

Title: Combining images and anatomical knowledge to improve automated vein segmentation in MRI

Authors: Phillip G. D. Ward^{1,2,3}, Nicholas J. Ferris^{1,4}, Parnesh Raniga⁵, David L. Dowe², Amanda C. L. Ng⁶, David G. Barnes⁷, Gary F. Egan^{1,3}.

¹ Monash Biomedical Imaging, Monash University, Clayton VIC, Australia.

² Faculty of Information Technology, Clayton, Monash University, Clayton VIC, Australia.

³ ARC Centre of Excellence for Integrative Brain Function, Melbourne VIC, Australia.

⁴ Monash Imaging, Monash Health, Clayton VIC Australia.

⁵ The Australian eHealth Research Centre, CSIRO Health and Biosecurity, Herston, QLD, Australia.

⁶ Department of Anatomy and Neuroscience, The University of Melbourne, Melbourne, VIC, Australia.

⁷ Monash eResearch Centre, Monash University, Clayton VIC, Australia.

Corresponding author: Phillip G. D. Ward. phillipgdward@gmail.com

Word Count: 4232

Running title: Composite imaging for vein segmentation.

Abstract

Purpose: To improve the accuracy of automated vein segmentation by combining susceptibility-weighted images (SWI), quantitative susceptibility maps (QSM), and a vein atlas to produce a resultant image called a composite vein image (CV image).

Method: An atlas was constructed in common space from 1072 manually traced 2D-slices. The composite vein image was derived for each subject as a weighted sum of three inputs; a SWI image, a QSM image and the vein atlas. The weights for each input and each anatomical location, called template priors, were derived by assessing the accuracy of each input over an independent data set. The accuracy of venograms derived automatically from each of the CV image, SWI, and QSM image sets was assessed by comparison with manual tracings. Three different automated vein segmentation techniques were used, and ten performance metrics evaluated.

Results: Vein segmentations using the CV image were comprehensively better than those derived from SWI or QSM images (mean Cohen's $d = 1.1$). Sixty permutations of performance metric and automated segmentation technique were evaluated. Vein identification improvements that were both large and significant (Cohen's $d > 0.80$, $p < 0.05$) were found in 77% of the permutations, compared to no improvement in 5%.

Conclusion: The accuracy of automated venograms derived from the composite vein image was overwhelmingly superior to venograms derived from SWI or QSM alone.

Keywords: MRI, QSM, SWI, vein atlas, vein segmentation, cerebral vasculature.

1. Introduction:

Mapping cerebral veins using magnetic resonance (MR) images has until recently been technically challenging. Cerebral venograms are increasingly important for advancing our knowledge of cerebral vascularisation, oxygenation, metabolism and studies of cerebrovascular topology. The use of venograms in clinical research applications is growing rapidly, including for quantifying oxygen saturation (Fan et al., 2014), measuring the metabolic rate of oxygen consumption (Rodgers et al., 2016) and analyzing possible fMRI confounders (Vigneau-Roy et al., 2014).

Traditional vein imaging techniques require invasive contrast agents, have potential arterial confounds, and are limited to the large vessels, due to the reduced volume and flow of smaller cerebrovasculature segments. However, magnetic susceptibility provides an intrinsic contrast mechanism that is exquisitely sensitive to the presence of iron, particularly deoxygenated iron-rich haemoglobin proteins within red blood cells, making susceptibility techniques very useful for imaging small as well as large veins. The magnetic susceptibility of blood is modulated by oxygen (Pauling and Coryell, 1936), which facilitates the separation of arteries and veins, whilst providing a mechanism to quantify oxygen saturation (Fan et al., 2014).

Susceptibility-weighted imaging (SWI) and quantitative susceptibility mapping (QSM) are MR techniques based on magnetic susceptibility that provide a non-invasive method of imaging the cerebral veins. QSM and SWI derive contrast from gradient-recalled echo (GRE) phase information and have been applied to stroke, multiple sclerosis, cerebrovascular disease, and

examined in clinical and preclinical studies (Fan et al., 2015; Fujima et al., 2011; Goodwin et al., 2015; Jain et al., 2010; Li et al., 2013; Liu and Li, 2016; Rodgers et al., 2013; Santhosh et al., 2009). The way in which SWI and QSM process the phase information is very different.

SWI multiplies a non-linear mapping of high-pass filtered GRE phase with the GRE magnitude image, compounding the effects of signal cancellation from incoherent signals within each voxel and phase accumulation due to local sources of magnetic susceptibility (Haacke et al., 2004). Non-local sources are also included, such as the extravascular phase information, resulting in the magnification of small veins. The presentation of non-local sources, and non-linear mapping, generates a non-quantitative image best suited to radiological interpretation.

QSM estimates the magnetic susceptibility distribution directly by inverting the magnetic field information captured in the phase image (Marques and Bowtell, 2005; Salomir et al., 2003). Mathematically, QSM involves a linear system inversion that is ill-posed and requires regularization or fitting (Li et al., 2015; Liu et al., 2016; Wang and Liu, 2015; Wharton et al., 2010). QSM has the benefits of being quantitative and is designed to resolve extravascular field effects, leaving only local sources of magnetic susceptibility contrast.

The differing approaches (QSM and SWI) have unique image contrasts, and each have their own vein-like confounders. SWI images, for instance, do not distinguish between signal cancellation due to venous blood, and low concentrations of free protons (Haacke et al., 2004). The lack of distinction is problematic when analyzing veins which reside near non-vein low signal structures, such as in the vicinity of the tentorium and in the interhemispheric fissure (due to the falx

cerebri). Both SWI and QSM also suffer different artefacts, such as cruciform artefacts in QSM images.

As neither QSM nor SWI isolate blood signal intrinsically, unlike spin-labelling or contrast agent-based techniques, venous voxels within the brain must be identified before the veins can be analysed. The process of identifying venous voxels in the brain, or vein segmentation, produces a vein mask (or venogram) that can then be used to extract the vein signal from an image, or examined directly for topographic analysis.

A number of algorithms for automatic segmentation of blood vessels in the body have been proposed, including shape-driven, intensity-driven and hybrid approaches (Lesage et al., 2009). A common approach in the analysis of SWI and QSM data is to employ a preliminary filtering step, such as Hessian-based filtering (Frangi et al., 1998), before applying a simple threshold classification method (Vigneau-Roy et al., 2014). Recent work has combined Hessian-based filtering into a segmentation framework with diffusion techniques to overcome noise and low vein visibility (Bazin et al., 2016; Manniesing et al., 2006). Statistical modeling of spatial relations has also been proposed to improve continuity and smoothness in vein segmentation (Bériault et al., 2014; Ward et al., 2017b).

The previously mentioned work focused upon SWI or QSM, and did not attempt to extract information from both images. Methods have been proposed that merge SWI with QSM (Ward et al., 2015), and $R2^*$ maps (Monti et al., 2015). Both approaches were globally homogeneous, i.e., they combined two images without consideration for anatomical location. As SWI and QSM

have differing image contrasts, and artefacts that are specific to anatomy, it is possible a superior segmentation could be achieved if the method for combining the two images was sensitive to location.

Prior anatomical knowledge has recently been incorporated into a vein segmentation technique to reduce false positives in specific brain regions (Bériault et al., 2015). However, this approach was limited to specific deep-brain regions, it did not directly address boundaries between tissue types and neural structures, and it was hand-tuned.

There are two anatomical factors that contribute to vein segmentation accuracy. The first is vein anatomy, i.e., expected vein occurrence, size and shape at an anatomical location. The second is image contrast, i.e., expected tissue signal relative to vein signal, which is specific to SWI and QSM. In this study, these two factors are exploited to improve cerebrovenous contrast and subsequent vein segmentation accuracy. We propose a vein identification and segmentation method that is based on a locally varying combination of SWI and QSM contrast which is informed by known vein anatomy in specific neuroanatomical structures. The proposed method derives a single composite vein image (CV image) that incorporates the strengths of SWI and QSM, with the anatomical knowledge of a vein atlas.

