



# Cross-modal coherent registration of whole mouse brains

Lei Qu<sup>1,2,3,17</sup>, Yuanyuan Li<sup>1,17</sup>, Peng Xie<sup>2</sup>, Lijuan Liu<sup>2,4</sup>, Yimin Wang<sup>1,2,5</sup>, Jun Wu<sup>1</sup>, Yu Liu<sup>1</sup>, Tao Wang<sup>1</sup>, Longfei Li<sup>1</sup>, Kaixuan Guo<sup>1</sup>, Wan Wan<sup>1</sup>, Lei Ouyang<sup>1</sup>, Feng Xiong<sup>2</sup>, Anna C. Kolstad<sup>1,6</sup>, Zhuhao Wu<sup>6,7</sup>, Fang Xu<sup>8</sup>, Yefeng Zheng<sup>1,9</sup>, Hui Gong<sup>10,11,12</sup>, Qingming Luo<sup>1,10,11,12,13</sup>, Guoqiang Bi<sup>13,8,14</sup>, Hongwei Dong<sup>15</sup>, Michael Hawrylycz<sup>1,16</sup>, Hongkui Zeng<sup>1,16</sup> and Hanchuan Peng<sup>1,2,16</sup>

Recent whole-brain mapping projects are collecting large-scale three-dimensional images using modalities such as serial two-photon tomography, fluorescence micro-optical sectioning tomography, light-sheet fluorescence microscopy, volumetric imaging with synchronous on-the-fly scan and readout or magnetic resonance imaging. Registration of these multi-dimensional whole-brain images onto a standard atlas is essential for characterizing neuron types and constructing brain wiring diagrams. However, cross-modal image registration is challenging due to intrinsic variations of brain anatomy and artifacts resulting from different sample preparation methods and imaging modalities. We introduce a cross-modal registration method, mBrainAligner, which uses coherent landmark mapping and deep neural networks to align whole mouse brain images to the standard Allen Common Coordinate Framework atlas. We build a brain atlas for the fluorescence micro-optical sectioning tomography modality to facilitate single-cell mapping, and used our method to generate a whole-brain map of three-dimensional single-neuron morphology and neuron cell types.

Recent advances in high-resolution light microscopy<sup>1–8</sup>, tissue clearing<sup>9–15</sup>, sparse labeling techniques<sup>2,16–19</sup> and neuron tracing methods<sup>20–23</sup> now make it feasible to map the whole mammalian brain at single-cell resolution. Large international efforts, such as the BRAIN Initiative Cell Census Network (BICCN)<sup>24</sup>, MouseLight project<sup>21</sup>, Allen Mouse Brain Connectivity Atlas<sup>25</sup> and Mouse Connectome project<sup>26–28</sup>, are engaged in cell typing, mapping long-range axonal projections and microcircuit connection analyses. Massive image datasets are acquired through a variety of high-resolution, high-throughput imaging techniques, such as serial two-photon tomography (STPT)<sup>2,4</sup>, fluorescence micro-optical sectioning tomography (fMOST)<sup>1</sup>, light-sheet fluorescence microscopy (LSFM)<sup>7,8</sup> or volumetric imaging with synchronous on-the-fly-scan and readout (ViSoR)<sup>5,6</sup>.

A critical technique required to use these data is registration (alignment) of brains, their compartments, and various neuronal structures including somas, dendrites and axon arbors. Data that are acquired across subjects, modalities, time or even species must be mapped onto a canonical coordinate space. Its importance is obvious in multiple ways. First, registration of datasets

of different modalities generated in different projects provides a valuable resource for ever-expanding studies of neuron wiring and functions. Second, the registration to a common standard atlas enables digital delineation and quantification of brain anatomy and neuron morphology and other attributes such as transcriptomics. Third, registration allows quantitative comparison of features of neurons across samples under different conditions, thus enabling analysis of distinct aspects of neurobiology in a coordinated manner.

Established brain atlases such as the Allen Common Coordinate Framework atlas (CCFv3) (using STPT imaging)<sup>29,30</sup> and its derived LSFM-domain atlas<sup>31</sup> have helped neurobiologists to study gene expression patterns, brain anatomy and neural circuits. Currently, fMOST imaging is increasingly used for whole-brain single-cell analysis. High-resolution images in this modality are being generated (for example in the BICCN program) to produce large single-neuron morphology databases<sup>32</sup>. It is crucial to register such data to the Allen brain atlas to compare mapped neurons and populations of neurons generated using different experimental conditions and modalities.

