http://eprints.gla.ac.uk/191840/

Stacked In-plane Histology for Quantitative Validation of Non-invasive Imaging Biomarkers: Application to an Infiltrative Brain Tumour Model

H. Al-Mubarak[a,b*], A. Vallatos[c], L. Gallagher[a], J. Birch[d], L. Gilmour[d], J. Foster[f], A.J. Chalmers[d], W.M. Holmes[a]

[a]Glasgow Experimental MRI centre, Institute of Neuroscience and Psychology, University of Glasgow, G61 1QH, U.K.

[b]Department of Physics, College of science, University of Misan, Iraq.

[c]Centre for Clinical Brain Sciences, University of Edinburgh, EH16 4SB ,UK.

[d]Wolfson Wohl Translational Cancer Research Centre, Institute of Cancer Sciences University of Glasgow, G61 1QH, UK.

[f]Department of Clinical Physics and Bioengineering, Greater Glasgow Health Board and University of Glasgow, B15 2GW,U.K.

*Haitham Al-Mubarak e-mail: h.al-mubarak.1@research.gla.ac.uk.

Antoine Vallatos e-mail: antoine.vallatos@ed.ac.uk.

Lindsay Gallagher e-mail: Lindsay.Gallagher@ glasgow.ac.uk

Joanna Birch e-mail : Joanna.Birch@ glasgow.ac.uk

Lesley Gilmour e-mail : Lesley.Gilmour@glasgow.ac.uk

John Foster e-mail: john.foster@ glasgow.ac.uk

Anthony Chalmers e-mail: Anthony.Chalmers@glasgow.ac.uk

William Holmes e-mail: William.Holmes@glasgow.ac.uk

Corresponding author: Haitham. Al-Mubarak*

Mailing address:

Mobile phone:++(44)7436852487

E-mail: h.al-mubarak.1@research.gla.ac.uk

Abstract

While it is generally agreed that histopathology is the gold standard for assessing non-invasive imaging biomarkers, most validation has been by qualitative visual comparison. . To date, the difficulties involved in accurately co-registering histology sections with imaging slices have prevented a voxel-by-voxel assessment of imaging modalities. By contrast with previous studies, which focus on improving the registration algorithms, we have taken the approach of improving the quality of the histological processing and analysis.

New method: To account for imaging slice orientation and thickness, multiple histology sections were cut in the MR imaging plane and averaged to produce stacked in-plane histology (SIH) maps. When combined with intensity sensitive staining this approach gives histopathology maps, which can be used as the gold standard to validate imaging biomarkers.

Results: We applied this pipeline to a patient-derived mouse model of glioblastoma multiforme (GBM). Increasing the number of stacked histology sections significantly increased SIH measured tumour volume. The SIH technique proposed here resulted in reduced variability of volume measurements and this allowed significant improvements in the quantitative volumetric assessment of multiple MRI modalities. Further, high quality registration enabled a voxel-wise comparison between MRI and histopathology maps.

Previous approaches to the validation of imaging biomarkers with histology, have been either qualitative or of limited accuracy. Here we propose a pipeline that allows for a more accurate validation via co-registration with SIH maps, potentially allowing validation in a voxel-wise mode.

Conclusion: This work demonstrates that methodically produced SIH maps facilitate the quantitative histopathologic assessment of imaging biomarkers.

Keywords: Imaging Biomarker, Registration, Mutual information, histology, MRI, validation.

1   Introduction

The use of non-invasive imaging modalities for clinical diagnosis continues to advance rapidly. A variety of methods are now available including Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Single Photon Emission Computerised Tomography (SPECT), Ultrasound (US), and x-ray Computed Tomography (x-ray CT). Often, the source of image contrast is related only indirectly to the underlying biology. This is especially true for MRI, where the signal intensity can depend upon many physical parameters including water content, local structure, tumbling rates, diffusion and hypoxia (Dominietto et al.,2014). There has been considerable interest in identifying whether such biologically indirect image contrasts can be used as non-invasive imaging biomarkers, either for normal biological functions, pathogenic processes or pharmacological responses to therapeutic interventions (Atkinson et al.,2001).

Histopathology is generally considered to be the ground thruth when considering the characterisation of diseased tissue (Kiessling et al.,2011). For histology, a post mortem or biopsy specimen is cut into thin sections to reveal its internal morphology and then stained to observe complex differentiated structures at the cellular level (Kiessling et al.,2011). The cutting process inherently yields 2D sections, which is the manner by which most histology is analysed. However, considerable work has been undertaken to reconstruct 3D histological volumes from serial 2D sections (Pichat et al.,2018). Though difficult, this allows knowledge of the 3D environment to be regained, while still accessing microscopic information (Stille et al.,2013).

When new imaging modalities are proposed as imaging biomarkers for particular diseases (Price,2011), it is difficult or impossible to validate them in human patients for ethical reasons. Validation against histopathology is limited to biopsy (Madabhushi et al.,2005) and later post-mortem comparisons (Kimt et al.,2000). In the case of biopsy, the size and number of samples taken is very limited and difficult to localise on images. Comparison of in-vivo non-invasive imaging and later post-mortem histology would be compromised by disease progression between imaging and death. Further, comparison of post-mortem imaging and post-mortem histology, would be compromised by the ex-vivo state of the tissue (Fagan et al.,2008).
Nevertheless, biomarker validation can be performed in preclinical disease models, where the animal can be terminated immediately following imaging for histological analysis. In principle, co-registration of imaging biomarkers with histopathology would allow direct validation. Indeed, there is a considerable literature describing such image registration algorithms and their application (Dauguet et al.,2007, Pichat et al.,2018). However, in practice, most preclinical validation is qualitative, limited to visual comparisons with sample histology sections, with little attempt made to match these to the corresponding imaging slice (Henning et al.,2007, Langer et al.,2009, Coquery et al.,2014). The reason for this is that accurate co-registration of non-invasive images with histology sections is challenging.

The processing, cutting and staining of histology sections can result in complex deformations such as fixation shrinkage, tears and cutting artefacts (Stille et al.,2013, Agarwal et al.,2018), which are difficult for registration algorithms to handle.

118   Examples of quantitative comparison of imaging biomarkers with histology include Stille et al.,

119   which used rigid registration and selected anatomical landmarks with a rodent stroke model (Stille

120   et al.,2013). This approach requires an expert to identify the control points, which can be a difficult

121   task due to internal distortions. Similarly, Ou et al. (2009) used non-rigid registration with selected

122   anatomical landmarks to register histology to MRI of prostate tumours, using two criteria:

123   maximization of landmark similarities and maximization of cancer region overlap. Jardim-Perassi

124   et al. (2019) used MRI-guided 3D printed tumour moulds to facilitate registration in a murine

125   breast tumour model. However, the accuracy of these approaches was limited by not accounting

126   for imaging slice thickness and often slice orientation too.

127

128   In this paper, we have tried to achieve high quality registration of histology with non-invasive

129   imaging data, not by improving on current image registration algorithms, but by focusing on

130   improving the quality of the histology used. This was done in five ways:

131

132   1. The use of a "flash-freeze" method for fixation instead of transcardial fixation with paraffin

133   embedding. With care and experience this helps to preserve tissue morphology, reducing

134   macroscopic distortions associated with extracting, cutting, and staining (Peters,2003, Ou

135   et al.,2009).

136   2. Histology was cut in relatively thick 20 µm sections, to reduce the risk of tears/distortions.

137   3. Histology sections were carefully cut in the image acquisition plane (e.g. the MRI plane),

138   guided by thin slice T2-weighted MRI. This is particularly important in order to maximise

139   spatial correlation between MRI and histology (cf. Figure 1).