The CV image is generated from three input images (SWI, QSM and atlas) that are combined using a weighted-sum. The weights are derived from template priors that capture the location-specific venous contrast of the three input images throughout the brain. Separate vein atlases and template priors were calculated for each subject within the study from an independent sample of

the cohort to ensure data independence. Future applications of the technique would use a single template prior and atlas calculated from the entire cohort. The CV image was compared to SWI and QSM images for the purpose of vein segmentation using automated techniques. Segmented CV images were compared with segmented SWI and QSM images using a broad array of accuracy measures and three automated segmentation techniques.

2. Methods:

All procedures were reviewed and approved by the local ethics committee. Informed consent was obtained from all volunteers. The code and data used in this study have been available to the public using github and figshare respectively (Ward et al., 2016a, 2017).

2.1 Data Acquisition

Ten healthy volunteers were scanned using a 3T Siemens Skyra with a 32-channel head and neck coil (6 females, mean age 56.2 years, standard deviation 25.2). The protocol was a single echo, flow-compensated, gradient-recalled echo (GRE) sequence (TE=20ms, TR=30ms, flip angle=15°, voxel=0.9×0.9×1.8mm anisotropic, matrix 256×232×72). Four of the subjects were acquired with a smaller voxel size (voxel=0.9mm isotropic, matrix=256×232×160). A T1-Weighted MPRAGE scan was also acquired and used for registration purposes (TE=2ms, TR=2300ms, TI=900ms, voxel=1.0mm isotropic, matrix=240×256×192, flip angle=9°). All registration was performed using FNIRT and FLIRT from the FSL toolkit (<https://fsl.fmrib.ox.ac.uk/>)(Jenkinson et al., 2012).

For all subjects, raw k-space data for the GRE acquisition was saved for each coil and retrospectively reconstructed to generate phase and magnitude images. Individual coil phase images were processed to remove phase wraps and background phase shifts using Laplacian unwrapping (Li et al., 2014) and V-SHARP (Wu et al., 2012). QSM maps were computed using LSQR in the STI-Suite (Li et al., 2014). The SWI images were taken directly from the scanner console.

Six of the ten subjects were healthy elderly subjects recruited for the ASPirin in Reducing Events in the Elderly (ASPREE) clinical trial (Group, 2013) and scanned at baseline as part of the ASPREE NEURO sub-study (S. A. Ward et al., 2016).

2.2 Manual vein tracing

A mask containing venous voxels was created for each subject by manually labelling voxels as vein or non-vein, using FSLView (Jenkinson et al., 2012). Tracing was performed by author PW under the supervision of author NF (a clinical radiologist). The venous voxels were initially identified based upon SWI contrast and prior anatomical knowledge, in transverse reconstructions of the 3D SWI acquisition, and refined by reference to sagittal and finally coronal reformats. Initial SWI-only masks were then overlaid side-by-side on SWI and QSM images for editing. Editing was performed slice by slice in the sagittal plane, followed by the axial and finally the coronal plane. An example ground-truth venogram is shown in Figure 1. In total 1072 transverse slices were examined.

*** Figure 1 appears near here ***

2.3 Composite vein image process

The composite vein (CV) image (\check{X}) was constructed by combining a vein atlas, with a SWI image and a QSM image. The three inputs are referred to as C , where $C = \{atlas = 1, SWI = 2, QSM = 3\}$, and are denoted X_c .

*** Figure 2 appears near here ***

The CV method combines the three inputs with weights based on the relative predictive power of each input in different regions of the brain. The relative predictive power was captured in a *template prior* (P_c) for each of the three inputs (Figure 2A). The inputs were normalized \check{X}_c , and then combined using a weighted-mean with the weights derived from the priors, P_c .

$$\check{X}(r) = \frac{1}{\sum_{c=1}^C P_c(r)} \cdot \sum_{c=1}^C (P_c(r) \cdot \check{X}_c(r)) \quad \text{Eq. 1}$$

The CV image, $\check{X}(r)$, for each subject used the atlas and priors pre-calculated from the training cohort (M) (Figure 2A). For each subject in this study, the training cohort consisted of the other nine subjects to ensure independence of the atlas and priors for each subject specific CV image.

The atlas construction, normalization process, and details of the template priors are explained in the following sub-sections.

2.4 Vein atlas

Manually traced vein masks (Y) were weighted to reflect the variance and uncertainty of the human observer (W).

$$W(r) = \begin{cases} 0.9, & Y(r) = vein \\ 0.1, & Y(r) \neq vein \end{cases} \quad \text{Eq. 2}$$

The weighted tracings (W_i) for each subject, i , were interpolated into 0.5mm MNI space (Montreal Neurological Institute standard brain atlas) and the average (mean) calculated for each voxel, r , to construct a vein atlas.

$$X_{Atlas}(r) = \frac{1}{|M|} \sum_{i=1}^M W_i(r) \quad \text{Eq. 3}$$

Visual inspection of the atlas was performed to explore the variability of vein location between subjects within the cohort.

2.5 Normalisation

The SWI and QSM images were processed separately to remove biases and to normalize their voxel intensities using a Gaussian mixture model (GMM) with two components (vein and non-vein). This approach has been explored previously (Ward et al., 2015, 2017b), and similar techniques have been used in blood vessel segmentation before (Bazin et al., 2016; Bériault et al., 2015, 2014; Hassouna et al., 2006).

For each image (SWI and QSM), a GMM was fit using a log-likelihood expectation-maximisation approach (Dempster et al., 1977). Both GMMs (one for SWI and one for QSM) used the same initial seed taken from the QSM images ($X_{QSM} > 0.05ppm$) to impart prior knowledge of the components. High-pass filtering was also applied to X_{SWI} .

The GMM process mapped the image contrasts to a unity space [0,1] and reduced the presence of subject specific biases. Normalized images (\dot{X}_{QSM} and \dot{X}_{SWI}) were produced from the mixture coefficient of the vein component for each voxel.

$$\dot{X}_{i,c} = \frac{\varpi_{c,V} Pr(V|X_{i,c}, \mu_{c,V}, \Sigma_{c,V})}{\varpi_{c,V} Pr(V|X_{i,c}, \mu_{c,V}, \Sigma_{c,V}) + \varpi_{c,\sim V} Pr(\sim V|X_{i,c}, \mu_{c,\sim V}, \Sigma_{c,\sim V})} \quad \text{Eq. 4}$$

where c was the image (SWI or QSM), i was the voxel, and $Pr(V|X_c, \mu_{c,V}, \Sigma_{c,V})$ was the posterior probability of being labelled vein (or not vein, $\sim V$) given the distribution parameters for the image-specific GMM: μ, Σ , relative abundance ϖ , and the voxel intensity, X . Supporting information on this process has been published previously (Ward, 2017).

The atlas (X_{Atlas}) is intrinsically normalized, and was interpolated into the subject space to provide the final input (\dot{X}_{Atlas}).

2.6 Template priors

Subject specific confidence maps, $p_{i,c}$, were calculated using log-loss scoring (Dowe, 2008; Good, 1952) for each input ($c \in C$) using Eq. 5. Voxel location (r) and subject (i) have been omitted from this equation for clarity.

$$p_c(\dot{X}_c, W) = -\log(W \cdot (1 - \dot{X}_c) + (1 - W) \cdot \dot{X}_c) \quad \text{Eq. 5}$$

The template priors (P_c) for SWI, QSM and the vein atlas were calculated by taking the cohort mean of the confidence maps interpolated into MNI space.

$$P_c(r) = \frac{1}{M} \sum_{i=1}^M p_{c,i}(\dot{X}_{c,i}(r), W_i(r)) \quad \text{Eq. 6}$$

The predictive power represented in the priors for each information source (P_{Atlas} , P_{SWI} , and P_{QSM}) was examined visually. In order to visualize all three values, each was normalized, $\dot{P}_c(r) = P_c(r) \cdot \min_d P_d^{-1}(r)$, and encoded in a colour channel of a colour image (blue, green, red for P_{Atlas} , P_{SWI} , and P_{QSM} respectively).