<sup>1</sup>Ministry of Education Key Laboratory of Intelligent Computation & Signal Processing, Information Materials and Intelligent Sensing Laboratory of Anhui Province, School of Electronics and Information Engineering, Anhui University, Hefei, China. <sup>2</sup>SEU-ALLEN Joint Center, Institute for Brain and Intelligence, Southeast University, Nanjing, China. <sup>3</sup>Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China. <sup>4</sup>Ministry of Education Key Laboratory of Developmental Genes and Human Disease, School of Life Science and Technology, Southeast University, Nanjing, China. <sup>5</sup>School of Computer Engineering and Science, Shanghai University, Shanghai, China. <sup>6</sup>Department of Cell, Developmental & Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>7</sup>Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>8</sup>CAS Key Laboratory of Brain Connectome and Manipulation, Interdisciplinary Center for Brain Information, The Brain Cognition and Brain Disease Institute, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen, China. <sup>9</sup>Tencent Jarvis Lab, Shenzhen, Guangdong, China. <sup>10</sup>Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, MoE Key Laboratory for Biomedical Photonics, Huazhong University of Science and Technology, Wuhan, China. <sup>11</sup>HUST-Suzhou Institute for Brainsmatics, JITRI Institute for Brainsmatics, Suzhou, China. <sup>12</sup>CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Science, Shanghai, China. <sup>13</sup>School of Biomedical Engineering, Hainan University, Haikou, China. <sup>14</sup>Center for Integrative Imaging, Hefei National Laboratory for Physical Sciences at the Microscale, and School of Life Sciences, University of Science and Technology of China, Hefei, China. <sup>15</sup>Center for Integrative Connectomics, Department of Neurobiology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>16</sup>Allen Institute for Brain Science, Seattle, WA, USA. <sup>17</sup>These authors contributed equally: Lei Qu, Yuanyuan Li. e-mail: [h@braintell.org](mailto:h@braintell.org)

Despite a number of existing studies on registration of whole-brain images within one single imaging modality, including for smaller nervous systems from *Caenorhabditis elegans*<sup>33</sup>, fruitfly<sup>34</sup>, to zebrafish<sup>35,36</sup>, there has been little reported work on cross-modal registration for prevailing bioimaging methods, especially considering the increasing diversity of intensity, texture and anatomy of the structures imaged. In particular, nonuniform characteristics of brain structures makes it challenging to coregister datasets using a number of existing registration tools developed in the medical imaging field<sup>37–39</sup> or tools adapted from off-the-shelf registration toolkits<sup>40–42</sup>. Recently, deep neural network (DNN)-based methods were proposed to improve registration efficiency and accuracy further<sup>43,44</sup>. Yet, application of DNNs to cross-modal bioimaging registration remains unrealized due to the large scale of the problem and difficulties in designing effective similarity loss functions.

In this study, we develop a coherent landmark mapping (CLM) method, which integrates deep learning for reliable cross-modal whole-brain registration. With CLM as the core local registration module, we built a comprehensive registration pipeline, named mBrainAligner, to facilitate the whole-brain mapping of images, as well as their digitally reconstructed neuronal structures. Different from the ‘detection-matching-screening’ strategy that was often used in landmark-based registration methods, CLM uses a ‘detection-mapping’ strategy to systematically tackle challenges of cross-modal registration, such as problematic landmark detection across modalities, nontrivial identification and screening of unreliable matches, and inefficient utilization of shape priors of brain regions that are embedded in atlases. In addition, we also used a DNN to boost CLM further in terms of registration accuracy and robustness. We show that mBrainAligner can tackle the challenging registration case of fMOST data to the Allen brain atlas with improved accuracy compared to existing pipelines, and provides a foundation for current large-scale single-neuron analysis efforts<sup>32</sup>. We have also validated the generalization ability and the performance of mBrainAligner using other modalities including LSF, VISO and magnetic resonance imaging (MRI). We also report an fMOST-space mouse brain atlas and showcase the utility of our method for analyzing single cell types and registering partially imaged brain.

## Results

**mBrainAligner algorithm and workflow.** We designed mBrainAligner to align three-dimensional (3D) mouse brain images of different modalities to the CCFv3 (Fig. 1). Considering the need, importance and technical challenges of registering rapidly accumulating fMOST data to the CCFv3, we used fMOST image alignment as a specific example to describe our general cross-modal registration strategy. mBrainAligner also enables mapping digitally reconstructed dendritic, axonal and soma distributions to the target space (Fig. 1f), thus mBrainAligner facilitates the direct visualization, comparison and analysis of various neuronal structures at different resolutions, such as at the population, single-neuron or subcellular level.

