Virtual Mouse Brain Histology from Multi-contrast MRI via Deep Learning

Zifei Liang[1], Choong H. Lee[1], Tanzil. M. Arefin[1], Zijun. Dong[1], Piotr Walczak[2], Song-Hai Shi[3], Florian Knoll[1], Yulin Ge[1], Leslie Ying[4], Jiangyang Zhang[1*]

[1] Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY 10016, USA

[2] Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland, Baltimore, MD

[3] Developmental Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA;

[4] Departments of Biomedical Engineering, Electrical Engineering, University at Buffalo, the State University of New York, Buffalo, NY, United States

Corresponding author:

Jiangyang Zhang, Ph.D.
Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine,
660 First Ave.
New York, NY 10016, USA
Jiangyang.zhang@nyulangone.org

Abstract

^{1}H MRI maps brain structure and function non-invasively through versatile contrasts that exploit inhomogeneity in tissue micro-environments. Inferring histopathological information from MRI findings, however, remains challenging due to absence of direct links between MRI signals and cellular structures. Here, we show that deep convolutional neural networks, developed using co-registered multi-contrast MRI and histological data of the mouse brain, can estimate histological staining intensity directly from MRI signals at each pixel. The results provide three-dimensional maps of axons and myelin with tissue contrasts that closely mimics target histology and enhanced sensitivity and specificity compared to conventional MRI markers. Furthermore, the relative contribution of each MRI contrast within the networks can be used to optimize multi-contrast MRI acquisition. We anticipate our method to be a starting point for translation of MRI results into easy-to-understand virtual histology for neurobiologists and provide resources for validating novel MRI techniques.

Introduction:

Magnetic resonance imaging (MRI) is one of a few techniques that can image the brain non-invasively and without ionizing radiation, and this advantage is further augmented by a large collection of versatile tissue contrasts. While MRI provides unparalleled insight into brain structures and functions at the macroscopic level (*1*), inferring the spatial organization of microscopic structures (e.g. axons and myelin) and their integrity from MR signals remains a challenging inverse problem. Without a thorough understanding of the link between MR signals and specific cellular structures, uncertainty often arises when determining the exact pathological events and their severities. The lack of specificity hinders direct translation of MRI findings into histopathology and limits its diagnostic value.

Tremendous efforts have been devoted to developing new mechanisms to amplify the affinity of MRI signals to target cellular structures in order to improve sensitivity and specificity. Recent progress in multi-modal MRI promises enhanced specificity by integrating multiple MR contrasts that target distinct aspects of a cellular structure (*2*). For example, magnetization transfer (MT), T_2, and diffusion MRI are sensitive to the physical and chemical compositions of myelin, and combining them can lead to more specific myelin measurements than individual contrast (*3*). Progress in this front, however, has been hindered by the lack of realistic tissue models for inference and ground truth histological data for validation.

The objective of this study is to test whether deep convolutional neural networks (CNNs), developed using co-registered histology and MRI data, can bypass the above-mentioned obstacles and enhance our ability to map key cellular structures from MR signals. With its capability to bridge data acquired with different modalities (*4-6*), the deep learning framework (*7*) has clear advantages over existing modeling approaches, as it is data-driven and not limited by particular models and associated assumptions. As MR signals are the ensemble average of all spins within each pixel, a typical set of three-dimensional (3D) MRI data with millions of pixels, thus provides ample instances to train deep CNNs. Through training, the networks can potentially reconstruct the link between MR signals and cellular structures in co-registered histology and translate multi-contrast MRI data into maps that mimic histology. Our results demonstrate that this approach offers enhanced specificity for detecting axons and myelin compared to existing MRI-

based markers. Furthermore, careful examinations of the networks allow us to measure the relative contribution of individual MR contrast, which can be used to optimize multi-contrast MRI strategy and evaluate novel imaging contrasts.

Results:

Prediction of auto-fluorescence images of the mouse brain from MR images using deep learning

We first demonstrated our method using co-registered three-dimensional (3D) MRI and auto-fluorescence (AF) data. MRI dataset from *ex vivo* C57BL/6 mouse brain (P60, n=6), each contained 67 3D MR (T_2, MT, and diffusion) images, were spatially normalized to the Allen Reference Atlas (ARA) (*8*) (**Fig. 1A**). We then selected 100 AF datasets from the Allen Mouse Brain Connectivity Atlas (AMBCA) (*9*) with minimal amounts of tracer signals in the forebrain. The contrast in the AF data is not specific to a particular structure, but a majority of hypo-intense regions co-localized with myelinated white matter tracts (*10*). These 3D AF data had already been normalized to the ARA and were down-sampled to the resolution of the MRI data (0.06 mm isotropic). Mismatches between the MRI and AF images were mostly within one to two pixels (**Supplementary Fig. 1A-B**).

A deep CNN, named MRH-AF, was trained using multiple 3x3 patches from the forebrain region of each MRI data (40,000 patches, N=6) as inputs and their corresponding patches in the co-registered AF data as targets (**Fig. 1B**) (Details on the network and training can be found in the Methods section). In order to determine the amount of training data sufficient to capture the relationship between these two modalities, we performed separate training sessions with target AF data ranging from randomly selected 60 subjects (i), 6 subjects (ii), single subject data (iii), down to 5,000 and 1,000 3x3 patches randomly selected within a single subject data (iv & v) (**Fig. 1C**). The 3x3 patch size was shown to accommodate residual mismatches between MRI and AF data (**Supplementary Fig. 1F-H**), and we chose such as small patch size instead of the entire image for training because we aim to define the local relationship between cellular structures within an MRI pixel and corresponding ensemble-averaged MRI signals.

The performance of MRH-AF was evaluated using the average 3D AF data in the ARA (CCF version 3, average of 1,675 mouse brains)(*9*) as the reference and MRI data from

a separate group of mice (P60, n=4) as the inputs. The MRH-AF results trained with 60-subject AF data as training targets (i) showed good agreement with the reference (**Fig. 1D**) and strong voxel-wise signal correlation (R^2=0.73, p<0.001, **Supplementary Fig. 3A**). The agreement was maintained for (ii) and (iii) both visually and quantitatively, as measured by the root mean square errors (RMSE) and structural similarity index (SSIM) (**Fig. 1D-E**). The specificity to hypo-intense regions in the reference defined by optimal thresholding was evaluated using receiver operating characteristic (ROC) analysis. The MRH-AFs trained with 60 and 6-subject AF data (i and ii) showed high specificity with areas under curve (AUCs) greater than 0.94, and the MRH-AF trained with 1-subject data (iii) had a slightly reduced average AUC of 0.937 (**Fig. 1F**). The variation in the ROC curves in (iii), caused by the inter-subject variations in AF signals among subjects chosen for training, was relatively small. Further reducing the size of training data (iv and v) resulted in declined performances (**Fig. 1D-F)**, emphasizing the need for sufficient training data.

