**Title:** General recipes for mapping diverse mouse cell type data into a common reference space

**Authors:** Min Chen, Yongsoo Kim, Lydia Ng, Nick Tustison, …… James C. Gee

**Institutions:**

**Abstract**

Recent advances in high resolution 3D imaging methods make it possible to examine single-cell resolution data within three dimensional volumes of the mouse brain. The diversity of data types across multiple modalities and scales generates a need to map such data into standardized reference frames, which provides spatial and structural information that can be associated with each cell type data. There is significant variance in the acquisition and imaging protocols across these modalities, and different tissue processing and imaging methods can potentially introduce unique tissue distortion and signal differences. Mapping such diverse data into a common reference brain presents unique challenges for each modality, and it is unlikely that a singular mapping method can be generally applicable to the wide variety of modalities currently being studied. This paper details a collection of recipes, tools, and common practices that enable a modular approach towards performing spatial mapping of diverse high resolution mouse imaging data into a common reference space. Our goal is to provide general guidance and a point of reference that allows investigators in this area to customize a pipeline that is tailored to the challenges of their novel data using openly available tools, without the need to develop specialized software. In addition, we describe standard preprocessing and semi-manual paradigms for addressing common challenges observed in high resolution mouse imaging data. As examples, we demonstrate how the presented recipes can be adapted into distinct pipelines for performing spatial mapping of mouse brain cell type data from three different modalities (FMOST, LSFM, and MERFISH) into a standard reference framework.

**1 Introduction**

Over the past two decades there has been a notable explosion of significant advancements in mesoscopic analysis of the mouse brain. It is now possible to track single cell neurons in three dimensions (3D) across full mouse brains (Keller and Ahrens 2015), observe whole brain developmental changes on a cellular level (La Manno et al. 2021), characterize brain regions and tissues by their genetic composition (Wen et al. 2022), and follow whole tract connectivity across the brain (Oh et al. 2014). Many of these exciting achievements have been enabled by breakthroughs in high resolution imaging techniques that now allows for submicron, 3D imaging of whole mouse brains. Imaging methods such as micro-optical sectioning tomography (Gong et al. 2013; Li et al. 2010), tissue clearing(Ueda et al. 2020; Keller and Ahrens 2015), spatial transcriptomics(Ståhl et al. 2016; Burgess 2019), [others?] have gained significant traction, and utilized by investigators globally to study mesoscale relationships in the brain that previously could not be observed. Many laboratories have developed in-house variations and adaptations of these techniques to meet the needs of their study and experimental goals.[XXX] The result is a diverse range of cell-type imaging modalities and protocols that target different tissue, cells, and genes.[xxx]

To aggregate this rich diversity of data, research groups have developed extensive computational references spaces that serve as atlases to map different types of single-cell mouse data into. Notable examples of these for mouse models include the Allen Brain Atlas and Coordinate Frameworks (Lein et al. 2007; Wang et al. 2020), the Waxholm Space(Johnson et al. 2010) and the XXX atlas. Mapping data into these reference spaces provide a common spatial framework for different data types to be associated with each other, and the reference space themselves each provide spatial and structural information that can be associated with each cell type data. Newly developed imaging modality benefits greatly from being aligned into these common reference spaces. The cross-modality associations that can be learned from being in the same spatial space are critical for improving our understanding of the complex interplay between structures, cells, and genetics in the brain.

One primary challenge associated with these developments is that mapping new cell type data into a reference space is a heterogeneous task. There is significant variance in the acquisition and imaging protocols across different cell type data, and different tissue processing and imaging methods can potentially introduce unique tissue distortion and signal differences. Due to different imaging techniques and/or protocols, there is usually imperfect intensity correspondence between the data and atlas images, and for studies targeting specific tissue or cell types, the data can be missing structural correspondences altogether. A survey of existing mapping techniques for mouse cell type data shows that most such methods are built in-house, each using specialized preprocessing and alignment techniques to target the specific challenges of their data.(Fürth et al. 2018; Leergaard and Bjaalie 2022; Niedworok et al. 2016; Newmaster et al. 2022; Carey et al. 2023; Jin et al. 2022; Koh and McCormick 2003; Ni et al. 2020; Qu et al. 2022) To date there is no single method that can address all the challenges of new cell-type imaging modalities.