Brain images of different modalities often vary substantially in their voxel intensity, image texture and brain anatomy (for example, due to uneven brain shrinkage) compared to the average template of the CCFv3 acquired using STPT (Fig. 1a). To address this challenge, we designed three modules in mBrainAligner: (1) an image preprocessing and robust point matching-based automatic global registration module (Fig. 1b); (2) a CLM-based automatic local registration module (Fig. 1c); and (3) an optional semiautomatic refinement module (Fig. 1d and Extended Data Fig. 1). Using mBrainAligner, we also translated the CCFv3 to a ‘normalized’ fMOST space to produce a fMOST-domain atlas (Fig. 1g).

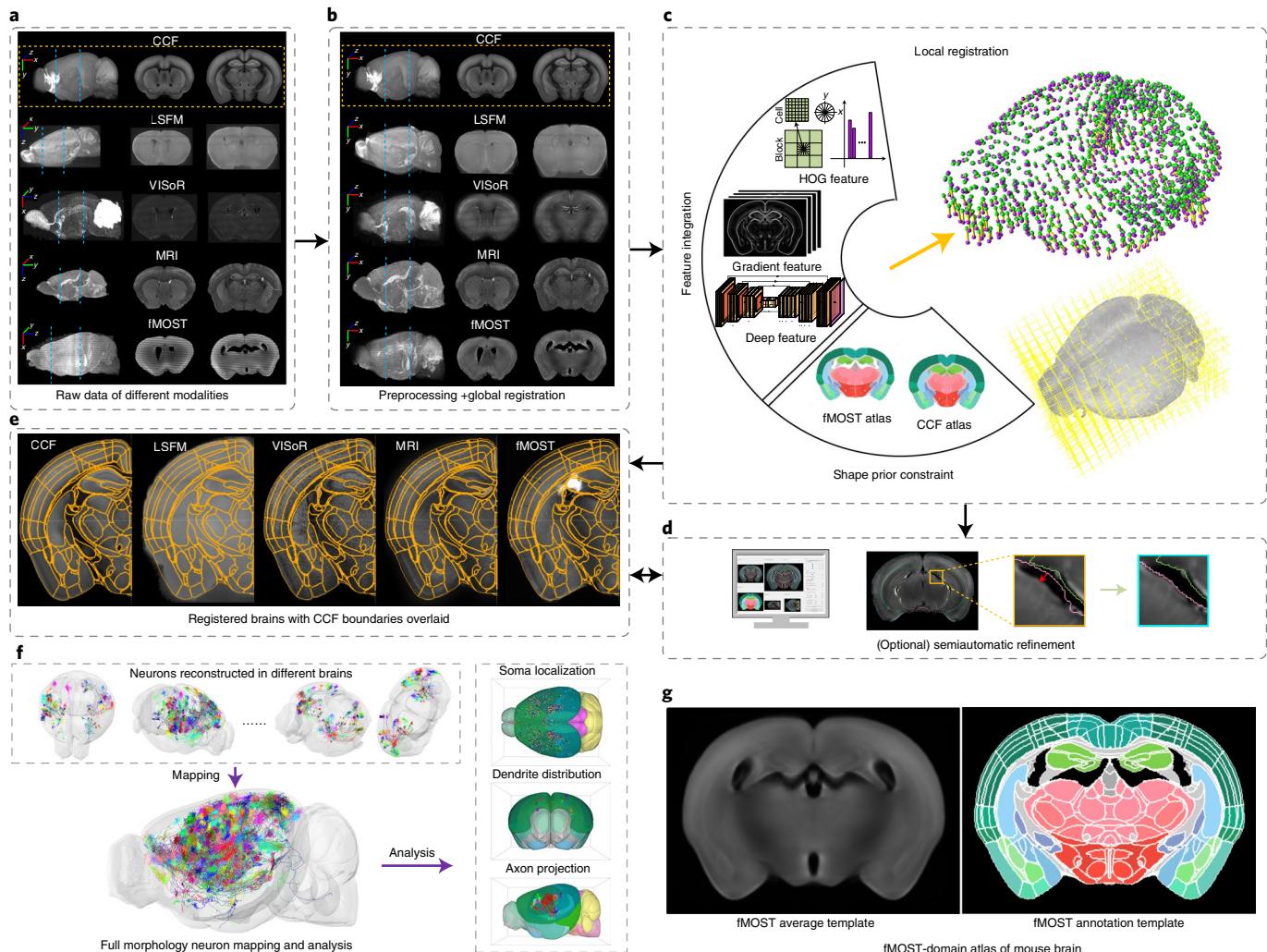
In an initial step, images are preprocessed to match in size and overall intensity profile. In the fMOST example, all images are first

downsampled anisotropically (XYZ multipliers are  $64 \times 64 \times 16$ ) roughly to match the size of the target image (CCFv3 average template, 25  $\mu\text{m}$ ), followed by contrast enhancement and noise filtering to remove stripe artifacts that were generated during the knife cutting and imaging processes (Fig. 1b and Supplementary Video 1). Next, we extract about 1,500 brain contour points in the subject and target brain, estimate an affine transform that maximizes the similarity of the brains’ outer contour, and register the global orientation and scale of these brains (Fig. 1b and Supplementary Video 2).