140   4. A protocol was developed to register and stack multiple in-plane histology sections in

141   order to reflect the imaging slice thickness. For example, the thickness of an MRI slice

142   (~1-2 mm) is approximately 100 times thicker than a histology section (~10 to 20 µm) (cf.

143   Figure 2). This is crucial where the pathology is heterogeneous, with variations occurring

144   on the length scale of the imaging slice thickness.

145   5. The use of histological stains that exhibit signal intensities proportional to the observed

146   phenomenon, in order to produce semi-quantitative maps. This facilitates intensity based

147   registration, thus avoiding overfitting limitations of commonly used affine transformations

148   (Wells et al.,1996).

149

150

151   To evaluate the ability of this overall approach to provide a quantitative histopathologic

152   assessment of in-vivo imaging biomarkers, we applied it to a patient-derived mouse model of

153   glioblastoma multiforme. In GBM, a major factor contributing to treatment failure is the ability of

154   tumour cells to infiltrate adjacent normal brain tissue (Price,2011), with low tumour cell density

155   extending far beyond the bulk of the tumour. Identifying the full extent of infiltration is important for

156   both radiotherapy planning and to achieve complete surgical resection. Here, we present the

157   different steps leading to the production of a 3D data matrix, from co-registration of multiple MRI

158   modalities with stacked in-plane histology. We show how the resulting matrix allows MRI

159   modalities to be assessed, both in terms of tumour volume detection and via direct voxel-wise

160 comparison. Such an approach should become the accepted gold standard for validating non-
161 invasive imaging biomarkers.
162

2   Materials and Methods

2.1   Animal preparation

165 Experiments were performed on ten immunocompromised CD1 nude mice (20 to 25g, Charles
166 River Laboratories), which were acclimatized at least one week prior to any experimental
167 procedure. Animals that did not exhibit tumour growth (n=1) were removed from the study. G7
168 human glioblastoma cells were cultured in stem-like conditions (Advanced DMEM:F12, containing
169 20µM EGF/FGF, 1% B27, 0.5% N2, heparin, 1% L-Glut) on Matrigel coated plates. The animals
170 were intracranially injected with G7 cells (10^5 cells per mouse) into the sub-ventricular zone using
171 stereotactic equipment (Gomez-Roman et al.,2017). This cell line produces a tumour bulk with
172 infiltrative edges in-vivo that replicates the human disease (Ahmed et al.,2015). To avoid
173 unnecessary animal use, this work is based on data produced by a brain tumour infiltration study
174 where animals were scanned using MRI at weeks 9 and 12 after GBM injection (Vallatos et
175 al.,2018a). Experiments were carried out in accordance with the local ethical review panel, the UK
176 Home Office Animals (Scientific Procedures) Act 1986 and the United Kingdom National Cancer
177 Research Institute guidelines for the welfare of animals in cancer research (Workman et al.,2010).
178 Study outcomes are reported according to the ARRIVE guidelines (Kilkenny et al.,2011).
179

2.2   MRI set up

181 MRI experiments were performed on a 7Tesla Bruker Biospec Avance system (Bruker Biospin,
182 Ettlingen, Germany). Homogeneous radiofrequency excitation was achieved using a 72mm
183 birdcage volume resonator, with the signal detected using an actively decoupled 4-channel
184 phased array receive-only head surface coil (Rapid Biomedical, Wurzburg, Germany). The mice
185 were initially anaesthetized using 5% isoflurane and a 30:70 O_2/N_2O ratio and were positioned
186 prone on an MRI cradle. A hot water circulation jacket was used to regulate the animal
187 temperature (37±1 °C) , which was monitored using a rectal probe. The head was secured
188 laterally using conical ear rods and longitudinally by the nose cone used for anaesthetic gas
189 delivery. The animals breathed spontaneously through a facemask delivering a constant flow of
190 isoflurane mixed with a 40:60 ratio of O_2/N_2O (1 L min^{-1}). Isoflurane concentration varied from 1.5
191 % to 3 %, in order to maintain stable respiration rates within normal physiological ranges (40-70
192 bpm). Respiration was monitored throughout the experiment using a pressure sensor connected
193 to an air-filled balloon placed under the animal abdomen (Biotrig software, Bruker, Ettlingen,
194 Germany).
195

2.3   MRI experiments

197 Following a geometry correction scan, a series of MRI experiments were performed (field of view
198 2×2 cm, five 1.5 mm thick coronal imaging slices centred at 4 mm from the rhinal fissure). T2-
199 weighted imaging (T2W) was performed using a rapid acquisition with relaxation enhancement
200 (RARE) sequence (TE=47 ms, TR=4,300 ms, matrix=176×176, in-plane resolution 113 µm x113
201 µm, 9 min). A second set of T2-weighted images were acquired (T2W$_{Histology}$), with the same

parameters as above but with fifteen 0.5 mm thick slices, for the specific purpose of guiding the cutting of histology sections. Diffusion-weighted imaging (DW) was performed using a 4-shot spin-echo planar imaging DW scan (TE=37ms, TR=4,500 ms, matrix=128×128, 1.5 mm slice thickness, 6 directions, b-values = 0, 1000 s mm^{-2}, 10 min). Perfusion weighted imaging (PW) was performed using a multiple boli Arterial Spin Labelling sequence (mbASL) (Vallatos et al.,2018b), labelling with a train of twenty hyperbolic-secant inversion pulses (duration=3.3 ms, dimensionless amplitude parameter μ=8, angular modulation β=760 s^{-1}) evenly distributed over a 5 s labelling duration. The inversion slice width was 8.5 mm and the offset from the imaging slice was 15 mm. Image acquisition was achieved using an EPI module (TE=12 ms, TR=7 s, matrix=96×96, 4 acquisition segment, partial FT=1.4, 12 averages, 9 min). T1 weighted imaging was performed using a RARE acquisition (TE=12.3 ms, TR=800 ms, matrix=176×176,RT=4, 8 min. Following in-vivo scanning, a doped water phantom was scanned using the above sequences, for use in correcting the receiver coil bias. MRI data were exported in DICOM format.

2.4 Histology protocol

Following MRI scanning, the animal was taken to deep anaesthesia then removed from the MRI cradle to a nearby bench and decapitated. The skin was peeled back and the brain removed from the skull. The brain was then fresh-frozen for 2 minutes at -45 ˚C using an isopentane solution tube immersed in dry ice. The frozen brains were then embedded in Cryomatrix and protected in an M-1 embedding matrix to prevent dehydration (Thermo Fisher Scientific, UK). Freezing was favoured in order to avoid the unpredictable macroscopic tissue deformation related to perfusion-fixation and paraffin embedding (Petersen et al.,2001). Brain slicing was performed manually on an OTF 5000 Bright cryostat (-16 ˚C) equipped with a rotary knife (Bright, Criostato-OTF-5000). A relatively thick section thickness of 20 μm was chosen, to reduce the risk of tears/distortions. Care was taken to cut the sections in planes parallel to the MRI imaging plane. For this, sectioning was guided by T2W$_{Histology}$ images (slice thickness = 0.5 mm), see Figure 1. Common brain structures identified by an experienced neuroscience research technician (L. Gallagher) were used to iteratively orientate the sectioning plane parallel to the MRI plane. Five pairs of adjacent sections within the MRI thickness were cut at 300 μm intervals and then lifted onto to poly-L-lysine slides, see Figure 2. The sections were then stained using either haematoxylin and eosin (H&E) or Human Leukocyte Antigen (HLA) to identify the human tumour cells. The 20 μm cryosections were fixed in ice cold acetone and washed in PBS before blocking in 3% BSA/TBS-tween for 30 mins at room temperature. A 1:500 dilution HLA antibody (abcam ab70328) in blocking buffer was added and incubated for 2 hours at room temperature. Sections were washed three times with TBS-Tween before the addition of 1:1000 anti-mouse Alexa Fluor 647-conjugated secondary antibody (Thermo Fisher Scientific, UK - A21236) for 1 hour incubation in the dark. Sections were washed 3 times with TBS-Tween and mounted in a ProLong Diamond Antifade mount with DAPI (Thermo Fisher Scientific, UK - P36966). Whole brain section tile scans were conducted using a Zeiss 710 upright confocal microscope at x 10 magnification (Far red filters- 638 – 747. Beam splitters- MBS : MBS 488/561/633, MBS_InVis : Plate, DBS1 : Mirror). Histology images (~1300×1000 pixels) were exported as .tiff files.

