**Additional authors related to this work:**

Michael Kunst1 <michael.kunst@alleninstitute.org>

Lydia Ng1 <LydiaN@alleninstitute.org>

Rachel Dalley1 <racheld@alleninstitute.org>

Staci Sorensen1 <stacis@alleninstitute.org>

Quanxin Wang1 <quanxinw@alleninstitute.org>

1Allen Institute for Brain Science, Seattle, WA, USA

**1. fMOST**

**Define the task and detail the associated materials.**

*1. Overview*

We have developed a framework for mapping fluorescence micro-optical sectioning tomography (fMOST) mouse brain images into the Allen Common Coordinate Framework (CCFv3). Our approach uses an average fMOST atlas to serve as an intermediate registration target for mapping fMOST images from individual specimens into the CCFv3. First, we use a one-time annotation-driven registration to establish a canonical mapping between the fMOST atlas and the CCFv3. This step allows us to align expert determined landmarks to accurately map structures with large morphological differences between the modalities, which are difficult to address using standard approaches. Once this canonical mapping is established, standard intensity-based registration is used to align each new fMOST image to the fMOST specific atlas. This mapping is concatenated with the canonical atlas-to-CCFv3 mapping to further map each individual brain into the CCFv3 without the need to generate additional landmarks. Transformations learned through this mapping can be applied to single neuron reconstructions from the fMOST images to evaluate neuronal distributions across different specimens into the CCFv3 for the purpose of cell census analyses. Figure 1. Shows the general workflow for this registration pipeline.

**A diagram of a diagram of a human brain

Description automatically generated with medium confidence**

**Figure 1.** Block diagram of the ANTS pipeline for mapping fMOST imaging data into the Allen CCFv3 using a fMOST specific average atlas as an intermediate step.

*2. Data*

The high-throughput and high-resolution fluorescence micro-optical sectioning tomography (fMOST) [gong 2016, wang 2021] platform was used to image 55 mouse brains containing gene-defined neuron populations, with sparse transgenic expression [Rotolo 2008], detailed in Peng et al. (2021). In short, the fMOST imaging platform results in 3D images with voxel sizes of 0.35 × 0.35 × 1.0 μm and is a two-channel imaging system where the green channel displays the GFP labeled neuron morphology and the red channel is used to visualize the counterstained propidium iodide cytoarchitecture. Registrations in this work were performed using the red channel, which offered higher tissue contrast for alignment.

**Describe the step-by-step design and implementation of solution approach.**

*1. Preprocessing*

*1.1 Downsampling*

The first challenge when mapping fMOST images into the CCFv3 is addressing the resolution scale of the data. Native fMOST data from an individual specimen can range in the order of terabytes, which leads to two main problems. First, three-dimensional registration methods (particularly those estimating local deformation) have high computational complexity and typically cannot operate on such high-resolution data under reasonable memory and runtime constraints. Second, the resolution of the CCFv3 atlas is much lower than the fMOST data, thus the mapping process will cause much of the high-resolution information in the fMOST images to be lost regardless. Thus, we perform a cubic-bspline downsampling of the fMOST data to reduce the resolution of each image to 25 micron (µm) isotropic to match the 25 µm CCFv3 intensity atlas. An important detail to note is that while the fMOST images and atlas are downsampled, the mapping learned during the registration is assumed to be continuous. Thus, after establishing the mapping to the CCFv3, we can interpolate the learned mapping and apply it to the high-resolution native data directly to transform any spatially aligned data (such as the single-cell neuron reconstructions) into the CCFv3.

*1.2 Stripe Artifact Removal*

Repetitive pattern artifacts are a common challenge in fMOST imaging where inhomogeneity during the cutting and imaging of different sections can leave ‘stripes’ of hyper- and hypo-intensity across the image. These stripe artifacts can be latched onto by the registration algorithm as unintended features that are then misregistered to non-analogous structures in the CCFv3. We address these artifacts by fitting a 3D bandstop (notch) filter to target the frequency of the strip patterns and removing them prior to the image registration (Fig. 2).

A collage of images of a brain

Description automatically generated

**Figure 2.** Example of the ‘stripe’ artifact removal applied to our fMOST data. (A) and (C) show an fMOST image slice before and after the artifact removal, respectively. (B) and (D) show the frequency domain of images of (A) and (C), where the red arrows indicate the frequency peaks of the stripe artifacts that are targeted and removed by the bandstop filter.

