

1

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

4 Nicholas J. Tustison<sup>1</sup>, Min Chen<sup>2</sup>, Fae N. Kronman<sup>3</sup>, Jeffrey T. Duda<sup>2</sup>, Clare Gamlin<sup>4</sup>, Mia  
5 G. Tustison, Michael Kunst<sup>4</sup>, Rachel Dalley<sup>4</sup>, Staci Sorenson<sup>4</sup>, Quanxin Wang<sup>4</sup>, Lydia Ng<sup>4</sup>,  
6 Yongsoo Kim<sup>3</sup>, and James C. Gee<sup>2</sup>

7 <sup>1</sup>Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

8 <sup>2</sup>Department of Radiology, University of Pennsylvania, Philadelphia, PA

9 <sup>3</sup>Department of Neural and Behavioral Sciences, Penn State University, Hershey, PA

10 <sup>4</sup>Allen 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](mailto:ntustison@virginia.edu)

18

19 James C. Gee, PhD

20 Department of Radiology

21 University of Pennsylvania

22 [gee@upenn.edu](mailto:gee@upenn.edu)

<sup>23</sup> **Abstract**

<sup>24</sup> Large-scale, international collaborative efforts by members of the BRAIN Initiative Cell  
<sup>25</sup> Census Network (BICCN) consortium have recently begun aggregating the most compre-  
<sup>26</sup> hensive reference database to date for diverse cell type profiling of the mouse brain, which  
<sup>27</sup> encompasses over 40 different multi-modal profiling techniques from more than 30 research  
<sup>28</sup> groups. One central challenge for this integrative effort across different investigators and  
<sup>29</sup> laboratories has been the need to map these unique datasets into common reference spaces  
<sup>30</sup> such that the spatial, structural, and functional information from different cell types can be  
<sup>31</sup> jointly analyzed across modalities. However, significant variations in the acquisition, tissue  
<sup>32</sup> processing, and imaging techniques across data types makes mapping such diverse data a  
<sup>33</sup> multifarious problem. Different data types exhibit unique tissue distortion and signal char-  
<sup>34</sup> acteristics that precludes a single mapping strategy from being generally applicable across  
<sup>35</sup> all cell type data. Diverse, and often specialized, mapping approaches are needed to address  
<sup>36</sup> the unique barriers present in each modality. This work highlights modular atlas mapping  
<sup>37</sup> strategies developed across three separate BICCN studies using the Advanced Normalization  
<sup>38</sup> Tools Ecosystem (ANTsX) to map spatial transcriptomic (MERFISH) and high-resolution  
<sup>39</sup> morphology (fMOST) mouse brain data into the Allen Common Coordinate Framework  
<sup>40</sup> (AllenCCFv3), and developmental (MRI and LSFM) data into the Developmental Common  
<sup>41</sup> Coordinate Framework (DevCCF). We discuss both common mapping strategies that can be  
<sup>42</sup> shared across modalities, and targeted strategies driven by specific challenges from each data  
<sup>43</sup> type. Novel open-source contributions, made publicly available through ANTSX, include a  
<sup>44</sup> generic velocity flow-based approach for continuously mapping developmental trajectories  
<sup>45</sup> such as that characterizing the DevCCF as well as an automated framework for determining  
<sup>46</sup> structural morphology made possible through the leveraging of public resources such as the  
<sup>47</sup> AllenCCFv3 and the DevCCF. Finally, we provide general guidance to aid investigators in  
<sup>48</sup> their efforts to tailor these strategies to address unique challenges in their data without the  
<sup>49</sup> need to develop additional specialized software.

## 50 1 Introduction

51 Over the past decade there have been significant advancements in mesoscopic single-cell anal-  
52 ysis of the mouse brain. It is now possible to track single neurons in mouse brains<sup>1</sup>, observe  
53 whole brain developmental changes on a cellular level<sup>2</sup>, associate brain regions and tissues  
54 with their genetic composition<sup>3</sup>, and locally characterize neural connectivity<sup>4</sup>. Much of these  
55 scientific achievements have been made possible due to breakthroughs in high resolution cell  
56 profiling and imaging techniques that permit submicron, multi-modal, 3D characterizations  
57 of whole mouse brains. Among these include advanced techniques such as micro-optical  
58 sectioning tomography<sup>6</sup>, tissue clearing<sup>1,7</sup>, spatial transcriptomics<sup>9</sup>, and single-cell genomic  
59 profiling<sup>10</sup>, which have greatly expanded the resolution and specificity of single-cell measure-  
60 ments in the brain.

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

76 **1.1 Mouse brain mapping**

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

91 Existing solutions to address mapping cell type data into the AllenCCFv3 falls broadly into  
92 three main categories. The first consists of integrated processing platforms that directly  
93 provide mapped data to the users. These include the Allen Brain Cell Atlas<sup>24</sup> for the Allen  
94 Reference Atlas (ARA) and associated data, the Brain Architecture Portal<sup>25</sup> for combined  
95 ex vivo radiology and histology data, OpenBrainMap<sup>26</sup> for connectivity data, and the Image  
96 and Multi-Morphology Pipeline<sup>27</sup> for high resolution morphology data. These platforms  
97 provide users online access to pre-processed, multi-modal cell type data that are already  
98 mapped to the AllenCCFv3. The platforms are designed such that the data is interactively  
99 manipulated by users through integrated visualization software that allow users to spatially  
100 manipulate and explore each dataset within the mapped space. While highly convenient  
101 for investigators who are interested in studying the specific modalities provided by these  
102 platforms, these systems can be limited in flexibility and more general applicability. The  
103 mapping software and pipelines are typically developed specifically with the data type and  
104 platform in mind, and the software is limited public availability. Investigators will find it

<sup>105</sup> difficult to apply the same mapping to their own data without direct collaboration with the  
<sup>106</sup> platform owners.

<sup>107</sup> The second category are specialized approaches specifically designed for mapping one or  
<sup>108</sup> more modalities into a CCF. These approaches use combinations of specialized manual and  
<sup>109</sup> automated processes that address specific challenges in each modality. Examples include ap-  
<sup>110</sup> proaches for mapping histology<sup>28–30</sup>, magnetic resonance imaging (MRI)<sup>37</sup>, micro-computed  
<sup>111</sup> tomography (microCT)<sup>35,37</sup>, light-sheet fluorescence microscopy (LSFM)<sup>34,36–39</sup>, fluorescence  
<sup>112</sup> micro-optical sectioning tomography (fMOST)<sup>15,40</sup> and transcriptomic data<sup>41–43</sup>. As special-  
<sup>113</sup> ized approaches, these techniques tend to boast higher mapping accuracy, robustness, and  
<sup>114</sup> ease of use when ran with applicable modalities. Conversely, their specialized designs often  
<sup>115</sup> rely on base assumptions regarding the data type that can make them rigid and difficult  
<sup>116</sup> to adapt for new modalities or unexpected artifacts and distortions in the data. Retooling  
<sup>117</sup> these specialize software to use with new data can require significant development, validation  
<sup>118</sup> time, and engineering expertise that may not be readily available for all investigators.

<sup>119</sup> The last category are modular mapping approaches constructed using general image analy-  
<sup>120</sup> sis toolkits, which are software packages that include varied collections of image processing,  
<sup>121</sup> segmentation and registration tools that have been previously developed, and validated for  
<sup>122</sup> multiple application areas. Examples of such toolkits include elastix<sup>44</sup>, Slicer3D<sup>45</sup>, ANTsX<sup>46</sup>,  
<sup>123</sup> and several others which have all been applied towards mouse brain spatial mapping. The  
<sup>124</sup> main challenge, in these mouse-specific study scenarios, is that tailored pipelines often need  
<sup>125</sup> be constructed from available software components. Investigators must therefore be familiar  
<sup>126</sup> with the these tools for formulating new or adapting existing pipelines. However, in com-  
<sup>127</sup> parison to previously described specialized mapping approaches, these approaches are often  
<sup>128</sup> easier to create and prone to robustness, being typically constructed from pipelin compo-  
<sup>129</sup> nents which have been previously vetted in other contexts. In this work, we highlight such  
<sup>130</sup> mapping strategies designed using the ANTsX framework to map three distinct mouse cell  
<sup>131</sup> type data with different characteristics into existing CCFs.

<sup>132</sup> **1.2 Advanced Normalization Tools (ANTsX)**

<sup>133</sup> The Advanced Normalization Tools Ecosystem (ANTsX) has been used in a number of  
<sup>134</sup> applications for mapping mouse brain data as part of core processing steps in various  
<sup>135</sup> workflows<sup>30,47–50</sup>, particularly its pairwise, intensity-based image registration capabilities and  
<sup>136</sup> bias field correction. Historically, ANTsX development is originally based on fundamental  
<sup>137</sup> approaches to image mapping<sup>51–53</sup>, particularly in the human brain, which has resulted  
<sup>138</sup> in core contributions to the field such as the widely-used Symmetric Normalization (SyN)  
<sup>139</sup> algorithm<sup>54</sup>. Since its development, various independent platforms have been used to eval-  
<sup>140</sup> uate ANTsX image registration capabilities in the context of different application foci which  
<sup>141</sup> include multi-site brain MRI data<sup>55</sup>, pulmonary CT data<sup>56</sup>, and most recently, multi-modal  
<sup>142</sup> brain registration in the presence of tumors<sup>57</sup>.