*** Figure1 and 2 to appears near here***

6

2.5   Data processing pipeline

Both MRI and histology data were processed using Matlab code developed in-house (Matlab R2015a , MathWorks Ltd., U.K.). The overall processing pipeline is summarised in Figure 3.

Figure 3 appears near here

2.6   MRI data analysis

To remove any bias that could arise due to differences in the image intensity values for the different modalities, the DICOM images were normalised (0-1 range). Furthermore, non-uniform detection sensitivity associated with the use of a surface receiver coil was corrected, as it can adversely affect the registration processes: T1W, T2W and DW images were normalised (removing the sensitivity of the surface coil) using corresponding phantom MRI images acquired using the same parameters (Axel et al.,1987). Apparent diffusion coefficient (ADC) maps were calculated by fitting the DWI data to the mono-exponential equation of the Stejskal and Tanner model (Stejskal et al.,1965). Normalized relative perfusion maps (PW) were produced from the MRI signal of the control image ($M_{control}$) and labelled image (M_{label}), using the equation ($M_{control}$ - M_{label})/$M_{control}$. Prior to comparison, all data (T1W, DWI, ADC, ASL, and SIH) were resized to the T2W in-plane resolution (176 x 176).To reduce processing time, the brain region was separated from the background by the application of an active contour method following manual delineation (Caselles et al.,1997).

2.7   Histology data analysis

Following digital scanning of whole brain histology sections, the histology images were rotated by a small angle to remove differences in orientation of the brain due to the laying out on glass microscope slides. Histology images were then resampled from their original resolution using the cubic spline method to match the resolution of the T2W images (176x176). Signal intensity inhomogeneity due to a difference in staining across the image were automatically corrected for each section by using a histogram equalization method (Belsare,2012). Further, a threshold value was selected, creating a brain mask to remove the background signal. Stacked In-plane Histology (SIH) maps were generated first by co-registration of multiple histology sections, then by taking a voxel-wise average of the signal intensities. Registration used non-rigid Mutual Information (MI) based transformation with global translation, rotation, scaling and shearing for optimal registration. In one histology section with greater tissue deformations, a B-spline method was applied to improve the registration.

2.8   Histology to MRI co-registration and 3D matrices production

Registration of histology sections with MR images is typically challenging due to a significant variation of image properties, such as resolution, field of view and contrast (Madabhushi et

al.,2005). Here, the SIH maps, allowed intensity based registration with MRI images to be undertaken using the Mutual Information registration method. For consistent registration, the histology sections were transformed so they had the same resolution and dimensions as the MR images.

2.9 Segmentation protocol

For both histology and MRI, tumour related abnormal regions of interest (ROI) were manually drawn by 3 observers (with more than 3 years' experience). Histology ROIs were selected on the basis of HLA stain intensity on SIH maps. MRI ROI delineation was performed without prior knowledge of the histology data, to avoid selection bias. Care was taken not to include non-infiltration related enhancement (e.g. ventricle compression). Inter-observer reproducibility of ROI selection was evaluated using the coefficient of variation (100 ×standard deviation/mean), resulting in 10% for histology and 12-21% for the different MRI modalities.

2.10 Statistical analysis

ROC curve analysis (Garcia-Lorenzo et al.,2013) was used to compare tumour volume dependence on number of sections used to produce the SIH maps, with 5-section SIH maps being the gold standard: voxels correctly identified as tumour are true positive (TP): voxels incorrectly identified as tumour are false positive (FP): voxels correctly identified as non-tumour are true negative (TN) and voxels identified incorrectly as non-tumour are false negative (FN). These values are then used to calculate Sensitivity, Specificity, Accuracy and Dice,

$$Sensitivity = \frac{TP}{TP+FN} \quad..........(1) \qquad Specificity = \frac{TN}{TP+TN} \quad..........(2)$$

$$Accuracy = \frac{TP+TN}{TP++FP+FN+TN} \quad.....(3) \qquad Dice = \frac{2*TP}{FP+FN+2*TP} \quad............(4)$$

Additional comparison between single- and multiple-section histology was conducted using Bland Altman plots. The two-tailed Student's t-test was used for comparisons between MR measurements of tumour volume and histology measurements of tumour volume (SSH or SIH), using a Bonferroni correction (Graph Pad prism 6, Ver.6.01, 2012). All values are reported as means ± standard deviation. * Statistically significant $p<0.01$, ** statistically significant $p<0.001$, *** statistically significant $p<0.0001$ and NS not statistically significant. Statistical power analysis was performed using G-Power (version 3.1) software (Erdfelder et al.,1996).

3 Results & Discussion

We present a quantitative method for validating imaging-based biomarkers by registration with stacks of in-plane histology.

While it is generally agreed that histopathology is the gold standard for assessment, in practice most preclinical validation is limited to visual comparisons with sample histology sections, with little attempt made to spatially match the histology section to the corresponding imaging slice. By improving the quality of the histology processing and analysis, we have been able to produce stacked in-plane histology (SIH) maps. These high quality SIH maps can then be co-registered with non-invasive imaging modalities, allowing more direct and quantitative validation of imaging biomarkers than has previously been possible.

To demonstrate this methodology, we applied it to a patient-derived mouse model of glioblastoma multiforme (GBM). In the case of GBM patients, an imaging biomarker capable of identifying the full extent of GBM cell infiltration would be valuable for both radiotherapy planning and in achieving complete surgical resection. Below, we assess the optimal number of histology sections for SIH maps and the quality of SIH with MRI registration. Finally, potential MRI biomarkers are assessed, both by volumetric and voxel-wise comparison with SIH maps.

3.1 Tumour volume measurement via single-section Histology (SSH)

Figure 4A shows examplary H&E and HLA stained sections, obtained from within the 1.5 mm MRI slice. The heterogeneity in tumour shape and size is readily apparent at this length scale. The commonly applied method of arbitrarily selecting a single-section of histology (SSH) to estimate tumour volume inevitably leads to significant measurement variation. For example, the percentage difference between the minimum tumour volume (V_{min}) and maximum tumour volume (V_{max}) for each series of sections was found to be 46% for H&E and 50% for HLA (Figure 4B). While tumours can be identified on H&E sections due to a much higher density of cell nuclei, there is less sensitivity in detecting regions of low-density GBM infiltration. However, Human Leukocyte Antigen (HLA) staining is very specific in the mouse model, as it only stains cells that originated from the implanted human tumour cells. Hence, in the following analysis we define tumour volume as the maximum extent of GBM cell infiltration identified using the HLA sections.

*** Figure 4 appears near here***

3.2 Determining optimal number of histology sections for SIH maps

It is clear that increasing the number of histology sections (20 µm thick) used to generate a SIH map will make the SIH map more representative of the corresponding imaging slice (1500 µm thick). However, this comes at the expense of longer processing time (histology preparation / analysis time). In applying the SIH method, the optimal number of sections will be disease specific, depending on the length scale of heterogeneities and the corresponding imaging slice thickness.