2.7 Performance evaluation

Vein contrast in the CV image was assessed in comparison to SWI and QSM images using automated vein segmentation techniques. Three vein segmentation techniques were employed to compute venograms for each of the three image sets (the CV images, SWI images and QSM

images). The first segmentation technique was a Hessian-based vesselness filter followed by an Otsu threshold (VN) (Frangi et al., 1998; Otsu, 1975). The second was a statistical method based on an Ising model Markov random field using an anisotropic graph (MRF) (Bériault et al., 2014). The third was a recursive ridge-based filter (RR) (Bazin et al., 2016).

The accuracy of the venograms from each image set was evaluated with standard metrics (Table 1). Many standard overlap metrics are not informative due to the high surface-to-area ratio of cerebral vein masks. To overcome this limitation, dilated versions of many of the metrics were used (Bazin et al., 2016). The metric values for the SWI and QSM image based venograms were calculated as benchmarks, and the differences between these reference values and the venograms computed from the CV image were examined.

Two techniques had parameters that required training (VN and MRF). A leave-one-out approach to training was taken, with the performance of the left-out subject recorded for comparison purposes. The parameters (θ) were optimized in a standardized space to minimize a composite cost function (Z), which included a regularization term (Λ).

$$Z = (1 - DSS) + \frac{MHD}{2} + \frac{AVD}{2} + \Lambda \quad \text{Eq. 7}$$

$$\Lambda = \sqrt{\sum_i^{N_p} (\theta_i - 0.5)^2} \quad \text{Eq. 8}$$

where N_p is the number of parameters to train. Dice similarity score (DSS) (Dice, 1945), mean Hausdorff distance (MHD) (Shonkwiler, 1989) and average volume difference (AVD) are

described in Table 1. Scaling factors were chosen to normalize metric ranges as DSS is constrained to a unity range, whilst MHD and AVD are unbounded.

A gradient descent algorithm was used to search the parameter space from an initial parameter estimate. The initial parameter estimate was the optimal parameter values from a uniform random sample of the entire parameter space (1000 samples). Each iteration of the search algorithm used 32 randomly sampled potential steps. Samples were taken from a hypercube with dimensions equal to 5% of the parameter space. In all cases the cost function was found to have converged before 50 iterations. All parameters are reported in Table 2. All operations were performed in MATLAB 2015b using the MASSIVE supercomputer (Goscinski et al., 2014).

Ten sets of parameters were trained for each technique (Table 2), using a different subset of nine subjects from the ten available. The mean and standard deviation of the results are shown in Table 2. The difference in cost function value between the mean of the training set (nine subjects) and the left-out validation subject was low. In the majority of cases the validation score was better than the worst individual score in the training set.

The difference in performance was assessed between the CV images and SWI images, and CV images and QSM images. The magnitude of performance difference was quantified using Cohen's d (Cohen, 2013) and interpreted on a qualitative scale (Sawilowsky, 2009). The significance of the difference was tested using a paired two-tailed Wilcoxon signed-rank test (Wilcoxon, 1945). A description of all metrics used can be found in Table 1.

3. Results:

All permutations of vein segmentation technique, performance metric and benchmark image (SWI or QSM) were explored, resulting in 60 comparisons. When using the CV image, 77% of the permutations showed a large or higher improvement (Cohen's $d > 0.80$) that was statistically significant ($p < 0.05$), compared to a negative effect in 5% of the permutations. The mean effect size across all permutations was 1.1 (large) and very large effects in favor of the CV image were found in 65% of comparisons. The results are shown pictographically in Figure 3.

*** Figure 3 appears near here ***

Negative results (5%) occurred in paired-metrics and were not observed in comprehensive balanced metrics, such as DSS, MCC and MHD, which overwhelmingly displayed performance improvements with the CV image. In paired-metrics, a corresponding positive effect was found for each negative effect in the metric pair, such as negative specificity and positive sensitivity for the QSM image when using the VN technique. In these paired cases the effect size was comparable for both positive and negative results.

Inconclusive results were most common for average-volume difference (AVD) (50% of comparisons). AVD measures bias in the errors of the final masks, rather than a direct measure of performance. Greater improvement was shown in comparison to the QSM based venograms relative to SWI based venograms.

*** Figure 4 appears near here ***

The similarity of vein locations between subjects was represented in the vein atlas (Figure 4). High atlas values indicated consistent vein locations between subjects. The highest value voxels were found near the major veins, including in the superior sagittal sinus, dural sinuses, straight sinus and internal cerebral veins (green regions in Figure 4). The deep gray matter structures, the inferior frontal and inferior temporal regions showed lower values.

*** Figure 5 appears near here ***

The relative predictive power of the atlas, SWI and QSM template priors was observed to be heterogeneous across brain regions (Figure 5). QSM was found to have comparatively higher power in the falx cerebri and lower power in the deep-gray matter structures, relative to SWI and the atlas. The atlas was highest in the deep-gray matter, particularly on the edge of structures. SWI had higher predictive power on the superior surface of the cortex, and lower power on the inferior surface of the brain.

4. Discussion

In this work a composite vein (CV) imaging technique was proposed that combined three sources of vein information, an atlas, a SWI image, and a QSM image. The CV image showed a large improvement in vein segmentation accuracy when compared with SWI and QSM images. A robust improvement was observed in the majority of permutations across ten performance metrics and three segmentation methods.

The CV image was found to combine the complementary strengths of SWI and QSM, and produce an image with significantly improved vein contrast by incorporating the relative predictive power of SWI and QSM in a weighted-average approach. A comprehensive analysis was performed using over 1000 MRI slices, including anisotropic and isotropic acquisitions, multiple automated segmentation techniques, and multiple performance metrics.

The template priors characterized the anatomically heterogeneous value of the three inputs. The atlas template prior (P_{Atlas}) had higher relative values to SWI and QSM in the deep gray matter structures possibly due to non-venous iron deposits in tissue. However, the relative predictive power of SWI increased towards the center of these structures. High-pass filtering in SWI may be the cause of this effect by reducing the low-frequency spatial contrast of these iron sources and increasing the sharpness of structure boundaries. A common trend observed in the larger veins, particularly those in the interhemispheric region, was higher SWI predictive power in the center and higher QSM predictive power at the vessel wall, extending into the surrounding tissue. Extravascular enhancement on SWI images may be the source of the decrease in SWI predictive power at the vessel wall. Two exceptions to the greater predictive power of SWI in the center of larger veins were the superior sagittal sinus and the transverse sinus, where the predictive power of the atlas was higher. Decreased anatomical variability and hyper-intense GRE signal due to imperfect flow compensation are possible causes of the reduction in SWI predictive power relative to the atlas.

There was minimal discrepancy between performance of the training and validation datasets. The tuned parameter values for each of the ten-independent datasets had low variance (Table 2) despite the stochastic nature of the initial parameter values. These two results indicate minimal over-fitting occurred when optimizing the technique parameters.

Comparative analysis of published automated vein segmentation techniques is difficult for a number of reasons. Manually traced ground truth vein masks typically cover small manually traced regions and/or minimum-intensity projections (Bazin et al., 2016; Bériault et al., 2014; Monti et al., 2015). The small regions may not be indicative of performance across the entire brain, and are a source of variability between studies that cannot be controlled for. Studies also use different performance metrics including sensitivity and specificity (Monti et al., 2015), positive-predictor value, negative predictor value and overlap (Bazin et al., 2016), and accuracy (Bériault et al., 2014). The use of different metrics may be due to the specific application that each technique has been designed for. However, the selective use of metrics can result in one-sided conclusions being drawn, and can frustrate meta-analysis efforts.