<sup>143</sup> Apart from its registration capabilities, ANTsX comprises additional functionality such  
<sup>144</sup> as template generation<sup>58</sup>, intensity-based segmentation<sup>59</sup>, preprocessing<sup>60,61</sup>, deep learning  
<sup>145</sup> networks<sup>46</sup>, and other utilities relevant to brain mapping (see Table 1). The use of the toolkit  
<sup>146</sup> has demonstrated high performance in multiple application areas (e.g., consensus labeling<sup>62</sup>,  
<sup>147</sup> brain tumor segmentation<sup>63</sup>, and cardiac motion estimation<sup>64</sup>). Importantly, ANTsX is built  
<sup>148</sup> on the Insight Toolkit (ITK)<sup>65</sup> deriving benefit from the open-source community of sci-  
<sup>149</sup> entists and programmers as well as providing an important resource for algorithmic develop-  
<sup>150</sup> ment, evaluation, and improvement. In this paper we demonstrate how ANTsX provides a  
<sup>151</sup> comprehensive toolset provides the basis to develop modular frameworks for mapping di-  
<sup>152</sup> verse mouse cell type data into common coordinate frameworks (CCFs). Specifically, we  
<sup>153</sup> highlight its application for mapping data from three separate BICCN projects focused on  
<sup>154</sup> distinct data types: morphology data using fluorescence micro-optical sectioning tomog-  
<sup>155</sup> raphy (fMOST), spatial transcriptomics from multiplexed error-robust fluorescence in situ  
<sup>156</sup> hybridization (MERFISH) data, and time-series developmental data using light sheet fluores-  
<sup>157</sup> cence microscopy (LSFM) and magnetic resonance imaging (MRI). We describe both shared  
<sup>158</sup> and targeted strategies developed to address the specific challenges of these modalities.

<sup>159</sup> **1.3 Novel ANTsX-based open-source contributions**

<sup>160</sup> We introduce two novel inclusions to the ANTsX toolset that were developed as part of  
<sup>161</sup> the MRI mapping and analysis pipeline for the Developmental Common Coordinate Frame-  
<sup>162</sup> work (DevCCF). Consistent with previous ANTsX development, newly introduced capa-  
<sup>163</sup> bilities introduced below are available through ANTsX (specifically, via R and Python  
<sup>164</sup> ANTsX packages), and illustrated through self-contained examples in the ANTsX tuto-  
<sup>165</sup> rial (<https://tinyurl.com/antsxtutorial>) with a dedicated GitHub repository specific to this  
<sup>166</sup> work (<https://github.com/ntustison/ANTsXMouseBrainMapping>). To complement stan-  
<sup>167</sup> dard preprocessing steps (e.g., bias correction, brain masking), additional mouse brain spe-  
<sup>168</sup> cific tools have also been introduced to the ANTsX ecosystem, such as section reconstruction  
<sup>169</sup> and landmark-based alignment with corresponding processing scripts (<https://github.com/>  
<sup>170</sup> [dontminchenit/CCFAignmentToolkit](https://github.com/dontminchenit/CCFAignmentToolkit)).

<sup>171</sup> **1.3.1 Continuously mapping the DevCCF detjectory with a velocity flow**  
<sup>172</sup> **model**

<sup>173</sup> Recently, the Developmental Common Coordinate Framework (DevCCF) was introduced to  
<sup>174</sup> the mouse brain research community as a public resource<sup>16</sup> comprising symmetric atlases of  
<sup>175</sup> multimodal image data and anatomical segmentations defined by developmental ontology.  
<sup>176</sup> These templates sample the mouse embryonic days E11.5, E13.5, E15.5, E18.5 and postna-  
<sup>177</sup> tal days P4, P14, and P56. Modalities include light sheet floourescence miscroscopy (LSFM)  
<sup>178</sup> and at least four MRI contrasts per developmental stage. Anatomical parcellations are also  
<sup>179</sup> available for each time point and were generated from ANTsX-based mappings of gene ex-  
<sup>180</sup> pression and other cell type data. Additionally, the P56 template was integrated with the  
<sup>181</sup> AllenCCFv3 to further enhance the practical utility of the DevCCF. These processes, specif-  
<sup>182</sup> ically template generation and multi-modal image mapping, were performed using ANTsX  
<sup>183</sup> functionality in the presence of image mapping difficulties such as missing data and tissue  
<sup>184</sup> distortion.

<sup>185</sup> Given the temporal gaps in the discrete set of developmental atlases, we also provide an

<sup>186</sup> open-source framework for inferring correspondence within the temporally continuous do-  
<sup>187</sup> main sampled by the existing set of embryonic and postnatal atlases of the DevCCF. This  
<sup>188</sup> recently developed functionality permits the generation of a diffeomorphic velocity flow trans-  
<sup>189</sup> formation model<sup>66</sup>, influenced by previous work<sup>67</sup>. The resulting time-parameterized veloc-  
<sup>190</sup> ity field spans the stages of the DevCCF where mappings between any two continuous time  
<sup>191</sup> points within the span bounded by the E11.5 and P56 atlases is determined by integration  
<sup>192</sup> of the optimized velocity field.

### <sup>193</sup> 1.3.2 Automated structural parcellations of the mouse brain

<sup>194</sup> In contrast to the pipeline development in human data<sup>46</sup>, limited tools exist yet to cre-  
<sup>195</sup> ate adequate training data for automated parcellations of the mouse brain. In addition,  
<sup>196</sup> mouse brain data acquisition often has unique issues, such as lower data quality or sampling  
<sup>197</sup> anisotropy which limits its applicability to high resolution resources (e.g., AllenCCFv3, De-  
<sup>198</sup> vCCF), specifically with respect to the corresponding granular brain parcellations derived  
<sup>199</sup> from numerous hours of expert annotation leveraging multimodal imaging resources.

<sup>200</sup> Herein, we introduce a mouse brain parcellation pipeline for T2-weighted (T2-w) MRI com-  
<sup>201</sup> prising two novel deep learning components: two-shot learning brain extraction from data  
<sup>202</sup> augmentation of two ANTsX templates generated from two open datasets<sup>68,69</sup> and single-  
<sup>203</sup> shot brain parcellation derived from the AllenCCFv3 labelings mapped to the corresponding  
<sup>204</sup> DevCCF P56 T2-w component. Although we anticipate that this pipeline will be benefi-  
<sup>205</sup> cial to the research community, this work demonstrates more generally how one can leverage  
<sup>206</sup> ANTsX tools for developing quantitative mouse brain morphological tools using only publicly  
<sup>207</sup> available resources. Evaluation is performed on an independent open dataset<sup>70</sup> comprising  
<sup>208</sup> longitudinal acquisitions of multiple specimens.

209 **2 Results**

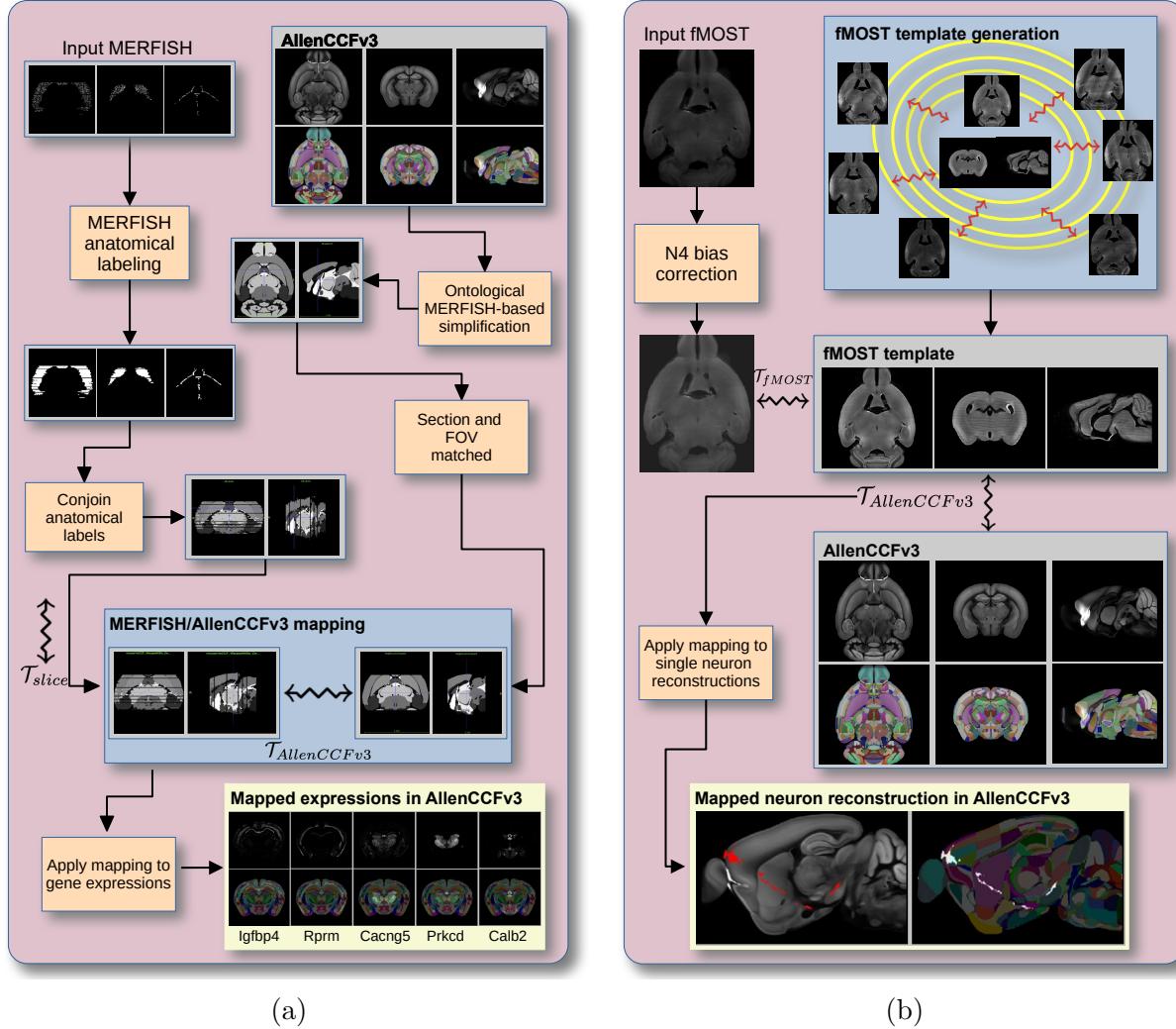


Figure 1: Diagram of the two ANTsX-based pipelines for mapping (a) MERFISH and (b)fMOST data into the space of AllenCCFv3. Each generates the requisite transforms,  $\mathcal{T}$ , to map individual images.

