

1

2 **Modular strategies for spatial mapping of**
3 **diverse cell type data of the mouse brain**

4 Nicholas J. Tustison¹, Min Chen², Fae N. Kronman³, Jeffrey T. Duda², Clare Gamlin⁴, Mia
5 G. Tustison, Michael Kunst⁴, Rachel Dalley⁴, Staci Sorenson⁴, Quanxin Wang⁴, Lydia Ng⁴,
6 Yongsoo Kim³, and James C. Gee²

7 ¹Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

8 ²Department of Radiology, University of Pennsylvania, Philadelphia, PA

9 ³Department of Neural and Behavioral Sciences, Penn State University, Hershey, PA

10 ⁴Allen Institute for Brain Science, Seattle, WA

11 _____

12 Corresponding authors:

13

14 Nicholas J. Tustison, DSc

15 Department of Radiology and Medical Imaging

16 University of Virginia

17 ntustison@virginia.edu

18

19 James C. Gee, PhD

20 Department of Radiology

21 University of Pennsylvania

22 gee@upenn.edu

²³ **Abstract**

²⁴ Large-scale, international collaborative efforts by members of the BRAIN Initiative Cell Census
²⁵ Network (BICCN) consortium have recently begun aggregating the most comprehensive
²⁶ reference database to date for diverse cell type profiling of the mouse brain, which encompasses
²⁷ over 40 different multi-modal profiling techniques from more than 30 research groups. One
²⁸ central challenge for this integrative effort across different investigators and laboratories has
²⁹ been the need to map these unique datasets into common reference spaces such that the
³⁰ spatial, structural, and functional information from different cell types can be jointly analyzed
³¹ across modalities. However, significant variations in the acquisition, tissue processing, and
³² imaging techniques across data types makes mapping such diverse data a multifarious problem.
³³ Different data types exhibit unique tissue distortion and signal characteristics that precludes
³⁴ a single mapping strategy from being generally applicable across all cell type data. Diverse,
³⁵ and often specialized, mapping approaches are needed to address the unique barriers present
³⁶ in each modality. This work highlights modular atlas mapping strategies developed across
³⁷ three separate BICCN studies using the ANTsX framework to map spatial transcriptomic
³⁸ (MERFISH) and high-resolution morphology (fMOST) mouse brain data into the Allen
³⁹ Common Coordinate Framework (AllenCCFv3), and developmental (MRI and LSFM) data
⁴⁰ into the Developmental Common Coordinate Framework (DevCCF). We discuss both common
⁴¹ mapping strategies that can be shared across modalities, and targeted strategies driven by
⁴² specific challenges from each data type. Novel contributions include velocity flow-based
⁴³ approaches for mapping longitudinal trajectory in the DevCCF, and automated structural
⁴⁴ morphological approaches for determining cortical parcellations. Finally, we provide general
⁴⁵ guidance and open source tools to aid investigators in this effort to tailor these strategies to
⁴⁶ address unique challenges in their data without the need to develop additional specialized
⁴⁷ software.

48 1 Introduction

49 Over the past decade there have been significant advancements in mesoscopic single-cell
50 analysis of the mouse brain. It is now possible to track single neurons in mouse brains¹,
51 observe whole brain developmental changes on a cellular level², associate brain regions
52 and tissues with their genetic composition³, and locally characterize neural connectivity⁴.

53 Much of these scientific achievements have been made possible due to breakthroughs in high
54 resolution cell profiling and imaging techniques that permit submicron, multi-modal, 3D
55 characterizations of whole mouse brains. Among these include advanced techniques such
56 as micro-optical sectioning tomography⁶, tissue clearing^{1,7}, spatial transcriptomics⁹, and
57 single-cell genomic profiling¹⁰, which have greatly expanded the resolution and specificity of
58 single-cell measurements in the brain.

59 Recent efforts by the National Institutes of Health’s Brain Research Through Advancing
60 Innovative Neurotechnologies (BRAIN) Initiative has pushed for large-scale, international
61 collaborative efforts to utilize these advanced single cell techniques to create a comprehensive
62 reference database for high-resolution transcriptomic, epigenomic, structural and imaging
63 data of the mouse brain. This consortium of laboratories and data centers, known as the
64 BRAIN Initiative Cell Census Network (BICCN), has to date archived datasets encompassing
65 over 40 different multi-modal profiling techniques from more than 30 research groups, each
66 providing unique characterizations of distinct cell types in the brain.¹¹ Several of these
67 modalities have been further developed into reference atlases to facilitate spatial alignment
68 of individual brains and different data types into a common coordinate framework (CCF),
69 thus allowing diverse single-cell information to be integrated and analyzed in tandem. The
70 most notable of these atlases is the Allen Mouse Brain Common Coordinate Framework
71 (AllenCCFv3)¹², which serves as the primary target coordinate space that the majority of
72 BICCN mouse data are mapped to. Other atlases include modality specific atlases^{13–15}, and
73 spatiotemporal atlases^{16,17} for the developing mouse brain.

⁷⁴ 1.1 Mouse brain mapping

⁷⁵ The cross-modality associations that can be learned from mapping different cell type data
⁷⁶ into a CCF is critical for improving our understanding of the complex relationships between
⁷⁷ cellular structure, morphology, and genetics in the brain. However, finding an accurate
⁷⁸ mapping between each individual mouse brain and a CCF is a challenging and heterogeneous
⁷⁹ task. There is significant variance in the acquisition, fixation and imaging protocols across
⁸⁰ different cell type data, and different tissue processing and imaging methods can potentially
⁸¹ introduce modality specific tissue distortion and signal differences.^{18,19} Certain modalities
⁸² can have poor intensity correspondence with the CCF, making image alignment less robust.
⁸³ Studies targeting specific regions or cell types can lead to missing anatomical correspondences.
⁸⁴ Other considerations include artifacts such as tissue distortion, holes, bubbles, folding, tears,
⁸⁵ and missing sections in the data that need to be addressed on a per-case basis.^{20–23} Given
⁸⁶ the diversity of these challenges, it is unlikely any single mapping approach can be generally
⁸⁷ applicable across all cell type data. Diverse, and often specialized, strategies are needed to
⁸⁸ address the unique barriers present for mapping each modality.

⁸⁹ Existing solutions to address mapping cell type data into the AllenCCFv3 falls broadly into
⁹⁰ three main categories. The first consists of integrated processing platforms that directly
⁹¹ provide mapped data to the users. These include the Allen Brain Cell Atlas²⁴ for the Allen
⁹² Reference Atlas (ARA) and associated data, the Brain Architecture Portal²⁵ for combined
⁹³ ex vivo radiology and histology data, Openbrainmap²⁶ for connectivity data, and the Image
⁹⁴ and Multi-Morphology Pipeline²⁷ for high resolution morphology data. These platforms
⁹⁵ provide users online access to pre-processed, multi-modal cell type data that are already
⁹⁶ mapped to the AllenCCFv3. The platforms are designed such that the data is interactable by
⁹⁷ users through integrated visualization software that allow users to spatially manipulate and
⁹⁸ explore each dataset within the mapped space. While highly convenient for investigators who
⁹⁹ are interested in studying the specific modalities provided by these platforms, these system
¹⁰⁰ are limited in flexibility and general applicability. The mapping software and pipeline are
¹⁰¹ typically developed specifically with the data type and platform in mind, and the software
¹⁰² are often not openly accessible to the public. Investigators will find it difficult to apply the

103 same mapping to their own data without direct collaboration with the platform owners.

104 The second category are specialized approaches specifically designed for mapping one or
105 more modalities into a CCF. These approaches use combinations of specialized manual
106 and automated processes that address specific challenges in each modality. Examples
107 include approaches for mapping histology^{28–30}, magnetic resonance imaging (MRI)^{30–37}, micro-
108 computed tomography (microCT)^{35,37}, light-sheet fluorescence microscopy (LSFM)^{34,36–39},
109 fluorescence micro-optical sectioning tomography (fMOST)^{15,40} and transcriptomic^{41–43} data.
110 As specialized approaches, these techniques tend to boast higher mapping accuracy, robustness,
111 and ease of use when ran with applicable modalities. Conversely, their specialized designs
112 often rely on base assumptions regarding the data type that can make them rigid and difficult
113 to adapt for new modalities or unexpected artifacts and distortions in the data. Retooling
114 these specialize software to use with new data can require significant development, validation
115 time, and engineering expertise that may not be readily available for all investigators.

116 The last category are modular mapping approaches constructed using general image analysis
117 toolkits, which are software packages that include varied collections of image processing,
118 segmentation and registration tools that have been previously developed, and validated for
119 general use. Examples of such toolkits include elastix⁴⁴, slicer3D⁴⁵, and ANTsX⁴⁶, which
120 have all been applied towards mapping mouse cell type data. The main challenge for using
121 these approaches is that the individual tools in the toolbox are not data type specific. Thus,
122 investigators must construct pipelines that link together a variety of tools to address data
123 specific problem, and certain tools may still require specialized input data such as landmarks
124 or annotations to operate. Investigators must be familiar with the toolkits and how to use
125 them to build or adapt pipelines for new data types. However, in comparison to previously
126 described specialized mapping approaches, toolbox driven pipelines are often easier to create,
127 making them more accessible for the general investigator, and individual pieces of the pipeline
128 have already been validated from other efforts. Using a general toolkit allows for modular
129 mapping strategies that can handle a wide array of different data types by piecing together
130 distinct solutions for modality specific problems. In this work, we highlight such mapping
131 strategies designed using the ANTsX framework to map three distinct mouse cell type data

¹³² with different characteristics into CCFs.

¹³³ 1.2 Advanced Normalization Tools (ANTsX)

¹³⁴ The Advanced Normalization Tools (ANTsX) framework has been used in a number of
¹³⁵ applications for mapping mouse brain data as part of core processing steps in various
¹³⁶ workflows^{30,47–50}, particularly its pairwise, intensity-based image registration capabilities and
¹³⁷ bias field correction. Historically, ANTsX development is originally based on fundamental
¹³⁸ approaches to image mapping^{51–53}, particularly in the human brain, which has resulted
¹³⁹ in core contributions to the field such as the widely-used Symmetric Normalization (SyN)
¹⁴⁰ algorithm⁵⁴. Since its development, various independent platforms have been used to evaluate
¹⁴¹ ANTsX image registration capabilities in the context of different application foci which
¹⁴² include multi-site brain MRI data⁵⁵, pulmonary CT data⁵⁶, and most recently, multi-modal
¹⁴³ brain registration in the presence of tumors⁵⁷.

¹⁴⁴ Apart from its registration capabilities, ANTsX comprises additional functionality such
¹⁴⁵ as template generation⁵⁸, intensity-based segmentation⁵⁹, preprocessing^{60,61}, deep learning
¹⁴⁶ networks⁴⁶, and other utilities relevant to brain mapping (see Methods - Table 1). The use of
¹⁴⁷ the toolkit has demonstrated high performance in multiple application areas (e.g., consensus
¹⁴⁸ labeling⁶², brain tumor segmentation⁶³, and cardiac motion estimation⁶⁴). Importantly,
¹⁴⁹ ANTs is built on the Insight Toolkit (ITK)⁶⁵ deriving benefit from the open-source community
¹⁵⁰ of scientists and programmers as well as providing an important resource for algorithmic
¹⁵¹ development, evaluation, and improvement. In this paper we demonstrate how ANTs'
¹⁵² comprehensive toolset provides the basis to develop modular frameworks for mapping diverse
¹⁵³ mouse cell type data into common coordinate frameworks (CCFs). Specifically, we highlight
¹⁵⁴ its application for mapping data from three separate BICCN projects focused on distinct data
¹⁵⁵ types: morphology data using fluorescence micro-optical sectioning tomography (fMOST),
¹⁵⁶ spatial transcriptomics from multiplexed error-robust fluorescence *in situ* hybridization
¹⁵⁷ (MERFISH) data, and time-series developmental data using light sheet fluorescence microscopy
¹⁵⁸ (LSFM) and magnetic resonance imaging (MRI). We describe both shared and targeted
¹⁵⁹ strategies developed to address the specific challenges of these modalities.