The quantification of vein segmentation performance is a contentious issue (Gerig et al., 2001). To capture compensatory behavior, where one metric is optimized at the expense of another, metrics are often reported in quasi-orthogonal pairs, such as specificity and sensitivity. Although, when one label (vein) is less numerous in abundance than the other (non-vein) the trade-off between pairs will not be even due to the disparate magnitude of the denominators. Non-paired metrics, such as overlap and volume difference metrics, are more robust in these scenarios, albeit

at the expense of interpretability. Comprehensive reporting of multiple metrics should be adopted to enhance both interpretability and transparency.

The template priors, vein atlas, and manual vein tracings are publicly available (Ward et al., 2016a). The manual vein tracing required hundreds of hours to complete, and the release of the segmented data may facilitate future work in vein segmentation techniques. The data sharing may result in a large, publicly available set of cross-validated ground truth vein images for future collaborative studies.

4.1 Limitations

A distinction should be considered between accurate vein segmentation and accurate imitation of manually traced vein masks. The tracings are not a direct measure of veins, but a subjective radiological interpretation (Drew et al., 2013), and are produced in the presence of the artifacts that occur in susceptibility-based MRI. A ground truth that is independent of MRI-based artifacts would be required to directly quantify vein segmentation accuracy.

Furthermore, only binary masks have been addressed in this work. However, vein geometry does not conform to a cubic grid. Recent work has found significant error associated with binary representations of veins in simulated QSM images (Ward et al., 2017a). Further work will explore non-binary manual tracings to incorporate both partial volume and marker uncertainty.

5. Conclusions

A large improvement in venogram accuracy was achieved using the composite vein image technique. The composite vein image technique incorporates the heterogeneous vein contrast profile across the brain to extract the complementary information available from SWI and QSM images, and a vein atlas. The technique's performance was evaluated with multiple segmentation techniques and metrics. The accuracy provided by the composite vein image allows improved quantification of cerebrovenous topology and cerebrovascular oxygenation using MRI.

Acknowledgements:

We thank the investigators of the ASPREE study and the ASPREE-NEURO sub-study for the coordination and recruitment of the volunteers. We thank the volunteers for their time and willingness to participate.

The Alzheimer's Australia Dementia Research Foundation (AADRF), the Victorian Life Sciences Computation Initiative (VLSCI), the Multi-model Australian ScienceS Imaging and Visualisation Environment (MASSIVE) and the National Health and Medical Research Council (NHMRC) supported this work (NHMRC Grant APP1086188).

We would like to thank Dr. Pierre-Louis Bazin and Dr. Silvain Bériault for their assistance in implementing and applying their vein segmentation methods.

Initial work and preliminary findings related to this study were previously presented at annual meetings of the International Society for Magnetic Resonance in Medicine (Ward et al., 2015, 2016b, 2017d).

References:

- Bazin, P.L., Plessis, V., Fan, A.P., Villringer, A., Gauthier, C.J., 2016. Vessel segmentation from quantitative susceptibility maps for local oxygenation venography, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). Presented at the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1135–1138. doi:10.1109/ISBI.2016.7493466
- Bériault, S., Archambault-Wallenburg, M., Sadikot, A., Louis Collins, D., Bruce Pike, G., 2014. Automatic Markov Random Field Segmentation of Susceptibility-Weighted MR Venography, in: Clinical Image-Based Procedures. Translational Research in Medical Imaging, Lecture Notes in Computer Science. pp. 39–47.
- Bériault, S., Xiao, Y., Collins, D.L., Pike, G.B., 2015. Automatic SWI Venography Segmentation Using Conditional Random Fields. *IEEE Trans. Med. Imaging* 34, 2478–2491. doi:10.1109/TMI.2015.2442236
- Cohen, J., 2013. *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Ser. B Methodol.* 39, 1–38.
- Dice, L.R., 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology* 26, 297–302. doi:10.2307/1932409
- Dowe, D.L., 2008. Foreword re C. S. Wallace. *Computer J.* 51, 523–560. doi:10.1093/comjnl/bxm117
- Drew, T., Vo, M.L.H., Wolfe, J.M., 2013. “The invisible gorilla strikes again: Sustained inattentive blindness in expert observers.” *Psychol. Sci.* 24, 1848–1853. doi:10.1177/0956797613479386
- Fan, A.P., Bilgic, B., Gagnon, L., Witzel, T., Bhat, H., Rosen, B.R., Adalsteinsson, E., 2014. Quantitative oxygenation venography from MRI phase. *Magn. Reson. Med.* 72, 149–159. doi:10.1002/mrm.24918
- Fan, A.P., Govindarajan, S.T., Kinkel, R.P., Madigan, N.K., Nielsen, A.S., Benner, T., Tinelli, E., Rosen, B.R., Adalsteinsson, E., Mainero, C., 2015. Quantitative oxygen extraction fraction from 7-Tesla MRI phase: reproducibility and application in multiple sclerosis. *J. Cereb. Blood Flow Metab.* 35, 131–139. doi:10.1038/jcbfm.2014.187
- Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering, in: Wells, W.M., Colchester, A., Delp, S. (Eds.), *Medical Image Computing and Computer-Assisted Intervention — MICCAI’98*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 130–137. doi:10.1007/BFb0056195
- Fujima, N., Kudo, K., Terae, S., Ishizaka, K., Yazu, R., Zaito, Y., Tha, K.K., Yoshida, D., Tsukahara, A., Haacke, M.E., Sasaki, M., Shirato, H., 2011. Non-invasive measurement of oxygen saturation in the spinal vein using SWI: Quantitative evaluation under

- conditions of physiological and caffeine load. *Neuroimage* 54, 344–349. doi:10.1016/j.neuroimage.2010.08.020
- Gerig, G., Jomier, M., Chakos, M., 2001. Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation, in: Niessen, W.J., Viergever, M.A. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 516–523. doi:10.1007/3-540-45468-3_62
- Good, I.J., 1952. Rational Decisions. *J. R. Stat. Soc. Ser. B Methodol.* 14, 107–114.
- Goodwin, J.A., Kudo, K., Shinohe, Y., Higuchi, S., Uwano, I., Yamashita, F., Sasaki, M., 2015. Susceptibility-Weighted Phase Imaging and Oxygen Extraction Fraction Measurement during Sedation and Sedation Recovery using 7T MRI. *J. Neuroimaging* 25, 575–581. doi:10.1111/jon.12192
- Goscinski, W.J., McIntosh, P., Felzmann, U.C., Maksimenko, A., Hall, C.J., Gureyev, T., Thompson, D., Janke, A., Galloway, G., Killeen, N.E.B., Raniga, P., Kaluza, O., Ng, A., Poudel, G., Barnes, D., Nguyen, T., Bonnington, P., Egan, G.F., 2014. The multi-modal Australian ScienceS Imaging and Visualization Environment (MASSIVE) high performance computing infrastructure: applications in neuroscience and neuroinformatics research. *Front. Neuroinformatics* 8, 30. doi:10.3389/fninf.2014.00030
- Group, A.I., 2013. Study design of Aspirin in Reducing Events in the Elderly (ASPREE): A randomized, controlled trial. *Contemp. Clin. Trials* 36, 555–564.
- Haacke, E.M., Cheng, Y.-C.N., Reichenbach, J.R., 2004. Susceptibility weighted imaging (SWI). *Magn Reson Med* 52, 612–8. doi:10.1002/mrm.20198
- Hassouna, M.S., Farag, A.A., Hushek, S., Moriarty, T., 2006. Cerebrovascular segmentation from TOF using stochastic models. *Med. Image Anal.* 10, 2–18. doi:10.1016/j.media.2004.11.009
- Jain, V., Langham, M.C., Wehrli, F.W., 2010. MRI estimation of global brain oxygen consumption rate. *J. Cereb. Blood Flow Metab.* 30, 1598–1607. doi:10.1038/jcbfm.2010.49
- Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. *NeuroImage, 20 YEARS OF fMRI* 20 YEARS OF fMRI 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015
- Lesage, D., Angelini, E.D., Bloch, I., Funke-Lea, G., 2009. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Med. Image Anal., Includes Special Section on Computational Biomechanics for Medicine* 13, 819–845. doi:10.1016/j.media.2009.07.011
- Li, M., Hu, J., Miao, Y., Shen, H., Tao, D., Yang, Z., Li, Q., Xuan, S.Y., Raza, W., Alzubaidi, S., Haacke, E.M., 2013. In Vivo Measurement of Oxygenation Changes after Stroke Using Susceptibility Weighted Imaging Filtered Phase Data. *Plos One* 8, e63013. doi:10.1371/journal.pone.0063013
- Li, W., Avram, A.V., Wu, B., Xiao, X., Liu, C., 2014. Integrated Laplacian-based phase unwrapping and background phase removal for quantitative susceptibility mapping. *NMR Biomed.* 27, 219–227. doi:10.1002/nbm.3056
- Li, W., Wang, N., Yu, F., Han, H., Cao, W., Romero, R., Tantiwongkosi, B., Duong, T.Q., Liu, C., 2015. A method for estimating and removing streaking artifacts in quantitative susceptibility mapping. *NeuroImage* 108, 111–122. doi:10.1016/j.neuroimage.2014.12.043