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 a  
216 high-resolution transcriptomic atlas of the mouse brain<sup>50</sup>, performs mappings by first gen-  
217 erating 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  
224 2) 2D global and deformable mapping between each MERFISH section and matched AllenC-  
225 CFv3 section. Mappings learned via each step in the pipeline are preserved and concatenated  
226 to provide point-to-point correspondence between the original MERFISH data and AllenC-  
227 CFv3, thus allowing individual gene expressions to be transferred into the AllenCCFv3.

228 **Data.** MERFISH mouse brain data was acquired using a previously detailed procedure<sup>50</sup>.  
229 Briefly, a brain of C57BL/6 mouse was dissected according to standard procedures and  
230 placed into an optimal cutting temperature (OCT) compound (Sakura FineTek 4583) in  
231 which it was stored at -80°C. The fresh frozen brain was sectioned at 10 $\mu$ m on Leica 3050  
232 S cryostats at intervals of 200 $\mu$ m to evenly cover the brain. A set of 500 genes were imaged  
233 that had been carefully chosen to distinguish the ~5200 clusters of our existing RNAseq  
234 taxonomy. For staining the tissue with MERFISH probes, a modified version of instructions  
235 provided by the manufacturer was used<sup>50</sup>. Raw MERSCOPE data were decoded using  
236 Vizgen software (v231). Cells were segmented based on DAPI and PolyT staining using  
237 Cellpose<sup>71,72</sup>. Segmentation was performed on a median z-plane (fourth out of seven) and  
238 cell borders were propagated to z-planes above and below. To assign cluster identity to each  
239 cell in the MERFISH dataset, we mapped the MERFISH cells to the scRNA-seq reference  
240 taxonomy.

241 **Evaluation.** Alignment of the MERFISH data into the AllenCCFv3 was qualitatively as-  
242 sessed by an expert anatomist at each iteration of the registration using known correspon-  
243 dence of gene markers and their associations with the AllenCCFv3. As previously reported<sup>50</sup>,

<sup>244</sup> further assessment of the alignment showed that, of the 554 terminal regions (gray matter  
<sup>245</sup> only) in the AllenCCFv3, only seven small subregions were missed from the MERFISH  
<sup>246</sup> dataset: frontal pole, layer 1 (FRP1), FRP2/3, FRP5; accessory olfactory bulb, glomerular  
<sup>247</sup> layer (AOBgl); accessory olfactory bulb, granular layer (AOBgr); accessory olfactory bulb,  
<sup>248</sup> mitral layer (AOBmi); and accessory supraoptic group (ASO).

### <sup>249</sup> 2.1.2 Mapping fluorescence micro-optical sectioning tomography (fMOST) data

<sup>250</sup> **Overview.** We developed a pipeline for mapping fluorescence micro-optical sectioning to-  
<sup>251</sup> mography (fMOST) mouse brain images into the AllenCCFv3 (see Figure 1(b)). The pipeline  
<sup>252</sup> is adapted from previously developed frameworks for human brain mapping<sup>58</sup>, and uses a  
<sup>253</sup> modality specific (fMOST) average atlas to assist in the image registration and mapping.  
<sup>254</sup> This approach has been well validated in human studies<sup>73–75</sup>, and successfully used in other  
<sup>255</sup> mouse data<sup>12,15,34</sup>. Briefly, we construct an intensity- and shape-based average fMOST atlas  
<sup>256</sup> using 30 fMOST images to serve as an intermediate registration target for mapping fMOST  
<sup>257</sup> images from individual specimens into the AllenCCFv3. Preprocessing steps include down-  
<sup>258</sup> sampling to match the 25 $\mu$ m isotropic AllenCCFv3, acquisition-based stripe artifact removal,  
<sup>259</sup> and inhomogeneity correction<sup>61</sup>. Preprocessing also includes a single annotation-driven reg-  
<sup>260</sup> istration to establish a canonical mapping between the fMOST atlas and the AllenCCFv3.  
<sup>261</sup> This step allows us to align expert determined landmarks to accurately map structures  
<sup>262</sup> with large morphological differences between the modalities, which are difficult to address  
<sup>263</sup> using standard approaches. Once this canonical mapping is established, standard intensity-  
<sup>264</sup> based registration is used to align each new fMOST image to the fMOST specific atlas.  
<sup>265</sup> This mapping is concatenated with the canonical fMOST atlas-to-AllenCCFv3 mapping to  
<sup>266</sup> further map each individual brain into the latter without the need to generate additional  
<sup>267</sup> landmarks. Transformations learned through this mapping can be applied to single neuron  
<sup>268</sup> reconstructions from the fMOST images to evaluate neuronal distributions across different  
<sup>269</sup> specimens into the AllenCCFv3 for the purpose of cell census analyses.

<sup>270</sup> **Data.** The high-throughput and high-resolution fluorescence micro-optical sectioning to-  
<sup>271</sup> mography (fMOST)<sup>76,77</sup> platform was used to image 55 mouse brains containing gene-defined

neuron populations, with sparse transgenic expression<sup>78,79</sup>. 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 Continuously mapping the DevCCF trajectory with a velocity flow model

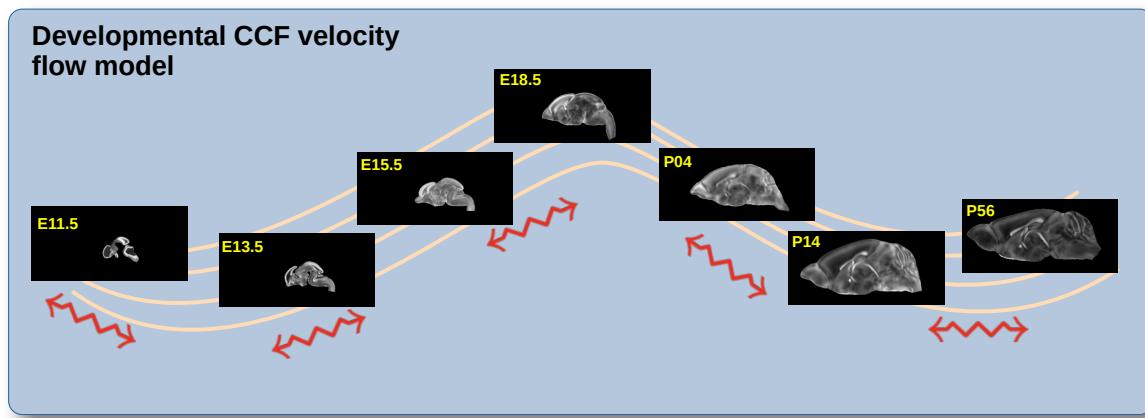


Figure 2: The spatial transformation between any two time points within the continuous DevCCF longitudinal developmental trajectory is available through the use of ANTsX functionality for generating a velocity flow model.

288 The DevCCF is an openly accessible resource for the mouse brain research community<sup>16</sup>. It  
289 consists of multi-modal MRI and LSFM symmetric ANTsX-generated templates<sup>58</sup> sampling  
290 the mouse brain developmental trajectory, specifically the embryonic (E) and postnatal (P)  
291 days E11.5, E13.5, E15.5, E18.5 P4, P14, and P56. Each template space includes structural  
292 labels defined by a developmental ontology. Its utility is also enhanced by a coordinated  
293 construction with AllenCCFv3. Although this work represents a significant contribution, the  
294 gaps between timepoints potentially limit its applicability which could be addressed through  
295 the development of the ability to map not only between timepoints but also within and  
296 across timepoints.

297 To continuously generate transformations between the different stages of the DevCCF atlases,  
298 we developed a general velocity flow model approach which we apply to DevCCF-derived  
299 data. We also introduce this functionality into both the ANTsR and ANTsPy packages (for  
300 the latter, see `ants.fit_time_varying_transform_to_point_sets(...)`) for potential  
301 application to this and other analogous scenarios (e.g., modeling the cardiac and respiratory  
302 cycles). ANTsX, being built on top of ITK, uses an ITK image data structure for the 4-D  
303 velocity field where each voxel contains the  $x$ ,  $y$ ,  $z$  components of the field at that point.