160 **1.3 Novel ANTsX-based open-source contributions**

161 We introduce two novel inclusions to the ANTsX toolset that were developed as part of the
162 MRI mapping and analysis pipeline for the Developmental Common Coordinate Framework
163 (DevCCF). Consistent with previous ANTsX development, newly introduced capabilities
164 introduced below are available through ANTsX (specifically, via R and Python ANTsX
165 packages), and illustrated through self-contained examples in the ANTsX tutorial (<https://tinyurl.com/antsxtutorial>) with a dedicated GitHub repository specific to this work
166 (<https://github.com/ntustison/ANTsXMouseBrainMapping>). To complement standard
167 preprocessing steps (e.g., bias correction, brain masking), additional mouse brain specific
168 tools have also been introduced to the ANTsX ecosystem, such as section reconstruction
169 and landmark-based alignment with corresponding processing scripts available at <https://github.com/dontminchenit/CCFAAlignmentToolkit>.
170
171

172 **1.3.1 The DevCCF velocity flow model**

173 Recently, the Developmental Common Coordinate Framework (DevCCF) was introduced to
174 the mouse brain research community as a public resource¹⁶ comprising symmetric atlases of
175 multimodal image data and anatomical segmentations defined by developmental ontology.
176 These templates sample the mouse embryonic days (E) 11.5, E13.5, E15.5, E18.5 and postnatal
177 day (P) 4, P14, and P56. Modalities include light sheet fluorescence microscopy (LSFM)
178 and at least four MRI contrasts per developmental stage. Anatomical parcellations are
179 also available for each time point and were generated from ANTsX-based mappings of gene
180 expression and other cell type data. Additionally, the P56 template was integrated with
181 the AllenCCFv3 to further enhance the practical utility of the DevCCF. These processes,
182 specifically template generation and multi-modal image mapping, were performed using
183 ANTsX functionality in the presence of image mapping difficulties such as missing data and
184 tissue distortion.¹⁶

185 Given the temporal gaps in the discrete set of developmental atlases, we also provide an
186 open-source framework for inferring correspondence within the temporally continuous domain
187 sampled by the existing set of embryonic and postnatal atlases of the DevCCF. This recently

¹⁸⁸ developed functionality permits the generation of a diffeomorphic velocity flow transformation
¹⁸⁹ model⁶⁶, influenced by previous work⁶⁷. The resulting time-parameterized velocity field spans
¹⁹⁰ the stages of the DevCCF where mappings between any two continuous time points within
¹⁹¹ the span bounded by the E11.5 and P56 atlases is determined by integration of the optimized
¹⁹² velocity field.

¹⁹³ **1.3.2 Structural morphology and parcellations of the mouse brain**

¹⁹⁴ In contrast to the pipeline development in human data⁴⁶, limited tools exist yet to create
¹⁹⁵ adequate training data for automated parcellations of the mouse brain. In addition, mouse
¹⁹⁶ brain data acquisition often has unique issues, such as lower data quality or sampling
¹⁹⁷ anisotropy which limits its applicability to high resolution resources (e.g., AllenCCFv3,
¹⁹⁸ DevCCF), specifically with respect to the corresponding granular brain parcellations derived
¹⁹⁹ from numerous hours of expert annotation leveraging multimodal imaging resources.

²⁰⁰ Herein, we introduce a mouse brain parcellation pipeline for T2-weighted (T2-w) MRI
²⁰¹ comprising two novel deep learning components: two-shot learning brain extraction from data
²⁰² augmentation of two ANTsX templates generated from two open datasets^{68,69} and single-shot
²⁰³ brain parcellation derived from the AllenCCFv3 labelings mapped to the corresponding
²⁰⁴ DevCCF P56 T2-w component. Although we anticipate that this pipeline will be beneficial
²⁰⁵ to the research community, this work demonstrates more generally how one can leverage
²⁰⁶ ANTsX tools for developing tailored brain parcellation schemes using publicly available
²⁰⁷ resources. Evaluation is performed on an independent open dataset⁷⁰ comprising longitudinal
²⁰⁸ acquisitions of multiple specimens.

209 **2 Results**

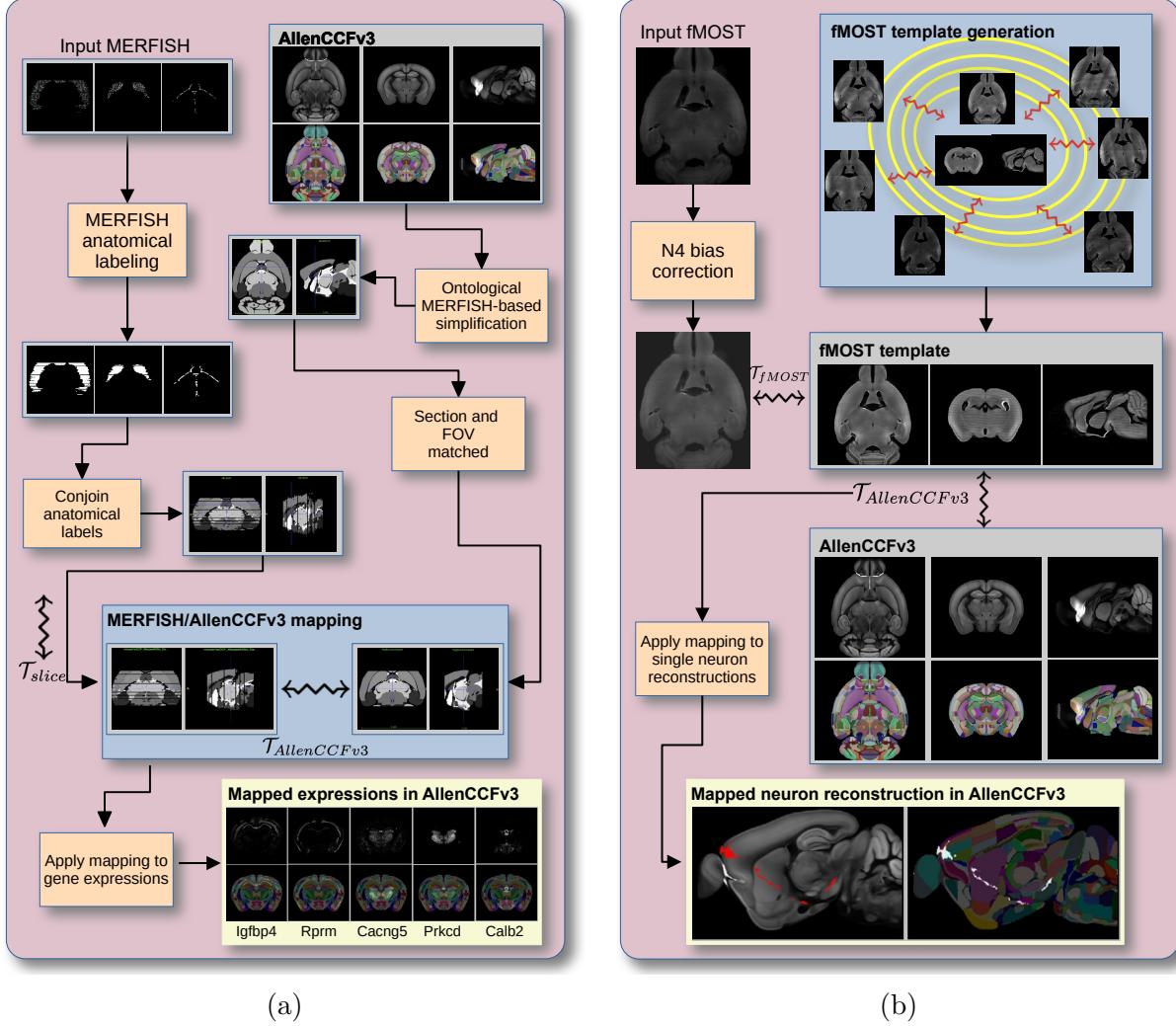


Figure 1: Diagram of the two ANTsX-based pipelines for mapping (a) MERFISH and (b)fMOST data into the space of AllenCCFv3. Each generates the requisite transforms, \mathcal{T} , to map individual images.

210 **2.1 AllenCCFv3 brain image mapping**

211 **2.1.1 Mapping multiplexed error-robust fluorescence *in situ* hybridization
212 (MERFISH) data**

213 **Overview.** The ANTsX framework was used to develop a pipeline for mapping multiplexed
214 error-robust fluorescence *in situ* hybridization (MERFISH) spatial transcriptomic mouse
215 data onto the AllenCCFv3 (see Figure 1(a)). This pipeline, used recently in creating

216 a high-resolution transcriptomic atlas of the mouse brain⁵⁰, performs mappings by first
217 generating anatomical labels from tissue related gene expressions in the MERFISH data, and
218 then spatially matching these labels to corresponding anatomical tissue parcellations in the
219 AllenCCFv3. The pipeline consists of MERFISH data specific preprocessing which includes
220 section reconstruction, mapping corresponding anatomical labels between AllenCCFv3 and
221 the spatial transcriptomic maps of the MERFISH data, and matching MERFISH sections to
222 the atlas space. Following pre-processing, two main alignment steps were performed: 1) 3-D
223 global affine mapping and section matching of the AllenCCFv3 into the MERFISH data and 2)
224 2D global and deformable mapping between each MERFISH section and matched AllenCCFv3
225 section. Mappings learned via each step in the pipeline are preserved and concatenated to
226 provide point-to-point correspondence between the original MERFISH data and AllenCCFv3,
227 thus allowing individual gene expressions to be transferred into the AllenCCFv3.

228 **Data.** MERFISH mouse brain data was acquired using a previously detailed procedure⁵⁰.
229 Briefly, a brain of C57BL/6 mouse was dissected according to standard procedures and placed
230 into an optimal cutting temperature (OCT) compound (Sakura FineTek 4583) in which it
231 was stored at -80°C. The fresh frozen brain was sectioned at 10 μm on Leica 3050 S cryostats
232 at intervals of 200 μm to evenly cover the brain. A set of 500 genes were imaged that had
233 been carefully chosen to distinguish the ~ 5200 clusters of our existing RNAseq taxonomy.
234 For staining the tissue with MERFISH probes, a modified version of instructions provided by
235 the manufacturer was used⁵⁰. Raw MERSCOPE data were decoded using Vizgen software
236 (v231). Cell segmentation was performed⁷¹. In brief, cells were segmented based on DAPI
237 and PolyT staining using Cellpose⁷². Segmentation was performed on a median z-plane
238 (fourth out of seven) and cell borders were propagated to z-planes above and below. To assign
239 cluster identity to each cell in the MERFISH dataset, we mapped the MERFISH cells to the
240 scRNA-seq reference taxonomy.

241 **Evaluation.** Alignment of the MERFISH data into the AllenCCFv3 was qualitatively assessed
242 by an expert anatomist at each iteration of the registration using known correspondence of
243 gene markers and their associations with the AllenCCFv3. As previously reported⁵⁰, further
244 assessment of the alignment showed that, of the 554 terminal regions (gray matter only)

²⁴⁵ in the AllenCCFv3, only seven small subregions were missed from the MERFISH dataset:
²⁴⁶ frontal pole, layer 1 (FRP1), FRP2/3, FRP5; accessory olfactory bulb, glomerular layer
²⁴⁷ (AOBgl); accessory olfactory bulb, granular layer (AOBgr); accessory olfactory bulb, mitral
²⁴⁸ layer (AOBmi); and accessory supraoptic group (ASO).