A key advance in mBrainAligner is the robust and accurate nonlinear local registration that warps a subject image to the target via a dense point-by-point correspondence between two images. As the cross-modal variations in intensity, texture and shape of globally registered brains often remain severe (Fig. 1b), we used a CLM algorithm to coherently deform the target landmark points obtained via two and a half dimensional (2.5D) corner detector to fit the subject brain (Fig. 1c and Supplementary Video 3), so that each landmark and its neighbors in the target image find their best matches in a coordinated manner. As landmarks before and after deformation naturally form matching pairs, CLM intrinsically avoids error-prone landmark detection and identification of unreliable matches across modalities. To ensure a robust and coherent deformation of landmarks, the deformation process is driven by three complementary features (gradient, histogram of oriented gradient and segmentation probability) that encode the edge, shape context and semantic information, respectively, while being regularized by the shape priors of brain anatomy embedded in the CCFv3 (or the fMOST atlas). We used a DNN to estimate the segmentation probability of six main brain regions, including hypothalamus (HY), caudoputamen (CP), hippocampal formation (HPF), cerebral cortex (CTX), cerebellar cortex (CBX) and brain stem (BS) (Extended Data Fig. 2). As brain delineation or segmentation is a major application of atlas-based registration, we show that the registration task can also benefit from segmentation by incorporating the semantic information as a discriminative feature (Extended Data Fig. 2c,d). Using CLM, the joint efforts of a coherent mapping strategy, discriminative feature integration and effective shape prior utilization ensure the robustness and accuracy of mBrainAligner in challenging cross-modal scenarios. The computation time for CLM-based local registration was about 6 hours per brain with 2214 target landmarks on a 2.3 GHz dual-10-core Xeon workstation with 32 G RAM.

**Registration accuracy.** We first evaluated mBrainAligner’s performance in registering the border of six brain regions of interest, that is, HY, HPF, CTX, CBX, BS (thalamus (TH), midbrain (MB) and hindbrain (HB)) and CP, which three neuroanatomists annotated independently (Fig. 2). The registration quality of mBrainAligner was visually evident for test images from different modalities including fMOST, LSF, VISO and MRI (Fig. 2a). For 31 fMOST whole mouse brain images, the region-wise median Dice score achieved by mBrainAligner’s local alignment ranges from 0.82 to 0.90, while the average median Dice score for all brain regions of interest reaches 0.88 (Fig. 2b). We compared mBrainAligner with the mouse brain registration pipelines MIRACL<sup>42</sup>, ClearMap<sup>41</sup> and aMAP<sup>40</sup>. In our tests, mBrainAligner outperformed these methods in all six regions, and achieved 0.11 higher average median Dice score than the second-ranked method (for example, 0.77 for aMAP) (Fig. 2b). We also compared the performance of mBrainAligner and alternative methods in aligning LSF, VISO and MRI modality images to the CCFv3 (Extended Data Fig. 3), where we observed consistent superiority of mBrainAligner.

The anatomical location of somas is often used as a key feature to assign neuron types<sup>45</sup>. We considered 1,708 neurons that were reconstructed from 31 fMOST mouse brains and that had all somas manually annotated (Fig. 3a). These somas distribute mainly in the cortex and thalamus. Five brain regions, ventral posteromedial