These challenges present two limitations for the community. First, many research groups are not equipped with the expertise to develop specialized image mapping tools for their data, resulting in either suboptimal mappings using tools not developed for their data, or more likely, deferring to manual solutions for the mapping, slowing down the scalability of their studies. Second, the development of specialized tools is time consuming, and it is difficult to keep up with the rapidly expanding number of new cell type imaging data. By the time a method is fully developed and validated, new modalities have arisen that require its own mapping solution. Thus, even for groups with strong expertise in this area, mapping approaches are rarely kept up to date with the latest available modalities. General methods for mapping similar data have been proposed, such as by X et al, who have imbedded questionnaire into their pipeline designs so that other investigators can tune the pipelines to their data. Unfortunately, such approaches generally cannot predict the unique artifacts and distortions of new data modalities, nor can they extend too far away from the range of data types that the pipeline were designed for. Practically, due to the wide possible variations in the protocol that can occur during preparation, and it is unlikely that a single general technique or pipeline can be applicable to the multitude of different modalities currently being studied and introduced.

The goal of this paper is to generalize and highlight common patterns of pipeline components that can be found across different mapping approaches. Using these pipeline components, we present a collection of computational recipes that we believe can serve as a starting resource for investigators with novel cell type modalities who aim to set up a specialized mapping pipeline for their data. To aid in this endeavor, we include pre-packaged workflows and general guidelines for the mapping of diverse cell type data into a common coordinate frame built using the Advanced Neuroimaging Tools (ANTS) (Avants et al. 2011) package. The tools and recipes described in this paper are all open sources and provided openly to the community.

**2 Methods**

**2.1 Recipes for Data Preprocessing and Intensity Standardization**

The first challenge in mapping mouse cell type data into a common reference space is the diversity of image intensity appearance across different data type. Rarely do cell type data match the intensity patterns of the target atlas intensity. Thus, almost all mapping to an atlas is a cross-modality registration problem. To address these challenges, different preprocessing methods have been developed that target specific artifacts from each modality. In addition, intensity normalization techniques are used to better harmonize the intensity of the data with the target atlas. These steps, referred to as *preprocessing steps*, are placed at the start of the pipeline, prior to any estimation for mapping between image spaces. We describe below several common preprocessing steps, and their application to our test cases.

**2.1.1 Resampling**

An immediate challenge for high resolution cell type data is the absolute scale of such data. Data from an individual specimen can range on the order of the terabytes, thus all mapping methods presented to date perform some form of downsampling to reduce the resolution of the source data prior to estimating a spatial mapping into the atlas. There’s two primary reasons for this. First, three dimension registration methods, particularly those estimating local deformation, have high computational complexity and cannot operate on such high resolution data under reasonable memory and runtime constraints. Two, the target atlas spaces themselves are generally much lower in resolution, thus the mapping process will lose the high resolution information. This step is necessary to generate an atlas mapping in a reasonable time. These data are typically reduced to the resolution of the atlas space, which is on the order of 10-50 micron. [] Something to note is that while the image spaces are downsampled, the transformations learned are often presumed to be continuous. Thus, after establishing the mapping, it is possible to apply such transformations to the high resolution native data to transform

**2.1.1 Artifact removal**

Repetitive pattern artifacts are a common challenge in novel cell type data that is introduced as part of the acquisition process. For example, in FMOST imaging, ‘stripe’ patterns manifest as a function of the inhomogeneity in the slice imaging process, leaving stripes of higher or lower intensity across the image (figure X). A standard solution that helps address such artifacts is to apply a common bandstop filters targeting the frequency of the strip patterns (see fig). Other examples of such artifacts includes .. in opt images, and ….

Figure 3a, we show an example of stripe artifacts which can be commonly seen in the FMOST images, generated as

**2.1.2 Inhomogeneity correction**

Intensity inhomogeneity can occur within and between sections due to staining or irregularity in the lighting during acquisition. They can get misconstrued as features during the image mapping and result the matching of non-corresponding structures.

**2.1.3 Brain extraction**

<Yongsoo – I have in my notes this is something done in your pipeline, what does it entail? >

**2.2 Recipes for Image Registration**

Methods for performing the spatial mapping between two image spaces are referred to as *image registration*. Such methods have been a main stay in the field medical image analysis, and has acquired a large body of literature spanning decades[]. Since many different registration approaches are applicable, here we attempt to describe their application for cell type mapping as generally as possible, without delving into the specifics of any single registration approach. For the purpose of examples, we employ the ANTS package for performing these alignments, but equivalent methods have been used with great success though out the literature.[]

**2.2.1 Global Mapping**

Registration approaches can be broadly classified by the transformation constraints on the mapping. The most standard and first step is a global transform applied to the full brain, defined by an affine matrix. This transformation accounts for rotation, translation, scale and skew (obliqueness) of the brain. Global transformations are used universally as the first step of the alignment, and serves an initial mapping between the image and the atlas.