The way that MRH-AF in (iii) translated individual 3x3 MR patches into AF signals was visualized in a 2D feature space derived by *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) analysis(*11*) (**Fig. 1G**). Patches in the MRI data that were assigned with hypo-intense AF signals (orange) mostly clustered at the lower right corner, well separated from patches that were assigned with hyper-intense AF signals (turquoise) or near the brain surface (gray). Representative patches from the first two categories showed distinctive signal profiles (**Fig. 1H**). Overall, the result demonstrates that the ability of MRH-AF to translate multi-contrast MRI data into maps that mimic the tissue AF contrast in the AMBCA.

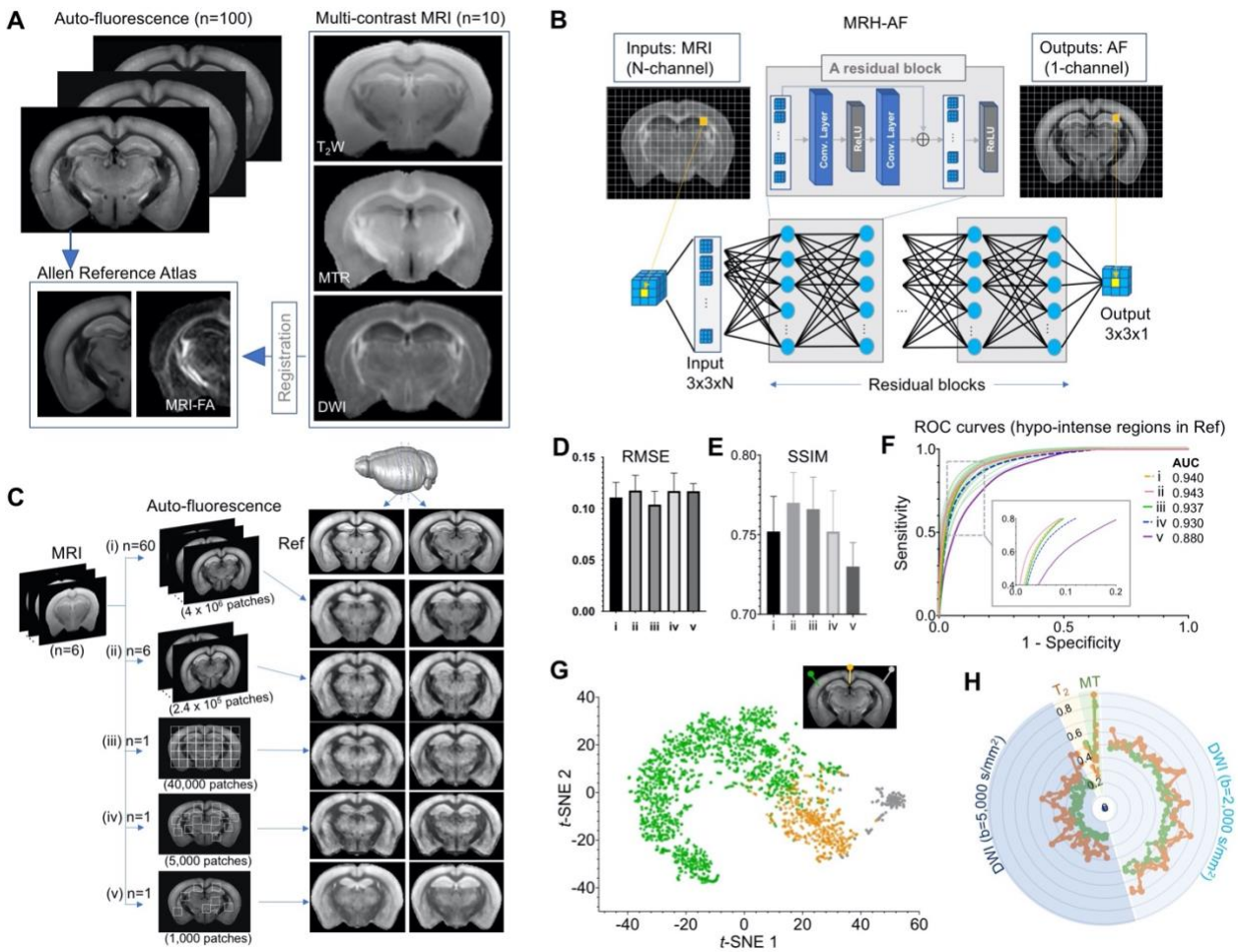

Fig. 1: *Connect multi-contrast MRI and auto-fluorescence (AF) data of the mouse brain using deep learning.* ***A:*** *T_2-weighted (T_2W), magnetization transfer ratio (MTR), and diffusion weighted images (DWI) were registered to the ARA space, from which 100 already registered AF data were selected and down-sampled to the same resolution of the MRI data.* ***B:*** *The deep CNN contained 64 layers and was trained using multiple 3x3 MRI patches as inputs and corresponding 3x3 patches from histology as targets.* ***C:*** *The CNN was trained using the MRI data (n=6) and different amounts of randomly selected AF data (i – v). The results generated by applying the CNN to a separate set of MRI data (n=4) were shown on the right for visual comparison with the reference (Ref: average AF data from 1,675 subject).* ***D-E:*** *Quantitative evaluation of the results in* ***C*** *with respect to the reference using RMSE and SSIM. The error bars indicate the standard deviations due to random selections of AF data used to train the network.* ***F:*** *The receiver operating characteristic curves of the results in* ***C*** *in identifying hypo-intense structures in the*

reference and their areas under the curve (AUCs). The ROC curves from 25 separate experiments in (iii) (light green) show the variability with respect to the mean ROC curve (dark green) due to inter-subject variations in AF intensity. ***G:*** *The distribution of randomly selected 3x3 MRI patches in the network's 2D feature space, defined using the t-SNE analysis based on case (iii) in* ***C****, shows three clusters of patches based on the intensity of their corresponding patches in the reference AF data (turquoise: hyper-intense, orange: hypo-intense; gray: brain surfaces).* ***H:*** *MRI signals from two representative patches with hyper-intense AF signals (turquoise) and two patches with hypo-intense AF signals (orange). The orange profiles show higher diffusion-weighted imaging (DWI) signals and larger oscillation among them than the turquoise profiles (both at b=2,000 s/mm^2 and 5,000 s/mm^2).*