For the GBM tumours, we assessed how tumour volume measurements were improved by using more HLA sections to generate the SIH maps. In 6 out of 9 mice, where five quality HLA sections were available, SIH maps were produced with one, two, three, four and five sections (Figure 5A-E), using all possible combinations of the sections. The measured tumour volume reaches a plateau when 3 or more sections are used to produce the SIH map, with no significant difference found between using 3, 4 or 5 sections (Figure 5F). These volume measurements are analogous to numerical integration where the more sections used to calculate a volume, the more accurate the result will be, eventually converging at the true value.

To further investigate the effect of using multiple histology sections, we performed a ROC curve analysis (Garcia-Lorenzo et al.,2013) comparing SIH maps produced with 1, 2, 3 and 4 HLA sections to maps produced with 5 HLA sections. This assumed the 5-section map was the 'ground thruth' for the assessment of the other maps, To avoid bias in the selection of sections, maps were produced from all possible combinations of sections for each mouse.). It should be noted that the values of Dice, sensitivity, and accuracy indices will be dependent on the number of sections used in the gold standard. However, Figure 5G-J does show a diminishing increase of these indices with the number of sections used, and their standard deviation decreases markedly. As expected, specificity measures were not affected by this evaluation, as smaller numbers of sections tended to underestimate the tumour region. A Bland-Altman analysis showed that the number of sections required is inversely proportional to the tumour size, see supplemental information (Figure 1S).

Given the above analysis, as a trade-off between improved accuracy and expanded processing time, we settled on using three histology sections for the remaining analysis. After excluding poor quality sections, the three sections with the largest tumour area were selected. It is crucial to note that the choice of three sections is very specific to this particular disease model (mouse model of glioblastoma). If the SIH method is to be applied to different disease models or different species, then the optimum number of slices will be different and will need to be assessed.

Figure 5 appears near here

3.3 SIH to MRI registration quality

Registration of histology with MRI was qualitatively and quantitatively evaluated at each stage of the process. The qualitative evaluation consisted of a visual inspection of the overlay of the inner and outer contours of the T2W image and histology section (Figure 6). Accurate alignment was observed between borders and internal structures. Excellent post-registration overlays were found; with Dice values above (0.96±0.011). The resized and co-registered data were used to

create a 3-dimensional data matrix, which allowed the MRI modalities to be assessed against histology, both in terms of tumour volume detection and via direct voxel-wise comparison.

3.4   Volumetric assessment of MR biomarkers.

The proposed SIH approach was used to quantitatively validate different MRI modalities as imaging biomarkers of GBM infiltration. For this, both the SSH and SIH approaches were used to measure the "ground thruth" tumour volume (i.e. the volume of GBM infiltrated tissue) for the same dataset. Given the analysis in section 3.3, as a trade-off between improved accuracy and expanded processing time, we settled on using three HLA sections to produce the SIH maps. In all 9 mice, after excluding poor quality sections, the three sections with the largest tumour area were selected.

Figure 7A shows representative manual ROI selections for each imaging modality in the same animal. Figure 7B shows tumour volumes obtained from manual delineation of the various MRI modalities, compared with tumour volumes measured from five individual single sections of histology (SSH). Clearly, the large standard deviation of the SSH tumour volume measurements ($\pm$6.022) makes it a poor 'ground thruth' for validating the MR tumour volume measurements, with no significant differences found between any of the MRI modalities and the SSH measurements.

However, tumour volume measurements made using SIH maps show a much lower standard deviation ($\pm$0.81) (Figure 7C), allowing better validation of the different MRI modalities. A statistically significant difference was found between the SIH measured tumour volume and those measured with T2W, DW and ADC, whereas both T1W and ASL measurements showed no significant difference and, hence, can be evaluated as better biomarkers for tumour cell infiltration in this animal model.

The lower standard deviation of SIH tumour volume measurements has important implications, as it allows statistical significance to be achieved without requiring an increase in the number of animals used. To further examine this, we performed power analysis using the results presented in Figure 7 (Figure 2S). For example, to achieve a statistically significant difference ($p<0.05$) between T2W and SSH tumour volume measurements, would require between 72 and 800 mice. By contrast, using SIH maps as the ground thruth, required only 9 mice to achieve $p<0.01$. Such an impressive reduction in animal usage is a stated aim of the UK government, via its policy of Replacement, Refinement and Reduction of Animals in Research (NC3Rs) and should arguably feature in future guidelines for reporting in-vivo experiments (Kilkenny et al.,2011).

11

3.5 Towards voxel-by-voxel assessment

The ideal approach to validating imaging biomarkers would involve voxel-by-voxel comparison with co-registered histology. To date, the difficulties involved in accurately co-registering histology sections with imaging slices have prevented this. However, we believe that the methodological pipeline we have outlined overcomes many of these difficulties, yielding a co-registered multi-dimensional data matrix (T1W, T2W, DWI, ADC, ASL and SIH map).

In the case of HLA, the staining intensity is proportional to the density of tumour cell membranes. Therefore, by averaging multiple histology sections, the resulting SIH maps represent a semi-quantitative measure of tumour cell density in the MRI slice. This allows the MRI modalities to be more accurately evaluated against histology in a direct voxel-by-voxel analysis. For example, Figure 8A-C shows scatter plots of different MRI modalities against SIH intensity. Here the pathogenic regions identified by the 'ground thruth' histology ROIs can be highlighted, allowing the relationship between the MRI signal and the underlying histopathology to be assessed. Furthermore, it enables the assessment of multi-spectral analysis approaches on a voxel-by-voxel basis, investigating and validating combinations of MRI parameters against histology.