**2.2.1 Section matching through piecewise mapping**

Certain cell type data imaging acquisition still require sectioning of the data into slices. These sections can have tilts that are not fulling accounted for in the reconstruction (due to the cutting angle) or can have independent movement due to variance in the slicing mechanism. This leads to obliqueness artifacts in the slice that may give the image the appearance of expansion or contraction. Also, structural correspondence may be lost if a section is cut with significant slant such that structures one may typically expect in the slice may not appear. To address this, one can adjust the orientation of the section relative to the atlas space. However, in practice, this section is matching is often done in reverse, where the atlas is sampled at particular sections to match the original data. This limits the transformation that has to be applied to the image, and the atlases tend to be more dense

**2.3 Manual initialization**

There are circumstances where manual initialization of the sections is necessary to, this involves simple tasks such as matching the midline, placing landmarks, or parcellation of large structures that can serve as an initial placement of the brain.

**2.4 Local Registration**

Once a global registration is established, the next step is to establish local correspondence. This is typically performed by matching the intensity between the image and the atlas, or by establishing explicit correspondence between labeled structures.

**2.4.1 Intensity driven Atlas Alignment**

Intensity based alignment uses a cost function to determine how similar the intensity of the registered image and the atlas are to each other. These can be direct intensity comparisons (e.g. cross-correlation[]) or statistic measures of similarity (e.g. mutual information[]). The benefit of intensity based approaches is they do not require us to provide correspondence between the images or

**2.4.1 Parcellation and landmark driven alignments**

Parcellation based alignment are performed by generating segmentations of structures that are analogous to structures in the atlas. These segmentations are then registered directly used between each other. Many approaches have been proposed for how to generate these segmentations, both through automated and semi-automatic approaches.[]

**2.4.3 Modality specific atlas as an intermediate target**

There are times when cross-modality alignment is insufficient to generate an accurate mapping due significant intensity or morphological differences between the image and the atlas, and corresponding parcellations can’t be established for images from every subject. In these cases, one option is to create a modality-specific atlas using the data to serve as an intermediate registration point. This has the advantage of only needing to generate one set of corresponding annotations between the two atlases, and then future images can be aligned to just the modality specific atlas.

**2.3 Post mapping processing**

* + Tools for transforming/mapping single cell neuron reconstruction data (.swc) into CCF
  + Multi-modal registration:
    - High resolution LSFM/STPT to MRI templates
    - Importing anatomical labels into the new templates
  + Tools for transforming/mapping signals into new Dec CCF

**2.4 Example recipes for mapping three distinct cell type data**

This section describes example recipes we have prebuilt for the alignment of 3 different data modalities into a common references frame using combinations of the steps. Each recipe was tailored to the specific challenges of each data type. The goal is to show how a pipeline can be adapted depending on the needs of the data.

**2.4.1 Data**

2.1.1 FMOST

<Get data description from Lydia>

* + Average FMOST Atlas
  + CCF
  + 12 Label corresponding annotations between FMOST and CCF structures
  + ~25 Individual fmost brains with neuron reconstructions

2.1.2 MERFISH, Spatial Transcriptomics

<Get data description from Lydia>

* + 1-2 mouse
  + ~20 corresponding landmarks between MERFISH and CCF data

2.1.3 Lightsheet and MRI

<Get data and pipeline description from Yongsoo>

* + - Population average (intensity and morphology) at each age
    - Multi-modal registration, potentially cross-time registration between different age groups.
    - *Creating 3D labels (this can be optional).*

**3 Recipes for Validation**

Challenges due to the diversity of cell type data do not end at the registration step. While registration methods can be evaluated use standard alignment metrics (such as the Dice’s coefficient) to determine accurate alignment of the structures. This rarely tells the whole story. The alignment of the landmarks or parcellations does not necessarily translate to alignment of the cell type data.

