaki Y, Seki F, Shibata S, Okano H, Tanaka KF. Correlative study using structural MRI and super-resolution microscopy to detect structural alterations induced by long-term optogenetic stimulation of striatal medium spiny neurons. *Neurochemistry International*. 2019;125:163-174. doi:10.1016/j.neuint.2019.02.017

[13] Mollink J, Hiemstra M, Miller KL, et al. White matter changes in the perforant path area in patients with amyotrophic lateral sclerosis. *Neuropathol Appl Neurobiol*. 2019;45(6):570-585. doi:10.1111/nan.12555

[14] Lazari A, Lipp I. Can MRI measure myelin? Systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology. *NeuroImage*. 2021;230:117744. doi:10.1016/j.neuroimage.2021.117744

[15] van der Weijden CWJ, García DV, Borra RJH, et al. Myelin quantification with MRI: A systematic review of accuracy and reproducibility. *Neuroimage*. 2021;226:117561. doi:10.1016/j.neuroimage.2020.117561

[16] Mancini M, Karakuzu A, Cohen-Adad J, Cercignani M, Nichols TE, Stikov N. An interactive meta-analysis of MRI biomarkers of myelin. Jbabdi S, Baker CI, Jbabdi S, Does M, eds. *eLife*. 2020;9:e61523. doi:10.7554/eLife.61523

[17] De Barros A, Arribarat G, Combis J, Chaynes P, Péran P. Matching ex vivo MRI With Iron Histology: Pearls and Pitfalls. *Frontiers in Neuroanatomy*. 2019;13. Accessed January 18, 2022. https://www.frontiersin.org/article/10.3389/fnana.2019.00068

[18] Seewann A, Vrenken H, van der Valk P, et al. Diffusely Abnormal White Matter in Chronic Multiple Sclerosis: Imaging and Histopathologic Analysis. *Archives of Neurology*. 2009;66(5):601-609. doi:10.1001/archneurol.2009.57

[19] Bulk M, Kenkhuis B, van der Graaf LM, Goeman JJ, Natté R, van der Weerd L. Postmortem T2*- Weighted MRI Imaging of Cortical Iron Reflects Severity of Alzheimer's Disease. *J Alzheimers Dis*. 2018;65(4):1125-1137. doi:10.3233/JAD-180317

[20] Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*. 2001;23(4):291-299.

[21] Wiggermann V, Hametner S, Hernández-Torres E, et al. Susceptibility-sensitive MRI of multiple sclerosis lesions and the impact of normal-appearing white matter changes. *NMR Biomed*. 2017;30(8). doi:10.1002/nbm.3727

[22] Hametner S, Endmayr V, Deistung A, et al. The influence of brain iron and myelin on magnetic susceptibility and effective transverse relaxation - A biochemical and histological validation study. *Neuroimage*. 2018;179:117-133. doi:10.1016/j.neuroimage.2018.06.007

[23] Bagnato F, Hametner S, Boyd E, et al. Untangling the R2* contrast in multiple sclerosis: A combined MRI-histology study at 7.0 Tesla. *PLoS One*. 2018;13(3):e0193839. doi:10.1371/journal.pone.0193839

[24] van der Loos CM. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. *J Histochem Cytochem*. 2008;56(4):313-328. doi:10.1369/jhc.2007.950170

[25] Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological Image Analysis: A Review. *IEEE Rev Biomed Eng*. 2009;2:147-171. doi:10.1109/RBME.2009.2034865

[26] McCann MT, Ozolek JA, Castro CA, Parvin B, Kovacevic J. Automated Histology Analysis: Opportunities for signal processing. *IEEE Signal Processing Magazine*. 2015;32(1):78-87. doi:10.1109/MSP.2014.2346443

[27] Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Anal*. 2016;33:170-175. doi:10.1016/j.media.2016.06.037

[28] Farahani N, Parwani AV, Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *PLMI*. 2015;7:23-33. doi:10.2147/PLMI.S59826