Figure 8 appears near here

4 Conclusion

We have introduced a novel methodological pipeline to improve the validation of non-invasive imaging biomarkers. In contrast to most previous studies, which focus on improving the registration algorithms, we have taken the approach of improving the quality of the histology processing and analysis. In an infiltrative brain tumour model we have demonstrated how stacked in-plane histology (SIH) maps, co-registered with multiple MRI modalities, provide a 'ground truth' for quantitative comparisons. Our results demonstrate that, in cases of small and heterogeneous tumours the use of this multi-section approach is crucial, as conventional assessment using single-section histology is prone to significant errors. Finally, the development of robust three-dimensional registration with non-invasive imaging modalities could lead to the emergence of voxel-by-voxel histopathologic assessment of new imaging biomarkers.
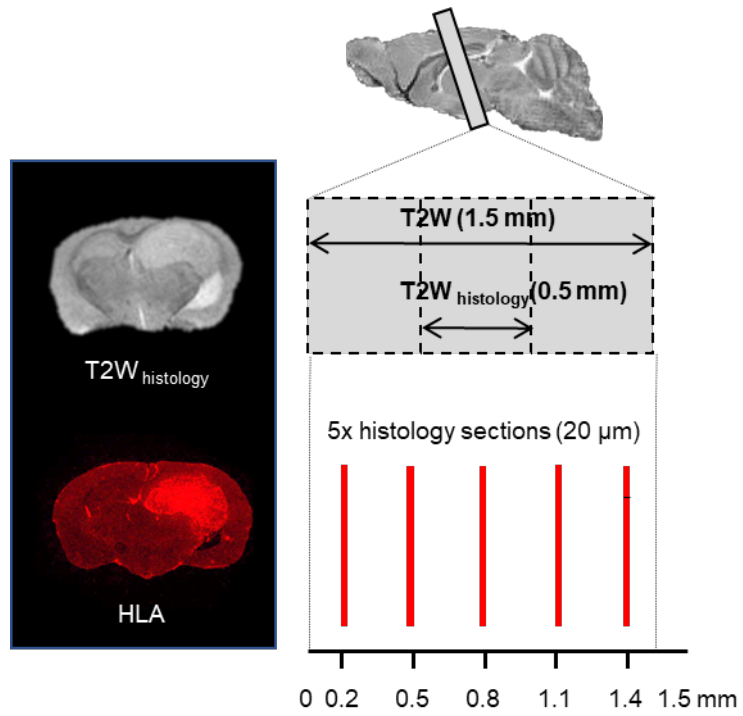
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
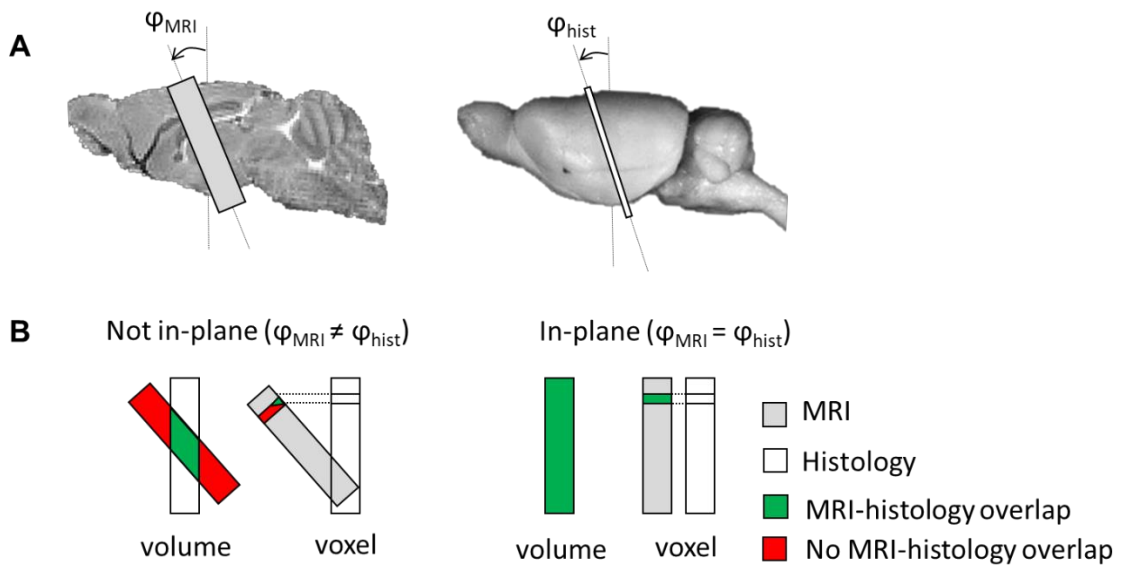
Figure 1. Effect of cutting angle (φ) on MRI to histology comparison: (A) MRI (1.5 mm thickness) and histology (20 μm thickness) cutting angles. (B) The effect of cutting angle discrepancies on the overall volume and voxel-wise overlap between MRI and stacked histology.
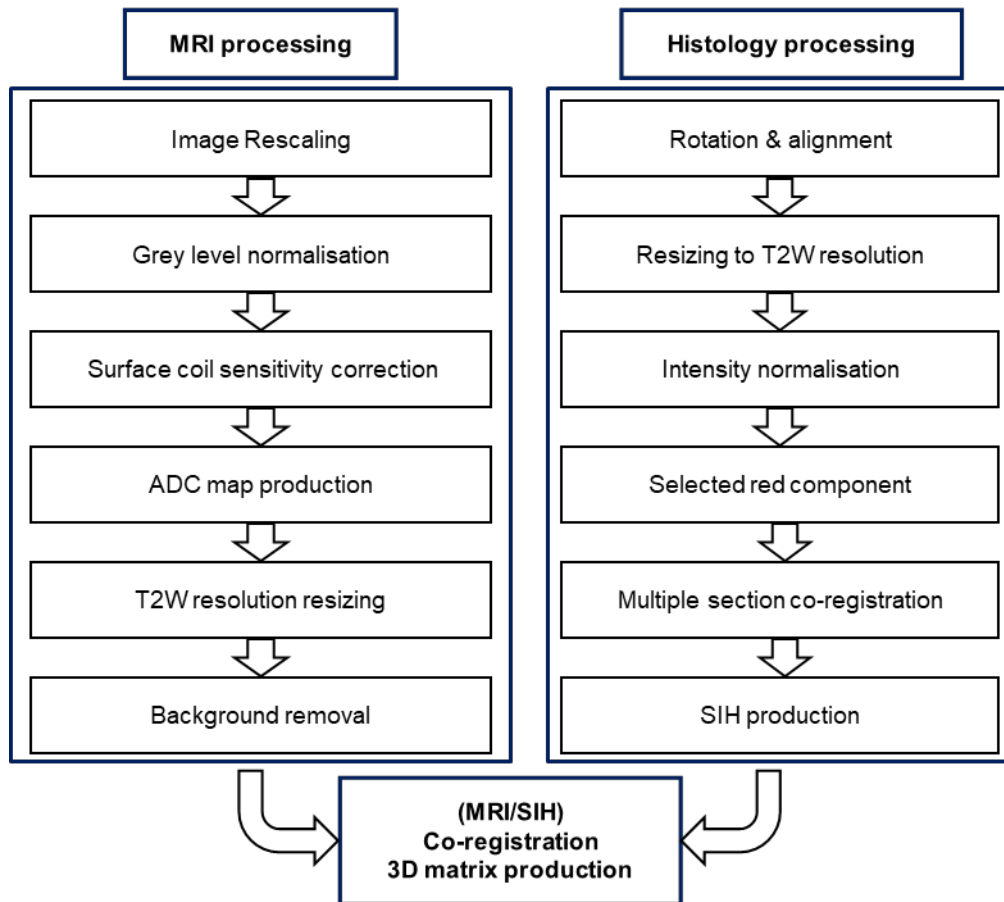
513
514
515
516

517
518 *Figure 2. The cutting of histology section was guided by 0.5 mm thick T2-weighted images*
519 *(T2W$_{Histology}$), matching the cryo-section plane to the MRI acquisition plane. Five evenly distributed*
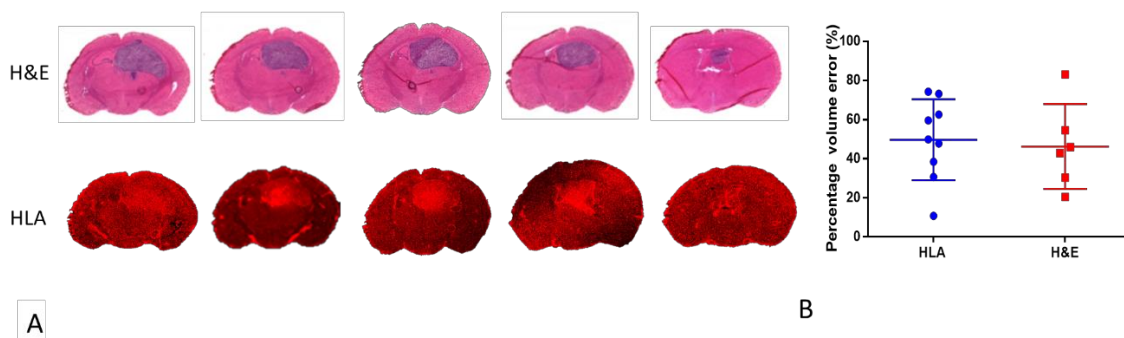520 *histology sections (20 µm) were cut (red) to cover the 1.5 mm thickness of the T2W scans.*
521
522

Figure 3. *Simplified diagram of the image processing pipeline leading to the production of 3D matrices combining MRI modalities and SIH data.*