### 304 2.2.1 Data

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

312 Prior to velocity field optimization, all data were rigidly transformed to DevCCF P56 using  
313 the centroids of the common label sets. In order to determine the landmark correspondence  
314 across DevCCF stages, the multi-metric capabilities of `ants.registration(...)` were used.  
315 Instead of performing intensity-based pairwise registration directly on these multi-label im-

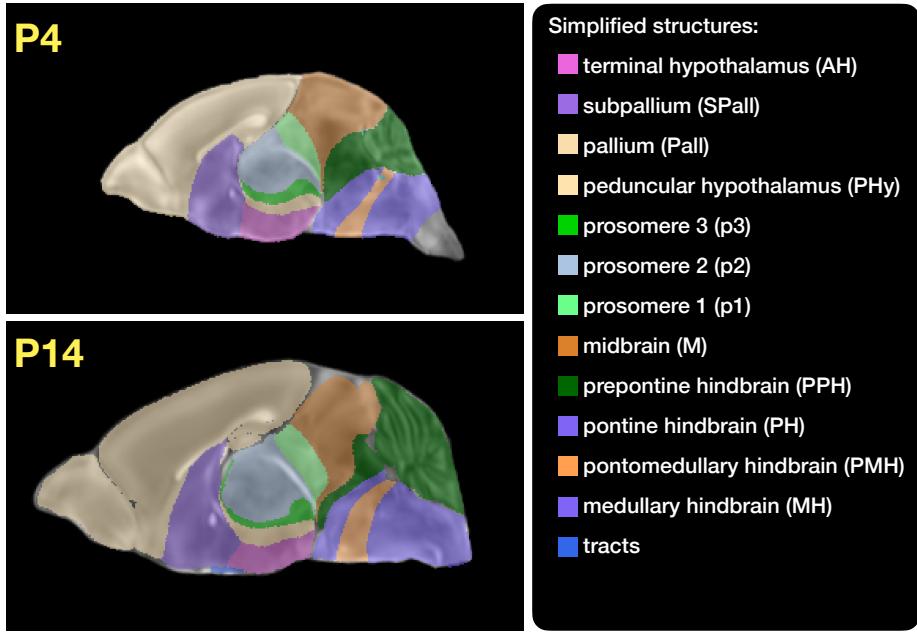


Figure 3: Annotated regions representing common labels across developmental stages which are illustrated for both P4 and P14.

316 ages, each label was used to construct a separate fixed and moving image pair resulting in a  
 317 multi-metric registration optimization scenario involving 24 binary image pairs (each label  
 318 weighted equally) for optimizing diffeomorphic correspondence between neighboring time  
 319 point atlases using the mean squares metric and the symmetric normalization transform<sup>54</sup>.

320 To generate the set of common point sets across all seven developmental atlases, the label  
 321 boundaries and whole regions were sampled in the P56 atlas and then propagated to each  
 322 atlas using the transformations derived from the pairwise registrations. We selected a sam-  
 323 pling rate of 10% for the contour points and 1% for the regional points for a total number  
 324 of points being per atlas being 173303 ( $N_{contour} = 98151$  and  $N_{region} = 75152$ ). Regional  
 325 boundary points were weighted twice as those of non-boundary points during optimization.

### 326 2.2.2 Optimization

327 The velocity field was optimized using the input composed of the seven corresponding point  
 328 sets and their associated weight values, the selected number of integration points for the  
 329 velocity field ( $N = 11$ ), and the parameters defining the geometry of the spatial dimensions

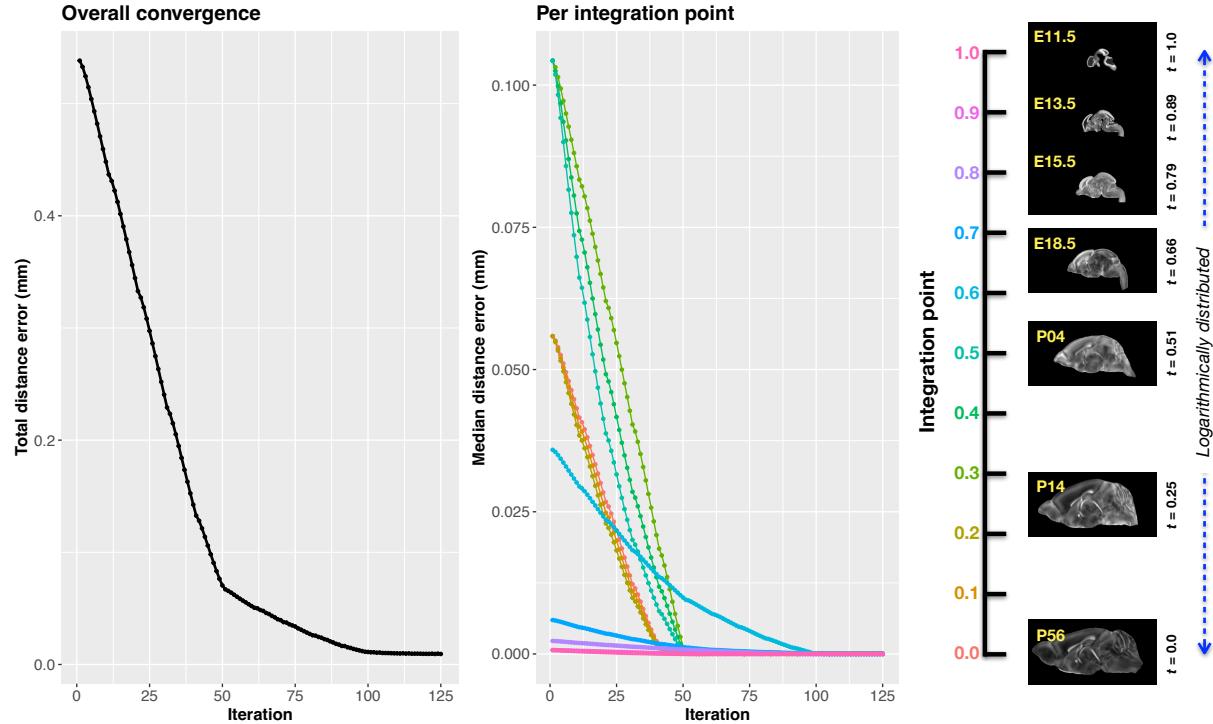


Figure 4: Convergence of the optimization of the velocity field for describing the transformation through the developmental stages from E11.5 through P56. Integration points in diagram on the right are color-coordinated with the center plot and placed in relation to the logarithmically situated temporal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

330 of the velocity field. Thus, the optimized velocity field described here is of size [256, 182, 360]  
 331 ( $50\mu\text{m}$  isotropic)  $\times$  11 integration points for a total compressed size of a little over 2 GB.  
 332 This choice represented weighing the trade-off between tractability, portability, and accuracy.  
 333 However, all data and code to reproduce the results described (with possible variation in the  
 334 input parameters) are available in the dedicated GitHub repository.

335 The normalized time point scalar value for each atlas/point-set in the temporal domains [0, 1]  
 336 was also defined. Given the increasingly larger gaps in the postnatal timepoint sampling, we  
 337 made two adjustments. Based on known mouse brain development, we used 28 days for the  
 338 P56 data. We then computed the log transform of the adjusted set of time points prior to  
 339 normalization between 0 and 1 (see the right side of Figure 4). This log transform, as part  
 340 of the temporal normalization, significantly improved data spacing.

341 The maximum number of iterations was set to 200 with each iteration taking approximately  
 342 six minutes on a 2020 iMac (processor, 3.6 GHz 10-Core Intel Core i9; memory, 64 GB 2667

343 MHz DDR4) At each iteration we looped over the 11 integration points. At each integration  
 344 point, the velocity field estimate was updated by warping the two immediately adjacent  
 345 point sets to the integration time point and determining the regularized displacement field  
 346 between the two warped point sets. As with any gradient-based descent algorithm, this field  
 347 was multiplied by a small step size ( $\delta = 0.2$ ) before adding to the current velocity field.  
 348 Convergence is determined by the average displacement error over each of the integration  
 349 points. As can be seen in the left panel of Figure 4, convergence occurred around 125  
 350 iterations when the average displacement error over all integration points is minimized. The  
 351 median displacement error at each of the integration points also trends towards zero but at  
 352 different rates.

### 353 2.2.3 The transformation model

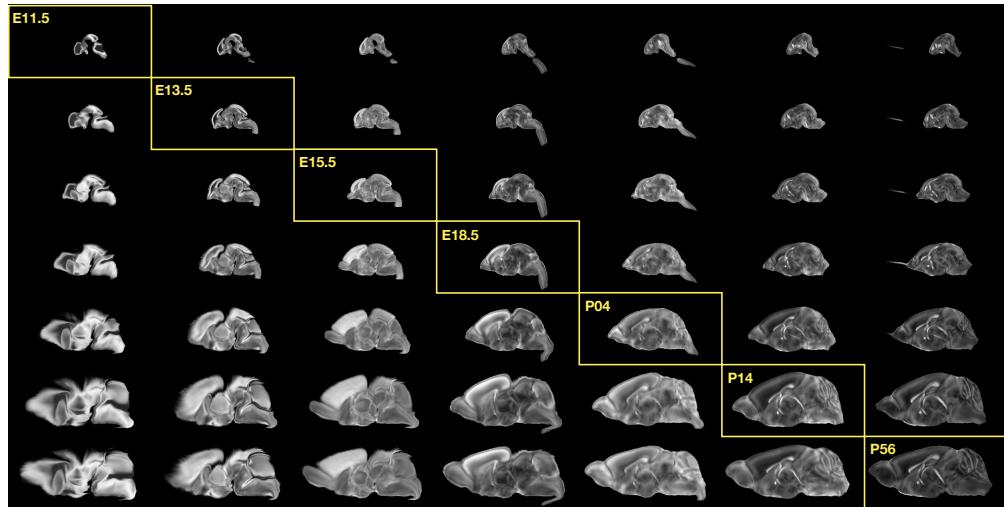


Figure 5: Mid-sagittal visualization of the effects of the transformation model in warping every developmental stage to the time point of every other developmental stage. The original images are located along the diagonal. Columns correspond to the warped original image whereas the rows represent the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

354 Once optimized, the resulting velocity field can be used to generate the deformable transform  
 355 between any two continuous points within the time interval bounded by E11.5 and P56. As  
 356 a demonstration, in Figure 5, we transform each atlas to the space of every other atlas  
 357 using the DevCCF transform model. Additionally, one can use this transformation model  
 358 to construct virtual templates in the temporal gaps of the DevCCF. Given an arbitrarily