Estimation of the relative contribution of individual MRI contrast

Based on the local ensemble average property of MR signals, we probed the inner-working of MRH-AF by adding random noises to each of the 67 MR images, one at a time, as perturbations to the network (*12*). We then measured the effect on network outcomes with respect to noise-free results (**Fig. 2A**), which reflected how each MR image influenced the outcome of MRH-AF or its relative contribution in the network. Similar information can also be obtained by training the networks with different subsets of the MRI contrasts and comparing the network predictions, but the perturbation method allows us to probe the existing network without retraining. We found that adding noises to a few images (e.g., T_2 and MT images) produced noticeably larger effects, in terms of output image quality and the ability of the network to separate different tissue types, than adding a comparable level of noises to other images (**Supplementary Figs. 4-5**). This information can be used to accelerate MRI acquisition by prioritizing the acquisition of images or contrasts with high relative contributions. The top 4, 17, and 38 images ranked based on their contributions accounted for 28%, 50%, and 75% of the total contribution to the final result respectively (**Fig. 2B**). Results from training the network with the top 38 MR images as inputs showed comparable visual quality (**Fig. 2B-C**) and diagnostic power (**Fig. 2D**) as the results based on the full dataset, but only required 57% the imaging time.

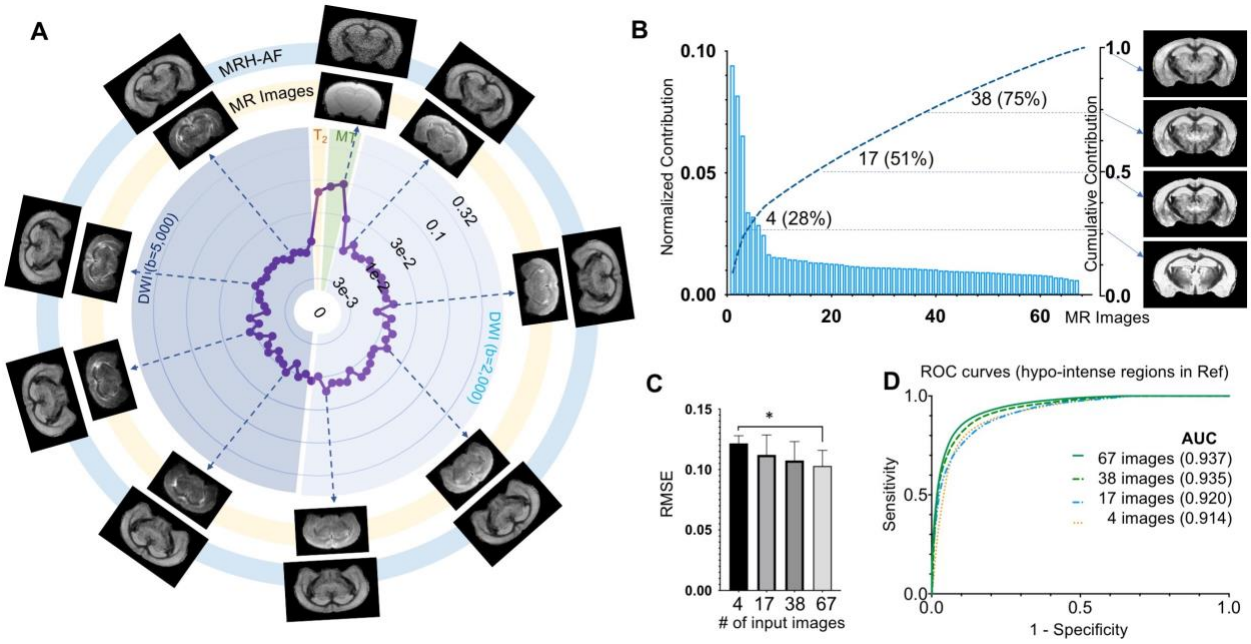

Fig. 2: *Estimating how each MR image influences the final result.* ***A:*** *Plots of the relative contribution of individual MRI images, normalized by the total contribution of all MR images. Images displayed on the outer ring (light blue, MRH-AF) show the network outcomes after adding 10% random noises to a specific MR image on the inner ring (light yellow).* ***B:*** *The relative contributions of all 67 MR images arranged in descending order and their cumulative contribution. The images on the right show the MRH-AF results with the network trained using only the top 4, 17, 38, and all images as inputs.* ***C:*** *RMSE measurements of images in* ***B*** *(n=4) with respect to the reference AF data. Lower RMSE values indicate better image quality. * indicates statistically significant difference (p=0.028, t-test).* ***D:*** *ROC curves of MRH-AF results in* ***B*** *and the AUC values.*

<u>Use deep learning to generate virtual maps of axon and myelin and enhance specificity</u>

Next, we trained our network using serial histological sections immuno-stained for neurofilament (NF) and myelin basic protein (MBP), two commonly used markers for axons and myelin, from the Allen mouse brain atlas. These images were down-sampled, normalized to the ARA (**Supplementary Fig. 6**). Part of the images were used for training and the rest were used as references. Due to limited histological images stained for myelin and axons, we adopted the transfer learning strategy (*13*). Using the MRH-AF

network as a starting point, we fixed most of its network layers while leaving the last three convolutional layers as trainable with MBP and NF stained histological images.

The MRH results (**Fig. 3A**) showed closer visual congruence with the histological references than commonly used MRI-based markers for axons (fractional anisotropy, FA) and myelin (magnetization transfer ratio, MTR, and radial diffusivity (*14*), D_R). Even though MRH was trained using coronal sections, it can generate maps along other axes when applied to 3D MRI data **(Fig. 3B)**. The MRH-NF/MBP results also showed strong signal correlations with the reference data (R^2 = 0.61/0.73, respectively, **Supplementary Fig. 3B-C**). ROC analyses (**Fig. 3B**) on detecting axon and myelin rich structures demonstrate improved specificity, while *t*-SNE analyses visualize how the two networks separate the patches in MRI data that corresponded to NF and MBP rich structures from the rest (**Fig. 3C**). Applying the MRH-MBP network to MRI data that were collected from dysmyelinating *shiverer* and control mouse brains (n=5/5) generated maps that resembled the MBP-stained histology (**Fig. 3D**). In the corpus callosum, the MRH-MBP results showed similar contrasts between *shiverer* and control moue brains as MTR (**Fig. 3E**).