- Liu, Z., Kee, Y., Zhou, D., Wang, Y., Spincemaille, P., 2016. Preconditioned total field inversion (TFI) method for quantitative susceptibility mapping. *Magn. Reson. Med.* n/a-n/a. doi:10.1002/mrm.26331
- Liu, Z., Li, Y., 2016. Cortical Cerebral Blood Flow, Oxygen Extraction Fraction, and Metabolic Rate in Patients with Middle Cerebral Artery Stenosis or Acute Stroke. *Am. J. Neuroradiol.* 37, 607–614. doi:10.3174/ajnr.A4624
- Manniesing, R., Viergever, M.A., Niessen, W.J., 2006. Vessel enhancing diffusion: A scale space representation of vessel structures. *Med. Image Anal.* 10, 815–825. doi:10.1016/j.media.2006.06.003
- Marques, J.P., Bowtell, R., 2005. Application of a Fourier-based method for rapid calculation of field inhomogeneity due to spatial variation of magnetic susceptibility. *Concepts Magn. Reson. Part B Magn. Reson. Eng.* 25B, 65–78. doi:10.1002/cmr.b.20034
- Matthews, B.W., 1975. Matthews, B.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405(2), 442–451. *Biochim. Biophys. Acta* 405, 442–51. doi:10.1016/0005-2795(75)90109-9
- Monti, S., Palma, G., Borrelli, P., Tedeschi, E., Cocozza, S., Salvatore, M., Mancini, M., 2015. A multiparametric and multiscale approach to automated segmentation of brain veins, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3041–3044. doi:10.1109/EMBC.2015.7319033
- Otsu, N., 1975. A threshold selection method from gray-level histograms. *Automatica* 11, 23–27.
- Pauling, L., Coryell, C.D., 1936. The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proc. Natl. Acad. Sci.* 22, 210–216. doi:10.1073/pnas.22.4.210
- Rodgers, Z.B., Detre, J.A., Wehrli, F.W., 2016. MRI-based methods for quantification of the cerebral metabolic rate of oxygen. *J. Cereb. Blood Flow Metab.* 0271678X16643090. doi:10.1177/0271678X16643090
- Rodgers, Z.B., Jain, V., Englund, E.K., Langham, M.C., Wehrli, F.W., 2013. High temporal resolution MRI quantification of global cerebral metabolic rate of oxygen consumption in response to apneic challenge. *J. Cereb. Blood Flow Metab.* 33, 1514–1522. doi:10.1038/jcbfm.2013.110
- Salomir, R., de Senneville, B.D., Moonen, C.T., 2003. A fast calculation method for magnetic field inhomogeneity due to an arbitrary distribution of bulk susceptibility. *Concepts Magn. Reson. Part B Magn. Reson. Eng.* 19B, 26–34. doi:10.1002/cmr.b.10083
- Santhosh, K., Kesavadas, C., Thomas, B., Gupta, A.K., Thamburaj, K., Kapilamoorthy, T.R., 2009. Susceptibility weighted imaging: a new tool in magnetic resonance imaging of stroke. *Clin Radiol* 64, 74–83. doi:10.1016/j.crad.2008.04.022
- Sawilowsky, S., 2009. New Effect Size Rules of Thumb. *Theor. Behav. Found. Educ. Fac. Publ.*
- Shonkwiler, R., 1989. An image algorithm for computing the Hausdorff distance efficiently in linear time. *Inf. Process. Lett.* 30, 87–89. doi:10.1016/0020-0190(89)90114-2
- Vigneau-Roy, N., Bernier, M., Descoteaux, M., Whittingstall, K., 2014. Regional variations in vascular density correlate with resting-state and task-evoked blood oxygen level-dependent signal amplitude. *Hum. Brain Mapp.* 35, 1906–1920. doi:10.1002/hbm.22301

- Wang, Y., Liu, T., 2015. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. *Magn. Reson. Med.* 73, 82–101. doi:10.1002/mrm.25358
- Ward, P.G.D., 2017. Composite vein imaging for susceptibility-based magnetic resonance imaging [WWW Document]. URL <https://github.com/philgd/CVI-MRI>
- Ward, P.G.D., 2017. Automated characterization of cerebral veins (PhD Thesis). Monash University, Clayton, Australia. doi:10.4225/03/58d09aaa98b87
- Ward, P.G.D., Fan, A.P., Raniga, P., Barnes, D.G., Dowe, D.L., Ng, A.C.L., Egan, G.F., 2017a. Improved quantification of cerebral vein oxygenation using partial volume correction. *Front. Neurosci.* 11. doi:10.3389/fnins.2017.00089
- Ward, P.G.D., Ferris, N.J., Ng, A.C.L., Barnes, D.G., Dowe, D.L., Egan, G.F., Raniga, P., 2015. Venous segmentation using Gaussian mixture models and Markov random fields, in: *Proceedings of the 23rd Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Presented at the ISMRM, Toronto, Canada, p. 3759. doi:10.13140/RG.2.2.16479.87202
- Ward, P.G.D., Ferris, N.J., Raniga, P., Dowe, D.L., Ng, A.C.L., Barnes, D.G., Egan, G.F., 2016a. Combined magnetic susceptibility contrast for vein segmentation from a single MRI acquisition using a vein frequency atlas. *figshare*. doi:10.4225/03/57B6B3D884B80
- Ward, P.G.D., Ferris, N.J., Raniga, P., Ng, A.C.L., Barnes, D.G., Dowe, D.L., Egan, G.F., 2017b. Vein segmentation using shape-based Markov Random Fields, in: *Proceedings of the 14th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. Presented at the The 14th IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, Melbourne, Australia, pp. 1133–1136.
- Ward, P.G.D., Ferris, N.J., Raniga, P., Ng, A.C.L., Dowe, D.L., Barnes, D.G., Egan, G.F., 2017c. Regional variations in cerebral venous contrast using susceptibility-based MRI, in: *Proceedings of the 25th Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Presented at the ISMRM, Honolulu, USA, p. 1450.
- Ward, P.G.D., Raniga, P., Ferris, N.J., Ng, A.C.L., Barnes, D.G., Dowe, D.L., Storey, E., Woods, R.L., Egan, G.F., 2016b. Venous metrics in a large cohort of healthy elderly individuals from susceptibility-weighted images and quantitative susceptibility maps, in: *Proceedings of the 24th Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Singapore, p. 3795. doi:10.13140/RG.2.2.19835.31529
- Ward, S.A., Raniga, P., Ferris, N.J., Woods, R.L., Storey, E., Bailey, M.J., Brodtmann, A., Yates, P.A., Donnan, G.A., Trevaks, R.E., Wolfe, R., Egan, G.F., McNeil, J.J., (Authorship), (Joint Senior, Group, on behalf of the A. investigator, 2016. ASPREE-NEURO study protocol: A randomized controlled trial to determine the effect of low-dose aspirin on cerebral microbleeds, white matter hyperintensities, cognition, and stroke in the healthy elderly. *Int. J. Stroke* 1747493016669848. doi:10.1177/1747493016669848
- Wharton, S., Schäfer, A., Bowtell, R., 2010. Susceptibility mapping in the human brain using threshold-based k-space division. *Magn. Reson. Med.* 63, 1292–1304. doi:10.1002/mrm.22334
- Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biom. Bull.* 1, 80–83.
- Wu, B., Li, W., Guidon, A., Liu, C., 2012. Whole brain susceptibility mapping using compressed sensing. *Magn. Reson. Med.* 67, 137–147. doi:10.1002/mrm.23000