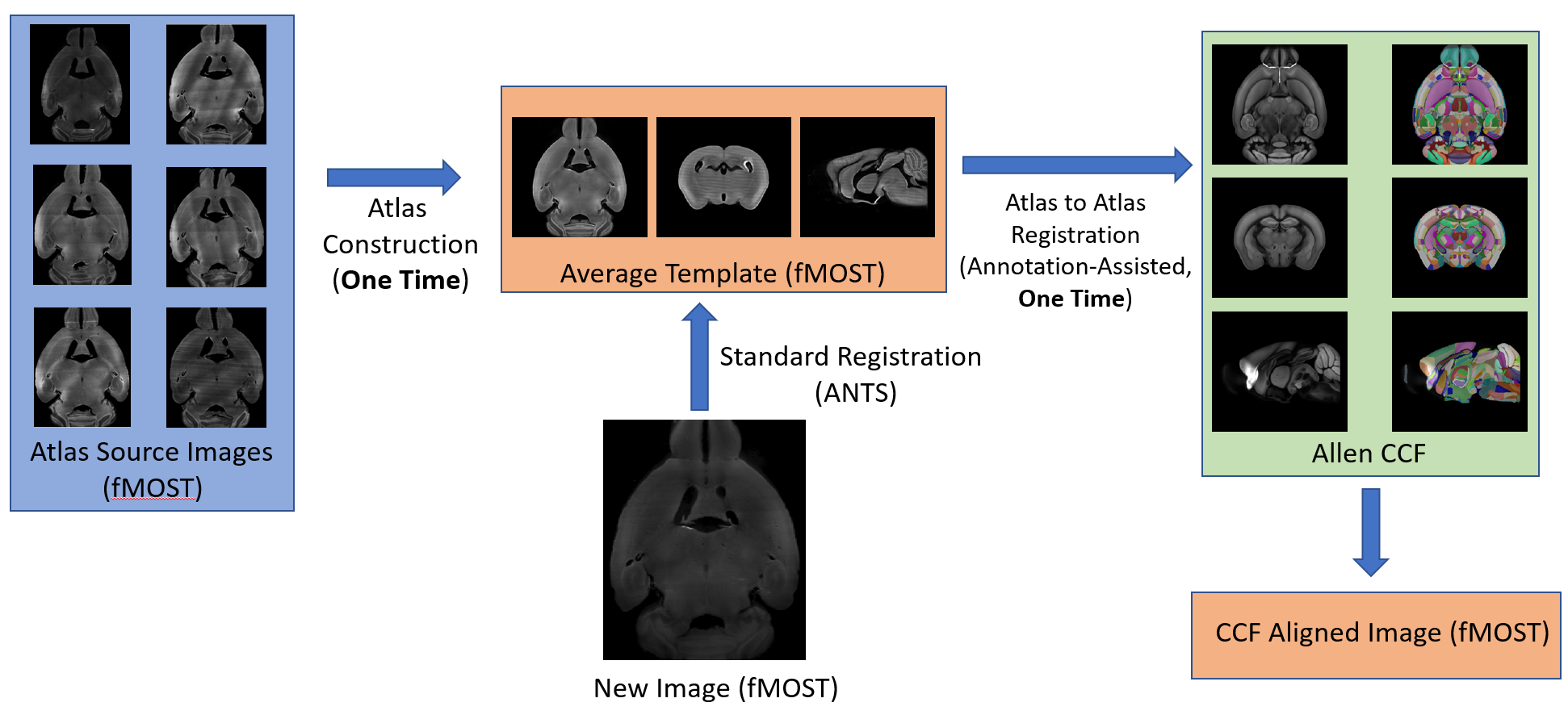
* UPenn/ABIS – FMOST
  + Alignment of corresponding annotations in average FMOST Atlas and CCFF
    - Registration accuracy of canonical transform (Dice’s Coeff)
  + Single Cell soma locations in individual brain
    - Anatomical assessment of where soma lands after alignment into CCF
* Penn State
  + Alignment between average LSFM/STPT and MRI
    - Registration accuracy of canonical transform (Dice’s Coeff)
  + Cell type data or gene expression alignment with anatomical labels.