Figure 4. *Examples of histology sections for a GBM mouse and volume error comparison: (A) Five corresponding histology sections (H&E and HLA) cut within the 1.5 mm thickness of one MRI slice (B) Percentage volume error between maximum and minimum tumour volumes (Vmax and Vmin respectively) in the five sections for each animal, calculated using (Vmax - Vmin)/Vmax.)*100.*

539

540

541

542
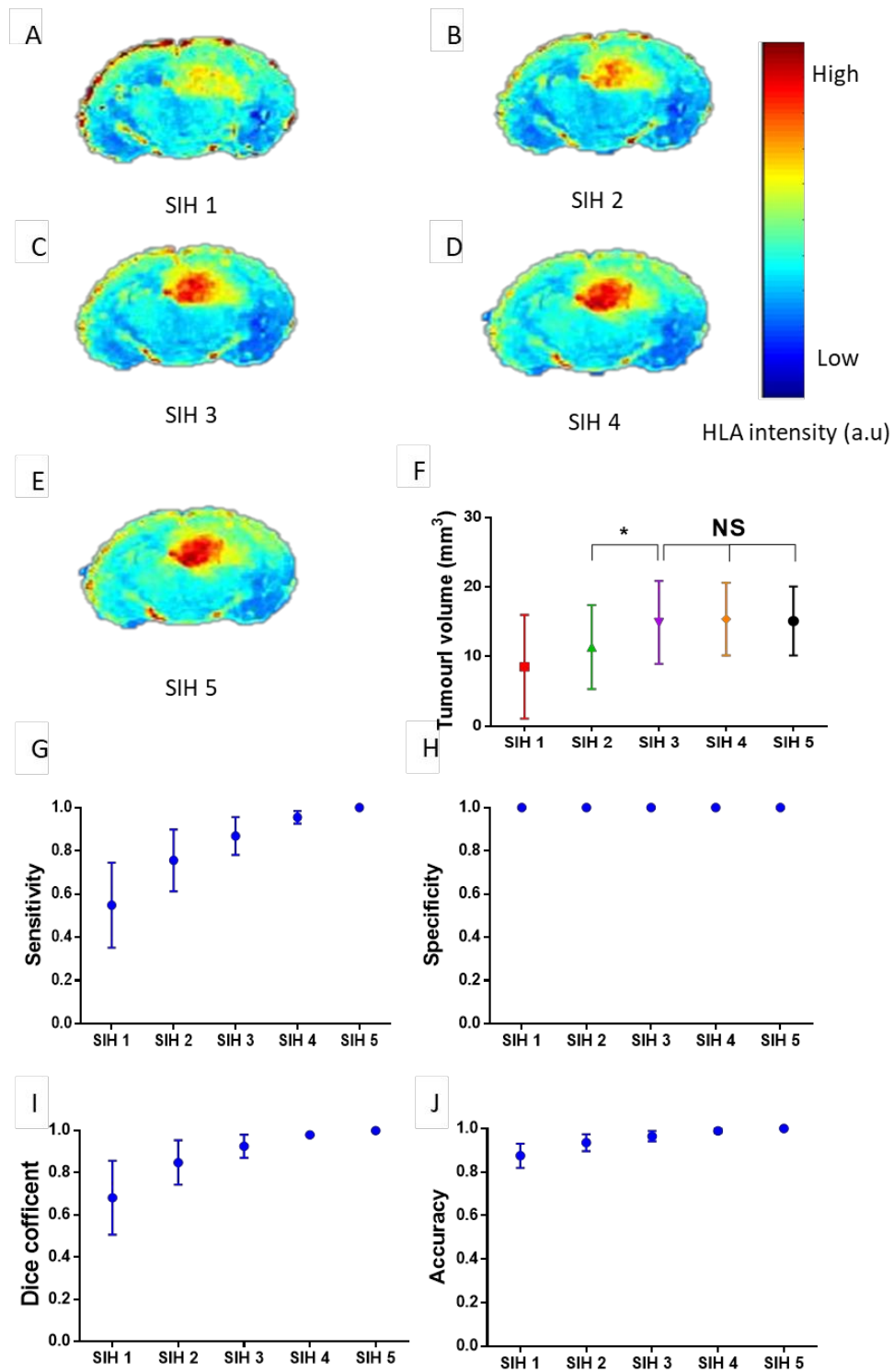
Figure 5. (A,B,C,D and E): SIH maps generated using one section (SIH1), two sections (SIH2), three sections (SIH3), four sections (SIH4) and five sections of HLA (SIH5). (F) Measured tumour volume against number of sections used to produce the SIH map (n=6). Evaluation of the ability of SIH maps to probe the tumour related abnormal area in comparison with the 5-section SIH map: (G) sensitivity, (H) specificity, (I) Dice similarity coefficient and (J) accuracy indices.
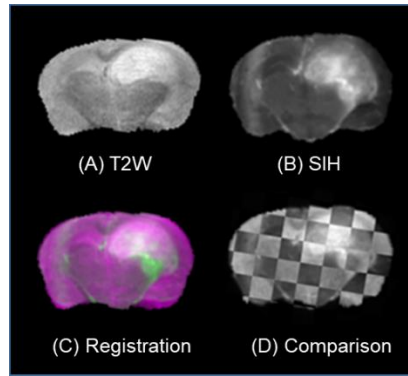
Figure 6. Example of non-rigid co-registration of histology with MRI using the Mutual Information method: (A) a T2-weighted image (T2W) (B) Stacked in-plane histology (SIH). (C) co-registration fusion image with false colour showing similarities (purple) and difference (green). (D) checkerboard comparison between registered MRI and SIH maps.
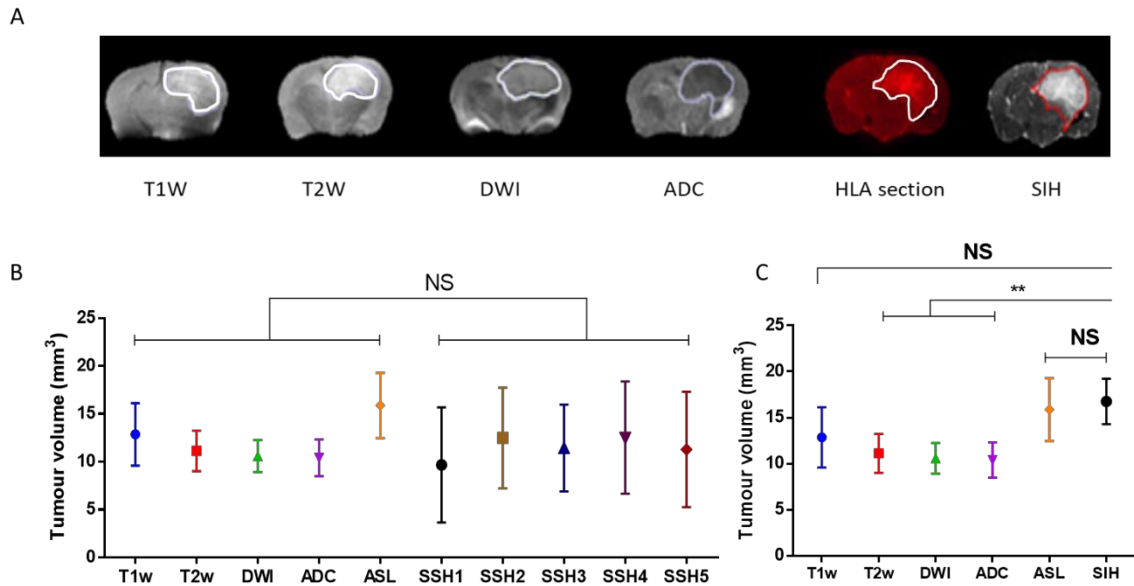
Figure7. Volumetric analysis: (A) Examples of Regions of interest (ROIs) for MRI modalities and histology from the same animal. (B) Comparison of tumour volume measurements made using MRI modalities and single sections of histology (SSH). (C) Comparison of tumour volume measurements made using MRI modalities and stacked in-plane histology (SIH) maps generated using 3 sections.
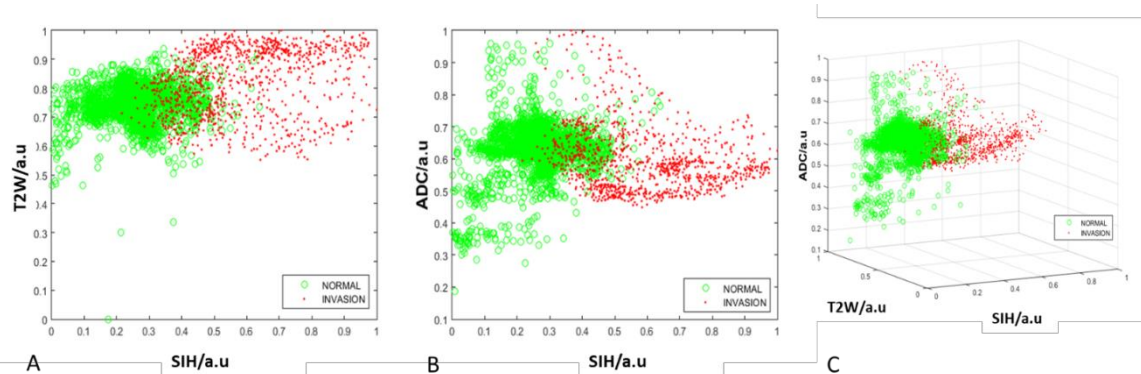
*Figure*8. *Voxel-wise scatter plots of MRI and SIH normalised data, showing the overlap between normal tissue and voxels infiltrated by GBM cells. (A) T2W against SIH and (B) ADC against SIH. (C) 3-D plot of T2W and ADC against SIH map.*