Table 1 Venogram accuracy metrics. The metrics compare the manual mask of veins (V) and non-veins (N) with an automated estimate (V' and N'). A mask morphologically dilated by one-voxel is preceded by a δ , e.g. δV is the dilated manual vein mask. Dilation was used as the boundaries of small vein may be uncertain due to bloom and partial volume. The Hausdorff distance (41), $\mathfrak{D}(X, Y)$, is the mean value in a minimum distance map, i.e., the minimum distance from each surface voxel in mask X to a surface voxel in mask Y . The surface is defined as all voxels removed by a one-voxel morphological erosion.

Name	Equation
Number of true-positives (veins)	$TP = V \cap V' $ $\delta TP = \frac{1}{2} (\delta V \cap V' + V \cap \delta V')$
Number of true-negatives (non-veins)	$TN = N \cap N' $ $\delta TN = \frac{1}{2} (\delta N \cap N' + N \cap \delta N')$
Number of false-positives	$FP = N \cap V' $
Number of false-negatives	$FN = V \cap N' $
Accuracy	$ACC = \frac{\delta TP + \delta TN}{ V \cup N }$
Sensitivity	$SE = \frac{ \delta V' \cap V }{ V }$
Specificity	$SP = \frac{ \delta N' \cap N }{ N }$
Positive-predictor value	$PPV = \frac{ \delta V \cap V' }{ V' }$
Negative-predictor value	$NPV = \frac{ \delta N \cap N' }{ N' }$
Dice similarity score (Dice, 1945)	$DSS = \frac{2 \cdot \delta TP}{ V + V' }$
Matthews correlation coefficient (Matthews, 1975)	$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
Mean Hausdorff distance (Shonkwiler, 1989)	$MHD = \frac{1}{2} (\mathfrak{D}(V, V') + \mathfrak{D}(V', V))$
Average volume difference	$AVD = \frac{ FP - FN }{ V }$

Table 2. Tuned parameter descriptions for each automated segmentation technique. The trained value represents the mean and standard deviation across the 10 parameter sets trained.

Technique	Parameter	Description	Trained value		
			CV	SWI-Only	QSM-Only
MRF	β_{ratio}	Weighting factor between vein and non-vein in neighbours in clique potential calculation.	0.10 ± 0.02	0.58 ± 0.02	0.48 ± 0.01
VN	α	Non-plane like factor	0.24 ± 0.01	0.19 ± 0.02	0.61 ± 0.13
	β	Non-blob like factor	0.77 ± 0.09	0.34 ± 0.21	0.16 ± 0.02
	γ	Intensity factor	1.61 ± 0.08	3.89 ± 0.22	7.65 ± 1.23
	scale	Maximum scale explored	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01
RR	Recommended parameters specified by creator suitable for contrasts used (none trained).				

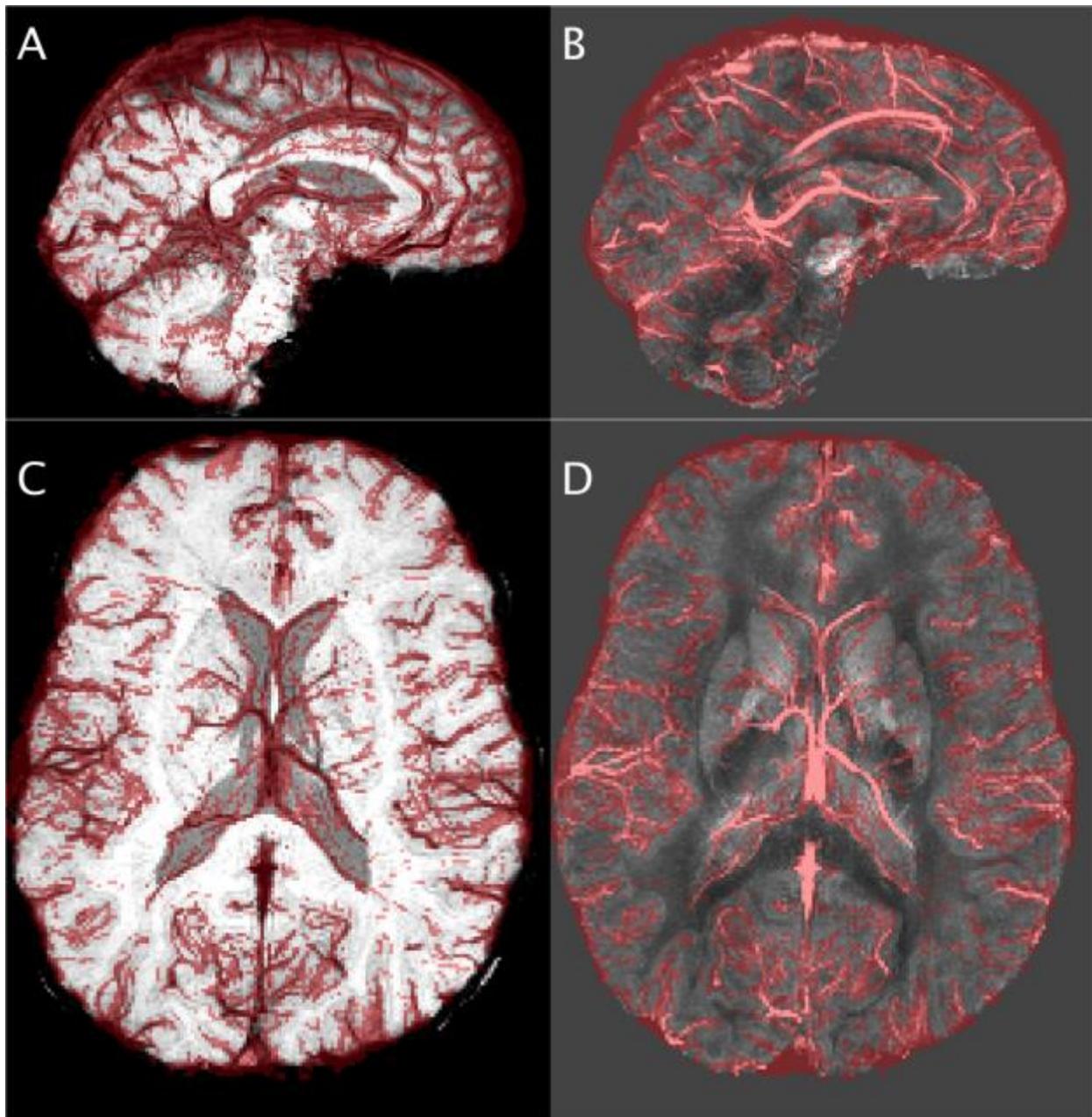


Figure 1. Example images of the manual vein tracings. Minimum-intensity projections for SWI (A and C) and maximum-intensity projections for QSM (B and D), with a semi-transparent vein mask overlay in red. Sagittal (9mm slab, A and B) and axial (18mm slab, C and D) slices are shown from a single subject.