*1.3 Inhomogeneity correction*

Regional intensity inhomogeneity can also occur within and between sections in fMOST imaging due to staining or lighting irregularity during acquisition. Similar to stripe artifacts, intensity gradients due to inhomogeneity can be misconstrued as features during the mapping and result in matching of non-corresponding structures. Our pipeline addresses these intensity inhomogeneities using the N4 bias field correction[Tustison 2010] approach built into ANTS.

*2. Spatial mapping*

The spatial mapping of the fMOST image into the CCFv3 in our pipeline is separated into three main steps: 1) First, we construct an fMOST average atlas, which is registered into the CCFv3 using a one-time, annotation-driven registration. 2) Individual fMOST images are then registered to the fMOST average atlas 3) finally, the atlas-to-CCFv3 mapping and individual-to-atlas mapping are concatenated to generate a final mapping from each individual image into the CCFv3.

*2.2. Average fMOST atlas as an intermediate target*

Due to the preparation of the mouse brain for fMOST imaging, the resulting structure in the mouse brain has several large morphological deviations from the CCFv3 atlas. Most notable of these is an enlargement of the ventricles, and compression of cortical structures. In addition, there is poor intensity correspondence for the same anatomic features due to the difference in imaging modalities. We’ve found that standard intensity-base registration is insufficient to capture the significant deformations required to map these structures correctly into the CCFv3. We address this challenge in ANTS by using explicitly corresponding parcellations of the brain, ventricles and surrounding structures to directly map these large morphological differences. However, generating these parcellations for each individual mouse brain is a labor-intensive task. Our solution is to create an average atlas that encapsulates these large morphological differences to serve as an intermediate registration point. This has the advantage of only needing to generate one set of corresponding annotations which is used to register between the two atlas spaces. New images are then aligned to the fMOST average atlas, which shares common intensity and morphological features and thus can be achieved through standard intensity-based registration.

*2.3 Average fMOST atlas construction*

To construct our average atlas, we use the ANTS atlas construction script (*antsMultivariateTemplateConstruction2.sh*), with 30 fMOST images. To ensure the atlas is symmetric, we also include a left-to-right flip of these 30 images, resulting in 60 images used as inputs for the script. We ran 3 iterations of the atlas construction using the default settings. Additional iterations (up to 6) were evaluated and showed minimal changes to the final atlas construction, suggesting a convergence of the algorithm.

*2.3 fMOST atlas to CCFv3 alignment*

Alignment between the fMOST average atlas and CCFv3 was performed using a one-time annotation-driven approach. Label-to-label registration is used to align 7 corresponding annotations in both atlases in the following: 1) Brain mask/ventricles, 2) caudate/putamen, 3) Fimbria, 4) posterior choroid plexus, 5) optic chiasm, 6) anterior choroid plexus, and 7) habenular commissure. The alignments were performed sequentially, with the largest, most relevant structures being aligned first using coarse registration parameters, followed by other structures using finer parameters. This approach allows us to address large morphological differences (such as brain shape and ventricle expansion) at the start of registration and then refine the mapping using the smaller structures. The overall ordering of these structures was determined manually by an expert anatomist, where anatomical misregistration after each step of the registration was evaluated and used to determine which structure should be used in the subsequent iteration to best improve the alignment. The transformation from this one-time alignment is preserved and used as the canonical fMOST atlas to CCFv3 mapping in the pipeline.

*2.4 Alignment of individual fMOST mouse brains*

The canonical transformation between the fMOST atlas and CCFv3 greatly simplifies the registration of new individual fMOST mouse brains into the CCFv3. Each new image is first registered into the fMOST average atlas, which shares intensity, modality, and morphological characteristics. This allows us to use standard, intensity-based alignment (SyN[Avants 2008]) in ANTS to perform this alignment. The *ants.apply\_transforms()* function is then used to concatenate this transformation with the canonical atlas-to-atlas transformation, which is then applied to the original fMOST image to move it into the CCFv3 space. Figure 3 shows an example of this alignment.