References

Uncategorized References

Agarwal, N., Xu, X. and Gopi, M., 2018. Geometry processing of conventionally produced mouse brain slice images. J Neurosci Methods 306: 45-56, DOI: 10.1016/j.jneumeth.2018.04.008.

Ahmed, S. U., Carruthers, R., Gilmour, L., Yildirim, S., Watts, C. and Chalmers, A. J., 2015. Selective Inhibition of Parallel DNA Damage Response Pathways Optimizes Radiosensitization of Glioblastoma Stem-like Cells. Cancer Res 75(20): 4416-4428, DOI: 10.1158/0008-5472.CAN-14-3790.

Atkinson, A., Colburn, W., Degruttola, V., Demets, D., J, D. G., Hoth, D., Oates , J. A., Peck, C. C., Schooley, R., Spiker, B., Woodecock, J. and Zeger, S., 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 69(3): 89-95, DOI: 10.1067/mcp.2001.113989.

Axel, L., Costantini, J. and Listerud, J., 1987. Intensity correction in surface-coil MR imaging. AJR Am J Roentgenol 148(2): 418-420, DOI: 10.2214/ajr.148.2.418.

Belsare, A. D., 2012. Histopathological Image Analysis Using Image Processing Techniques: An Overview. Signal & Image Processing : An International Journal 3(4): 23-36, DOI: 10.5121/sipij.2012.3403.

Caselles, V., Kimmel, R. and Sapiro, G., 1997. Geodesic active contours. International Journal of Computer Vision 22(1): 61-79, DOI: Doi 10.1023/A:1007979827043.

Coquery, N., Francois, O., Lemasson, B., Debacker, C., Farion, R., Remy, C. and Barbier, E. L., 2014. Microvascular MRI and unsupervised clustering yields histology-resembling

images in two rat models of glioma. J Cereb Blood Flow Metab 34(8): 1354-1362, DOI: 10.1038/jcbfm.2014.90.

Dauguet, J., Delzescaux, T., Conde, F., Mangin, J. F., Ayache, N., Hantraye, P. and Frouin, V., 2007. Three-dimensional reconstruction of stained histological slices and 3D non-linear registration with in-vivo MRI for whole baboon brain. J Neurosci Methods 164(1): 191-204, DOI: 10.1016/j.jneumeth.2007.04.017.

Dominietto, M. and Rudin, M., 2014. Could magnetic resonance provide in vivo histology? Front Genet 4(4): 298, DOI: 10.3389/fgene.2013.00298.

Erdfelder, E., Faul, F. and Buchner, A., 1996. GPOWER: A general power analysis program. Behavior Research Methods Instruments & Computers 28(1): 1-11, DOI: Doi 10.3758/Bf03203630.

Fagan, A. J., Mullin, J. M., Gallagher, L., Hadley, D. M., Macrae, I. M. and Condon, B., 2008. Serial postmortem relaxometry in the normal rat brain and following stroke. J Magn Reson Imaging 27(3): 469-475, DOI: 10.1002/jmri.21246.

Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L. and Collins, D. L., 2013. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med Image Anal 17(1): 1-18, DOI: 10.1016/j.media.2012.09.004.

Gomez-Roman, N., Stevenson, K., Gilmour, L., Hamilton, G. and Chalmers, A. J., 2017. A novel 3D human glioblastoma cell culture system for modeling drug and radiation responses. Neuro Oncol 19(2): 229-241, DOI: 10.1093/neuonc/now164.

Henning, E. C., Azuma, C., Sotak, C. H. and Helmer, K. G., 2007. Multispectral quantification of tissue types in a RIF-1 tumor model with histological validation. Part I. Magn Reson Med 57(3): 501-512, DOI: 10.1002/mrm.21161.

Jardim-Perassi, B. V., Huang, S., Dominguez-Viqueira, W., Poleszczuk, J., Budzevich, M. M., Abdalah, M. A., Pillai, S. R., Ruiz, E., Bui, M. M., Zuccari, D., Gillies, R. J. and Martinez, G. V., 2019. Multiparametric MRI and co-registered histology identify tumor habitats in breast cancer mouse models. Cancer Res, DOI: 10.1158/0008-5472.CAN-19-0213.

Kiessling, F. and Pichler, B. J. 2011. Small animal imaging basics and practical guide. Germany, Springer.

Kilkenny, C., Browne, W., Cuthill, I. C., Emerson, M., Altman, D. G. and National Centre for the Replacement, R., and Reduction of Amimals in Research, 2011. Animal research: reporting in vivo experiments--the ARRIVE guidelines. J Cereb Blood Flow Metab 31(4): 991-993, DOI: 10.1038/jcbfm.2010.220.

Kimt, T. S., Singh, M. and Sungkara, W., 2000. Automatic registration of postmortem brain slices to MRI reference volume. ICEE transactions on nuclear science 47(4): 1607-1613.