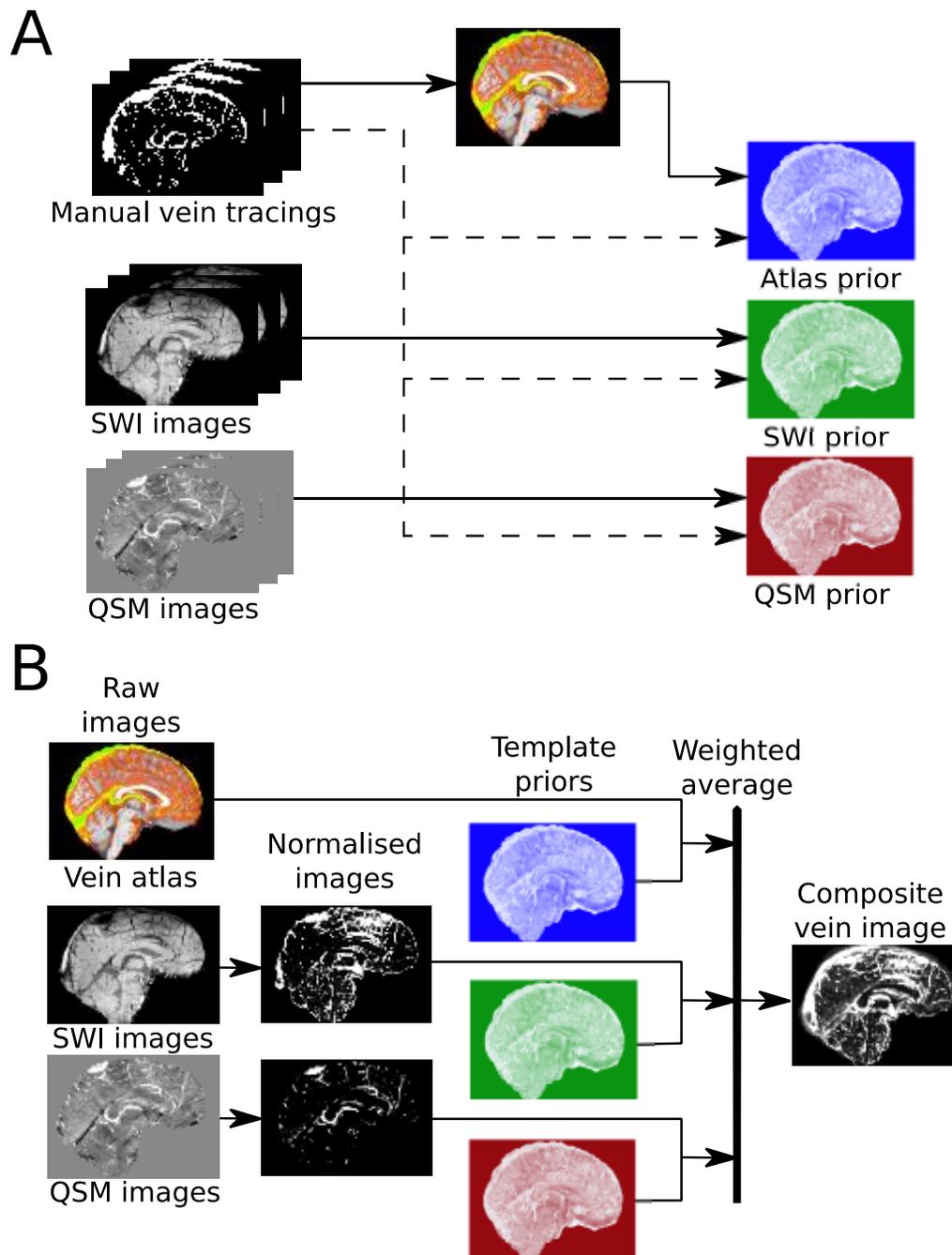


Figure 2. Schematic describing data flow and process for training the priors and atlas (A) and producing a composite vein image (B). The training data sets (manual vein tracings, SWI images, QSM images) in (A) are used to calculate the vein atlas and template priors. Operations in (A) occur within MNI template space. The vein atlas, atlas prior, SWI prior and QSM prior in (A) are inputs to (B) once they are interpolated into subject space. The training data set in (A) does not include the subject images in (B) to ensure independence.

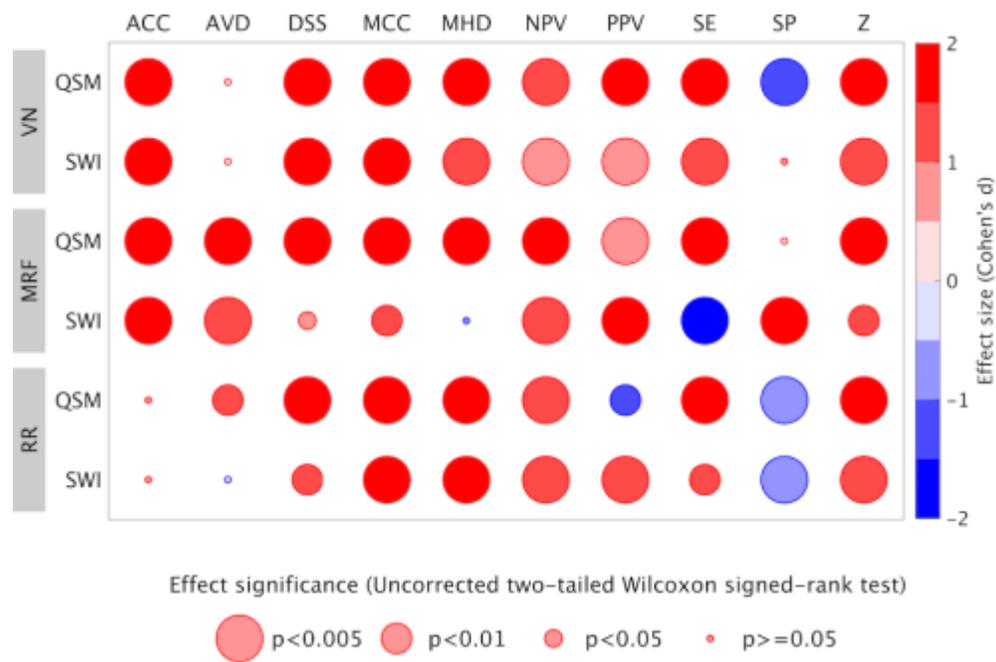


Figure 3. The improvement in performance of three automated segmentation techniques when using the CV image compared to two alternative images. The size of the improvement (Cohen's d) and the statistical significance of the improvement (p-values) are displayed in colour and size respectively. Red circles indicate superior performance using the CV image. Large circles indicate more significant results. Each row corresponds to an automated segmentation technique (RR, MRF, or VN) and input image (SWI or QSM). Each column denotes a different performance metric. All metric abbreviations are provided in Table 1. Statistical significance was measured using a two-tailed Wilcoxon signed-rank test and is uncorrected.