**4 Discussion**

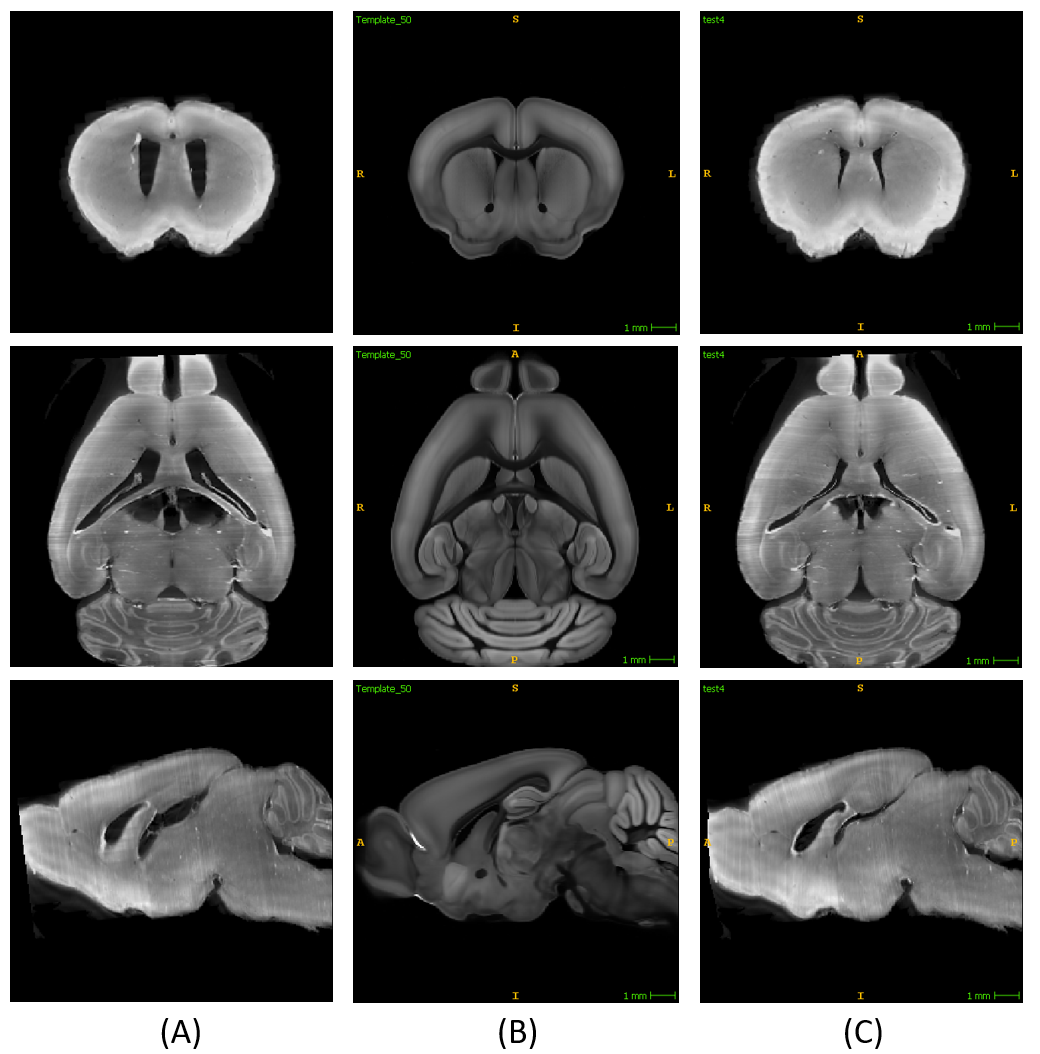
**5 Conclusion**

In this work we have described several recipes for the design of cell type mapping across diverse data sets. We presented applications of these recipes for three different imaging modalities. These recipes can be flexibly adapted to create new template from different tissue processing and imaging methods, or other non-traditional species.