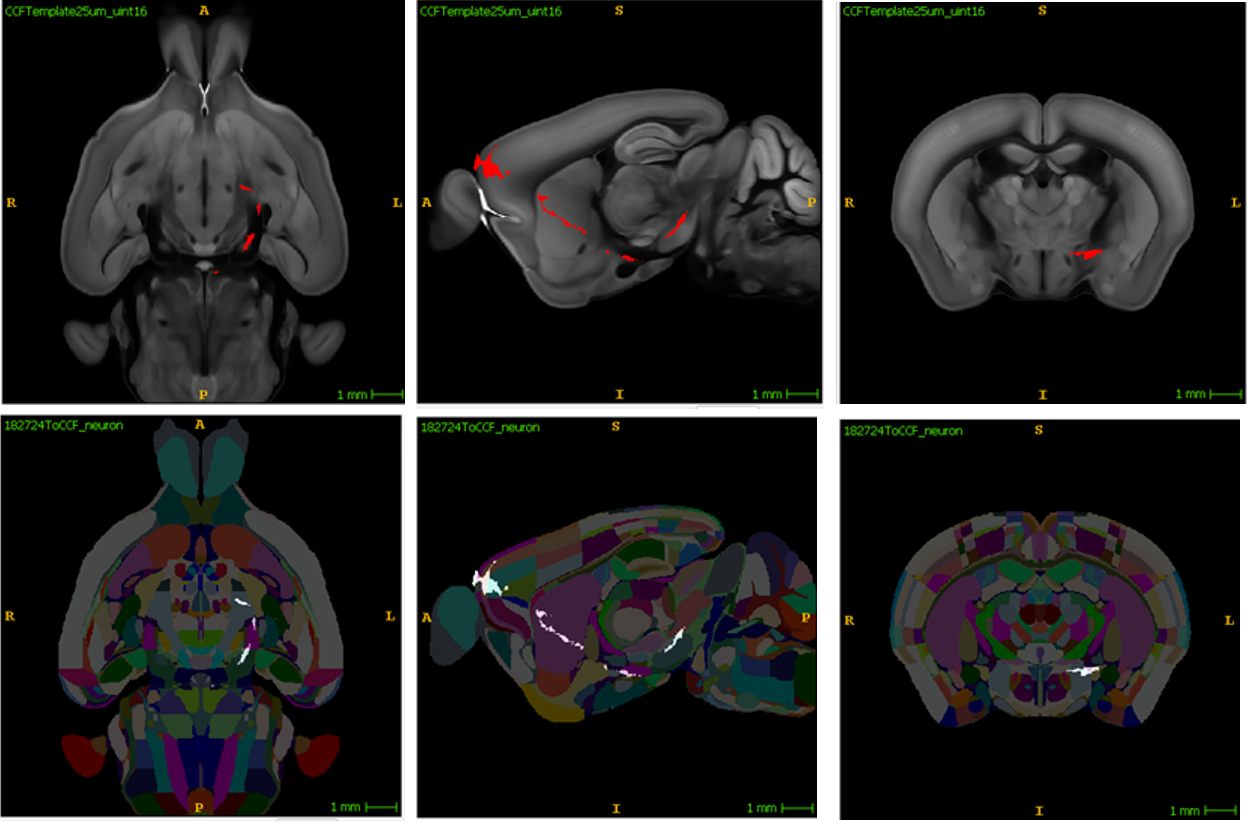
*3 Transformation of Single Cell Neurons*

A key feature of fMOST imaging is the ability to reconstruct and examine whole-brain single neuron projections, as described in Peng et al. (2021). Spatial mapping of these neurons from individual brains into the CCFv3 allows investigators to study different neuron types within the same space and characterize their morphology with respect to their transcriptomics. Mappings found between the fMOST image and the CCFv3 using our pipeline can be applied to fMOST neuron reconstruction data (.swc files) using the *ants.apply\_transforms\_to\_points()* function in ANTS. Figure 4 shows an example single reconstructed neuron (shown in red/white) mapped into the CCFv3 using this approach with the mapping produced from our pipeline.

A collage of images of a brain

Description automatically generated

**Figure 3.** Example mapping of a fMOST image into the Allen CCFv3. Shown are the triplanar view of (A) unregistered fMOST image, (B) Allen CCFv3 atlas, and (C) the FMOST image after the mapping.