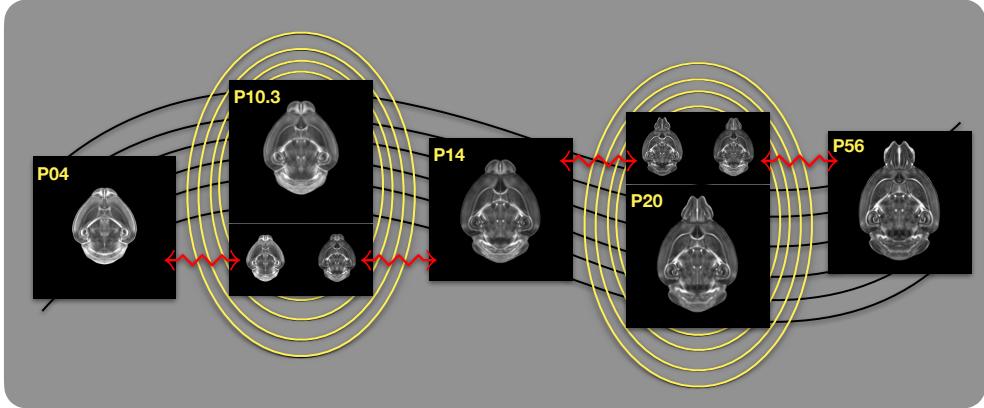


Figure 6: Illustration of the use of the velocity flow model for creating virtual templates at continuous time points not represented in one of the existing DevCCF time points. For example, FA templates at time point P10.3 and P20 can be generated by warping the existing temporally adjacent developmental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

359 chosen time point within the normalized time point interval, the existing adjacent DevCCF  
 360 atlases on either chronological side can be warped to the desired time point. A subsequent  
 361 call to one of the ANTsX template building functions then permits the construction of the  
 362 template at that time point. In Figure 6, we illustrate the use of the DevCCF velocity flow  
 363 model for generating two such virtual templates for two arbitrary time points. Note that  
 364 both of these usage examples can be found in the GitHub repository previously given.

### 365 2.3 Automated structural parcellations of the mouse brain

366 Brain parcellation strategies for the mouse brain are pivotal for understanding the complex  
 367 organization and function of murine nervous system<sup>80</sup>. By dividing the brain into distinct  
 368 regions based on anatomical, physiological, or functional characteristics, researchers can  
 369 investigate specific areas in isolation and identify their roles in various behaviors and pro-  
 370 cesses. For example, such parcellation schemes can help elucidate the spatial distribution of  
 371 gene expression patterns<sup>81</sup> as well as identify functional regions involved in specific cognitive  
 372 tasks<sup>82</sup>.

373 Although deep learning techniques have been used to develop useful parcellation tools for  
 374 human brain research (e.g., SynthSeg<sup>83</sup>, ANTsXNet<sup>46</sup>), development for the mouse brain is

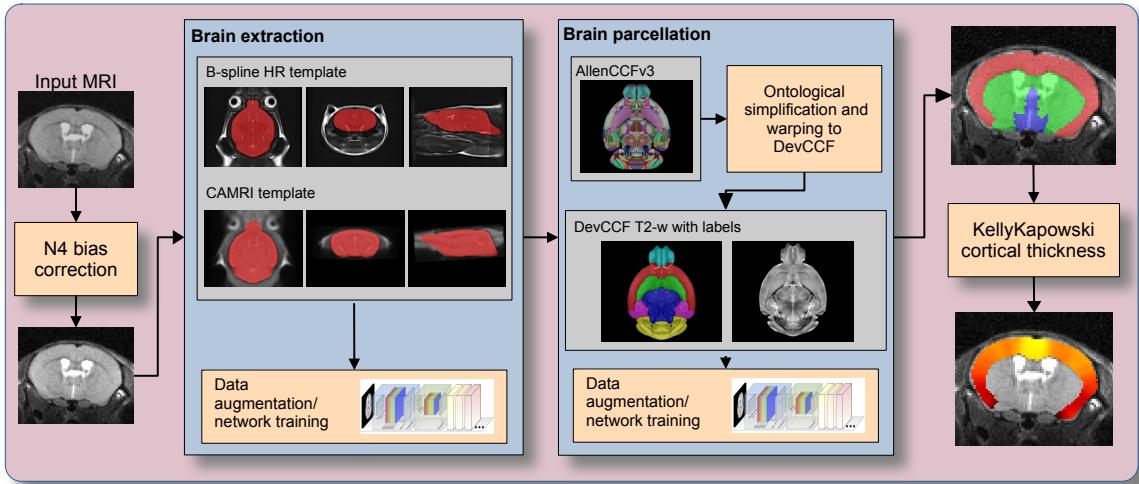


Figure 7: The mouse brain cortical parcellation pipeline integrating two deep learning components for brain extraction and brain parcellation prior to estimating cortical labels. Both deep learning networks rely heavily on aggressive data augmentation on templates built from open data and provide an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

not as extensive. In addition, mouse data is often characterized by unique imaging issues such as extreme anisotropic sampling which are often in sharp contrast to the high resolution template-based resources available within the community, e.g., AllenCCFv3 and DevCCF. We demonstrate how one can use the ANTsX tools to develop a complete mouse brain structural morphology pipeline as illustrated in Figure 7 and detailed below.

### 2.3.1 Two-shot mouse brain extraction network

In order to create a generalized mouse brain extraction network, we built whole-head templates from two publicly available datasets. The Center for Animal MRI (CAMRI) dataset<sup>68</sup> from the University of North Carolina at Chapel Hill consists of 16 T2-weighted MRI volumes of voxel resolution  $0.16 \times 0.16 \times 0.16 mm^3$ . The second high-resolution dataset<sup>69</sup> comprises 88 specimens each with three spatially aligned canonical views with in-plane resolution of  $0.08 \times 0.08 mm^2$  with a slice thickness of  $0.5 mm$ . These three orthogonal views were used to reconstruct a single high-resolution volume per subject using a B-spline fitting algorithm developed in ANTsX<sup>84</sup>. From these two datasets, two symmetric isotropic ANTsX templates<sup>58</sup>

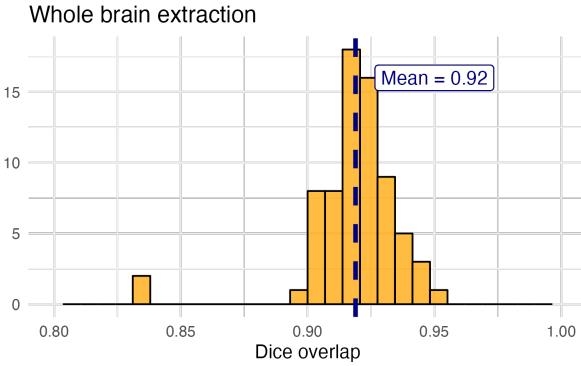
were generated analogous to the publicly available ANTsX human brain templates used in previous research<sup>85</sup>. Bias field simulation, intensity histogram warping, noise simulation, random translation and warping, and random anisotropic resampling in the three canonical directions were used for data augmentation in training a T2-weighted brain extraction network.

### 2.3.2 Single-shot mouse brain parcellation network

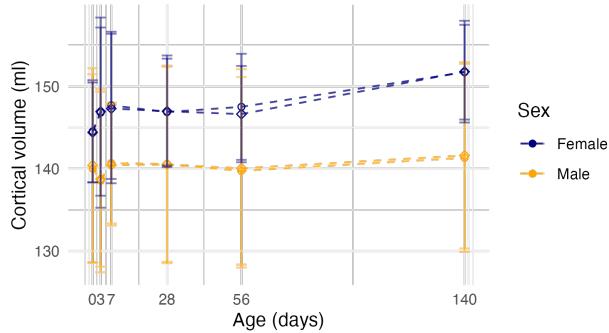
To create the network for generating a brain parcellation consistent with cortical thickness estimation, we used the AllenCCFv3 and the associated `allensdk` Python library. Using `allensdk`, a gross parcellation labeling was generated from the fine Allen CCFv3 labeling which includes the cerebral cortex,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

### 2.3.3 Evaluation

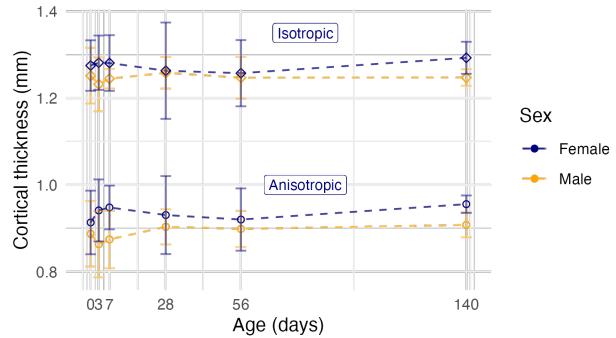
For evaluation, we used an additional publicly available dataset<sup>70</sup> which is completely independent from the data used in training the brain extraction and parcellation networks. Data includes 12 specimens each imaged at seven time points (Day 0, Day 3, Week 1, Week 4, Week 8, Week 20) with available brain masks. In-plane resolution is  $0.1 \times 0.1 mm^2$  with a slice thickness of  $0.5 mm$ . Since the training data is isotropic and data augmentation includes downsampling in the canonical directions, each of the two networks learns mouse brain-specific interpolation such that one can perform prediction on thick-sliced images, as, for example, in these evaluation data, and return isotropic probability and thickness maps (a choice available to the user). Figure 8 summarizes the results of the evaluation and comparison between isotropic and anisotropic cortical measurements in male and female specimens.



(a)



(b)



(c)

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.