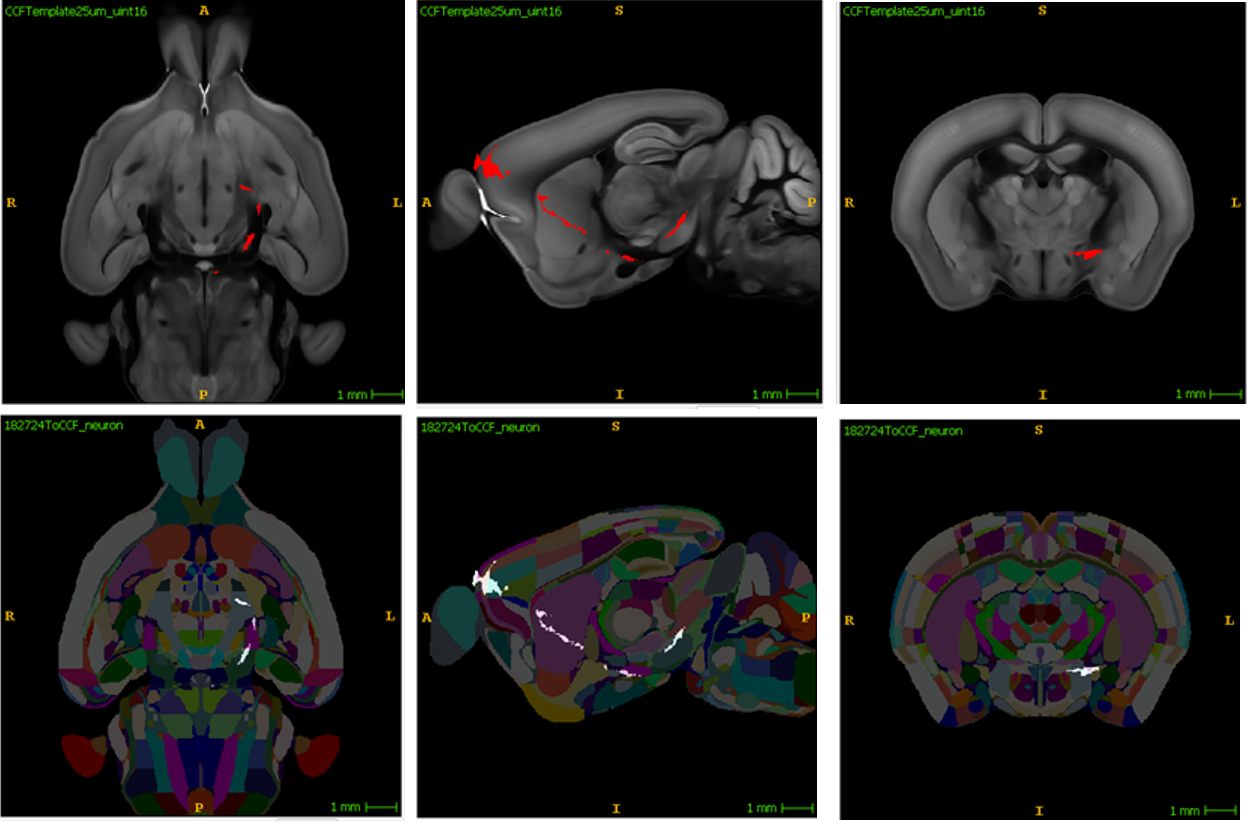
**Figures**

****

**Figure 1.** Recipe demonstrating a general pipeline for performing mapping using a modality specific atlas as an intermediate step for mapping fMOST data into the Allen CCF.



**Figure 2.** Example mapping of a fMOST image into the Allen CCF. Shown are the triplanar view of (A) unregistered fMOST image, (B) Allen CCF atlas, and (C) the FMOST image after alignment.



**Figure 3.** Mapping of a reconstructed neuron (shown in red) generated from the fMOST image (top row) into the Allen CCF (bottom row).