The MRH results, in combination with the structural labels in ARA, provided insights into how the networks balanced multiple MRI contrasts to map axons and myelin in brain regions with distinct microstructural compositions. In the cortex, MRH-NF and MTR showed similar contrasts and comparable specificities to axons (**Fig. 3D**), while in the whole brain, MTR had a noticeable lower specificity than MRH-NF and FA (**Fig. 3B**). This suggests that MRH-NF assigned additional weightings on MTR when processing cortical patches. Similarly, in ROC analysis for pixels within the hippocampus, the curve of MRH-MBP closely followed the curve of D_R at b=5,000 s/mm^2, in a departure from the whole brain result (**Fig. 3B**). Visual inspections of the MRH-MBP results revealed a layer structure in the hippocampus, which was not obvious in the MTR map but visible in the radial diffusivity (D_R) map at b=5,000 s/mm^2 (**Fig. 3E**). Relative contributions of T_2W and MT signals were significantly higher in the cortex and hippocampus than in white matter regions for both MRH-NF and MRH-MBP (**Fig. 3F**).

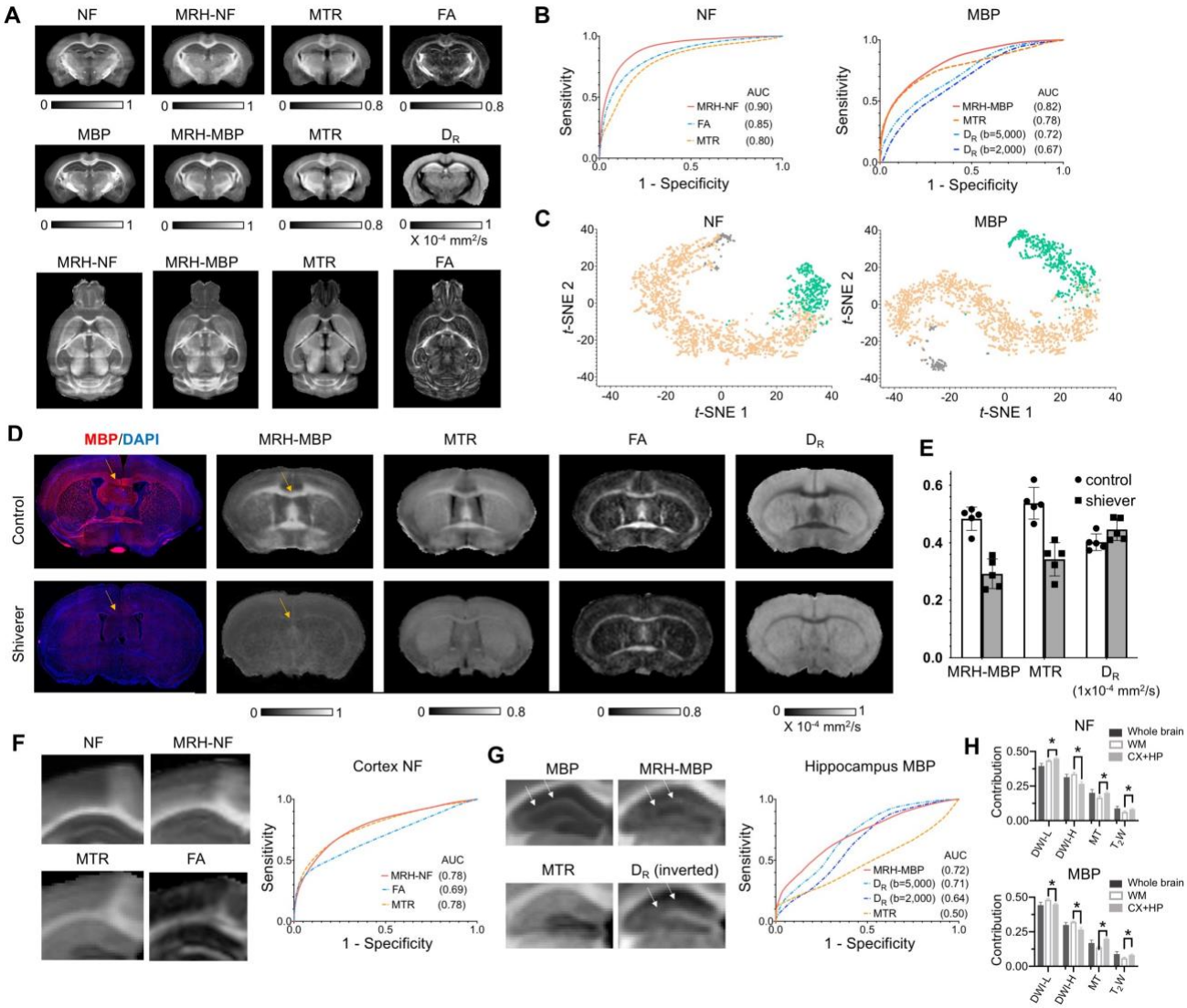

Fig. 3: *Inferring maps of neurofilament (NF) and myelin basic protein (MBP) from multi-contrast MRI data.* ***A:*** *Comparisons of MRH-NF/MBP results with reference histology and MRI-based markers that are commonly used to characterize axon and myelin in the brain (MTR: magnetization transfer ratio; FA: fractional anisotropy;* D_R*: radial diffusivity). Even though MRH-NF/MBP were trained using coronal sections, they were able to generate maps for orthogonal sections (e.g. horizontal sections in the bottom row) from 3D MRI, as expected from the local ensemble average property. The results show general agreements with structures in comparable horizontal MTR and FA maps but distinct tissue contrasts.* ***B:*** *ROC analyses of MRH-NF and MRH-MBP show enhanced specificity to their target structures defined in the reference data than MTR, FA, and* D_R*. Here,* D_R

values from DWIs with b-values of 2,000 and 5,000 s/mm^2 are examined separately. ***C:*** *The distribution of randomly selected 3x3 MRI patches in the network's 2D feature spaces of MRH-NF and MRH-MBP defined using the t-SNE analyses.* ***D:*** *Representative MRH-MBP results from dysmyelinated shiverer and control mouse brains, which were not included in training, show better agreement with histology than maps of MTR, FA, and D_R.* ***E:*** *Differences in MRH-MBP, MTR, and D_R values of the corpus callosum (t-test, n=5 in each group, p=0.00018/0.0061/0.475, respectively).* ***F-G:*** *Enlarged maps of the cortical (F) and hippocampal (G) regions of normal C57BL6 mouse brains comparing the tissue contrasts in MRH-NF/MBP with histology and MRI. In* ***(G)****, white arrows point to a layer structure in the hippocampus. ROC analyses performed within the cortex and hippocampus show that MRH-NF/MBP have higher specificity than FA, MTR, and D_R, but with lower AUCs than in* ***B*** *due to distinct tissue properties.* ***H:*** *Relative contributions of T_2W, MT, diffusion MRI (DWI-L: b=2,000 s/mm^2; DWI-H: b=5,000 s/mm^2) for the whole brain, white matter, and cortex/hippocampus. *: p<0.005 (paired t-test, n=4, from left to right, p=0.0043 / 0.000021 / 0.00072 / 0.0014 for NF, p=0.000058 / 0.000035 / 0.000002 / 0.00392 for MBP, respectively). Details on the contributions of each MRI contrast can be found in* ***Supplementary Fig. 7****.*