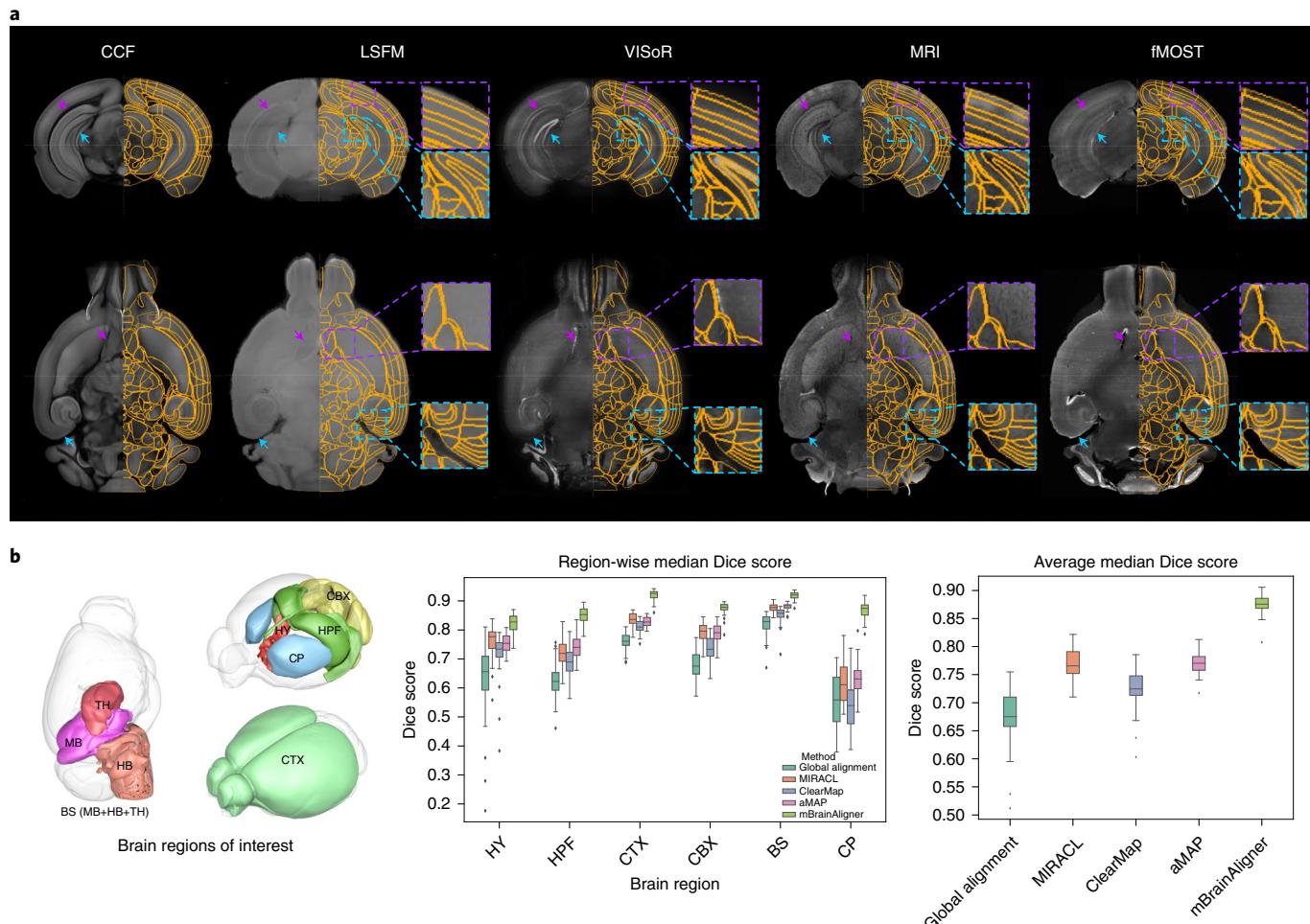
**Fig. 1 | Overview of mBrainAligner.** **a**, Left: maximum intensity projection of original brain images acquired with different modalities, center and right: coronal slices. **b**, Preprocessed and globally aligned brains. For fMOST images, the stripe artifacts were removed before applying global registration. **c**, Overview of CLM-based local registration. HOG, histogram of oriented gradients. **d**, Overview of the optional semiautomatic registration module. **e**, Registered brains of different modalities, overlaid with brain region boundaries defined in the CCFv3. **f**, Illustration of the mapping of dendritic, axonal and soma distributions into a canonical coordinate space for visualization, comparison and analysis. **g**, fMOST-domain atlas of the mouse brain built with mBrainAligner.

nucleus (VPM), CP, ventral posterolateral nucleus (VPL), dorsal part of the lateral geniculate complex (LGd) and medial geniculate complex (MG), contain more than 40 somas. We computationally identified the anatomical locations of these somas by mapping them to the CCFv3. Compared with the manually annotated results, mBrainAligner achieves an average soma mapping accuracy of 76% (Fig. 3b), which is at least 19% higher than other methods. A similar conclusion can be drawn by taking a closer look at the results of individual brain regions (Fig. 3c). For those somas whose locations mismatch between automatic registration and manual annotation, we calculated their shortest distance to the manually assigned region as distance error, and found mBrainAligner achieved an average distance error of  $47.88\text{ }\mu\text{m}$  for all somas which is half the value achieved with the second-ranked method (Fig. 3d). Visualization of registered somas in the VPM (in thalamus) and primary somatosensory area nose (SSp-n) (in cortex) shows that mBrainAligner produces results more consistent with annotators' expectation (Fig. 3e). We analyzed the distribution of wrongly localized somas (Fig. 3f and Extended Data Fig. 4) and observed that most errors occurred in localizing VPL as VPM, which are two adjacent small

regions without a clear discriminative boundary between them in fMOST imaging (Fig. 3g).