414 **3 Discussion**

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

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

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

439 In contrast to the MERFISH pipeline, our fMOST pipeline was designed to be a more  
440 general solution that can be employed in other modalities. The pipeline primarily uses  
441 previously developed ANTsX preprocessing and atlasing tools to map fMOST data into the

442 AllenCCFv3. The key component of the pipeline is the use of a fMOST specific average  
443 atlas to greatly simplify the image registration problem. This average atlas, also constructed  
444 using pre-existing ANTsX tools, allows for a one-time canonical alignment from the fMOST  
445 atlas to the AllenCCFv3 to be transferred and used for mapping new fMOST images. Lastly,  
446 ANTsX provides point set transformation tools to allow the mappings found through the  
447 pipeline to be directly applied to associated single-cell reconstructions from the fMOST data  
448 to study neuronal morphology.

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

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

471 can potentially enhance our understanding of intermediate developmental stages.

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

482 The ANTsX ecosystem is a powerful framework that has demonstrated applicability to di-  
483 verse cell type data in the mouse brain. This is further evidenced by the many software  
484 packages that use various ANTsX components in their own mouse-specific workflows. In  
485 and of itself, the extensive functionality of ANTsX makes it possible to create complete pro-  
486 cessing pipelines without requiring the integration of multiple packages or lengthy software  
487 development. These open-source components not only perform well but are available across  
488 multiple platforms which facilitates the construction of tailored pipelines for individual study  
489 solutions. These components are also supported by years of development not only by the  
490 ANTsX development team but by the larger ITK community.

491 **4 Methods**

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

497 **4.1 General ANTsX utilities**

498 Although they focus on distinct data types, the three pipelines presented share common  
499 components that are generally applicable when mapping mouse cell type data. These include,  
500 addressing intensity biases and noise in the data, image registration to solve the mapping,  
501 creating custom templates and atlases from the data, and visualization of the results. Table  
502 1 provides a brief summary of key general functionalities in ANTsX for addressing these  
503 challenges.

504 **4.1.1 Preprocessing: bias field correction and denoising**

505 Bias field correction and image denoising are standard preprocessing steps in improving over-  
506 all image quality in mouse brain images. The bias field, a gradual spatial intensity variation  
507 in images, can arise from various sources such as magnetic field inhomogeneity or acquisition  
508 artifacts, leading to distortions that can compromise the quality of brain images. Correct-  
509 ing for bias fields ensures a more uniform and consistent representation of brain structures,  
510 enabling more accurate quantitative analysis. Additionally, brain images are often suscep-  
511 tible to various forms of noise, which can obscure subtle features and affect the precision  
512 of measurements. Denoising techniques help mitigate the impact of noise, enhancing the  
513 signal-to-noise ratio and improving the overall image quality. The well-known N4 bias field  
514 correction algorithm<sup>61</sup> has its origins in the ANTs toolkit which was implemented and intro-  
515 duced into the ITK toolkit, i.e. `ants.n4_bias_field_correction(...)`. Similarly, ANTsX

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>).

516 contains an implementation of a well-performing patch-based denoising technique<sup>60</sup> and is  
517 also available as an image filter to the ITK community, `ants.denoise_image(...)`.

518 **4.1.2 Image registration**

519 The ANTs registration toolkit is a complex framework permitting highly tailored solutions  
520 to pairwise image registration scenarios<sup>86</sup>. It includes innovative transformation models  
521 for biological modeling<sup>54,67</sup> and has proven capable of excellent performance<sup>55,87</sup>. Vari-  
522 ous parameter sets targeting specific applications have been packaged with the different  
523 ANTsX packages, specifically ANTs, ANTsPy, and ANTsR<sup>46</sup>. In ANTsPy, the function  
524 `ants.registration(...)` is used to register a pair of images or a pair of image sets where  
525 `type_of_transform` is a user-specified option that invokes a specific parameter set. For ex-  
526 ample `type_of_transform='antsRegistrationSyNQuick[s]'` encapsulates an oft-used pa-  
527 rameter set for quick registration whereas `type_of_transform='antsRegistrationSyN[s]'`  
528 is a more detailed alternative. Transforming images using the derived transforms is performed  
529 via the `ants.apply_transforms(...)` function.

530 Initially, linear optimization is initialized with center of (intensity) mass alignment typically  
531 followed by optimization of both rigid and affine transforms using the mutual information  
532 similarity metric. This is followed by diffeomorphic deformable alignment using symmetric  
533 normalization (SyN) with Gaussian<sup>54</sup> or B-spline regularization<sup>67</sup> where the forward trans-  
534 form is invertible and differentiable. The similarity metric employed at this latter stage is  
535 typically either neighborhood cross-correlation or mutual information. Note that these pa-  
536 rameter sets are robust to input image type (e.g., light sheet fluorescence microscopy, Nissl  
537 staining, and the various MRI modalities) and are adaptable to mouse image geometry and  
538 scaling. Further details can be found in the various documentation sources for these ANTsX  
539 packages.

540 **4.1.3 Template generation**

541 ANTsX provides functionality for constructing templates from a set (or multi-modal sets) of  
542 input images as originally described<sup>58</sup> and recently used to create the DevCCF templates<sup>16</sup>.

543 An initial template estimate is constructed from an existing subject image or a voxelwise  
544 average derived from a rigid pre-alignment of the image population. Pairwise registration  
545 between each subject and the current template estimate is performed using the Symmetric  
546 Normalization (SyN) algorithm<sup>54</sup>. The template estimate is updated by warping all subjects  
547 to the space of the template, performing a voxelwise average, and then performing a “shape  
548 update” of this latter image by warping it by the average inverse deformation, thus yielding  
549 a mean image of the population in terms of both intensity and shape. The corresponding  
550 ANTsPy function is `ants.build_template(...)`.

#### 551 4.1.4 Visualization

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

## 556 4.2 Mapping fMOST data to AllenCCFv3

### 557 4.2.1 Preprocessing

- 558 • *Downsampling*. The first challenge when mapping fMOST images into the AllenCCFv3  
559 is addressing the resolution scale of the data. Native fMOST data from an individual  
560 specimen can range in the order of terabytes, which leads to two main problems. First,  
561 volumetric registration methods (particularly those estimating local deformation) have  
562 high computational complexity and typically cannot operate on such high-resolution  
563 data under reasonable memory and runtime constraints. Second, the resolution of  
564 the AllenCCFv3 atlas is much lower than the fMOST data, thus the mapping process  
565 will cause much of the high-resolution information in the fMOST images to be lost  
566 regardless. Thus, we perform a cubic B-spline downsampling of the fMOST data to  
567 reduce the resolution of each image to match the isotropic  $25 \mu\text{m}$  voxel resolution of the  
568 AllenCCFv3 intensity atlas using `ants.resample_image(...)`. An important detail

569 to note is that while the fMOST images and atlas are downsampled, the mapping  
570 learned during the registration is assumed to be continuous. Thus, after establishing  
571 the mapping to the AllenCCFv3, we can interpolate the learned mapping and apply it  
572 directly to the high-resolution native data directly to transform any spatially aligned  
573 data (such as the single-cell neuron reconstructions) into the AllenCCFv3.

- 574 • *Stripe artifact removal.* Repetitive pattern artifacts are a common challenge in fMOST  
575 imaging where inhomogeneity during the cutting and imaging of different sections can  
576 leave stripes of hyper- and hypo-intensity across the image. These stripe artifacts  
577 can be latched onto by the registration algorithm as unintended features that are  
578 then misregistered to non-analogous structures in the AllenCCFv3. We address these  
579 artifacts by fitting a 3-D bandstop (notch) filter to target the frequency of the stripe  
580 patterns and removing them prior to the image registration.
- 581 • *Inhomogeneity correction.* Regional intensity inhomogeneity can also occur within  
582 and between sections in fMOST imaging due to staining or lighting irregularity during  
583 acquisition. Similar to stripe artifacts, intensity gradients due to inhomogeneity  
584 can be misconstrued as features during the mapping and result in matching of non-  
585 corresponding structures. Our pipeline addresses these intensity inhomogeneities using  
586 N4 bias field correction<sup>61</sup>, `ants.n4_bias_field_correction(...)`.

#### 587 4.2.2 Steps for spatial normalization to AllenCCFv3

- 588 1. *Average fMOST atlas as an intermediate target.* Due to the preparation of the mouse  
589 brain for fMOST imaging, the resulting structure in the mouse brain has several large  
590 morphological deviations from the AllenCCFv3 atlas. Most notable of these is an en-  
591 largement of the ventricles, and compression of cortical structures. In addition, there is  
592 poor intensity correspondence for the same anatomic features due to intensity dissim-  
593 ilarity between imaging modalities. We have found that standard intensity-base reg-  
594 istration is insufficient to capture the significant deformations required to map these  
595 structures correctly into the AllenCCFv3. We address this challenge in ANTsX by  
596 using explicitly corresponding parcellations of the brain, ventricles and surrounding

597 structures to directly recover these large morphological differences. However, generating  
598 these parcellations for each individual mouse brain is a labor-intensive task. Our  
599 solution is to create an average atlas whose mapping to AllenCCFv3 encapsulates these  
600 large morphological differences to serve as an intermediate registration point. This has  
601 the advantage of only needing to generate one set of corresponding annotations which  
602 is used to register between the two atlas spaces. New images are first aligned to the  
603 fMOST average atlas, which shares common intensity and morphological features and  
604 thus can be achieved through standard intensity-based registration.

605 2. *Average fMOST atlas construction.* An intensity and shape-based contralaterally sym-  
606 metric average of the fMOST image data is constructed from 30 images and their  
607 contralateral flipped versions. We ran three iterations of the atlas construction using  
608 the default settings. Additional iterations (up to six) were evaluated and showed mini-  
609 mal changes to the final atlas construction, suggesting a convergence of the algorithm.