[29] Tendler BC, Foxley S, Hernandez-Fernandez M, et al. Use of multi-flip angle measurements to account for transmit inhomogeneity and non-Gaussian diffusion in DW-SSFP. *NeuroImage*. 2020;220:117113. doi:10.1016/j.neuroimage.2020.117113

[30] Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J*. 1994;66(1):259-267. doi:10.1016/S0006-3495(94)80775-1

[31] Huszar IN, Pallebage-Gamarallage M, Foxley S, et al. *Tensor Image Registration Library: Automated Non-Linear Registration of Sparsely Sampled Histological Specimens to Post-Mortem MRI of the Whole Human Brain*.; 2019:849570. doi:10.1101/849570

[32] Landini G, Martinelli G, Piccinini F. Colour deconvolution: stain unmixing in histological imaging. *Bioinformatics*. 2021;37(10):1485-1487. doi:10.1093/bioinformatics/btaa847

[33] Clarke EL, Treanor D. Colour in digital pathology: a review. *Histopathology*. 2017;70(2):153-163. doi:10.1111/his.13079

[34] Geijs DJ, Intezar M, Laak JAWM van der, Litjens GJS. Automatic color unmixing of IHC stained whole slide images. In: *Medical Imaging 2018: Digital Pathology*. Vol 10581. SPIE; 2018:165-171. doi:10.1117/12.2293734

[35] Yuan X cui, Wu L shen, Peng Q. An improved Otsu method using the weighted object variance for defect detection. *Applied Surface Science*. 2015;349:472-484. doi:10.1016/j.apsusc.2015.05.033

[36] Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. 1979;9(1):62-66. doi:10.1109/TSMC.1979.4310076

[37] Butterworth, S. On the Theory of Filter Amplifiers. *Experimental Wireless and the Wireless Engineer*. 1930;**7**: 536–541.

[38] Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 2002;17(2):825-841. doi:10.1016/s1053-8119(02)91132-8

[39] Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging*. 2001;20(1):45-57. doi:10.1109/42.906424

[40] Huber PJ. Robust Statistics. In: Lovric M, ed. *International Encyclopedia of Statistical Science*. Springer; 2011:1248-1251. doi:10.1007/978-3-642-04898-2_594.

[41] Grömping U. Variable importance in regression models. *WIREs Computational Statistics*. 2015;7(2):137-152. doi:10.1002/wics.1346

[42] Grussu F, Schneider T, Tur C, et al. Neurite dispersion: a new marker of multiple sclerosis spinal cord pathology? *Ann Clin Transl Neurol*. 2017;4(9):663-679. doi:10.1002/acn3.445

[43] Khodanovich M, Pishchelko A, Glazacheva V, et al. Quantitative Imaging of White and Gray Matter Remyelination in the Cuprizone Demyelination Model Using the Macromolecular Proton Fraction. *Cells*. 2019;8(10):1204. doi:10.3390/cells8101204

[44] Praet J, Manyakov NV, Muchene L, et al. Diffusion kurtosis imaging allows the early detection and longitudinal follow-up of amyloid-β-induced pathology. *Alzheimer's Research & Therapy*. 2018;10(1):1. doi:10.1186/s13195-017-0329-8

[45] Peters JM, Struyven RR, Prohl AK, et al. White matter mean diffusivity correlates with myelination in tuberous sclerosis complex. *Ann Clin Transl Neurol*. 2019;6(7):1178-1190. doi:10.1002/acn3.793

[46] Kozlowski C, Weimer RM. An Automated Method to Quantify Microglia Morphology and Application to Monitor Activation State Longitudinally In Vivo. *PLOS ONE*. 2012;7(2):e31814. doi:10.1371/journal.pone.0031814

[47] Stüber C, Morawski M, Schäfer A, et al. Myelin and iron concentration in the human brain: a quantitative study of MRI contrast. *Neuroimage*. 2014;93 Pt 1:95-106. doi:10.1016/j.neuroimage.2014.02.026

[48] Hakkarainen H, Sierra A, Mangia S, et al. MRI relaxation in the presence of fictitious fields correlates with myelin content in normal rat brain. *Magnetic Resonance in Medicine*. 2016;75(1):161-168. doi:10.1002/mrm.25590