²⁴⁹ 2.1.2 Mapping fluorescence micro-optical sectioning tomography (fMOST) data

²⁵⁰ **Overview.** We developed a pipeline for mapping fluorescence micro-optical sectioning
²⁵¹ tomography (fMOST) mouse brain images into the AllenCCFv3 (see Figure 1(b)). The
²⁵² pipeline is adapted from previously developed frameworks for human brain mapping⁵⁸, and
²⁵³ uses a modality specific (fMOST) average atlas to assist in the image registration and
²⁵⁴ mapping. This approach has been well validated in human studies^{73–75}, and successfully
²⁵⁵ used in other mouse data^{12,15,34}. Briefly, we construct an intensity- and shape-based average
²⁵⁶ fMOST atlas using 30 fMOST images to serve as an intermediate registration target for
²⁵⁷ mapping fMOST images from individual specimens into the AllenCCFv3. Preprocessing
²⁵⁸ steps include downsampling to match the $25\mu m$ isotropic AllenCCFv3, acquisition-based
²⁵⁹ stripe artifact removal, and inhomogeneity correction⁶¹. Preprocessing also includes a single
²⁶⁰ annotation-driven registration to establish a canonical mapping between the fMOST atlas and
²⁶¹ the AllenCCFv3. This step allows us to align expert determined landmarks to accurately map
²⁶² structures with large morphological differences between the modalities, which are difficult to
²⁶³ address using standard approaches. Once this canonical mapping is established, standard
²⁶⁴ intensity-based registration is used to align each new fMOST image to the fMOST specific
²⁶⁵ atlas. This mapping is concatenated with the canonical fMOST atlas-to-AllenCCFv3 mapping
²⁶⁶ to further map each individual brain into the latter without the need to generate additional
²⁶⁷ landmarks. Transformations learned through this mapping can be applied to single neuron
²⁶⁸ reconstructions from the fMOST images to evaluate neuronal distributions across different
²⁶⁹ specimens into the AllenCCFv3 for the purpose of cell census analyses.

²⁷⁰ **Data.** The high-throughput and high-resolution fluorescence micro-optical sectioning tomog-
²⁷¹ raphy (fMOST)^{76,77} platform was used to image 55 mouse brains containing gene-defined
²⁷² neuron populations, with sparse transgenic expression^{78,79}. In short, the fMOST imaging

platform results in 3-D images with voxel sizes of $0.35 \times 0.35 \times 1.0 \mu\text{m}^3$ and is a two-channel imaging system where the green channel displays the green fluorescent protein (GFP) labeled neuron morphology and the red channel is used to visualize the counterstained propidium iodide cytoarchitecture. The spatial normalizations described in this work were performed using the red channel, which offered higher tissue contrast for alignment, although other approaches are possible including multi-channel registration.

Evaluation. Evaluation of the canonical fMOST atlas to Allen CCFv3 mapping was performed via quantitative comparison at each step of the registration and qualitative assessment of structural correspondence after alignment by an expert anatomist. Dice values were generated for the following structures: whole brain, 0.99; fimbria, 0.91; habenular commissure, 0.63; posterior choroid plexus, 0.93; anterior choroid plexus, 0.96; optic chiasm, 0.77; caudate putamen, 0.97. Similar qualitative assessment was performed for each fMOST specimen including the corresponding neuron reconstruction data.

2.2 The DevCCF velocity flow model

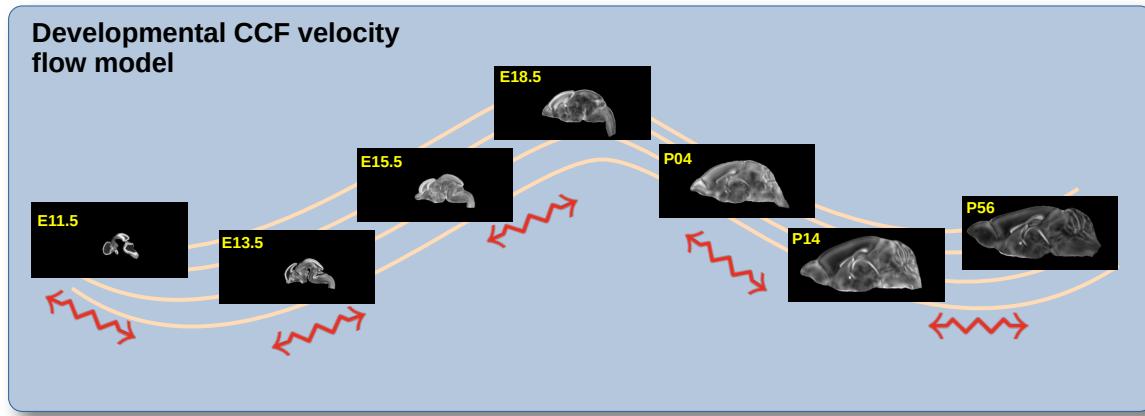


Figure 2: The spatial transformation between any two time points within the DevCCF longitudinal developmental trajectory is available through the use of ANTsX functionality for generating a velocity flow model.

To continuously interpolate transformations between the different stages of the DevCCF atlases, a velocity flow model was constructed using DevCCF derived data and functionality recently introduced into both the ANTsR and ANTsPy packages. Both platforms include

290 a complete suite of functions for determining dense correspondence from sparse landmarks
 291 based on a variety of transformation models ranging from standard linear models (i.e., rigid,
 292 affine) to deformable diffeomorphic models (e.g, symmetric normalization⁵⁴). The latter set
 293 includes transformation models for both the pairwise scenario and for multiple sets, as in the
 294 case of the DevCCF. ANTsX, being built on top of ITK, uses an ITK image data structure
 295 for the 4-D velocity field where each voxel contains the x , y , z components of the field at
 296 that point.

297 **2.2.1 Data**

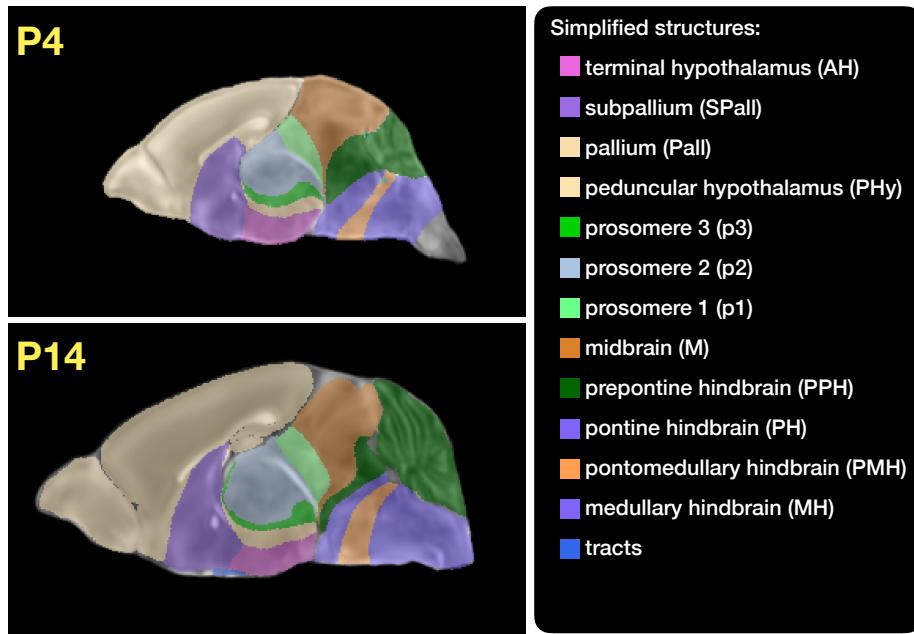


Figure 3: Annotated regions representing common labels across developmental stages which are illustrated for both P4 and P14.

298 Labeled annotations are available as part of the original DevCCF and reside in the space
 299 of each developmental template which range in resolution from $31.5 - 50\mu\text{m}$. Across all
 300 atlases, the total number of labeled regions exceeds 2500. From these labels, a common set
 301 of 26 labels (13 per hemisphere) across all atlases were used for optimization and evaluation.
 302 These simplified regions include: terminal hypothalamus, subpallium, pallium, peduncular
 303 hypothalamus, prosomere, prosomere, prosomere, midbrain, prepontine hindbrain, pontine
 304 hindbrain, pontomedullary hindbrain, medullary hindbrain, and tracts (see Figure 3).

305 Prior to velocity field optimization, all data were rigidly transformed to DevCCF P56 using
306 the centroids of the common label sets. In order to determine the landmark correspondence
307 across DevCCF stages, the multi-metric capabilities of `ants.registration(...)` were used.
308 Instead of performing intensity-based pairwise registration directly on these multi-label
309 images, each label was used to construct a separate fixed and moving image pair resulting in
310 a multi-metric registration optimization scenario involving 24 binary image pairs (each label
311 weighted equally) for optimizing diffeomorphic correspondence between neighboring time
312 point atlases using the mean squares metric and the symmetric normalization transform⁵⁴.
313 To generate the set of common point sets across all seven developmental atlases, the label
314 boundaries and whole regions were sampled in the P56 atlas and then propagated to each atlas
315 using the transformations derived from the pairwise registrations. We selected a sampling
316 rate of 10% for the contour points and 1% for the regional points for a total number of points
317 being per atlas being 173303 ($N_{contour} = 98151$ and $N_{region} = 75152$). Regional boundary
318 points were weighted twice as those of non-boundary points during optimization.

319 2.2.2 Optimization

320 The velocity field was optimized using the input composed of the seven corresponding point
321 sets and their associated weight values, the selected number of integration points for the
322 velocity field ($N = 11$), and the parameters defining the geometry of the spatial dimensions
323 of the velocity field. Thus, the optimized velocity field described here is of size [256, 182, 360]
324 (50 μ m isotropic) \times 11 integration points for a total compressed size of a little over 2 GB.
325 This choice represented weighing the trade-off between tractability, portability, and accuracy.
326 However, all data and code to reproduce the results described (with possible variation in the
327 input parameters) are available in the dedicated GitHub repository.
328 The normalized time point scalar value for each atlas/point-set in the temporal domains [0, 1]
329 was also defined. Given the increasingly larger gaps in the postnatal timepoint sampling, we
330 made two adjustments. Based on known mouse brain development, we used 28 days for the
331 P56 data. We then computed the log transform of the adjusted set of time points prior to
332 normalization between 0 and 1 (see the right side of Figure 4). This log transform, as part of

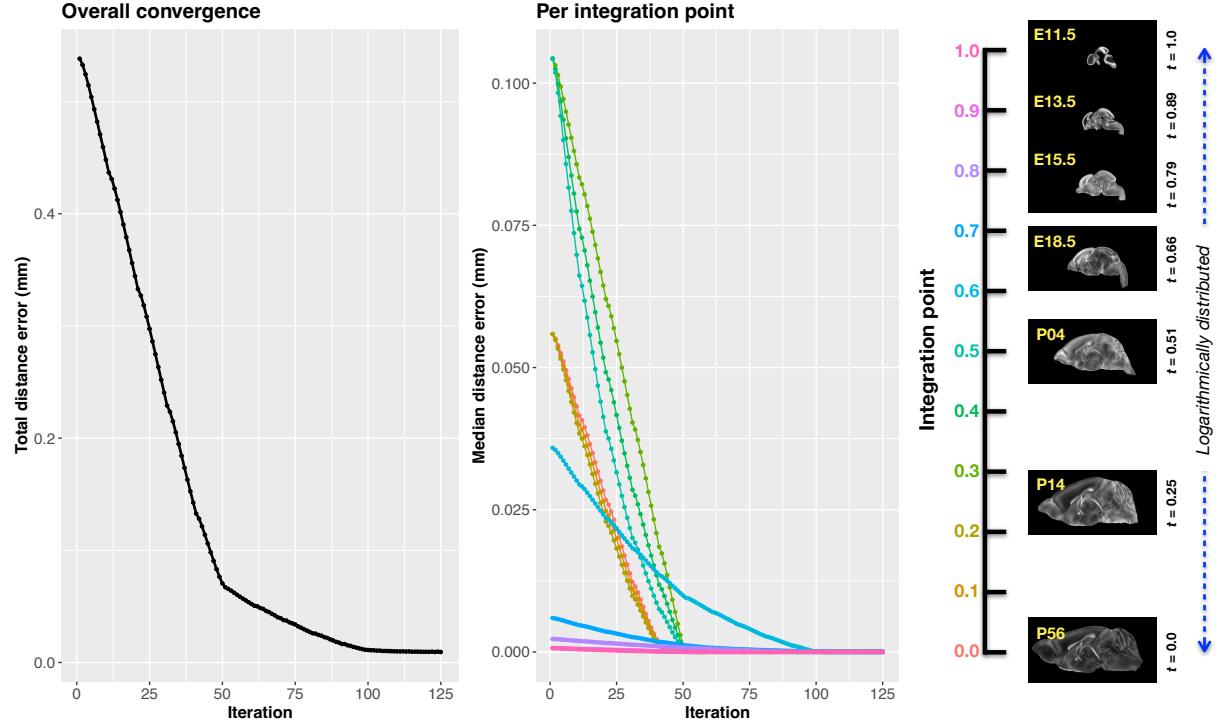


Figure 4: Convergence of the optimization of the velocity field for describing the transformation through the developmental stages from E11.5 through P56. Integration points in diagram on the right are color-coordinated with the center plot and placed in relation to the logarithmically situated temporal placement of the individual DevCCF atlases.

the temporal normalization, significantly improved data spacing.
 The maximum number of iterations was set to 200 with each iteration taking approximately six minutes on a 2020 iMac (processor, 3.6 GHz 10-Core Intel Core i9; memory, 64 GB 2667 MHz DDR4) At each iteration we looped over the 11 integration points. At each integration point, the velocity field estimate was updated by warping the two immediately adjacent point sets to the integration time point and determining the regularized displacement field between the two warped point sets. As with any gradient-based descent algorithm, this field was multiplied by a small step size ($\delta = 0.2$) before adding to the current velocity field. Convergence is determined by the average displacement error over each of the integration points. As can be seen in the left panel of Figure 4, convergence occurred around 125 iterations when the average displacement error over all integration points is minimized. The median displacement error at each of the integration points also trends towards zero but at different rates.