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

- 626 4. *Alignment of individual fMOST mouse brains.* The canonical transformation between  
627 the fMOST atlas and AllenCCFv3 greatly simplifies the registration of new individ-  
628 ual fMOST mouse brains into the AllenCCFv3. Each new image is first registered  
629 into the fMOST average atlas, which shares intensity, modality, and morphologi-  
630 cal characteristics. This allows us to leverage standard, intensity-based registration  
631 functionality<sup>86</sup> available in ANTsX to perform this alignment. Transformations are  
632 then concatenated to the original fMOST image to move it into the AllenCCFv3 space  
633 using `ants.apply_transforms(...)`.
- 634 5. *Transformation of single cell neurons.* A key feature of fMOST imaging is the ability  
635 to reconstruct and examine whole-brain single neuron projections<sup>79</sup>. Spatial mapping  
636 of these neurons from individual brains into the AllenCCFv3 allows investigators to  
637 study different neuron types within the same space and characterize their morphology  
638 with respect to their transcriptomics. Mappings found between the fMOST image  
639 and the AllenCCFv3 using our pipeline can be applied in this way to fMOST neuron  
640 reconstruction point set data using `ants.apply_transforms_to_points(..)`.

## 641 4.3 Mapping MERFISH data to AllenCCFv3

### 642 4.3.1 Preprocessing

- 643 • *Initial volume reconstruction.* Alignment of MERFISH data into a 3-D atlas space  
644 requires an estimation of anatomical structure within the data. For each section,  
645 this anatomic reference image was created by aggregating the number of detected  
646 genetic markers (across all probes) within each pixel of a  $10 \times 10 \mu\text{m}^2$  grid to match  
647 the resolution of the  $10 \mu\text{m}$  AllenCCFv3 atlas. These reference image sections are then  
648 coarsely reoriented and aligned across sections using manual annotations of the most  
649 dorsal and ventral points of the midline. The procedure produces an anatomic image  
650 stack that serves as an initialization for further global mappings into the AllenCCFv3.
- 651 • *Anatomical correspondence labeling.* Mapping the MERFISH data into the AllenCCFv3  
652 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 regions: cerebellum, CTXsp, hindbrain, HPF, hypothalamus, isocortex, LSX, midbrain, OLF, PAL, sAMY, STRd, STRv, thalamus and hindbrain. A label map of each section was generated for each region by aggregating the cells assigned to that region within a  $10 \times 10\mu m^2$  grid. The same approach was used to generate more fine grained region specific landmarks (i.e., cortical layers, habenula, IC). Unlike the broad labels which cover large swaths of the section these regions are highly specific to certain parts of the section. Once cells in the MERFISH data are labeled, morphological dilation is used to provide full regional labels for alignment into the AllenCCFv3.

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

679 **4.3.2 2.5D deformable, landmark-driven alignment to AllenCCFv3**

680 After global alignment of the AllenCCFv3 into the MERFISH dataset, 2D per-section de-  
681 formable refinements are used to address local differences between the MERFISH sections  
682 and the resampled AllenCCFv3 sections. Nine registrations were performed in sequence us-  
683 ing a single label at each iteration in the following order: 1) brain mask, 2) isocortex (layer  
684 2+3), 3) isocortex (layer 5), 4) isocortex (layer 6), 5) striatum, 6) medial habenula, 7) lateral  
685 habenula, 8) thalamus, and 9) hippocampus. This ordering was determined empirically by  
686 an expert anatomist who prioritized which structure to use in each iteration by evaluat-  
687 ing the anatomical alignment from the previous iteration. Global and local mappings are  
688 then all concatenated (with appropriate inversions) to create the final mapping between the  
689 MERFISH data and AllenCCFv3. This mapping is then used to provide a point-to-point  
690 correspondence between the original MERFISH coordinate space and the AllenCCFv3 space,  
691 thus allowing mapping of individual genes and cell types located in the MERFISH data to  
692 be directly mapped into the AllenCCFv3.

693 **4.4 DevCCF velocity flow transformation model**

694 Given multiple, linearly or non-linearly ordered point sets where individual points across the  
695 sets are in one-to-one correspondence, we developed an approach for generating a velocity  
696 flow transformation model to describe a time-varying diffeomorphic mapping as a variant of  
697 the landmark matching solution. Integration of the resulting velocity field can then be used  
698 to describe the displacement between any two time points within this time-parameterized  
699 domain. Regularization of the sparse correspondence between point sets is performed using  
700 a generalized B-spline scattered data approximation technique<sup>84</sup>, also created by the ANTsX  
701 developers and contributed to ITK.

702 **4.4.1 Velocity field optimization**

703 To apply this methodology to the developmental templates<sup>16</sup>, we coalesced the manual an-  
704 notations of the developmental templates into 26 common anatomical regions (see Figure 3).

705 We then used these regions to generate invertible transformations between successive time  
706 points. Specifically each label was used to create a pair of single region images resulting in 26  
707 pairs of “source” and “target” images. The multiple image pairs were simultaneously used to  
708 iteratively estimate a diffeomorphic pairwise transform. Given the seven atlases E11.5, E13.5,  
709 E15.5, E18.5, P4, P14, and P56, this resulted in 6 sets of transforms between successive time  
710 points. Approximately  $10^6$  points were randomly sampled labelwise in the P56 template  
711 space and propagated to each successive atlas providing the point sets for constructing the  
712 velocity flow model. Approximately 125 iterations resulted in a steady convergence based  
713 on the average Euclidean norm between transformed point sets. Ten integration points were  
714 used and point sets were distributed along the temporal dimension using a log transform for  
715 a more evenly spaced sampling. For additional information a help menu is available for the  
716 ANTsPy function `ants.fit_time_varying_transform_to_point_sets(...)`.

## 717 4.5 ANTsXNet mouse brain applications

### 718 4.5.1 General notes regarding deep learning training

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

729 Data augmentation is crucial for generalizability and accuracy of the trained networks.  
730 Intensity-based data augmentation consisted of randomly added noise (i.e., Gaussian, shot,  
731 salt-and-pepper), simulated bias fields based on N4 bias field modeling, and histogram warp-

732 ing for mimicking well-known MRI intensity nonlinearities<sup>46,88</sup>. These augmentation tech-  
733 niques are available in ANTsXNet (only ANTsPyNet versions are listed with ANTsRNet  
734 versions available) and include:

- 735 • image noise: `ants.add_noise_to_image(...)`,
- 736 • simulated bias field: `antspynet.simulate_bias_field(...)`, and
- 737 • nonlinear intensity warping: `antspynet.histogram_warp_image_intensities(...)`.

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

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

#### 743 4.5.2 Brain extraction

744 Similar to human neuroimage processing, brain extraction is a crucial preprocessing step  
745 for accurate brain mapping. We developed similar functionality for T2-weighted mouse  
746 brains. This network uses a conventional U-net architecture<sup>89</sup> and, in ANTsPyNet, this  
747 functionality is available in the program `antspynet.mouse_brain_extraction(...)`.  
748 For the two-shot T2-weighted brain extraction network, two brain templates were gen-  
749 erated along with their masks. One of the templates was generated from orthogonal  
750 multi-plane, high resolution data<sup>69</sup> which were combined to synthesize isotropic volu-  
751 metric data using the B-spline fitting algorithm<sup>84</sup>. This algorithm is encapsulated in  
752 `ants.fit_bspline_object_to_scattered_data(...)` where the input is the set of voxel  
753 intensity values and each associated physical location. Since each point can be assigned  
754 a confidence weight, we use the normalized gradient value to more heavily weight edge  
755 regions. Although both template/mask pairs are available in the GitHub repository  
756 associated with this work, the synthesized volumetric B-spline T2-weighted pair is available  
757 within ANTsXNet through the calls:

- 758     • template: `antspynet.get_antsxnet_data("bsplineT2MouseTemplate")` and  
759     • mask: `antspynet.get_antsxnet_data("bsplineT2MouseTemplateBrainMask")`.

760     4.5.3 Brain parcellation

761     The T2-weighted brain parcellation network is also based on a 3-D U-net architecture and the  
762     T2-w DevCCF P56 template component with extensive data augmentation, as described pre-  
763     viously. Intensity differences between the template and any brain extracted input image are  
764     minimized through the use of the rank intensity transform (`ants.rank_intensity(...)`).  
765     Shape differences are reduced by the additional preprocessing step of warping the brain ex-  
766     tracted input image to the template. Additional input channels include the prior probability  
767     images created from the template parcellation. These images are also available through the  
768     ANTsXNet `get_antsxnet_data(...)` interface.

769 **Data availability**

770 All data and software used in this work are publicly available. The DevCCF atlas is  
771 available at <https://kimlab.io/brain-map/DevCCF/>. ANTsPy, ANTsR, ANTsPyNet, and  
772 ANTsRNet are available through GitHub at the ANTsX Ecosystem ([https://github.com/](https://github.com/ANTsX)  
773 [ANTsX](#)). Training scripts for all deep learning functionality in ANTsXNet can also be  
774 found on GitHub (<https://github.com/ntustison/ANTsXNetTraining>). A GitHub reposi-  
775 tory specifically pertaining to the AllenCCFv3 mapping is available at <https://github.com/>  
776 [dontminchenit/CCFAAlignmentToolkit](#). For the other two contributions contained in this  
777 work, the longitudinal DevCCF mapping and mouse cortical thickness pipeline, we refer the  
778 interested reader to <https://github.com/ntustison/ANTsXMouseBrainMapping>.

<sup>779</sup> **Acknowledgments**

<sup>780</sup> Support for the research reported in this work includes funding from the National Institute  
<sup>781</sup> of Biomedical Imaging and Bioengineering (R01-EB031722) and National Institute of Mental  
<sup>782</sup> Health (RF1-MH124605 and U24-MH114827).