Different thalamic regions were hypothesized to have unique cortical projection patterns<sup>46</sup>. We examined this by computationally predicting the axon projections of 63 neurons from six different brains with somas in VPM or VPL. We found their axons terminated predominantly in layers 4 and 6 of the cortex and formed clear patterns (Fig. 3h). Compared with manual annotation by anatomists, mBrainAligner predicted 89.1% of the projection target regions when we used automatic local registration (Fig. 3i and Extended Data Fig. 5). The accuracy was further improved to 96.29% after we applied semiautomatic registration (Extended Data Fig. 1e). For 162 neurons with apical dendrites that distributed in the L1 layer of the cortex (Fig. 3j), the prediction accuracy is more than 90%.

**fMOST-domain atlas.** The Allen CCFv3 is an increasingly used standard in mouse brain research. This atlas was developed using iteratively averaged STPT TissueCytex imaging. In a similar manner we used mBrainAligner to build a fMOST-domain atlas of mouse brain (Fig. 4a,b and Extended Data Fig. 6), including an average



**Fig. 2 | Accuracy of mBrainAligner in brain delineation.** **a**, Registered brains of different modalities with brain region boundaries (orange) defined in the CCFv3 overlaid. The first and second rows show coronal and sagittal views of registered brains, respectively, with enlarged view of two subregions. The purple and cyan arrows in the first and second rows point to the same spatial coordinates. **b**, Quantitative evaluation of the registration accuracy of different methods in terms of region-wise and average median Dice score on 31 fMOST mouse brains. Left: six brain regions of interest that were manually annotated and used for quantifying the registration accuracy. Middle: region-wise median Dice score. Right: average median Dice score of different methods. Box plot: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers.

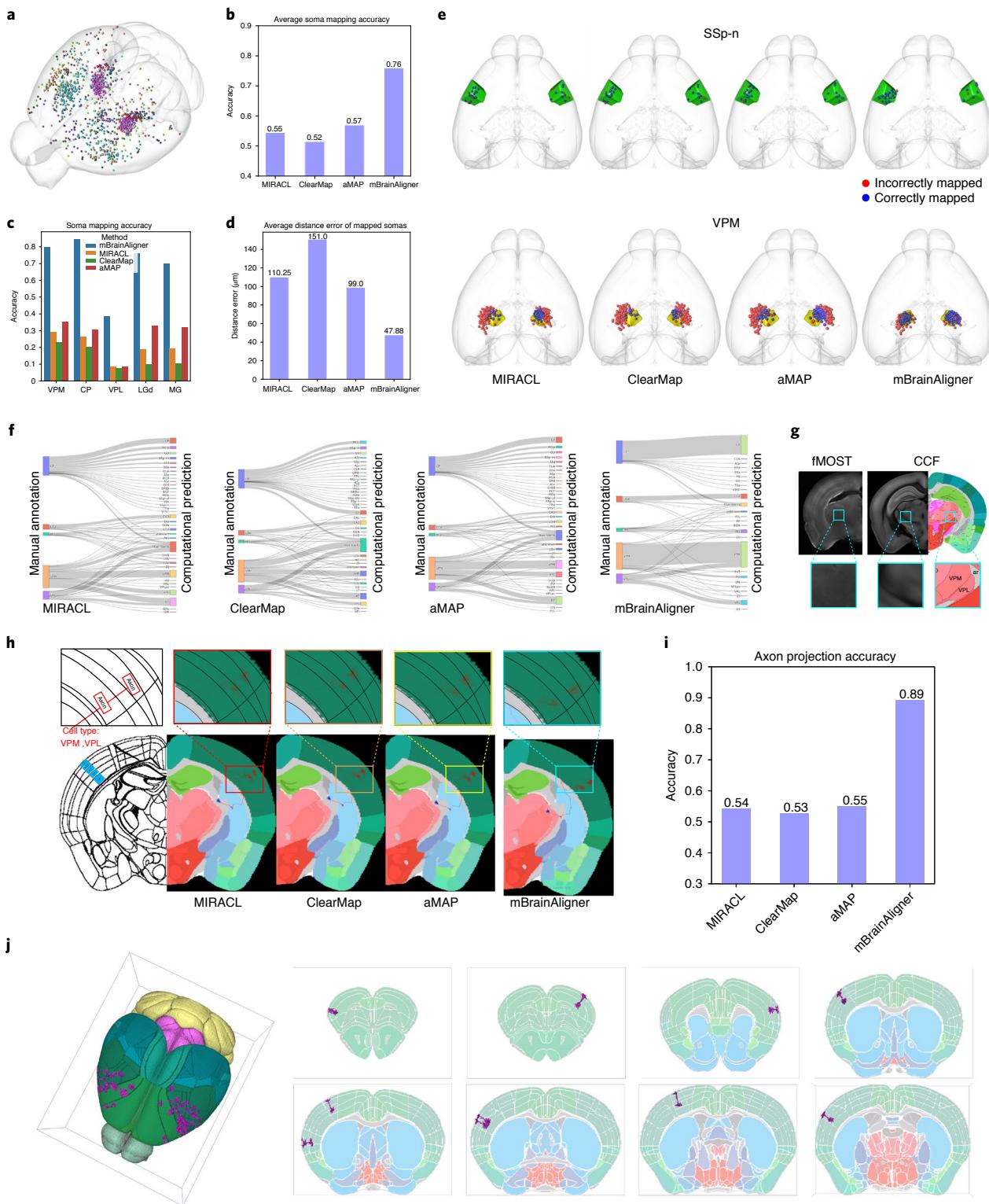
template and an annotation template (Fig. 4b). The fMOST atlas offers both biological and engineering advantages. First, we identified the morphology variation between two modalities in a quantitative way (Fig. 4c), which helps estimate the deformation caused by sample preparation (for example, different clearing methods) in various methods. Second, we established a reusable and invariant transformation between two atlases (Fig. 4b), thus brain annotation labels and reconstructed results provided in each modality can be effectively shared. Third, we reduced the technical difficulty to map images directly across these two modalities (Fig. 4d).