346 2.2.3 The transformation model

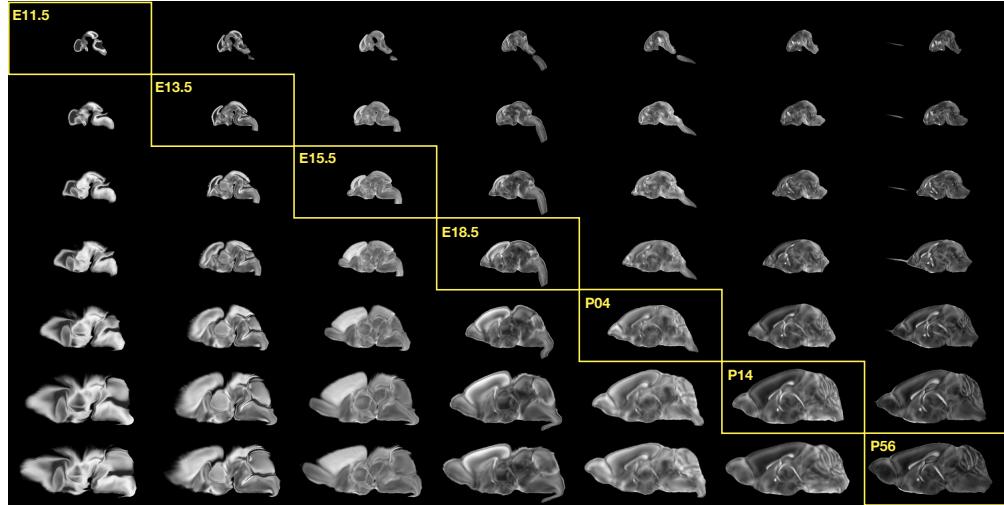


Figure 5: Mid-sagittal visualization of the effects of the transformation model in warping every developmental stage to the time point of every other developmental stage. The original images are located along the diagonal. Columns correspond to the warped original image whereas the rows represent the reference space to which each image is warped.

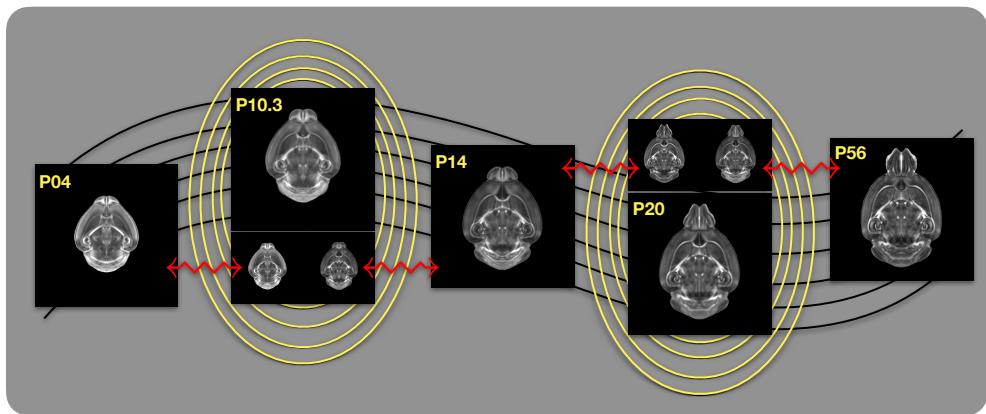


Figure 6: Illustration of the use of the velocity flow model for creating virtual templates at continuous time points not represented in one of the existing DevCCF time points. For example, FA templates at time point P10.3 and P20 can be generated by warping the existing temporally adjacent developmental templates to the target time point and using those images in the ANTsX template building process.

- 347 Once optimized, the resulting velocity field can be used to generate the deformable transform
 348 between any two continuous points within the time interval bounded by E11.5 and P56.
 349 In Figure 5, we transform each atlas to the space of every other atlas using the DevCCF
 350 transformation model. Additionally, one can use this transformation model to construct virtual

351 templates in the temporal gaps of the DevCCF. Given an arbitrarily chosen time point
 352 within the normalized time point interval, the existing adjacent DevCCF atlases on either
 353 chronological side can be warped to the desired time point. A subsequent call to one of
 354 the ANTsX template building functions then permits the construction of the template at
 355 that time point. In Figure 6, we illustrate the use of the DevCCF velocity flow model for
 356 generating two such virtual templates for two arbitrary time points. Note that both of these
 357 usage examples can be found in the GitHub repository previously given.

358 2.3 The Mouse Brain Parcellation Pipeline

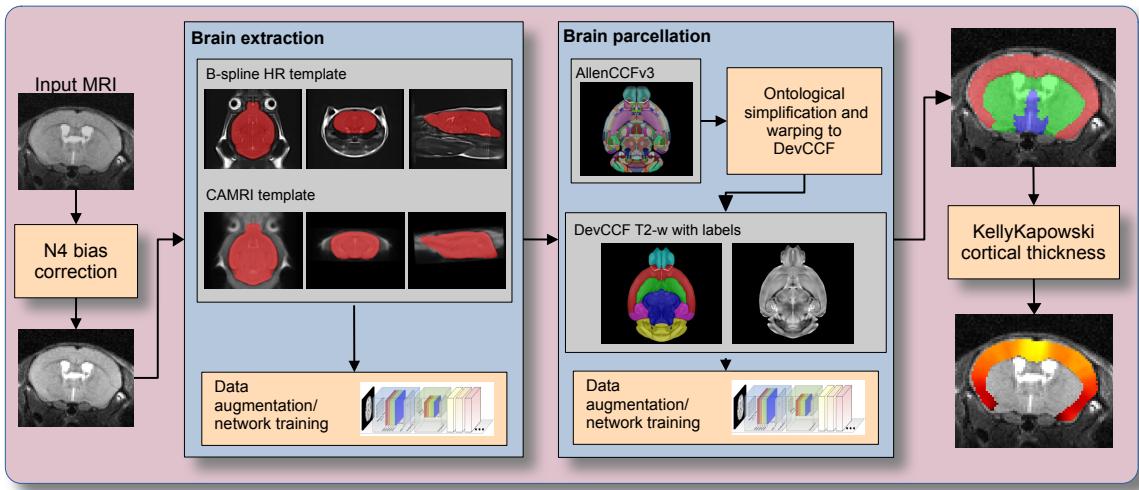


Figure 7: The mouse brain cortical parcellation pipeline integrating two deep learning components for brain extraction and brain parcellation prior to estimating cortical labels. Both deep learning networks rely heavily on data augmentation on templates built from open data and provide an outline for further refinement and creating alternative parcellations for tailored research objectives.

359 One of the most well-utilized pipelines in the ANTsX toolkit is the generation of cortical
 360 thickness maps in the human brain from T1-weighted MRI. Starting with the novel Diffeomor-
 361 phic Registration-based Cortical Thickness (DiReCT) algorithm⁸⁰, a complete algorithmic
 362 workflow was developed for both cross-sectional⁸¹ and longitudinal⁸² T1-weighted MR image
 363 data. This contribution was later refactored using deep learning⁴⁶ leveraging the earlier
 364 results⁸¹ for training data.

365 In the case of the mouse brain, the lack of training data and/or tools to generate training

³⁶⁶ data making analogous algorithmic development difficult. In addition, mouse data is often
³⁶⁷ characterized by unique issues such as frequent anisotropic sampling which are often in sharp
³⁶⁸ contrast to the high resolution resources available within the community, e.g., AllenCCFv3
³⁶⁹ and DevCCF. Using ANTsX and other publicly available data resources, we developed a
³⁷⁰ complete mouse brain structural morphology pipeline as illustrated in Figure 7 and detailed
³⁷¹ below.

³⁷² 2.3.1 Two-shot mouse brain extraction network

³⁷³ In order to create a generalized mouse brain extraction network, we built whole-head templates
³⁷⁴ from two publicly available datasets. The Center for Animal MRI (CAMRI) dataset⁶⁸ from
³⁷⁵ the University of North Carolina at Chapel Hill consists of 16 T2-weighted MRI volumes
³⁷⁶ of voxel resolution $0.16 \times 0.16 \times 0.16 mm^3$. The second high-resolution dataset⁶⁹ comprises
³⁷⁷ 88 specimens each with three spatially aligned canonical views with in-plane resolution of
³⁷⁸ $0.08 \times 0.08 mm^2$ with a slice thickness of $0.5 mm$. These three orthogonal views were used
³⁷⁹ to reconstruct a single high-resolution volume per subject using a B-spline fitting algorithm
³⁸⁰ developed in ANTsX⁸³. From these two datasets, two symmetric isotropic ANTsX templates⁵⁸
³⁸¹ were generated analogous to the publicly available ANTsX human brain templates used in
³⁸² previous research⁸¹. Bias field simulation, intensity histogram warping, noise simulation,
³⁸³ random translation and warping, and random anisotropic resampling in the three canonical
³⁸⁴ directions were used for data augmentation in training a T2-weighted brain extraction
³⁸⁵ network.

³⁸⁶ 2.3.2 Single-shot mouse brain parcellation network

³⁸⁷ To create the network for generating a brain parcellation consistent with cortical thickness
³⁸⁸ estimation, we used the AllenCCFv3 and the associated `allensdk` Python library. Using
³⁸⁹ `allensdk`, a gross parcellation labeling was generated from the fine Allen CCFv3 labeling
³⁹⁰ which includes the cerebral cortex, cerebral nuclei, brain stem, cerebellum, main olfactory
³⁹¹ bulb, and hippocampal formation. This labeling was mapped to the P56 component of
³⁹² the DevCCF. Both the T2-w P56 DevCCF and labelings, in conjunction with the data
³⁹³ augmentation described previously for brain extraction, was used to train a brain parcellation

394 network.