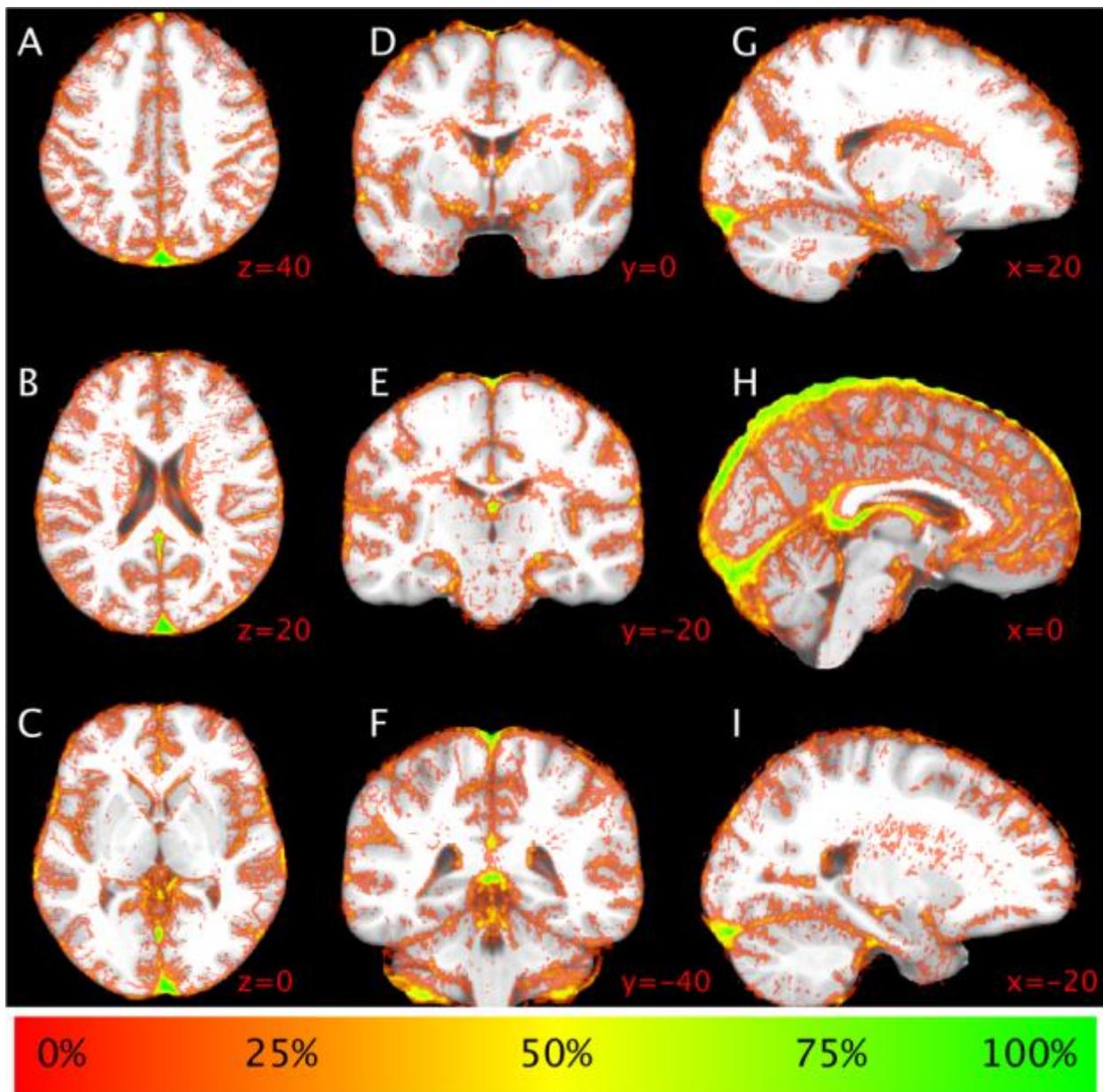


Figure 4. Vein frequency atlas (X_{Atlas}) demonstrating high reproducibility in vein location across subjects in the major veins (e.g. sagittal sinus, green/yellow) and low reproducibility in the deep grey matter structures. Values of 0% are transparent. Slice coordinates (red) are in MNI atlas space, with axial projections (A-C), coronal projections (D-F) and sagittal projections (G-I).

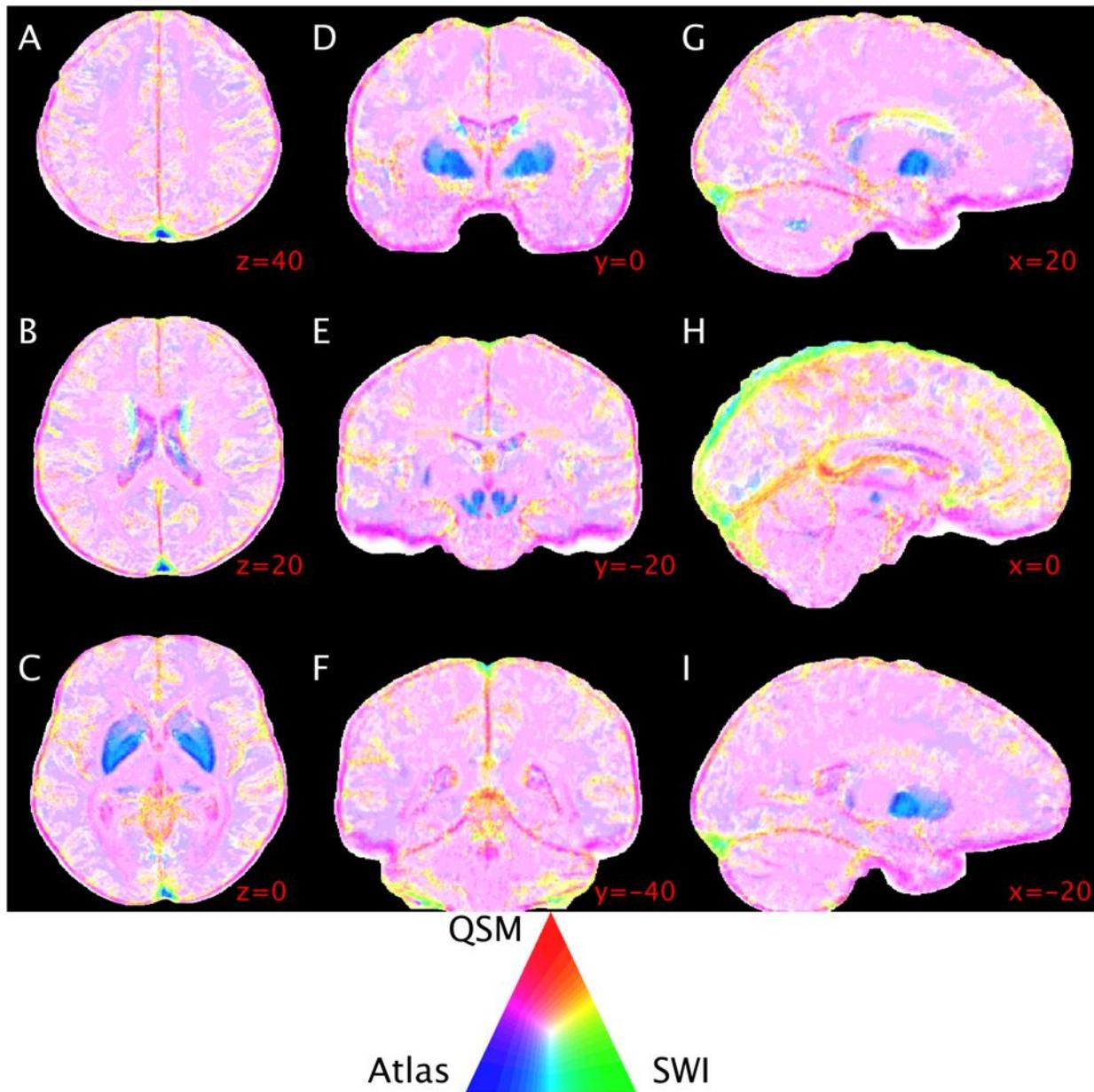


Figure 5. Slices from the atlas, SWI and QSM template priors (P_{Atlas} , P_{SWI} , and P_{QSM} respectively), color-coded to represent the relative weights of each. Blue regions show where P_{Atlas} is highest, green regions where P_{SWI} is highest and red regions where P_{QSM} is highest. A triangular colour-map is included. Slice coordinates (red) are in MNI atlas space, with axial projections (A-C), coronal projections (D-F) and sagittal projections (G-I).