<u>Use deep learning to generate maps that mimic Nissl staining:</u>

MRH networks can also be extended to other types of MR contrasts and histology. To demonstrate this, we used MRH to test whether cellularity in the mouse brain can be inferred from diffusion MRI signals, as our previous studies suggest that oscillating gradient spin echo (OGSE)(*15*) diffusion MRI can generate a contrast similar to Nissl-staining in both normal and injured mouse brains(*16, 17*). We separated the down-sampled single subject 3D Nissl data from ARA into two parts. One was used as the training target, and the rest was used as the reference for testing (**Fig. 4A**). The inputs to the so-called MRH-Nissl network included conventional pulsed gradient spin echo (PGSE) and recently developed OGSE diffusion MRI data. In the testing regions, the network that utilized all OGSE and PGSE data as inputs generated maps with good agreement with the ground truth Nissl data (**Fig. 4A**), showing higher sensitivity and specificity than PGSE

(**Fig. 4B**). In the 2D feature space from *t*-SNE analysis (**Fig. 4C**), the patches that correspond to regions with low Nissl signals were separated from other patches that correspond to regions with strong Nissl signals. Representative signal profiles from the three categories (**Fig. 4D**) revealed that signals in the high Nissl signal patches decreased as the oscillating frequency increased, whereas the other two types of patches showed no such pattern. Detailed analysis of contrast contribution showed that PGSE and OGSE data contribute equally (**Fig. 4E**), indicating the importance of OGSE data in generating the target tissue contrast. The MRH-Nissl map of the *sas4*$^{-/-}$*p53*$^{-/-}$ mouse brain, which contains a band of heterotopia consists of undifferentiated neurons(*18*), produced image contrasts that matched Nissl-stained histology (**Fig. 4F**).

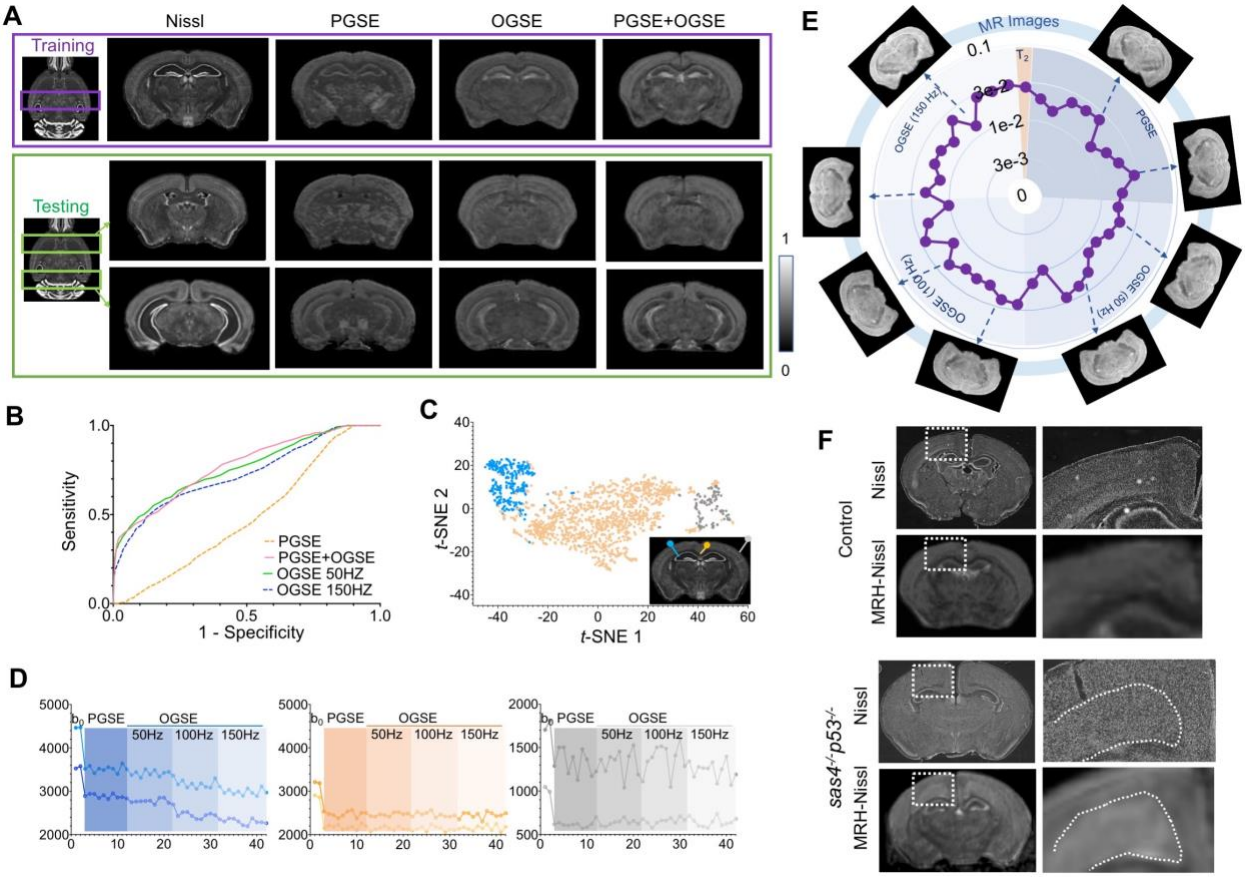

Fig. 4: *Generating maps that mimic Nissl stained histology from multi-contrast MRI data.* ***A:*** *Comparisons of reference Nissl histology and MRH-Nissl results with PGSE, OGSE, combined PGSE and OGSE diffusion MRI data in both training and testing datasets. The entire datasets consist of PGSE and OGSE data acquired with oscillating frequencies of 50, 100, and 150 Hz, a total of 42 images.* ***B:*** *ROC curves of MRH-Nissl show enhanced*

specificity for structures with high cellularity (strong Nissl staining) when both PGSE and OGSE data were included in the inputs than PGSE only. ***C:*** *The distribution of randomly selected 3x3 MRI patches in the network's 2D feature spaces of MRH-Nissl defined using t-SNE analyses. Green and orange dots correspond to regions with high and low cellularity, respectively, and gray dots represent patches on the brain surface.* ***D:*** *Representative signal profiles from different groups in* ***C****.* ***E:*** *Relative contributions of PGSE and three OGSE diffusion MRI datasets* ***F:*** *Representative MRH-Nissl results from sas4$^{-/-}$p53$^{-/-}$ and control mouse brains compared with Nissl stained sections. The location of the cortical heterotopia, consists of undifferentiated neurons, is indicated by the dashed lines in the mutant mouse brain image.*

Discussions:

The present study focused on inferring maps of key cellular structures in the mouse brain from multi-contrast MRI data. Previous works on this problem include: new MRI contrasts that capture specific aspects of cellular structures of interest (*19, 20*); carefully constructed tissue models for MR signals(*21*); statistical methods to extract relevant information from multi-contrast MRI(*2*); and techniques to register histology and MRI data(*22, 23*) for validation (*24, 25*). Here, we built on these efforts by demonstrating that deep learning networks trained by co-registered histological and MRI data can improve our ability to detect target cellular structures.