395 2.3.3 Evaluation

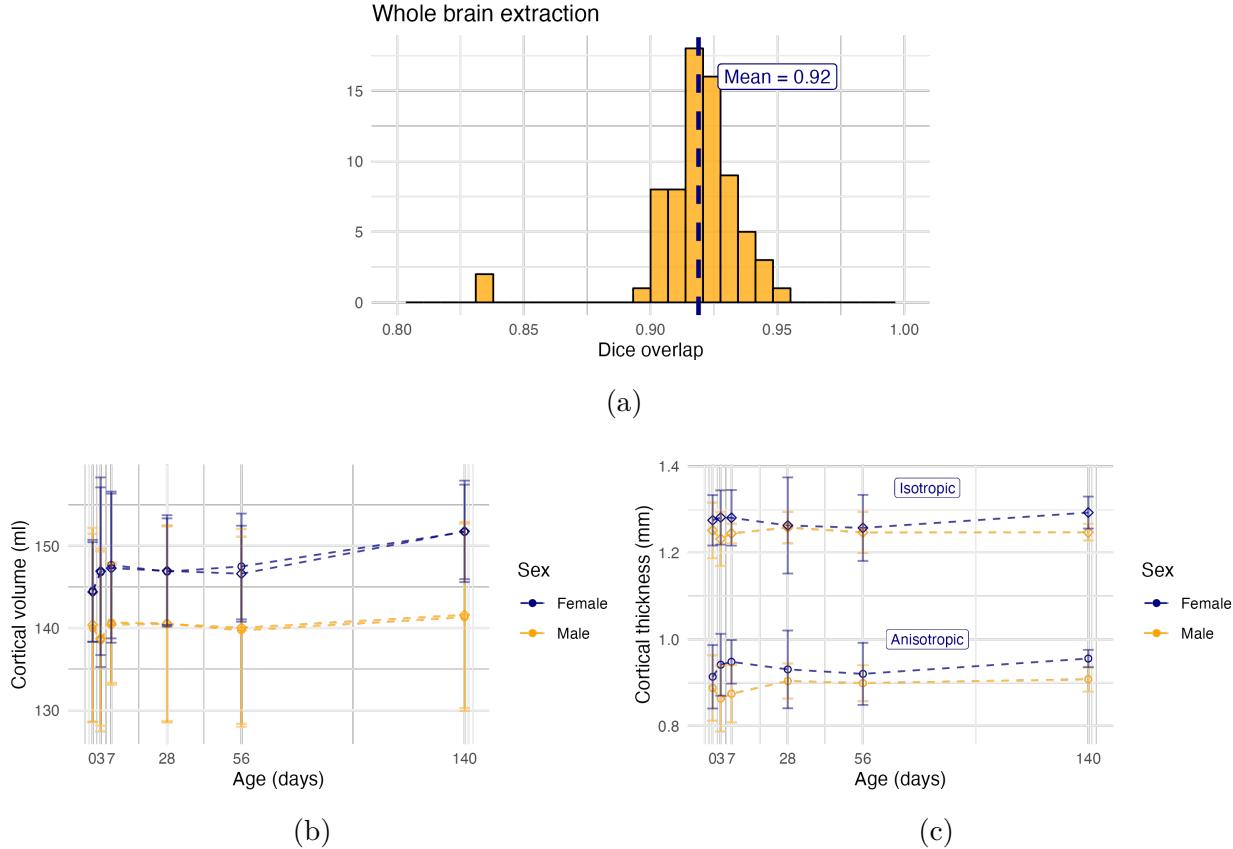


Figure 8: Evaluation of the ANTsX mouse brain extraction, parcellation, and cortical thickness pipeline on an independent dataset consisting of 12 specimens \times 7 time points = 84 total images. (a) Dice overlap comparisons with the provided brain masks provide generally good agreement with the brain extraction network. (b) Cortical volume measurements show similar average quantities over growth and development between the original anisotropic data and interpolated isotropic data. (c) These results contrast with the cortical thickness measurements which show that cortical thickness estimation in anisotropic space severely underestimates the actual values.

396 For evaluation, we used an additional publicly available dataset⁷⁰ which is completely
397 independent from the data used in training the brain extraction and parcellation networks.
398 Data includes 12 specimens each imaged at seven time points (Day 0, Day 3, Week 1, Week
399 4, Week 8, Week 20) with available brain masks. In-plane resolution is $0.1 \times 0.1 mm^2$ with a
400 slice thickness of $0.5 mm$. Since the training data is isotropic and data augmentation includes
401 downsampling in the canonical directions, each of the two networks learns mouse brain-specific

402 interpolation such that one can perform prediction on thick-sliced images, as, for example, in
403 these evaluation data, and return isotropic probability and thickness maps (a choice available
404 to the user). Figure 8 summarizes the results of the evaluation and comparison between
405 isotropic and anisotropic cortical measurements in male and female specimens.

406 **3 Discussion**

407 The diverse mouse brain cell type profiles gathered through BICCN and associated efforts
408 provides a rich multi-modal resource to the community. However, despite significant progress,
409 full integration of these valuable resources is not yet complete. Central to the data integration
410 is a continued need to accurately map each unique dataset into common coordinate frameworks
411 (CCFs) so that they can be accessed in connection with each other. Additionally, the ability
412 to map novel cell type data in the future to these existing BICCN resources is vital for
413 effective utilization of this endeavor and the continuation of its goals. To meet these needs,
414 tools for mapping mouse cell type data must be both generally accessible to a wide audience
415 of investigators, and still capable of handling distinct challenges unique to each data type.

416 In this work, we describe modular ANTsX-based pipelines developed to address the needs
417 of three BICCN projects that cover distinct cell type data, including spatial transcriptomic,
418 morphology, and developmental data. We highlight how a modular toolbox like ANTsX can
419 be tailored to address problems unique to each modality while still leveraging a variety of
420 ready-to-use powerful tools that have been externally validated.

421 Our MERFISH pipeline provides an example of how to map high-resolution spatial tran-
422 scriptomic data into the AllenCCFv3. Since full brain large-scale transcriptomics is still
423 rare and difficult to collect, the pipeline focuses on achieving the best possible anatomical
424 alignment and fully utilizing the available data. While the techniques employed for mapping
425 the sectioned data can be generally applicable to map other serial histology images, many
426 parts of the pipeline were designed to address very specific known alignment challenges in the
427 MERFISH data using a series of iterative registration steps. The pipeline shows how general
428 tools available in ANTsX can be adapted to target highly specialized problems in mouse cell
429 type data.

430 In contrast to the MERFISH pipeline, our fMOST pipeline was designed to be a more
431 general solution that can be employed in other modalities. The pipeline primarily uses
432 previously developed ANTsX preprocessing and atlasing tools to map fMOST data into the
433 AllenCCFv3. The key component of the pipeline is the use of a fMOST specific average atlas

434 to greatly simplify the image registration problem. This average atlas, also constructed using
435 pre-existing ANTsX tools, allows for a one-time canonical alignment from the fMOST atlas to
436 the AllenCCFv3 to be transferred and used for mapping new fMOST images. Lastly, ANTsX
437 provides point set transformation tools to allow the mappings found through the pipeline to
438 be directly applied to associated single-cell reconstructions from the fMOST data to study
439 neuronal morphology.

440 Our DevCCF pipeline shows the application of the toolkit for temporospatial developmental
441 data. ANTsX was crucial in providing necessary functionality for yielding high quality
442 output. For the generation of the individual developmental stage multi-modal, symmetric
443 templates, ANTsX is unique amongst image analysis software packages in providing existing
444 solutions for template generation which have been thoroughly vetted, including being used
445 in several studies over the years, and which continue to be under active refinement. At its
446 core, computationally efficient and quality template generation requires the use of precision
447 pairwise image mapping functionality which, historically, is at the origins of the ANTsX
448 ecosystem. Moreover, these mapping capabilities extend beyond template generation to the
449 mapping of other image data (e.g., gene expression maps) to a selected template for providing
450 further insight into the mouse brain.

451 With respect to the DevCCF, despite the significant expansion of available developmental age
452 templates beyond what existed previously, there are still temporal gaps in the DevCCF which
453 can be potentially sampled by future research efforts. However, pioneering work involving
454 time-varying diffeomorphic transformations allow us to continuously situate the existing
455 templates within a velocity flow model. This allows one to determine the diffeomorphic
456 transformation from any one temporal location to any other temporal location within the time
457 span defined by the temporal limits of the DevCCF. This functionality is built on multiple
458 ITK components including the B-spline scattered data approximation technique for field
459 regularization and velocity field integration. This velocity field model permits intra-template
460 comparison and the construction of virtual templates where a template can be estimated at
461 any continuous time point within the temporal domain. This novel application can potentially
462 enhance our understanding of intermediate developmental stages.

463 We also presented a mouse brain pipeline for brain extraction, parcellation, and cortical
464 thickness using single-shot and two-shot learning with data augmentation. This approach
465 attempts to circumvent (or at least minimize) the typical requirement of large training
466 datasets as with the human ANTsX pipeline analog. However, even given our initial success
467 on independent data, we fully anticipate that refinements will be necessary. Given that the
468 ANTsX toolkit is a dynamic effort undergoing continual improvement, we manually correct
469 cases that fail and use them for future training and refinement of network weights as we have
470 done for our human-based networks. Generally, these approaches provide a way to bootstrap
471 training data for manual refinement and future generation of more accurate deep learning
472 networks in the absence of other applicable tools.

473 The ANTsX ecosystem is a powerful framework that has demonstrated applicability to diverse
474 cell type data in the mouse brain. This is further evidenced by the many software packages
475 that use various ANTsX components in their own mouse-specific workflows. In and of itself,
476 the extensive functionality of ANTsX makes it possible to create complete processing pipelines
477 without requiring the integration of multiple packages or lengthy software development. These
478 open-source components not only perform well but are available across multiple platforms
479 which facilitates the construction of tailored pipelines for individual study solutions. These
480 components are also supported by years of development not only by the ANTsX development
481 team but by the larger ITK community.

482 **4 Methods**

483 The following methods are all available as part of the ANTsX ecosystem with analogous
484 elements existing in both ANTsR (ANTs in R) and ANTsPy (ANTs in Python) with an
485 ANTs/ITK C++ core. However, most of the development for the work described below was
486 performed using ANTsPy. For equivalent calls in ANTsR, please see the ANTsX tutorial at
487 <https://tinyurl.com/antsxtutorial>.

488 **4.1 General ANTsX utilities**

489 Although they focus on distinct data types, the three pipelines presented share common
490 components that are generally applicable when mapping mouse cell type data. These include,
491 addressing intensity biases and noise in the data, image registration to solve the mapping,
492 creating custom templates and atlases from the data, and visualization of the results. Table
493 1 provides a brief summary of key general functionalities in ANTsX for addressing these
494 challenges.

495 **4.1.1 Preprocessing: bias field correction and denoising**

496 Bias field correction and image denoising are standard preprocessing steps in improving
497 overall image quality in mouse brain images. The bias field, a gradual spatial intensity
498 variation in images, can arise from various sources such as magnetic field inhomogeneity or
499 acquisition artifacts, leading to distortions that can compromise the quality of brain images.
500 Correcting for bias fields ensures a more uniform and consistent representation of brain
501 structures, enabling more accurate quantitative analysis. Additionally, brain images are
502 often susceptible to various forms of noise, which can obscure subtle features and affect the
503 precision of measurements. Denoising techniques help mitigate the impact of noise, enhancing
504 the signal-to-noise ratio and improving the overall image quality. The well-known N4 bias
505 field correction algorithm⁶¹ has its origins in the ANTs toolkit which was implemented and
506 introduced into the ITK toolkit, i.e. `ants.n4_bias_field_correction(...)`. Similarly,
507 ANTsX contains an implementation of a well-performing patch-based denoising technique⁶⁰
508 and is also available as an image filter to the ITK community, `ants.denoise_image(...)`.