652 Langer, D. L., van der Kwast, T. H., Evans, A. J., Trachtenberg, J., Wilson, B. C. and
653 Haider, M. A., 2009. Prostate cancer detection with multi-parametric MRI: logistic
654 regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-
655 enhanced MRI. J Magn Reson Imaging 30(2): 327-334, DOI: 10.1002/jmri.21824.
656
657 Madabhushi, A., Feldman, M. D., Metaxas, D. N., Tomazeweski, J. and Chute, D., 2005.
658 Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI.
659 IEEE Trans Med Imaging 24(12): 1611-1625, DOI: 10.1109/TMI.2005.859208.
660
661 Ou, Y., Shen, D., Feldman, M., Tomaszewski, J. and Davatzikos, C.,2009. Non-rigid
662 registration between histological and MR images of the prostate: a joint segmentation and
663 registration framework. IEEE Computer Society Conference on Computer. Miami, FL,
664 USA, IEEE.
665
666 Peters, S. R., 2003. The art of embedding tissue for frozen section part I: a system for
667 precision face down cryoembedding of tissues using freezing temperture-embedding
668 wells. Journal of Histotechnology(26): 11-19.
669
670 Petersen, D. K. A., Nyengaard, J. R. and Gundersen, H. J. G., 2001. Tissue shrinkage and
671 unbiased stereological estimation of particle number and size. Journal of Microscopy
672 204(3).
673
674 Pichat, J., Iglesias, J. E., Yousry, T., Ourselin, S. and Modat, M., 2018. A Survey of
675 Methods for 3D Histology Reconstruction. Med Image Anal 46: 73-105, DOI:
676 10.1016/j.media.2018.02.004.
677
678 Price, S. J., Gillard, J. H., 2011. Imaging biomarkers of brain tumour margin and tumour
679 invasion. Br J Radiol 84 Spec No 2: S159-167, DOI: 10.1259/bjr/26838774.
680
681 Stejskal, E. O. and Tanner, J. E., 1965. Spin Diffusion Measurements: Spin Echoes in the
682 Presence of a Time-Dependent Field Gradient. Journal of Chemical Physics 42(1): 288-+,
683 DOI: Doi 10.1063/1.1695690.
684
685 Stille, M., Smith, E. J., Crum, W. R. and Modo, M., 2013. 3D reconstruction of 2D
686 fluorescence histology images and registration with in vivo MR images: application in a
687 rodent stroke model. J Neurosci Methods 219(1): 27-40, DOI:
688 10.1016/j.jneumeth.2013.06.003.
689
690 Vallatos, A., Al-Mubarak, H. F. I., Birch, J. L., Galllagher, L., Mullin, J. M., Gilmour, L.,
691 Holmes, W. M. and Chalmers, A. J., 2018a. Quantitative histopathologic assessment of
692 perfusion MRI as a marker of glioblastoma cell infiltration in and beyond the peritumoral
693 edema region. J Magn Reson Imaging: 1-12, DOI: 10.1002/jmri.26580.
694
695 Vallatos, A., Gilmour, L., Chalmers, A. J. and Holmes, W. M., 2018b. Multiple boli
696 arterial spin labeling for high signal-to-noise rodent brain perfusion imaging. Magnetic
697 Resonance in Medicine 79(2): 1020-1030, DOI: 10.1002/mrm.26706.
698
699 Wells, W. M., 3rd, Viola, P., Atsumi, H., Nakajima, S. and Kikinis, R., 1996. Multi-
700 modal volume registration by maximization of mutual information. Med Image Anal 1(1):
701 35-51, DOI: 10.1016/s1361-8415(01)80004-9.

702
703 Workman, P., Aboagye, E. O., Balkwill, F., Balmain, A., Bruder, G., Chaplin, D. J.,
704 Double, J. A., Everitt, J., Farningham, D. A., Glennie, M. J., Kelland, L. R., Robinson,
705 V., Stratford, I. J., Tozer, G. M., Watson, S., Wedge, S. R., Eccles, S. A. and Committee
706 of the National Cancer Research, I., 2010. Guidelines for the welfare and use of animals
707 in cancer research. Br J Cancer 102(11): 1555-1577, DOI: 10.1038/sj.bjc.6605642.
708
709

710 **Online Supplemental Information**

711
712 To evaluate the errors resulting from a single-section assessment in this case where at least three
713 sections where required, Bland-Altman plots were used to compare tumour volume
714 measurements in single-section and 3-section SIH maps. These show significant biases between
715 the two assessment methods (A). The average bias was of 4.77 mm^3 more for histologic maps,
716 about 35% of the observed volume. Interestingly, the bias appear to increase with decreasing
717 tumour volume. This highlights the fact that single-section approaches are more likely to fail when
718 characterising small tumours.

719
720 As arbitrary section selection for imaging assessment is common in the literature, it is interesting
721 to consider two extreme single-section cases where the sections with the greatest and smallest
722 tumour volumes are arbitrarily selected (Vmax and Vmin respectively): While bias reaches 54% of
723 the observed volume for sections with the smallest areas, a significant bias is found even if
724 histologic maps are compared with the single-sections showing the greatest abnormal area (9%).

725
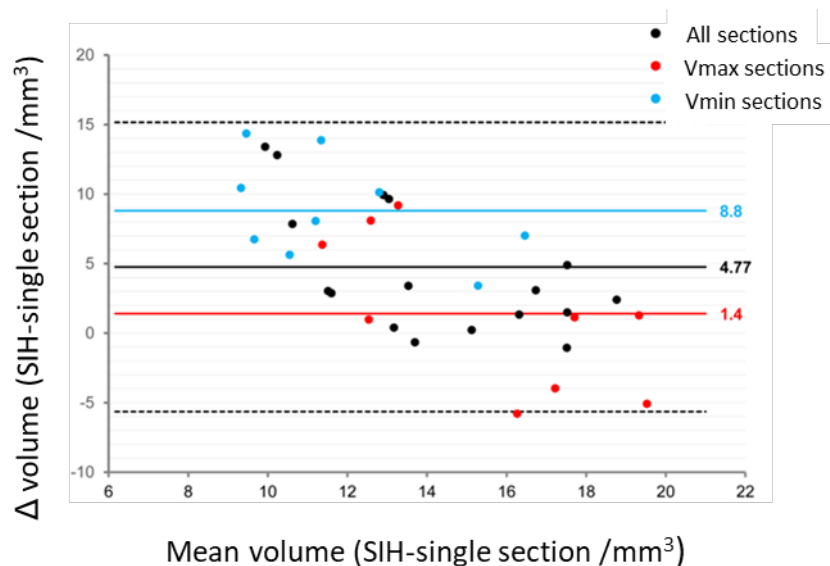
726
727
728 *Figure 1S. Bland–Altman plots for tumour volumes identified on single-section histology (SSH)*
729 *and stacked in-plane histology (SIH) maps using 3 sections.*

730
731
732

733
734
735

A
B
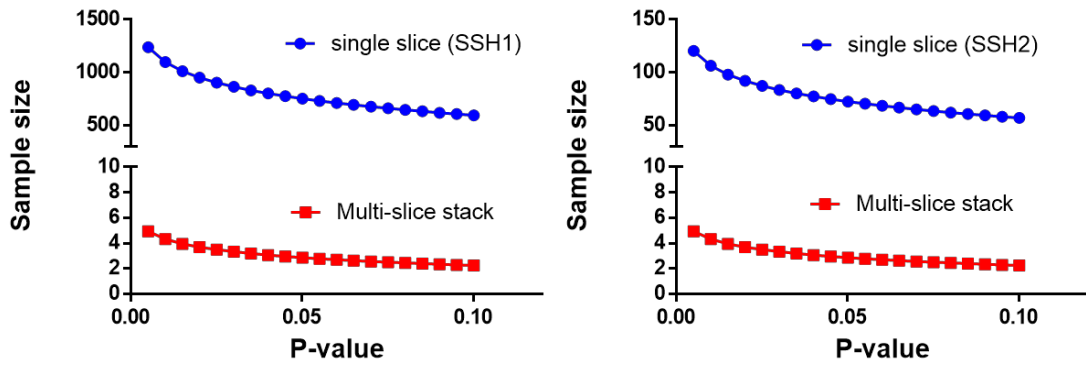
736
737
738   *Figure 2S: Power calculation of number of animals that would have been required for achieving a*
739   *given significance with the single-slice or the SIH approach for two different single slice groups*
740   *(A) section SSH1 (B) section SSH2.*
741
742