Previous studies on the relationship between histology and MRI signals focused on correlating histological and MRI markers as co-registered MRI and histological data as well as realistic tissue models are scarce (*21, 26*). Adopting similar approaches described by recent reports (*4, 6*) on using deep learning to generate histological labels from unlabeled light microscopy images, we demonstrate a proof of concept of using deep learning to solve the inverse problem of inferring histological information from MRI signals. Even though the resolution of the virtual histology is inevitably limited by the resolution of the input MRI data (~100 μm/pixel) (**Supplementary Fig. 8**), the presented approach has many potential applications in biomedical research involving MRI. It can enhance our ability to accurately map selected cellular structures and their pathology in mouse models of diseases using non-invasive MRI, with contrasts familiar to neurobiologists. Although

the networks cannot be applied to human MRI directly due to vast differences in tissue properties and scanning protocols, understanding how the networks improve specificity based on given MRI contrasts will guide the development of optimal imaging strategy in the clinics. In addition, the co-registered histology and MRI dataset provide a testbed for developing new MRI strategies. As it is relatively easy to normalize any new MRI data to our 3D multi-contrast MRI data and co-registered histology, the sensitivity and specificity of a new MRI contrast to target cellular structures can be evaluated. With quantitative information on the contributions of different MR contrasts, it is now straightforward to design accurate and efficient multi-contrast MRI strategy.

Perfect co-registration between MRI and histology is highly challenging, as conventional tissue preparations inevitably introduce tissue deformation and damages. In addition, differences in tissue contrasts between histology and MRI also limit the accuracy of registration. Serial two-photon tomography used by the Allen Institute and similar methods allow 3D uniform sampling of the entire brain, which facilitate registration using established registration pipelines(*27*). We expect recent advances in tissue clearing techniques can assist in this aspect once issues such as tissue shrinkage and penetration of antibodies for more target cellular structures are resolved. Remaining mismatches can be accommodated by choosing the appropriate patch size in the network as shown in our results and earlier studies(*28*).

There are several directions to further improve our work. First, it is important to curate a training dataset that covers a broad spectrum of conditions and, ideally, with MRI and histology data acquired from the same animal. The data included in this study were adult mouse brain and most white matter structures are myelinated. As a result, the network to predict axons place a substantial weight on MRI contrasts that reflect myelin content (e.g., MT). With the inclusion of unmyelinated embryonic or neonatal mouse brains, we anticipate that the contribution of myelin will be reduced. Inclusion of pathological examples, such as *shiverer* mouse brain data, for training will likely improve our ability to characterize pathological conditions. Secondly, the CNNs constructed in this study involved several common building blocks of deep learning, and new advances on network architecture design (e.g. (*29, 30*)) could further enhance the performance. While CNNs have been commonly treated as black boxes, several recently reported approaches, such

as deep Taylor decomposition(*31*) and Grad-CAM(*32*), can help explain the inner-working. Third, developing similar networks for *in vivo* MRI data and potential clinical application will require additional effort. The MRI data used in this study were collected from post-mortem mouse brain specimens, which are different from *in vivo* mouse brains due to death and chemical fixation. Differences in tissue properties between human and mouse brains also require additional steps. Finally, deep learning cannot replace the good understanding of the physics involved in MRI contrasts and the development of new MRI contrast that targets specific cellular structures.

Materials and Methods:

Animals and *ex vivo* MRI. All animal experiments have been approved by the Institute Animal Care and Use Committee at New York University, Memorial Sloan Kettering Cancer Center, and Johns Hopkins University. Adult C57BL/6 mice (P60, n=10, Charles River, Wilmington, MA, USA), *sas4*$^{-/-}$*p53*$^{-/-}$ (*18*) and littermate controls (n=4/4, P28), *rag2*$^{-/-}$ *shiverer* and littermate controls (n=5/5, P50) were perfusion fixed with 4% paraformaldehyde (PFA) in PBS. The samples were preserved in 4% PFA for 24 hours before transferring to PBS. *Ex vivo* MRI of mouse brain specimens was performed on a horizontal 7 Tesla MR scanner (Bruker Biospin, Billerica, MA, USA) with a triple-axis gradient system. Images were acquired using a quadrature volume excitation coil (72 mm inner diameter) and a receive-only 4-channel phased array cryogenic coil. The specimens were imaged with the skull intact and placed in a syringe filled with Fomblin (perfluorinated polyether, Solvay Specialty Polymers USA, LLC, Alpharetta, GA, USA) to prevent tissue dehydration. Three-dimensional diffusion MRI data were acquired using a modified 3D diffusion-weighted gradient- and spin-echo (DW-GRASE) sequence (*33*) with the following parameters: echo time (TE)/repetition time (TR) = 30/400ms; two signal averages; field of view (FOV) = 12.8 mm x 10 mm x 18 mm, resolution = 0.1 mm x 0.1 mm x 0.1 mm; two non-diffusion weighted images (b_0s); 30 diffusion encoding directions; and b = 2,000 and 5,000 s/mm^2, total 60 diffusion weighted images (DWIs). Co-registered T_2-weighted and magnetization transfer (MT) MRI data were acquired using a rapid acquisition with relaxation enhancement (RARE) sequence with the same FOV, resolution, and signal averages as the diffusion MRI acquisition and the following

parameters: T_2: TE/TR=50/3000 ms; MT: TE/TR=8/800ms, one baseline image (M_0) and one MT-weighted (M_t) images with offset frequency/power = -3 KHz/20 μT were acquired. The total imaging time was approximately 12 hours for each specimen. For the *sas4*$^{-/-}$ *p53*$^{-/-}$ and littermate controls (n=4/4, P28), PGSE and OGSE diffusion MRI data were acquired with the protocol described in (*17*) and a spatial resolution of 0.1 mm x 0.1 mm x 0.1 mm. All 3D MRI data were interpolated to a numerical resolution of 0.06 mm x 0.06 mm x 0.06 mm to match the resolution of our MRI-based atlas (*34*).