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

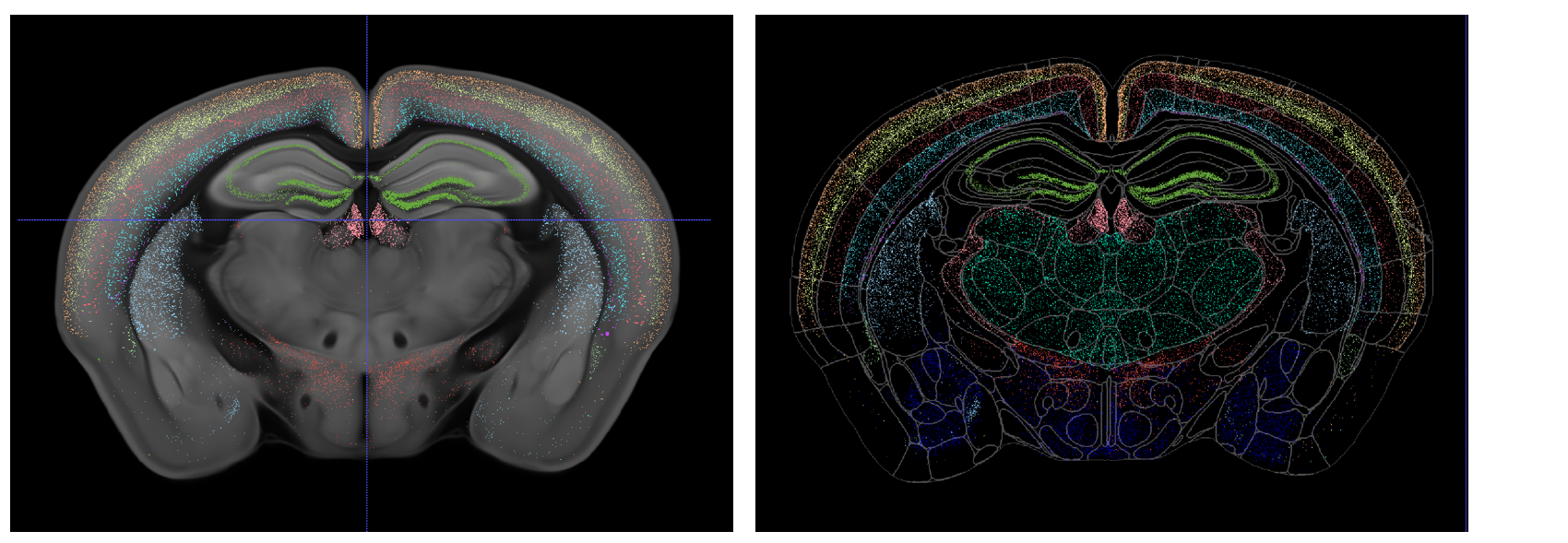
A line of black text

Description automatically generated

**Graphical user interface, timeline

Description automatically generated**

**Figure 4.** Recipe demonstrating how sectioned spatial transcriptomic (MERFISH) data can be mapped into the Allen CCF using a two-step alignment with section resampling.



**Figure 5.** Example of select genes from the MERFISH data mapped into the CCF.

**Bibliography**

Avants, Brian B., Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. 2011. “A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration.” *NeuroImage* 54 (3): 2033–44. https://doi.org/10.1016/j.neuroimage.2010.09.025.

Burgess, Darren J. 2019. “Spatial Transcriptomics Coming of Age.” *Nature Reviews Genetics* 20 (6): 317–317. https://doi.org/10.1038/s41576-019-0129-z.

Carey, Harry, Michael Pegios, Lewis Martin, Chris Saleeba, Anita J. Turner, Nicholas A. Everett, Ingvild E. Bjerke, Maja A. Puchades, Jan G. Bjaalie, and Simon McMullan. 2023. “DeepSlice: Rapid Fully Automatic Registration of Mouse Brain Imaging to a Volumetric Atlas.” *Nature Communications* 14 (1): 5884. https://doi.org/10.1038/s41467-023-41645-4.

Fürth, Daniel, Thomas Vaissière, Ourania Tzortzi, Yang Xuan, Antje Märtin, Iakovos Lazaridis, Giada Spigolon, et al. 2018. “An Interactive Framework for Whole-Brain Maps at Cellular Resolution.” *Nature Neuroscience* 21 (1): 139–49. https://doi.org/10.1038/s41593-017-0027-7.

Gong, Hui, Shaoqun Zeng, Cheng Yan, Xiaohua Lv, Zhongqin Yang, Tonghui Xu, Zhao Feng, et al. 2013. “Continuously Tracing Brain-Wide Long-Distance Axonal Projections in Mice at a One-Micron Voxel Resolution.” *NeuroImage* 74 (July): 87–98. https://doi.org/10.1016/j.neuroimage.2013.02.005.

Jin, Michelle, Joseph D. Nguyen, Sophia J. Weber, Carlos A. Mejias-Aponte, Rajtarun Madangopal, and Sam A. Golden. 2022. “SMART: An Open-Source Extension of WholeBrain for Intact Mouse Brain Registration and Segmentation.” *Eneuro* 9 (3): ENEURO.0482-21.2022. https://doi.org/10.1523/ENEURO.0482-21.2022.

Johnson, G. Allan, Alexandra Badea, Jeffrey Brandenburg, Gary Cofer, Boma Fubara, Song Liu, and Jonathan Nissanov. 2010. “Waxholm Space: An Image-Based Reference for Coordinating Mouse Brain Research.” *NeuroImage* 53 (2): 365–72. https://doi.org/10.1016/j.neuroimage.2010.06.067.

Keller, Philipp J., and Misha B. Ahrens. 2015. “Visualizing Whole-Brain Activity and Development at the Single-Cell Level Using Light-Sheet Microscopy.” *Neuron* 85 (3): 462–83. https://doi.org/10.1016/j.neuron.2014.12.039.