<sup>783</sup> **Author contributions**

<sup>784</sup> N.T., M.C., and J.G. wrote the main manuscript text and figures. M.C., M.K., R.D., S.S.,  
<sup>785</sup> Q.W., L.G., J.D., C.G., and J.G. developed the Allen registration pipelines. N.T. and F.K.  
<sup>786</sup> developed the time-varying velocity transformation model for the DevCCF. N.T. and M.T.  
<sup>787</sup> developed the brain parcellation and cortical thickness methodology. All authors reviewed  
<sup>788</sup> the manuscript.

789 **References**

- 790 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).
- 791 2. La Manno, G. *et al.* Molecular architecture of the developing mouse brain. *Nature*  
**596**, 92–96 (2021).
- 792 3. Wen, L. *et al.* Single-cell technologies: From research to application. *Innovation  
(Camb)* **3**, 100342 (2022).
- 793 4. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–14  
(2014).
- 794 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).
- 795 6. Li, A. *et al.* Micro-optical sectioning tomography to obtain a high-resolution atlas of  
the mouse brain. *Science* **330**, 1404–8 (2010).
- 796 7. Ueda, H. R. *et al.* Tissue clearing and its applications in neuroscience. *Nat Rev  
Neurosci* **21**, 61–79 (2020).
- 797 8. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by  
spatial transcriptomics. *Science* **353**, 78–82 (2016).
- 798 9. Burgess, D. J. Spatial transcriptomics coming of age. *Nat Rev Genet* **20**, 317 (2019).
- 799 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).
- 800 11. Hawrylycz, M. *et al.* A guide to the BRAIN initiative cell census network data  
ecosystem. *PLoS biology* **21**, e3002133 (2023).
- 801 12. Wang, Q. *et al.* The allen mouse brain common coordinate framework: A 3D reference  
atlas. *Cell* **181**, 936–953.e20 (2020).
- 802 13. Perens, J. *et al.* An optimized mouse brain atlas for automated mapping and quantification  
of neuronal activity using iDISCO+ and light sheet fluorescence microscopy.  
*Neuroinformatics* **19**, 433–446 (2021).
- 803 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).

- 804 15. Qu, L. *et al.* Cross-modal coherent registration of whole mouse brains. *Nature Methods* **19**, 111–118 (2022).
- 805 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).
- 806 17. Chuang, N. *et al.* An MRI-based atlas and database of the developing mouse brain. *Neuroimage* **54**, 80–89 (2011).
- 807 18. Dries, R. *et al.* Advances in spatial transcriptomic data analysis. *Genome research* **31**, 1706–1718 (2021).
- 808 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).
- 809 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).
- 810 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).
- 811 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).
- 812 23. Cahill, L. S. *et al.* Preparation of fixed mouse brains for MRI. *Neuroimage* **60**, 933–939 (2012).
- 813 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).
- 814 25. Kim, Y. *et al.* Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171**, 456–469 (2017).
- 815 26. Fürth, D. *et al.* An interactive framework for whole-brain maps at cellular resolution. *Nat Neurosci* **21**, 139–149 (2018).

- 816 27. Li, Y. *et al.* mBrainAligner-web: A web server for cross-modal coherent registration  
of whole mouse brains. *Bioinformatics* **38**, 4654–4655 (2022).
- 817 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).
- 818 29. Eastwood, B. S. *et al.* Whole mouse brain reconstruction and registration to a ref-  
erence atlas with standard histochemical processing of coronal sections. *Journal of  
Comparative Neurology* **527**, 2170–2178 (2019).
- 819 30. Ni, H. *et al.* A robust image registration interface for large volume brain atlas. *Sci  
Rep* **10**, 2139 (2020).
- 820 31. Pallast, N. *et al.* Processing pipeline for atlas-based imaging data analysis of struc-  
tural and functional mouse brain MRI (AIDAmri). *Front Neuroinform* **13**, 42 (2019).
- 821 32. Celestine, M., Nadkarni, N. A., Garin, C. M., Bougacha, S. & Dhenain, M. **Sammbar-  
MRI: A library for processing SmAll-MaMmal BrAin MRI data in python.** *Front  
Neuroinform* **14**, 24 (2020).
- 822 33. Ioanas, H.-I., Marks, M., Zerbi, V., Yanik, M. F. & Rudin, M. **An optimized regis-  
tration workflow and standard geometric space for small animal brain imaging.** *Neu-  
roimage* **241**, 118386 (2021).
- 823 34. Perens, J. *et al.* Multimodal 3D mouse brain atlas framework with the skull-derived  
coordinate system. *Neuroinformatics* **21**, 269–286 (2023).
- 824 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).
- 825 36. Goubran, M. *et al.* **Multimodal image registration and connectivity analysis for inte-  
gration of connectomic data from microscopy to MRI.** *Nat Commun* **10**, 5504 (2019).
- 826 37. Chandrashekhar, V. *et al.* CloudReg: Automatic terabyte-scale cross-modal brain  
volume registration. *Nature methods* **18**, 845–846 (2021).
- 827 38. Jin, M. *et al.* **SMART: An open-source extension of WholeBrain for intact mouse  
brain registration and segmentation.** *eNeuro* **9**, (2022).

- 828 39. Negwer, M. *et al.* FriendlyClearMap: An optimized toolkit for mouse brain mapping  
and analysis. *Gigascience* **12**, (2022).
- 829 40. Lin, W. *et al.* Whole-brain mapping of histaminergic projections in mouse brain.  
*Proceedings of the National Academy of Sciences* **120**, e2216231120 (2023).
- 830 41. Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by  
MERFISH. *Nature* **598**, 137–143 (2021).
- 831 42. Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution.  
*Nature* **622**, 552–561 (2023).
- 832 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).
- 833 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).
- 834 45. Fedorov, A. *et al.* 3D slicer as an image computing platform for the quantitative  
imaging network. *Magnetic resonance imaging* **30**, 1323–1341 (2012).
- 835 46. Tustison, N. J. *et al.* The ANTsX ecosystem for quantitative biological and medical  
imaging. *Sci Rep* **11**, 9068 (2021).
- 836 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).
- 837 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).
- 838 49. Allan Johnson, G. *et al.* Whole mouse brain connectomics. *Journal of Comparative  
Neurology* **527**, 2146–2157 (2019).
- 839 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).

- 840 51. Bajcsy, R. & Broit, C. Matching of deformed images. in *Sixth International Conference on Pattern Recognition (ICPR'82)* 351–353 (1982).
- 841 52. Bajcsy, R. & Kovacic, S. [Multiresolution elastic matching](#). *Computer Vision, Graphics, and Image Processing* **46**, 1–21 (1989).
- 842 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).
- 843 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).
- 844 55. Klein, A. *et al.* [Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration](#). *Neuroimage* **46**, 786–802 (2009).
- 845 56. Murphy, K. *et al.* [Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge](#). *IEEE Trans Med Imaging* **30**, 1901–20 (2011).
- 846 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).
- 847 58. Avants, B. B. *et al.* [The optimal template effect in hippocampus studies of diseased populations](#). *Neuroimage* **49**, 2457–66 (2010).
- 848 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).
- 849 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).
- 850 61. Tustison, N. J. *et al.* [N4ITK: Improved N3 bias correction](#). *IEEE Trans Med Imaging* **29**, 1310–20 (2010).
- 851 62. Wang, H. *et al.* [Multi-atlas segmentation with joint label fusion](#). *IEEE Trans Pattern Anal Mach Intell* **35**, 611–23 (2013).

- 852 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](https://doi.org/10.1007/s12021-014-9245-2).
- 853 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).
- 854 65. McCormick, M., Liu, X., Jomier, J., Marion, C. & Ibanez, L. **ITK: Enabling reproducible research and open science**. *Front Neuroinform* **8**, 13 (2014).
- 855 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).
- 856 67. Tustison, N. J. & Avants, B. B. **Explicit B-spline regularization in diffeomorphic image registration**. *Front Neuroinform* **7**, 39 (2013).
- 857 68. Hsu, L.-M. *et al.* CAMRI mouse brain MRI data.
- 858 69. Reshetnikov, V. *et al.* High-resolution MRI data of brain C57BL/6 and BTBR mice in three different anatomical views.
- 859 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).
- 860 71. Liu, J. *et al.* **Concordance of MERFISH spatial transcriptomics with bulk and single-cell RNA sequencing**. *Life Sci Alliance* **6**, (2023).
- 861 72. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. **Cellpose: A generalist algorithm for cellular segmentation**. *Nat Methods* **18**, 100–106 (2021).
- 862 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).
- 863 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).

- 864 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).
- 865 76. Gong, H. *et al.* High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nat Commun* **7**, 12142 (2016).
- 866 77. Wang, J. *et al.* Divergent projection patterns revealed by reconstruction of individual neurons in orbitofrontal cortex. *Neurosci Bull* **37**, 461–477 (2021).
- 867 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).
- 868 79. Peng, H. *et al.* Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021).
- 869 80. Chon, U., Vanselow, D. J., Cheng, K. C. & Kim, Y. Enhanced and unified anatomical labeling for a common mouse brain atlas. *Nat Commun* **10**, 5067 (2019).
- 870 81. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335–46 (2016).
- 871 82. Bergmann, E., Gofman, X., Kavushansky, A. & Kahn, I. Individual variability in functional connectivity architecture of the mouse brain. *Commun Biol* **3**, 738 (2020).
- 872 83. Billot, B. *et al.* SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Med Image Anal* **86**, 102789 (2023).
- 873 84. 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).
- 874 85. Tustison, N. J. *et al.* Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage* **99**, 166–79 (2014).
- 875 86. Avants, B. B. *et al.* The Insight ToolKit image registration framework. *Front Neuroinform* **8**, 44 (2014).

- 876 87. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance  
in brain image registration. *Neuroimage* **54**, 2033–44 (2011).
- 877 88. 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).
- 878 89. Falk, T. *et al.* U-net: Deep learning for cell counting, detection, and morphometry.  
*Nat Methods* **16**, 67–70 (2019).