**Figure 4.** Mapping of a reconstructed neuron (shown in red) generated from the fMOST image (top row) into the Allen CCFv3 (bottom row).

**Describe the validation performed to evaluate the developed solution.**

Validation of the canonical atlas to CCFv3 mapping was performed via quantitative Dice comparison at each step of the registration and qualitative assessment of structural correspondence after alignment by an expert anatomist. Table 1 shows the final Dice overlap of each of the structures after the canonical mapping.

**Table 1.** Dice overlap of landmark labels between corresponding structures in the fMOST average atlas and Allen CCFv3

A line of black text

Description automatically generated

**Provide a modality-specific configuration set of default (ANTs) parameters.**

ANTS Atlas Construction– 3 iterations, default parameters

ANTS SyN – default parameters, cross-correlation (intensity-based alignment), SSD (annotation driven alignment)

N4 – default parameters

**2. MERFISH**

***Define the task and detail the associated materials.***

*1. Overview*

We developed a full-scale ANTS pipeline for mapping multiplexed error-robust fluorescence in situ hybridization (MERFISH) spatial transcriptomic data into the Allen common coordinate framework (Allen CCFv3), as previously described in Yao et al. (2023). Mappings are performed by matching gene expression derived region labels from the MERFISH data to corresponding anatomical parcellations of the CCFv3. The pipeline consists of MERFISH data specific preprocessing and two main alignment steps: 1) 3D global affine mapping (12 dof) and section matching of the CCFv3 into the MERFISH data and 2) 2D global and deformable mapping between each MERFISH section and matched CCFv3 section. Mappings learned via each step in the pipeline are preserved and concatenated to provide point-to-point correspondence between the original MERFISH data and CCFv3, thus allowing individual gene expressions to be transferred into the CCFv3. Figure 5 shows an overview of the pipeline, which we describe in detail in the next section.

**Graphical user interface, timeline

Description automatically generated**

**Figure 5.** Pipeline for mapping sectioned spatial transcriptomic (MERFISH) data into the Allen CCFv3 using a two-step alignment with section resampling.

*2. Data*

MERFISH mouse brain data was acquired using the detailed procedure fully described in Yao et al. (2023). Briefly, a brain of C57BL/6 mouse was dissected according to standard procedures and placed into an optimal cutting temperature (OCT) compound (Sakura FineTek 4583) in which it was stored at -80°C. The fresh frozen brain was sectioned at 10 μm on Leica 3050 S cryostats at interval of 200 µm to evenly cover the brain. A set of 500 genes were imaged that had been carefully chosen to distinguish the ~5200 clusters of our existing RNAseq taxonomy. For staining the tissue with MERFISH probes a modified version of instructions provided by the manufacturer was used (Yao et al., 2023). Raw MERSCOPE data were decoded using Vizgen software (v231). Cell segmentation was performed as described previously (Liu, J. et al. (2023). In brief, cells were segmented based on DAPI and PolyT staining using Cellpose (Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods 18, 100–106 (2021).). Segmentation was performed on a median z-plane (4th out of 7) and cell borders were propagated to z-planes above and below. To assign cluster identity to each cell in the MERFISH dataset, we mapped the MERFISH cells to the scRNA-seq reference taxonomy.

**Describe the step-by-step design and implementation of solution approach.**

*2) Preprocessing*

*2.1) Section Reconstruction*

Alignment of MERFISH data into a 3D atlas space requires an estimation of anatomical structure within the data. For each section, this anatomic reference image was created by aggregating the number of detected genetic markers (across all probes) within each pixel of a 10μm x 10μm grid to match the resolution of the 10μm CCFv3 atlas. These reference image sections are then coarsely reoriented and aligned across sections using manual annotations of the most dorsal and ventral points of the midline. The procedure produces an anatomic image stack that serves as an initialization for further global mappings into the CCFv3 (Figure 5).