Magnetization transfer ratio (MTR) images were generated as $MTR=(M_0-M_t)/M_0$. From the diffusion MRI data, diffusion tensors were calculated using the log-linear fitting method implemented in MRtrix (http://www.mrtrix.org) at each pixel, and maps of mean and radial diffusivities and fractional anisotropy were generated, The mouse brain images were spatial normalized to an ex vivo MRI template (*34*) using the Large Deformation Diffeomorphic Metric Mapping (LDDMM) method (*35*) implemented in the DiffeoMap software (www.mristudio.org). The template images had been normalized to the Allen reference atlas using landmark-based image mapping and LDDMM.

Histological data. From the Allen moue brain atlas, single subject 3D Nissl data and 3D AF data (n=100), which were already registered to the ARA space, were down sampled to 0.06 mm isotropic resolution. The *sas4*$^{-/-}$*p53*$^{-/-}$, *rag2*$^{-/-}$ *shiverer* and control mouse brains were cryopreserved and cut into 30 μm coronal sections and processed for Nissl and immunofluorescence. For immunofluorescence, sections were first washed with PBS, blocked with 5% bovine serum albumin, and incubated overnight at 4 °C with primary antibodies: anti-myelin basic protein (AbD Serotec, MCA4095). Sections were rinsed with PBS and incubated with Alexa fluor secondary antibodies (Invitrogen) cover-slipped with anti-fade mounting medium containing DAPI (Vectrolabs, H-1200). Images were obtained and tile-stitched using an inverted microscope (Zeiss, Axio Observer.Z1) equipped with a motorized table.

Registration of MRI and histological data. Group average 3D MRI data in our previously published mouse brain atlas(*34*) were first spatially normalized to the ARA space. Briefly, fourteen major brain structures (e.g., cortex, hippocampus, striatum) in the atlas MRI data were manually segmented following the structural delineations in the ARA. Voxels that belong to these structures in the MRI and average 3D AF data in the ARA

(down sampled to 0.06 mm isotropic resolution) were assigned distinct intensity values, and a diffeomorphic mapping between the discretized atlas MRI and ARA AF data was computed using LDDMM. The mapping was then applied to the original atlas MRI data to generate an MRI template registered to the ARA space. Using dual-channel LDDMM (*35*) based on tissue contrasts in the average DWI and FA images and the MRI template, the 3D MRI data acquired in this study were accurately normalized to the ARA space.

NF and MBP stained images of the C57BL/6 mouse brain were downloaded from the ARA reference dataset. Images with major artifacts or tissue loss were excluded. Small tissue tearing and staining artifacts were removed using the Inpainting feature implemented in the photoshop heading brush tool (www.adobe.com), and dark pixels in the ventricles were replaced by the average intensity values of the cortex to match MRI data (Supplementary Fig. 4A). The repaired images were down-sampled to an in-plane resolution of 0.06 mm/pixel. For each 2D histological image, the best-matching MRI section in the MRI template was identified, and a coarse-to-fine alignment from histology to MRI using affine transform and landmark-based image warping tool in ImageJ (https://imageJ.nete/BUnwarpJ). The aligned 2D sections were then assembled into a 3D volume and mapped to the MRI template using LDDMM (between NF/MBP and FA) to further improve the quality of registration.

Evaluation of image resolution. The resolution of MRI, histological images, were evaluated using a parameter-free decorrelation analysis method (*36*), without the initial edge apodization step.

Design and training of the MRH networks. MRH-networks are constructed using a CNN model with convolutions from the MRI to histology space. The networks were implemented using the deep learning toolbox in Matlab (www.mathworks.com) using the directed acyclic graph architecture. To accommodate residual mismatches between MRI and histological data, the network consistently applied convolutional layers until the end layer that computed the distance loss to the target histology. The number of layers and neurons in each layer were determined empirically to balance performance and the risk of overfitting. We chose 64 hidden layers, each with 64 neurons, which applied a filter size of 3x3 and included a rectified linear unit (ReLu). Most of the hidden layers utilized skip connections to jump over layers to avoid vanishing gradients(*37*). The network was

initialized with orthogonal random weights and was trained using a stochastic gradient descent optimizer to minimize pixel-level absolute distance loss (**Supplementary Fig. 2A-B**). Several choices of mini batch sizes were tested and the size was set to 128 to attain a balance between training speed and performance. Stochastic gradient descent with a momentum beta of 0.9 was used for stochastic optimization, with an initial learning rate of 0.1 and a learning rate factor of 0.1. During training, the learning rate was reduced by a factor of 0.1 every 10 epochs. Maximum epoch number was set at 60, but early stopping was employed if the validation set loss did not decrease in 5 epochs. During hyper-parameter tuning, 1000 3x3 patches were randomly held out to as the validation dataset and isolated from the training dataset. The weights from the epoch with the lowest validation loss were selected for final testing.

When retraining the MRH-AF neural network using MBP/NF data, we refined the last three layers parameters while leave the rest layer parameters untouched in the Matlab deep learning toolbox using the directed acyclic graph architecture. The hyperprameters and training patches is the same as MRH-AF. Specifically, the network training initial learning rate was 0.0001, while learning rate factor was 0.1 to accomplish transfer learning. Using the stochastic gradient descent optimizer, our transfer learning converged as shown in **Supplementary Fig. 2C-D.**

***t*-Distributed Stochastic Neighbor Embedding(t-SNE) analysis.** The *t*-SNE cluster was performed using the Matlab t-SNE analysis function on the network prediction based on values in 2,000 randomly selected 3x3 patches in the mouse brain MRI data.

Contribution analysis. Following the perturbation method described by Olden et al.(*12*), Rician noises were added to one input MR image to the pre-trained MRH networks, and RMSE between the noise contaminated outputs and the original output without noise was recorded. By repeating this procedure for all MR images, the sensitivities of MRH to each of the 67 input MR image or their contributions were obtained.

Evaluate the effect of pixel mismatches. In the experiment that used the DWI and FA data of the mouse brains to train an MRH network, simulated pixel displacements were used to deform the FA data (target), which were perfectly co-registered to the DWI data (inputs), to test the effect of pixel mismatches between input and target data on network prediction. Gaussian random displacement fields were generated for pixels on a 1 mm by

1 mm grid in the coronal plane and propagated to other pixels by B-spline interpolations. The displacement fields followed a Chi distribution with 2 degrees of freedom and were adjusted to match the level of pixel mismatches observed between MRI and histological data.

Statistical analysis. Statistical significance was determined using unpaired Student's t-test with threshold set at 0.05. All statistical tests were performed with Prism (GraphPad). All values in bar graphs indicate mean + standard deviation.