The robust and accurate registration of mBrainAligner makes it possible to generate a high-quality average template using a small number of samples, that is 36 fMOST whole mouse brains in our case. This number is less than that used in building the CCFv3 (1675 STPT whole mouse brains) and the LSFV3 atlas (139 LSFV3 imaged brains). We built the fMOST annotation template by backward-mapping the annotation of the CCFv3 using the same warping parameters for image registration (Fig. 4b). In this way, we ensured consistent definition of brain anatomical regions in the two atlases, and facilitated the sharing of information across modalities. We also asked independent neuroanatomists to validate further the biological meaningfulness of the fMOST atlas, especially in the

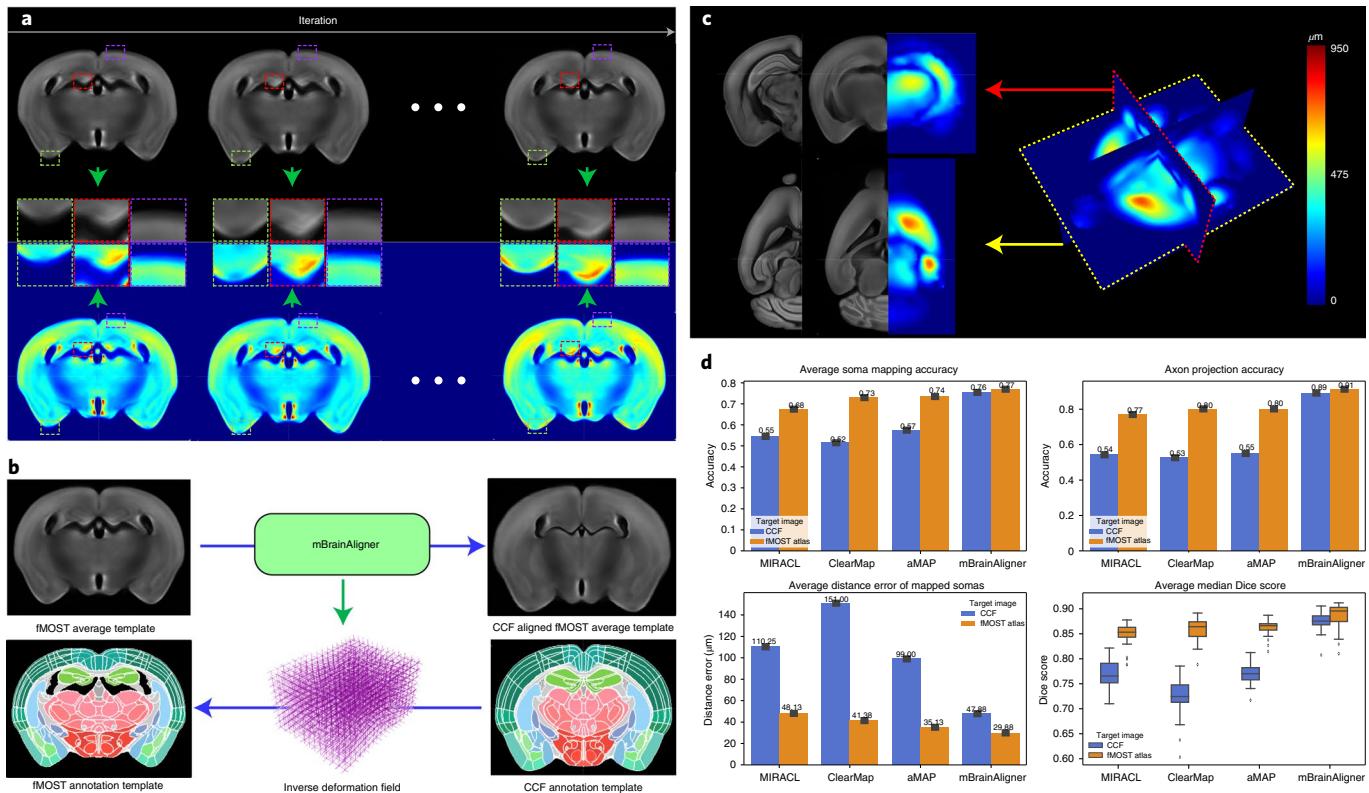
motor cortex, a region of recent focus in the BRAIN Initiative's consortium effort<sup>47</sup>.