[49] Warntjes JBM, Persson A, Berge J, Zech W. Myelin Detection Using Rapid Quantitative MR Imaging Correlated to Macroscopically Registered Luxol Fast Blue–Stained Brain Specimens. *AJNR Am J Neuroradiol*. 2017;38(6):1096-1102. doi:10.3174/ajnr.A5168

[50] Bachiller S, Jiménez-Ferrer I, Paulus A, et al. Microglia in Neurological Diseases: A Road Map to Brain-Disease Dependent-Inflammatory Response. *Frontiers in Cellular Neuroscience*. 2018;12:488. doi:10.3389/fncel.2018.00488

[51] Geloso MC, Corvino V, Marchese E, Serrano A, Michetti F, D'Ambrosi N. The Dual Role of Microglia in ALS: Mechanisms and Therapeutic Approaches. *Frontiers in Aging Neuroscience*. 2017;9:242. doi:10.3389/fnagi.2017.00242

[52] Chiang PL, Chen HL, Lu CH, et al. White matter damage and systemic inflammation in Parkinson's disease. *BMC Neurosci*. 2017;18(1):48. doi:10.1186/s12868-017-0367-y

[53] Samara A, Murphy T, Strain J, et al. Neuroinflammation and White Matter Alterations in Obesity Assessed by Diffusion Basis Spectrum Imaging. *Frontiers in Human Neuroscience*. 2020;13. Accessed January 19, 2022. https://www.frontiersin.org/article/10.3389/fnhum.2019.00464

[54] Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*. 2012;61(4):1000-1016. doi:10.1016/j.neuroimage.2012.03.072

[55] Rafipoor H, Zheng YQ, Griffanti L, Jbabdi S, Cottaar M. Identifying Microstructural Changes in Diffusion MRI; How to Break Parameter Degeneracy. Published online September 11, 2021:2021.09.09.459626. doi:10.1101/2021.09.09.459626

[56] Kwan JY, Jeong SY, Van Gelderen P, et al. Iron accumulation in deep cortical layers accounts for MRI signal abnormalities in ALS: correlating 7 tesla MRI and pathology. *PLoS One*. 2012;7(4):e35241. doi:10.1371/journal.pone.0035241

[57] Kaunzner UW, Kang Y, Zhang S, et al. Quantitative susceptibility mapping identifies inflammation in a subset of chronic multiple sclerosis lesions. *Brain*. 2019;142(1):133-145. doi:10.1093/brain/awy296

[58] Sugiyama M, Takao M, Hatsuta H, et al. Increased number of astrocytes and macrophages/microglial cells in the corpus callosum in amyotrophic lateral sclerosis. *Neuropathology*. 2013;33(6):591-599. doi:10.1111/neup.12027

[59] Langkammer C, Krebs N, Goessler W, et al. Quantitative MR Imaging of Brain Iron: A Postmortem Validation Study. *Radiology*. 2010;257(2):455-462. doi:10.1148/radiol.10100495

[60] Shepherd TM, Flint J, Thelwall PE, et al. Postmortem interval alters the water relaxation and diffusion properties of rat nervous tissue – Implications for MRI studies of human autopsy samples. *Neuroimage*. 2009;44(3):820-826. doi:10.1016/j.neuroimage.2008.09.054

[61] Shatil AS, Uddin MN, Matsuda KM, Figley CR. Quantitative Ex Vivo MRI Changes due to Progressive Formalin Fixation in Whole Human Brain Specimens: Longitudinal Characterization of Diffusion, Relaxometry, and Myelin Water Fraction Measurements at 3T. *Frontiers in Medicine*. 2018;5. Accessed January 19, 2022. https://www.frontiersin.org/article/10.3389/fmed.2018.00031

[62] Tan RH, Kril JJ, Fatima M, et al. TDP-43 proteinopathies: pathological identification of brain regions differentiating clinical phenotypes. *Brain*. 2015;138(Pt 10):3110-3122. doi:10.1093/brain/awv220