Table 1: Sampling of ANTsX functionality

<i>ANTsPy: Preprocessing</i>	
bias field correction	<code>n4_bias_field_correction(...)</code>
image denoising	<code>denoise_image(...)</code>
<i>ANTsPy: Registration</i>	
image registration	<code>registration(...)</code>
image transformation	<code>apply_transforms(...)</code>
template generation	<code>build_template(...)</code>
landmark registration	<code>fit_transform_to_paired_points(...)</code>
time-varying landmark reg.	<code>fit_time_varying_transform_to_point_sets(...)</code>
integrate velocity field	<code>integrate_velocity_field(...)</code>
invert displacement field	<code>invert_displacement_field(...)</code>
<i>ANTsPy: Segmentation</i>	
MRF-based segmentation	<code>atropos(...)</code>
Joint label fusion	<code>joint_label_fusion(...)</code>
diffeomorphic thickness	<code>kelly_kapowski(...)</code>
<i>ANTsPy: Miscellaneous</i>	
Regional intensity statistics	<code>label_stats(...)</code>
Regional shape measures	<code>label_geometry_measures(...)</code>
B-spline approximation	<code>fit_bspline_object_to_scattered_data(...)</code>
Visualize images and overlays	<code>plot(...)</code>
<i>ANTsPyNet: Mouse-specific</i>	
brain extraction	<code>mouse_brain_extraction(...modality="t2"...)</code> <code>mouse_brain_extraction(...modality="ex5"...)</code>
brain parcellation	<code>mouse_brain_parcellation(...)</code>
cortical thickness	<code>mouse_cortical_thickness(...)</code>
super resolution	<code>mouse_histology_super_resolution(...)</code>

ANTsX provides state-of-the-art functionality for processing biomedical image data. Such tools, including deep learning networks, support a variety of mapping-related tasks. A more comprehensive listing of ANTsX tools with self-contained R and Python examples is provided as a gist page on GitHub (<https://tinyurl.com/antsxtutorial>).

509 **4.1.2 Image registration**

510 The ANTs registration toolkit is a complex framework permitting highly tailored solutions
511 to pairwise image registration scenarios⁸⁴. It includes innovative transformation models
512 for biological modeling^{54,67} and has proven capable of excellent performance^{55,85}. Vari-
513 ous parameter sets targeting specific applications have been packaged with the different
514 ANTsX packages, specifically ANTs, ANTsPy, and ANTsR⁴⁶. In ANTsPy, the function
515 `ants.registration(...)` is used to register a pair of images or a pair of image sets where
516 `type_of_transform` is a user-specified option that invokes a specific parameter set. For ex-
517 ample `type_of_transform='antsRegistrationSyNQuick[s]'` encapsulates an oft-used pa-
518 rameter set for quick registration whereas `type_of_transform='antsRegistrationSyN[s]'`
519 is a more detailed alternative. Transforming images using the derived transforms is performed
520 via the `ants.apply_transforms(...)` function.

521 Initially, linear optimization is initialized with center of (intensity) mass alignment typically
522 followed by optimization of both rigid and affine transforms using the mutual information
523 similarity metric. This is followed by diffeomorphic deformable alignment using symmetric
524 normalization (SyN) with Gaussian⁵⁴ or B-spline regularization⁶⁷ where the forward transform
525 is invertible and differentiable. The similarity metric employed at this latter stage is typically
526 either neighborhood cross-correlation or mutual information. Note that these parameter
527 sets are robust to input image type (e.g., light sheet fluorescence microscopy, Nissl staining,
528 and the various MRI modalities) and are adaptable to mouse image geometry and scaling.
529 Further details can be found in the various documentation sources for these ANTsX packages.

530 **4.1.3 Template generation**

531 ANTsX provides functionality for constructing templates from a set (or multi-modal sets) of
532 input images as originally described⁵⁸ and recently used to create the DevCCF templates¹⁶.
533 An initial template estimate is constructed from an existing subject image or a voxelwise
534 average derived from a rigid pre-alignment of the image population. Pairwise registration
535 between each subject and the current template estimate is performed using the Symmetric
536 Normalization (SyN) algorithm⁵⁴. The template estimate is updated by warping all subjects

537 to the space of the template, performing a voxelwise average, and then performing a “shape
538 update” of this latter image by warping it by the average inverse deformation, thus yielding
539 a mean image of the population in terms of both intensity and shape. The corresponding
540 ANTsPy function is `ants.build_template(...)`.

541 **4.1.4 Visualization**

542 To complement the well-known visualization capabilities of R and Python, e.g., `ggplot2`
543 and `matplotlib`, respectively, image-specific visualization capabilities are available in the
544 `ants.plot(...)` function (Python). These are capable of illustrating multiple slices in
545 different orientations with other image overlays and label images.

546 **4.2 Mapping fMOST data to AllenCCFv3**

547 **4.2.1 Preprocessing**

- 548 • *Downsampling.* The first challenge when mapping fMOST images into the AllenCCFv3
549 is addressing the resolution scale of the data. Native fMOST data from an individual
550 specimen can range in the order of terabytes, which leads to two main problems. First,
551 volumetric registration methods (particularly those estimating local deformation) have
552 high computational complexity and typically cannot operate on such high-resolution
553 data under reasonable memory and runtime constraints. Second, the resolution of
554 the AllenCCFv3 atlas is much lower than the fMOST data, thus the mapping process
555 will cause much of the high-resolution information in the fMOST images to be lost
556 regardless. Thus, we perform a cubic B-spline downsampling of the fMOST data to
557 reduce the resolution of each image to match the isotropic $25 \mu\text{m}$ voxel resolution of the
558 AllenCCFv3 intensity atlas using `ants.resample_image(...)`. An important detail
559 to note is that while the fMOST images and atlas are downsampled, the mapping
560 learned during the registration is assumed to be continuous. Thus, after establishing
561 the mapping to the AllenCCFv3, we can interpolate the learned mapping and apply it
562 directly to the high-resolution native data directly to transform any spatially aligned
563 data (such as the single-cell neuron reconstructions) into the AllenCCFv3.

- 564 • *Stripe artifact removal.* Repetitive pattern artifacts are a common challenge in fMOST
565 imaging where inhomogeneity during the cutting and imaging of different sections can
566 leave stripes of hyper- and hypo-intensity across the image. These stripe artifacts
567 can be latched onto by the registration algorithm as unintended features that are
568 then misregistered to non-analogous structures in the AllenCCFv3. We address these
569 artifacts by fitting a 3-D bandstop (notch) filter to target the frequency of the stripe
570 patterns and removing them prior to the image registration.
- 571 • *Inhomogeneity correction.* Regional intensity inhomogeneity can also occur within
572 and between sections in fMOST imaging due to staining or lighting irregularity dur-
573 ing acquisition. Similar to stripe artifacts, intensity gradients due to inhomogeneity
574 can be misconstrued as features during the mapping and result in matching of non-
575 corresponding structures. Our pipeline addresses these intensity inhomogeneities using
576 N4 bias field correction⁶¹, `ants.n4_bias_field_correction(...)`.

577 **4.2.2 Steps for spatial normalization to AllenCCFv3**

578 1. *Average fMOST atlas as an intermediate target.* Due to the preparation of the mouse
579 brain for fMOST imaging, the resulting structure in the mouse brain has several large
580 morphological deviations from the AllenCCFv3 atlas. Most notable of these is an
581 enlargement of the ventricles, and compression of cortical structures. In addition,
582 there is poor intensity correspondence for the same anatomic features due to intensity
583 dissimilarity between imaging modalities. We have found that standard intensity-base
584 registration is insufficient to capture the significant deformations required to map these
585 structures correctly into the AllenCCFv3. We address this challenge in ANTsX by using
586 explicitly corresponding parcellations of the brain, ventricles and surrounding structures
587 to directly recover these large morphological differences. However, generating these
588 parcellations for each individual mouse brain is a labor-intensive task. Our solution
589 is to create an average atlas whose mapping to AllenCCFv3 encapsulates these large
590 morphological differences to serve as an intermediate registration point. This has the
591 advantage of only needing to generate one set of corresponding annotations which is

592 used to register between the two atlas spaces. New images are first aligned to the
593 fMOST average atlas, which shares common intensity and morphological features and
594 thus can be achieved through standard intensity-based registration.

595 2. *Average fMOST atlas construction.* An intensity and shape-based contralaterally
596 symmetric average of the fMOST image data is constructed from 30 images and their
597 contralateral flipped versions. We ran three iterations of the atlas construction using the
598 default settings. Additional iterations (up to six) were evaluated and showed minimal
599 changes to the final atlas construction, suggesting a convergence of the algorithm.

600 3. *fMOST atlas to AllenCCFv3 alignment.* Alignment between the fMOST average atlas
601 and AllenCCFv3 was performed using a one-time annotation-driven approach. Label-
602 to-label registration is used to align 7 corresponding annotations in both atlases in the
603 following: 1) brain mask/ventricles, 2) caudate/putamen, 3) fimbria, 4) posterior choroid
604 plexus, 5) optic chiasm, 6) anterior choroid plexus, and 7) habenular commissure. The
605 alignments were performed sequentially, with the largest, most relevant structures being
606 aligned first using coarse registration parameters, followed by other structures using
607 finer parameters. This coarse-to-fine approach allows us to address large morphological
608 differences (such as brain shape and ventricle expansion) at the start of registration
609 and then progressively refine the mapping using the smaller structures. The overall
610 ordering of these structures was determined manually by an expert anatomist, where
611 anatomical misregistration after each step of the registration was evaluated and used to
612 determine which structure should be used in the subsequent iteration to best improve
613 the alignment. The transformation from this one-time expert-guided alignment is
614 preserved and used as the canonical fMOST atlas to AllenCCFv3 mapping in the
615 pipeline.

616 4. *Alignment of individual fMOST mouse brains.* The canonical transformation between
617 the fMOST atlas and AllenCCFv3 greatly simplifies the registration of new individual
618 fMOST mouse brains into the AllenCCFv3. Each new image is first registered into the
619 fMOST average atlas, which shares intensity, modality, and morphological characteris-
620 tics. This allows us to leverage standard, intensity-based registration functionality⁸⁴

available in ANTsX to perform this alignment. Transformations are then concatenated to the original fMOST image to move it into the AllenCCFv3 space using `ants.apply_transforms(...)`.

5. *Transformation of single cell neurons.* A key feature of fMOST imaging is the ability to reconstruct and examine whole-brain single neuron projections⁷⁹. Spatial mapping of these neurons from individual brains into the AllenCCFv3 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 AllenCCFv3 using our pipeline can be applied in this way to fMOST neuron reconstruction point set data using `ants.apply_transforms_to_points(..)`.

4.3 Mapping MERFISH data to AllenCCFv3

4.3.1 Preprocessing

- *Initial volume reconstruction.* Alignment of MERFISH data into a 3-D 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 $10 \times 10 \mu\text{m}^2$ grid to match the resolution of the $10 \mu\text{m}$ AllenCCFv3 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 AllenCCFv3.
- *Anatomical correspondence labeling.* Mapping the MERFISH data into the AllenCCFv3 requires us to establish correspondence between the anatomy depicted in the MERFISH and AllenCCFv3 data. Intensity-based features in MERFISH data are not sufficiently apparent to establish this correspondence, so we need to generate instead corresponding anatomical labelings of both images with which to drive registration. These labels are already available as part of the AllenCCFv3; thus, the main challenge is deriving analogous labels from the spatial transcriptomic maps of the MERFISH data. Toward this end, we assigned each cell from the scRNA-seq dataset to one of the following major

649 regions: cerebellum, CTXsp, hindbrain, HPF, hypothalamus, isocortex, LSX, midbrain,
650 OLF, PAL, sAMY, STRd, STRv, thalamus and hindbrain. A label map of each section
651 was generated for each region by aggregating the cells assigned to that region within a
652 $10 \times 10 \mu\text{m}^2$ grid. The same approach was used to generate more fine grained region
653 specific landmarks (i.e., cortical layers, habenula, IC). Unlike the broad labels which
654 cover large swaths of the section these regions are highly specific to certain parts of the
655 section. Once cells in the MERFISH data are labeled, morphological dilation is used to
656 provide full regional labels for alignment into the AllenCCFv3.