Koh, Wonryull, and Bruce H. McCormick. 2003. “Registration of a 3D Mouse Brain Atlas with Brain Microstructure Data.” *Neurocomputing*, Computational Neuroscience: Trends in Research 2003, 52–54 (June): 307–12. https://doi.org/10.1016/S0925-2312(02)00793-2.

La Manno, Gioele, Kimberly Siletti, Alessandro Furlan, Daniel Gyllborg, Elin Vinsland, Alejandro Mossi Albiach, Christoffer Mattsson Langseth, et al. 2021. “Molecular Architecture of the Developing Mouse Brain.” *Nature* 596 (7870): 92–96. https://doi.org/10.1038/s41586-021-03775-x.

Leergaard, Trygve B., and Jan G. Bjaalie. 2022. “Atlas-Based Data Integration for Mapping the Connections and Architecture of the Brain.” *Science* 378 (6619): 488–92. https://doi.org/10.1126/science.abq2594.

Lein, Ed S., Michael J. Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F. Boe, et al. 2007. “Genome-Wide Atlas of Gene Expression in the Adult Mouse Brain.” *Nature* 445 (7124): 168–76. https://doi.org/10.1038/nature05453.

Li, Anan, Hui Gong, Bin Zhang, Qingdi Wang, Cheng Yan, Jingpeng Wu, Qian Liu, Shaoqun Zeng, and Qingming Luo. 2010. “Micro-Optical Sectioning Tomography to Obtain a High-Resolution Atlas of the Mouse Brain.” *Science* 330 (6009): 1404–8. https://doi.org/10.1126/science.1191776.

Newmaster, Kyra T., Fae A. Kronman, Yuan-ting Wu, and Yongsoo Kim. 2022. “Seeing the Forest and Its Trees Together: Implementing 3D Light Microscopy Pipelines for Cell Type Mapping in the Mouse Brain.” *Frontiers in Neuroanatomy* 15. https://www.frontiersin.org/articles/10.3389/fnana.2021.787601.

Ni, Hong, Chaozhen Tan, Zhao Feng, Shangbin Chen, Zoutao Zhang, Wenwei Li, Yue Guan, Hui Gong, Qingming Luo, and Anan Li. 2020. “A Robust Image Registration Interface for Large Volume Brain Atlas.” *Scientific Reports* 10 (1): 2139. https://doi.org/10.1038/s41598-020-59042-y.

Niedworok, Christian J., Alexander P. Y. Brown, M. Jorge Cardoso, Pavel Osten, Sebastien Ourselin, Marc Modat, and Troy W. Margrie. 2016. “aMAP Is a Validated Pipeline for Registration and Segmentation of High-Resolution Mouse Brain Data.” *Nature Communications* 7 (1): 11879. https://doi.org/10.1038/ncomms11879.

Oh, Seung Wook, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, et al. 2014. “A Mesoscale Connectome of the Mouse Brain.” *Nature* 508 (7495): 207–14. https://doi.org/10.1038/nature13186.

Qu, Lei, Yuanyuan Li, Peng Xie, Lijuan Liu, Yimin Wang, Jun Wu, Yu Liu, et al. 2022. “Cross-Modal Coherent Registration of Whole Mouse Brains.” *Nature Methods* 19 (1): 111–18. https://doi.org/10.1038/s41592-021-01334-w.

Ståhl, Patrik L., Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. “Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics.” *Science* 353 (6294): 78–82. https://doi.org/10.1126/science.aaf2403.

Ueda, Hiroki R., Ali Ertürk, Kwanghun Chung, Viviana Gradinaru, Alain Chédotal, Pavel Tomancak, and Philipp J. Keller. 2020. “Tissue Clearing and Its Applications in Neuroscience.” *Nature Reviews Neuroscience* 21 (2): 61–79. https://doi.org/10.1038/s41583-019-0250-1.

Wang, Quanxin, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, et al. 2020. “The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas.” *Cell* 181 (4): 936-953.e20. https://doi.org/10.1016/j.cell.2020.04.007.

Wen, Lu, Guoqiang Li, Tao Huang, Wei Geng, Hao Pei, Jialiang Yang, Miao Zhu, et al. 2022. “Single-Cell Technologies: From Research to Application.” *The Innovation* 3 (6): 100342. https://doi.org/10.1016/j.xinn.2022.100342.