A collage of images of the brain

Description automatically generated

**Figure 5** Example orientation and translation of four adjacent sections into an initial 3D MERFISH stack (top – Original, bottom- reconstructed)

*2.2) Corresponding Labels*

Mapping the MERFISH data into the CCFv3 requires us to establish correspondence between the MERFISH and CCFv3 anatomy. Intensity-based features in MERFISH data are not apparent enough to establish this correspondence, so we need to generate direct corresponding anatomical labeling of both images. These labels are already available as part of the CCFv3, thus the main challenge is deriving analogous labels from the spatial transcriptomic maps of the MERFISH data. To generate these labels, an we assigned each cell from the scRNA-seq dataset to one of these major regions: cerebellum, CTXsp, hindbrain, HPF, hypothalamus, isocortex, LSX, midbrain, OLF, PAL, sAMY, STRd, STRv, thalamus and hindbrain. A label map of each section was generated for each region by aggregating the cells assigned to that region within a 10×10 μm grid. The same approach was used to generate more fine grained region specific landmarks (i.e. cortical layers, habenula, IC). Unlike the broad labels which cover the entirety of the section these regions are highly specific to certain parts of the section. Once cells in the MERFISH data are labeled, morphological dilation is used to provide full regional labels for alignment into the CCFv3. Figure 6 shows examples of the anatomic labeling derived from this approach.

A picture containing text

Description automatically generated

**Figure 6**. Example MERFISH anatomic labels generated from cell type clustering. Each column corresponds with an existing label in the CCFv3 (left to right – isocortex, thalamus, ventricles, caudate-putamen, and all labels).

*4) Section Matching*

Since the MERFISH data is acquired as sections, its 3D orientation may not be fulling accounted for during the reconstruction step, due to the cutting angle. This can lead to obliqueness artifacts in the section where certain structures can appear to be larger or smaller, or outright missing from the section. To address this, we first use a global alignment to match the orientations of the MERFISH sections to the atlas space. In our pipeline, this section matching is performed in the reverse direction by performing a global affine transformation of the CCFv3 into the MERFISH data space, and then resample digital sections from the CCFv3 to match each MERFISH section. This approach limits the overall transformation that is applied to the MERFISH data, and, since the CCFv3 is densely sampled, it also reduces in-plane artifacts that result from missing sections or undefined spacing in the MERFISH data. Figure 7 shows the original MERFISH data and the CCFv3 after global alignment and section matching.

A collage of images of the brain

Description automatically generated

**Figure 7** Global and local mapping of a MERFISH dataset set into the CCFv3, starting from the reconstructed section labels.

**4) 2.5D Deformable, Landmark Driven Alignment**

After global alignment of the CCFv3 into the MERFISH dataset, 2D per-section deformable refinements are used to address local differences between the MERFISH sections and the resampled CCFv3 sections. Nine registrations were performed in sequence using a single label at each iteration in the following order: 1-Brain mask, 2-Isocortex (layer 2+3), 3-Isocortex (layer 5), 4-Isocortex (layer 6), 5-striatum, 6-medial habenula, 7-lateral habenula, 8-thalamus 9-hippocampus. This ordering was determined empirically by an expert anatomist who prioritized which structure to use in each iteration by evaluating the anatomical alignment from the previous iteration. Global and local mappings are then all concatenated (with appropriate inversions) to create the final mapping between the MERFISH data and CCFv3 (Figure 7). This mapping is then used to provide point-to-point correspondence between the original MERFISH coordinate space and the CCFv3 space, thus allowing mapping of individual genes and cell types located in the MERFISH data to be directly mapped into the CCFv3. (Figure 8)

A picture containing diagram

Description automatically generated

**Figure 8** Examples of individual expressions mapped into the CCFv3.

**Describe the validation performed to evaluate the developed solution.**

Alignment of the MERFISH data into the CCF was qualitatively assessed by an expert anatomist at each iteration of the registration using known correspondence of gene markers and their associations with the CCF. As previously reported in Yao et al. (2023), further assessment of the alignment showed that of the 554 terminal regions (GM only) in the CCF, only 7 small subregions were missed from the MERFISH dataset: frontal pole, layer 1 (FRP1), FRP2/3, FRP5, accessory olfactory bulb, glomerular layer (AOBgl), accessory olfactory bulb, granular layer (AOBgr), accessory olfactory bulb, mitral layer (AOBmi) and accessory supraoptic group (ASO).

**Provide a modality-specific configuration set of default (ANTs) parameters.**

ANTS affine – Default parameters

ANTS SyN – Default parameters, SSD