- 657 • *Section matching.* Since the MERFISH data is acquired as sections, its 3-D orientation
658 may not be fully accounted for during the volume reconstruction step, due to the
659 particular cutting angle. This can lead to obliqueness artifacts in the section where
660 certain structures can appear to be larger or smaller, or missing outright from the
661 section. To address this, we first use a global alignment to match the orientations of
662 the MERFISH sections to the atlas space. In our pipeline, this section matching is
663 performed in the reverse direction by performing a global affine transformation of the
664 AllenCCFv3 into the MERFISH data space, and then resampling digital sections from
665 the AllenCCFv3 to match each MERFISH section. This approach limits the overall
666 transformation and thus resampling that is applied to the MERFISH data, and, since
667 the AllenCCFv3 is densely sampled, it also reduces in-plane artifacts that result from
668 missing sections or undefined spacing in the MERFISH data.

669 4.3.2 2.5D deformable, landmark-driven alignment to AllenCCFv3

670 After global alignment of the AllenCCFv3 into the MERFISH dataset, 2D per-section
671 deformable refinements are used to address local differences between the MERFISH sections
672 and the resampled AllenCCFv3 sections. Nine registrations were performed in sequence using
673 a single label at each iteration in the following order: 1) brain mask, 2) isocortex (layer
674 2+3), 3) isocortex (layer 5), 4) isocortex (layer 6), 5) striatum, 6) medial habenula, 7) lateral
675 habenula, 8) thalamus, and 9) hippocampus. This ordering was determined empirically by an
676 expert anatomist who prioritized which structure to use in each iteration by evaluating the

677 anatomical alignment from the previous iteration. Global and local mappings are then all
678 concatenated (with appropriate inversions) to create the final mapping between the MERFISH
679 data and AllenCCFv3. This mapping is then used to provide a point-to-point correspondence
680 between the original MERFISH coordinate space and the AllenCCFv3 space, thus allowing
681 mapping of individual genes and cell types located in the MERFISH data to be directly
682 mapped into the AllenCCFv3.

683 4.4 DevCCF velocity flow transformation model

684 Given multiple, linearly or non-linearly ordered point sets where individual points across the
685 sets are in one-to-one correspondence, we developed an approach for generating a velocity
686 flow transformation model to describe a time-varying diffeomorphic mapping as a variant of
687 the landmark matching solution. Integration of the resulting velocity field can then be used
688 to describe the displacement between any two time points within this time-parameterized
689 domain. Regularization of the sparse correspondence between point sets is performed using a
690 generalized B-spline scattered data approximation technique⁸³, also created by the ANTsX
691 developers and contributed to ITK.

692 4.4.1 Velocity field optimization

693 To apply this methodology to the developmental templates¹⁶, we coalesced the manual
694 annotations of the developmental templates into 26 common anatomical regions (see Figure
695 3). We then used these regions to generate invertible transformations between successive time
696 points. Specifically each label was used to create a pair of single region images resulting in 26
697 pairs of “source” and “target” images. The multiple image pairs were simultaneously used to
698 iteratively estimate a diffeomorphic pairwise transform. Given the seven atlases E11.5, E13.5,
699 E15.5, E18.5, P4, P14, and P56, this resulted in 6 sets of transforms between successive time
700 points. Approximately 10^6 points were randomly sampled labelwise in the P56 template
701 space and propagated to each successive atlas providing the point sets for constructing the
702 velocity flow model. Approximately 125 iterations resulted in a steady convergence based on
703 the average Euclidean norm between transformed point sets. Ten integration points were

704 used and point sets were distributed along the temporal dimension using a log transform for
705 a more evenly spaced sampling. For additional information a help menu is available for the
706 ANTsPy function `ants.fit_time_varying_transform_to_point_sets(...)`.

707 4.5 ANTsXNet mouse brain applications

708 4.5.1 General notes regarding deep learning training

709 All network-based approaches described below were implemented and organized in the
710 ANTsXNet libraries comprising Python (ANTsPyNet) and R (ANTsRNet) analogs using the
711 Keras/Tensorflow libraries available as open-source in ANTsX GitHub repositories. For the
712 various applications, both share the identically trained weights for mutual reproducibility.
713 For all GPU training, we used Python scripts for creating custom batch generators which we
714 maintain in a separate GitHub repository for public availability (<https://github.com/ntustison/ANTsXNetTraining>). These scripts provide details such as batch size, choice of loss
715 function, and network parameters. In terms of GPU hardware, all training was done on a
716 DGX (GPUs: 4X Tesla V100, system memory: 256 GB LRDIMM DDR4).

718 Data augmentation is crucial for generalizability and accuracy of the trained networks.
719 Intensity-based data augmentation consisted of randomly added noise (i.e., Gaussian, shot,
720 salt-and-pepper), simulated bias fields based on N4 bias field modeling, and histogram warping
721 for mimicking well-known MRI intensity nonlinearities^{46,86}. These augmentation techniques
722 are available in ANTsXNet (only ANTsPyNet versions are listed with ANTsRNet versions
723 available) and include:

- 724 • image noise: `ants.add_noise_to_image(...)`,
- 725 • simulated bias field: `antspynet.simulate_bias_field(...)`, and
- 726 • nonlinear intensity warping: `antspynet.histogram_warp_image_intensities(...)`.

727 Shape-based data augmentation used both random linear and nonlinear deformations in
728 addition to anisotropic resampling in the three canonical orientations to mimic frequently
729 used acquisition protocols for mice brains:

- 730 • random spatial warping: `antspynet.randomly_transform_image_data(...)` and
731 • anisotropic resampling: `ants.resample_image(...)`.

732

4.5.2 Brain extraction

733 Similar to human neuroimage processing, brain extraction is a crucial preprocessing step for
734 accurate brain mapping. We developed similar functionality for T2-weighted mouse brains.
735 This network uses a conventional U-net architecture⁸⁷ and, in ANTsPyNet, this functionality is
736 available in the program `antspynet.mouse_brain_extraction(...)`. For the two-shot T2-
737 weighted brain extraction network, two brain templates were generated along with their masks.
738 One of the templates was generated from orthogonal multi-plane, high resolution data⁶⁹ which
739 were combined to synthesize isotropic volumetric data using the B-spline fitting algorithm⁸³.
740 This algorithm is encapsulated in `ants.fit_bspline_object_to_scattered_data(...)`
741 where the input is the set of voxel intensity values and each associated physical location.
742 Since each point can be assigned a confidence weight, we use the normalized gradient value
743 to more heavily weight edge regions. Although both template/mask pairs are available in the
744 GitHub repository associated with this work, the synthesized volumetric B-spline T2-weighted
745 pair is available within ANTsXNet through the calls:

- 746 • template: `antspynet.get_antsxnet_data("bsplineT2MouseTemplate")` and
747 • mask: `antspynet.get_antsxnet_data("bsplineT2MouseTemplateBrainMask")`.

748

4.5.3 Brain parcellation

749 The T2-weighted brain parcellation network is also based on a 3-D U-net architecture and
750 the T2-w DevCCF P56 template component with extensive data augmentation, as described
751 previously. Intensity differences between the template and any brain extracted input image
752 are minimized through the use of the rank intensity transform (`ants.rank_intensity(...)`).
753 Shape differences are reduced by the additional preprocessing step of warping the brain
754 extracted input image to the template. Additional input channels include the prior probability
755 images created from the template parcellation. These images are also available through the
756 ANTsXNet `get_antsxnet_data(...)` interface.

757 **Data availability**

758 All data and software used in this work are publicly available. The DevCCF atlas is available
759 at <https://kimlab.io/brain-map/DevCCF/>. ANTsPy, ANTsR, ANTsPyNet, and ANTsRNet
760 are available through GitHub at the ANTsX Ecosystem (<https://github.com/ANTsX>).
761 Training scripts for all deep learning functionality in ANTsXNet can also be found on GitHub
762 (<https://github.com/ntustison/ANTsXNetTraining>). A GitHub repository specifically
763 pertaining to the AllenCCFv3 mapping is available at <https://github.com/dontminchen>
764 [it/CCFAAlignmentToolkit](#). For the other two contributions contained in this work, the interested
765 longitudinal DevCCF mapping and mouse cortical thickness pipeline, we refer the interested
766 reader to <https://github.com/ntustison/ANTsXMouseBrainMapping>.

⁷⁶⁷ **Acknowledgments**

⁷⁶⁸ Support for the research reported in this work includes funding from the National Institute
⁷⁶⁹ of Biomedical Imaging and Bioengineering (R01-EB031722) and National Institute of Mental
⁷⁷⁰ Health (RF1-MH124605 and U24-MH114827).

771 Author contributions

772 N.T., M.C., and J.G. wrote the main manuscript text and figures. M.C., M.K., R.D., S.S.,
773 Q.W., L.G., J.D., C.G., and J.G. developed the Allen registration pipelines. N.T. and F.K.
774 developed the time-varying velocity transformation model for the DevCCF. N.T. and M.T.
775 developed the brain parcellation and cortical thickness methodology. All authors reviewed
776 the manuscript.

777 References

- 778 1. Keller, P. J. & Ahrens, M. B. Visualizing whole-brain activity and development at the
single-cell level using light-sheet microscopy. *Neuron* **85**, 462–83 (2015).
- 779 2. La Manno, G. *et al.* Molecular architecture of the developing mouse brain. *Nature*
596, 92–96 (2021).
- 780 3. Wen, L. *et al.* Single-cell technologies: From research to application. *Innovation
(Camb)* **3**, 100342 (2022).
- 781 4. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–14
(2014).
- 782 5. Gong, H. *et al.* Continuously tracing brain-wide long-distance axonal projections in
mice at a one-micron voxel resolution. *Neuroimage* **74**, 87–98 (2013).
- 783 6. Li, A. *et al.* Micro-optical sectioning tomography to obtain a high-resolution atlas of
the mouse brain. *Science* **330**, 1404–8 (2010).
- 784 7. Ueda, H. R. *et al.* Tissue clearing and its applications in neuroscience. *Nat Rev
Neurosci* **21**, 61–79 (2020).
- 785 8. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by
spatial transcriptomics. *Science* **353**, 78–82 (2016).
- 786 9. Burgess, D. J. Spatial transcriptomics coming of age. *Nat Rev Genet* **20**, 317 (2019).
- 787 10. Hardwick, S. A. *et al.* Single-nuclei isoform RNA sequencing unlocks barcoded exon
connectivity in frozen brain tissue. *Nature biotechnology* **40**, 1082–1092 (2022).
- 788 11. Hawrylycz, M. *et al.* A guide to the BRAIN initiative cell census network data
ecosystem. *PLoS biology* **21**, e3002133 (2023).
- 789 12. Wang, Q. *et al.* The allen mouse brain common coordinate framework: A 3D reference
atlas. *Cell* **181**, 936–953.e20 (2020).
- 790 13. Perens, J. *et al.* An optimized mouse brain atlas for automated mapping and quantifi-
cation of neuronal activity using iDISCO+ and light sheet fluorescence microscopy.
Neuroinformatics **19**, 433–446 (2021).
- 791 14. Ma, Y. *et al.* A three-dimensional digital atlas database of the adult C57BL/6J mouse
brain by magnetic resonance microscopy. *Neuroscience* **135**, 1203–1215 (2005).

- 792 15. Qu, L. *et al.* Cross-modal coherent registration of whole mouse brains. *Nature Methods* **19**, 111–118 (2022).
- 793 16. Kronman, F. A. *et al.* Developmental mouse brain common coordinate framework. *bioRxiv* (2023) doi:[10.1101/2023.09.14.557789](https://doi.org/10.1101/2023.09.14.557789).
- 794 17. Chuang, N. *et al.* An MRI-based atlas and database of the developing mouse brain. *Neuroimage* **54**, 80–89 (2011).
- 795 18. Dries, R. *et al.* Advances in spatial transcriptomic data analysis. *Genome research* **31**, 1706–1718 (2021).
- 796 19. Ricci, P. *et al.* Removing striping artifacts in light-sheet fluorescence microscopy: A review. *Progress in biophysics and molecular biology* **168**, 52–65 (2022).
- 797 20. Agarwal, N., Xu, X. & Gopi, M. Robust registration of mouse brain slices with severe histological artifacts. in *Proceedings of the tenth indian conference on computer vision, graphics and image processing* 1–8 (2016).
- 798 21. Agarwal, N., Xu, X. & Gopi, M. Automatic detection of histological artifacts in mouse brain slice images. in *Medical computer vision and bayesian and graphical models for biomedical imaging: MICCAI 2016 international workshops, MCV and BAMBI, athens, greece, october 21, 2016, revised selected papers* 8 105–115 (Springer, 2017).
- 799 22. Tward, D. *et al.* 3d mapping of serial histology sections with anomalies using a novel robust deformable registration algorithm. in *International workshop on multimodal brain image analysis* 162–173 (Springer, 2019).
- 800 23. Cahill, L. S. *et al.* Preparation of fixed mouse brains for MRI. *Neuroimage* **60**, 933–939 (2012).
- 801 24. Sunkin, S. M. *et al.* Allen brain atlas: An integrated spatio-temporal portal for exploring the central nervous system. *Nucleic acids research* **41**, D996–D1008 (2012).
- 802 25. Kim, Y. *et al.* Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171**, 456–469 (2017).
- 803 26. Fürth, D. *et al.* An interactive framework for whole-brain maps at cellular resolution. *Nat Neurosci* **21**, 139–149 (2018).