References

1. J. P. Lerch *et al.*, Studying neuroanatomy using MRI. *Nature Neuroscience* **20**, 314-326 (2017).
2. G. Mangeat, S. T. Govindarajan, C. Maineroc, J. Cohen-Adad, Multivariate combination of magnetization transfer, T-2* and B0 orientation to study the myelo-architecture of the in vivo human cortex. *Neuroimage* **119**, 89-102 (2015).
3. M. Cercignani, S. Bouyagoub, Brain microstructure by multi-modal MRI: Is the whole greater than the sum of its parts? *Neuroimage* **182**, 117-127 (2018).
4. E. M. Christiansen *et al.*, In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* **173**, 792-+ (2018).
5. A. P. Leynes *et al.*, Zero-Echo-Time and Dixon Deep Pseudo-CT (ZeDD CT): Direct Generation of Pseudo-CT Images for Pelvic PET/MRI Attenuation Correction Using Deep Convolutional Neural Networks with Multiparametric MRI. *J Nucl Med* **59**, 852-858 (2018).
6. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, G. R. Johnson, Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat Methods* **15**, 917-+ (2018).
7. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436-444 (2015).
8. L. Ng *et al.*, An anatomic gene expression atlas of the adult mouse brain. *Nat Neurosci* **12**, 356-362 (2009).
9. S. W. Oh *et al.*, A mesoscale connectome of the mouse brain. *Nature* **508**, 207-214 (2014).
10. P. C. Christensen *et al.*, High-resolution fluorescence microscopy of myelin without exogenous probes. *Neuroimage* **87**, 42-54 (2014).
11. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).
12. J. D. Olden, M. K. Joy, R. G. Death, An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol Model* **178**, 389-397 (2004).
13. K. Weiss, T. M. Khoshgoftaar, D. Wang, A survey of transfer learning. *J big Data* **3**, (2016).

14. S. K. Song *et al.*, Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *Neuroimage* **17**, 1429-1436 (2002).
15. M. D. Does, E. C. Parsons, J. C. Gore, Oscillating gradient measurements of water diffusion in normal and globally ischemic rat brain. *Magn Reson Med* **49**, 206-215 (2003).
16. M. Aggarwal, J. Burnsed, L. J. Martin, F. J. Northington, J. Zhang, Imaging neurodegeneration in the mouse hippocampus after neonatal hypoxia-ischemia using oscillating gradient diffusion MRI. *Magn Reson Med* **72**, 829-840 (2014).
17. M. Aggarwal, M. V. Jones, P. A. Calabresi, S. Mori, J. Zhang, Probing mouse brain microstructure using oscillating gradient diffusion MRI. *Magn Reson Med* **67**, 98-109 (2012).
18. R. Insolera, H. Bazzi, W. Shao, K. V. Anderson, S. H. Shi, Cortical neurogenesis in the absence of centrioles. *Nat Neurosci* **17**, 1528-1535 (2014).
19. N. Stikov *et al.*, In vivo histology of the myelin g-ratio with magnetic resonance imaging. *Neuroimage* **118**, 397-405 (2015).
20. J. Veraart *et al.*, Nonivasive quantification of axon radii using diffusion MRI. *Elife* **9**, (2020).
21. I. O. Jelescu, M. D. Budde, Design and Validation of Diffusion MRI Models of White Matter. *Front Phys-Lausanne* **5**, (2017).
22. D. Tward *et al.*, Diffeomorphic Registration With Intensity Transformation and Missing Data: Application to 3D Digital Pathology of Alzheimer's Disease. *Front Neurosci* **14**, 52 (2020).
23. J. Xiong, J. Ren, L. Luo, M. Horowitz, Mapping Histological Slice Sequences to the Allen Mouse Brain Atlas Without 3D Reconstruction. *Front Neuroinform* **12**, 93 (2018).
24. K. G. Schilling *et al.*, Histological validation of diffusion MRI fiber orientation distributions and dispersion. *Neuroimage* **165**, 200-221 (2018).
25. H. B. Stolp *et al.*, Voxel-wise comparisons of cellular microstructure and diffusion-MRI in mouse hippocampus using 3D Bridging of Optically-clear histology with Neuroimaging Data (3D-BOND). *Sci Rep* **8**, 4011 (2018).

26. D. S. Novikov, V. G. Kiselev, S. N. Jespersen, On modeling. *Magnetic Resonance in Medicine* **79**, 3172-3193 (2018).
27. L. Kuan *et al.*, Neuroinformatics of the Allen Mouse Brain Connectivity Atlas. *Methods* **73**, 4-17 (2015).
28. Y. Rivenson *et al.*, Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng* **3**, 466-477 (2019).
29. I. Goodfellow *et al.*, in *Conference on Neural Information Processing Systems*. (2014), vol. 3, pp. 2672-2680.
30. J. Zhu, T. Park, P. Isola, A. A. Efros, paper presented at the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.
31. G. Montavon, S. Lapuschkin, A. Binder, W. Samek, K. R. Muller, Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recogn* **65**, 211-222 (2017).
32. R. R. Selvaraju *et al.*, paper presented at the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.
33. D. Wu *et al.*, In vivo high-resolution diffusion tensor imaging of the mouse brain. *Neuroimage* **83**, 18-26 (2013).
34. N. Chuang *et al.*, An MRI-based atlas and database of the developing mouse brain. *NeuroImage* **54**, 80-89 (2011).
35. C. Ceritoglu *et al.*, Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. *NeuroImage* **47**, 618-627 (2009).
36. A. Descloux, K. S. Grussmayer, A. Radenovic, Parameter-free image resolution estimation based on decorrelation analysis. *Nat Methods* **16**, 918-+ (2019).
37. K. He, X. Zhang, S. Ren, J. Sun, paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 27-30 Jun 2016 2016.

Funding: This work was supported by NIH grants R01NS102904 and R01HD074593.

Author contributions:

The project was conceived by Z.L., J.Z., Y.G., L.Y., and F.K. and developed by Z.L., J.Z., C.L., T.A., Z.D., P.W., and S.H.S. Z.L. and J.Z. contributed most Figs. C.L. acquired most MR images. Z.L. and J.Z. wrote and prepared the paper, Z.D., F.K., L.Y., Y.G., C.L., T.A., P.W. and S.H.S. edited the paper.

Competing interests:

The authors declare no competing interest related to the work presented here.

Data and materials availability:

The analyses in this study were carried out on publically available datasets. The 3D auto-fluorescence data were obtained from the Allen mouse connectivity project (http://connectivity.brain-map.org). The neurofilament, myelin basic protein, and Nissl stained histological data were obtained from the Allen mouse brain atlas (https://connectivity.brain-map.org/). Code and data for training and histology prediction from MRI is available at https://github.com/liangzifei/MRH-Net.