- 804 27. Li, Y. *et al.* mBrainAligner-web: A web server for cross-modal coherent registration
of whole mouse brains. *Bioinformatics* **38**, 4654–4655 (2022).
- 805 28. Puchades, M. A., Csucs, G., Ledergerber, D., Leergaard, T. B. & Bjaalie, J. G. Spatial
registration of serial microscopic brain images to three-dimensional reference atlases
with the QuickNII tool. *PloS one* **14**, e0216796 (2019).
- 806 29. Eastwood, B. S. *et al.* Whole mouse brain reconstruction and registration to a
reference atlas with standard histochemical processing of coronal sections. *Journal of
Comparative Neurology* **527**, 2170–2178 (2019).
- 807 30. Ni, H. *et al.* A robust image registration interface for large volume brain atlas. *Sci
Rep* **10**, 2139 (2020).
- 808 31. Pallast, N. *et al.* Processing pipeline for atlas-based imaging data analysis of structural
and functional mouse brain MRI (AIDAmri). *Front Neuroinform* **13**, 42 (2019).
- 809 32. Celestine, M., Nadkarni, N. A., Garin, C. M., Bougacha, S. & Dhenain, M. **Sammba-
MRI: A library for processing SmAll-MaMmal BrAin MRI data in python**. *Front
Neuroinform* **14**, 24 (2020).
- 810 33. Ioanas, H.-I., Marks, M., Zerbi, V., Yanik, M. F. & Rudin, M. An optimized registration
workflow and standard geometric space for small animal brain imaging. *Neuroimage*
241, 118386 (2021).
- 811 34. Perens, J. *et al.* Multimodal 3D mouse brain atlas framework with the skull-derived
coordinate system. *Neuroinformatics* **21**, 269–286 (2023).
- 812 35. Aggarwal, M., Zhang, J., Miller, M. I., Sidman, R. L. & Mori, S. Magnetic resonance
imaging and micro-computed tomography combined atlas of developing and adult
mouse brains for stereotaxic surgery. *Neuroscience* **162**, 1339–1350 (2009).
- 813 36. Goubran, M. *et al.* Multimodal image registration and connectivity analysis for
integration of connectomic data from microscopy to MRI. *Nat Commun* **10**, 5504
(2019).
- 814 37. Chandrashekhar, V. *et al.* CloudReg: Automatic terabyte-scale cross-modal brain
volume registration. *Nature methods* **18**, 845–846 (2021).

- 815 38. Jin, M. *et al.* SMART: An open-source extension of WholeBrain for intact mouse
brain registration and segmentation. *eNeuro* **9**, (2022).
- 816 39. Negwer, M. *et al.* FriendlyClearMap: An optimized toolkit for mouse brain mapping
and analysis. *Gigascience* **12**, (2022).
- 817 40. Lin, W. *et al.* Whole-brain mapping of histaminergic projections in mouse brain.
Proceedings of the National Academy of Sciences **120**, e2216231120 (2023).
- 818 41. Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by
MERFISH. *Nature* **598**, 137–143 (2021).
- 819 42. Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution.
Nature **622**, 552–561 (2023).
- 820 43. Zhang, Y. *et al.* Reference-based cell type matching of in situ image-based spatial
transcriptomics data on primary visual cortex of mouse brain. *Scientific Reports* **13**,
9567 (2023).
- 821 44. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. Elastix: A
toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* **29**,
196–205 (2010).
- 822 45. Fedorov, A. *et al.* 3D slicer as an image computing platform for the quantitative
imaging network. *Magnetic resonance imaging* **30**, 1323–1341 (2012).
- 823 46. Tustison, N. J. *et al.* The ANTsX ecosystem for quantitative biological and medical
imaging. *Sci Rep* **11**, 9068 (2021).
- 824 47. Pagani, M., Damiano, M., Galbusera, A., Tsafaris, S. A. & Gozzi, A. Semi-automated
registration-based anatomical labelling, voxel based morphometry and cortical thick-
ness mapping of the mouse brain. *Journal of neuroscience methods* **267**, 62–73
(2016).
- 825 48. Anderson, R. J. *et al.* Small animal multivariate brain analysis (SAMBA) - a high
throughput pipeline with a validation framework. *Neuroinformatics* **17**, 451–472
(2019).
- 826 49. Allan Johnson, G. *et al.* Whole mouse brain connectomics. *Journal of Comparative
Neurology* **527**, 2146–2157 (2019).

- 827 50. Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the
whole mouse brain. *Nature* **624**, 317–332 (2023).
- 828 51. Bajcsy, R. & Broit, C. Matching of deformed images. in *Sixth International Conference
on Pattern Recognition (ICPR'82)* 351–353 (1982).
- 829 52. Bajcsy, R. & Kovacic, S. Multiresolution elastic matching. *Computer Vision, Graphics,
and Image Processing* **46**, 1–21 (1989).
- 830 53. Gee, J. C., Reivich, M. & Bajcsy, R. Elastically deforming 3D atlas to match anatomical
brain images. *J Comput Assist Tomogr* **17**, 225–36 (1993).
- 831 54. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic
image registration with cross-correlation: Evaluating automated labeling of elderly
and neurodegenerative brain. *Med Image Anal* **12**, 26–41 (2008).
- 832 55. Klein, A. *et al.* Evaluation of 14 nonlinear deformation algorithms applied to human
brain MRI registration. *Neuroimage* **46**, 786–802 (2009).
- 833 56. Murphy, K. *et al.* Evaluation of registration methods on thoracic CT: The EMPIRE10
challenge. *IEEE Trans Med Imaging* **30**, 1901–20 (2011).
- 834 57. Baheti, B. *et al.* The brain tumor sequence registration challenge: Establishing
correspondence between pre-operative and follow-up MRI scans of diffuse glioma
patients. (2021).
- 835 58. Avants, B. B. *et al.* The optimal template effect in hippocampus studies of diseased
populations. *Neuroimage* **49**, 2457–66 (2010).
- 836 59. Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A. & Gee, J. C. An open source
multivariate framework for n-tissue segmentation with evaluation on public data.
Neuroinformatics **9**, 381–400 (2011).
- 837 60. Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L. & Robles, M. Adaptive
non-local means denoising of MR images with spatially varying noise levels. *J Magn
Reson Imaging* **31**, 192–203 (2010).
- 838 61. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging*
29, 1310–20 (2010).

- 839 62. Wang, H. *et al.* Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Anal Mach Intell* **35**, 611–23 (2013).
- 840 63. Tustison, N. J. *et al.* Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. *Neuroinformatics* (2014) doi:10.1007/s12021-014-9245-2.
- 841 64. Tustison, N. J., Yang, Y. & Salerno, M. Advanced normalization tools for cardiac motion correction. in *Statistical atlases and computational models of the heart - imaging and modelling challenges* (eds. Camara, O. et al.) vol. 8896 3–12 (Springer International Publishing, 2015).
- 842 65. McCormick, M., Liu, X., Jomier, J., Marion, C. & Ibanez, L. ITK: Enabling reproducible research and open science. *Front Neuroinform* **8**, 13 (2014).
- 843 66. Beg, M. F., Miller, M. I., Trouvé, A. & Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision* **61**, 139–157 (2005).
- 844 67. Tustison, N. J. & Avants, B. B. Explicit B-spline regularization in diffeomorphic image registration. *Front Neuroinform* **7**, 39 (2013).
- 845 68. Hsu, L.-M. *et al.* CAMRI mouse brain MRI data.
- 846 69. Reshetnikov, V. *et al.* High-resolution MRI data of brain C57BL/6 and BTBR mice in three different anatomical views.
- 847 70. Rahman, N., Xu, K., Budde, M. D., Brown, A. & Baron, C. A. A longitudinal microstructural MRI dataset in healthy C57Bl/6 mice at 9.4 tesla. *Sci Data* **10**, 94 (2023).
- 848 71. Liu, J. *et al.* Concordance of MERFISH spatial transcriptomics with bulk and single-cell RNA sequencing. *Life Sci Alliance* **6**, (2023).
- 849 72. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: A generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100–106 (2021).
- 850 73. Jia, H., Yap, P.-T., Wu, G., Wang, Q. & Shen, D. Intermediate templates guided groupwise registration of diffusion tensor images. *NeuroImage* **54**, 928–939 (2011).

- 851 74. Tang, S., Fan, Y., Wu, G., Kim, M. & Shen, D. RABBIT: Rapid alignment of brains
by building intermediate templates. *NeuroImage* **47**, 1277–1287 (2009).
- 852 75. Dewey, B. E., Carass, A., Blitz, A. M. & Prince, J. L. Efficient multi-atlas registration
using an intermediate template image. in *Proceedings of SPIE—the international society
for optical engineering* vol. 10137 (NIH Public Access, 2017).
- 853 76. Gong, H. *et al.* High-throughput dual-colour precision imaging for brain-wide connec-
tome with cytoarchitectonic landmarks at the cellular level. *Nat Commun* **7**, 12142
(2016).
- 854 77. Wang, J. *et al.* Divergent projection patterns revealed by reconstruction of individual
neurons in orbitofrontal cortex. *Neurosci Bull* **37**, 461–477 (2021).
- 855 78. Rotolo, T., Smallwood, P. M., Williams, J. & Nathans, J. Genetically-directed, cell
type-specific sparse labeling for the analysis of neuronal morphology. *PLoS One* **3**,
e4099 (2008).
- 856 79. Peng, H. *et al.* Morphological diversity of single neurons in molecularly defined cell
types. *Nature* **598**, 174–181 (2021).
- 857 80. Das, S. R., Avants, B. B., Grossman, M. & Gee, J. C. Registration based cortical
thickness measurement. *Neuroimage* **45**, 867–79 (2009).
- 858 81. Tustison, N. J. *et al.* Large-scale evaluation of ANTs and FreeSurfer cortical thickness
measurements. *Neuroimage* **99**, 166–79 (2014).
- 859 82. Tustison, N. J. *et al.* Longitudinal mapping of cortical thickness measurements: An
Alzheimer’s Disease Neuroimaging Initiative-based evaluation study. *J Alzheimers Dis*
(2019) doi:[10.3233/JAD-190283](https://doi.org/10.3233/JAD-190283).
- 860 83. Tustison, N. J. & Amini, A. A. Biventricular myocardial strains via nonrigid registration
of anatomical NURBS model [corrected]. *IEEE Trans Med Imaging* **25**, 94–112 (2006).
- 861 84. Avants, B. B. *et al.* The Insight ToolKit image registration framework. *Front Neu-
roinform* **8**, 44 (2014).
- 862 85. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance
in brain image registration. *Neuroimage* **54**, 2033–44 (2011).

- 863 86. Nyúl, L. G., Udupa, J. K. & Zhang, X. New variants of a method of MRI scale
standardization. *IEEE Trans Med Imaging* **19**, 143–50 (2000).
- 864 87. Falk, T. *et al.* U-net: Deep learning for cell counting, detection, and morphometry.
Nat Methods **16**, 67–70 (2019).