The fMOST-domain atlas also affords improved registration accuracy. Using the same metrics described above, we re-evaluated the performance of mBrainAligner and alternative methods to register fMOST images to the fMOST atlas. Because the most notable differences of cross-modal anatomy and intensity are no longer present, all metrics, including the Dice score of brain delineation, soma localization accuracy, distance errors of mapped somas and axon projection prediction, show improvements especially for MIRACL, ClearMap and aMAP (Fig. 4d). mBrainAligner still outperforms the other methods in all metrics, and also outperforms its results achieved with the original CCFv3. With the fMOST atlas, mBrainAligner further reduces the average distance error of mapped somas from 47.88 μm to 29.88 μm, and improves the average median Dice score of brain regions to around 0.90. These results not only verified the quality and utility of the fMOST atlas, but also demonstrated the effectiveness of mBrainAligner in both intra and cross-modal registration.

**Cell type analysis enabled by mBrainAligner.** We used mBrainAligner in cell type analysis and also studied neuronal projection



**Fig. 3 | Accuracy of mBrainAligner in computational soma localization and axon projection prediction.** **a**, Distribution of 1708 somas reconstructed from 31 fMOST brains. Different colors represent different brain regions after manual annotation. **b**, Soma mapping accuracy of different methods (MIRACL, ClearMap, aMAP and mBrainAligner). **c**, Comparison of soma mapping accuracy in different methods in five brain regions (VPM, CP, VPL, LGd and MG). **d**, Average distance error of mapped somas in different methods. **e**, Visualization of mapped somas in two brain regions (SSp-n and VPM). The somas that were mapped outside of their manually annotated region are shown as red dots. Blue dots indicate the correctly mapped ones. **f**, Distribution of soma localization results generated by different methods. Left side of each diagram shows manually annotated results, and the right side shows their computationally predicted results. **g**, Enlarged view of two adjacent small regions (VPM and VPL) where most localization errors occurred. **h**, Visualization of axon mapping results of different methods. **i**, Prediction accuracy of axon projections by different methods on 63 neurons with somas in VPM and VPL. **j**, Registered results of 162 neurons with apical dendrites distributed in the L1 layer of the cortex. The left-most subpanel visualizes the distribution of neurons in the cortex, and other subpanels show the coronal slices with registered apical dendrites overlaid.



**Fig. 4 | fMOST-domain atlas of mouse brain.** **a**, Evolution of fMOST average template in the building process. Left to right shows results (coronal plane) generated in the first, second and seventh iteration, respectively. The top row and bottom row show the gray-scale and color-code visualized atlas, and the close-up views in the middle show the sharpness of the average template during iteration. **b**, Generation of fMOST annotation template. **c**, Visualization of anatomical variation between fMOST and the CCFv3. **d**, Re-evaluation of registration accuracy of different methods by registering 31 fMOST brains to the fMOST atlas. Box plot: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers.

pathways at single-neuron resolution. The topography of thalamic neurons was characterized previously at the neuron population level<sup>48</sup>. Here we used mBrainAligner to map neurons reconstructed via our BICCN consortium effort<sup>32</sup> to the CCFv3. We identified 1268 cortical axonal arbor clusters that projected to the major cortical areas (for example, primary somatosensory area barrel field/mouth/nose (SSp-bfd/m/n) for VPM neurons) and studied their correlation with soma locations (Fig. 5a and Extended Data Fig. 7). Following the convention of the CCFv3, larger values of  $x$ ,  $y$  and  $z$  dimensions represent more posterior, ventral and medial locations in the 3D brain atlas.

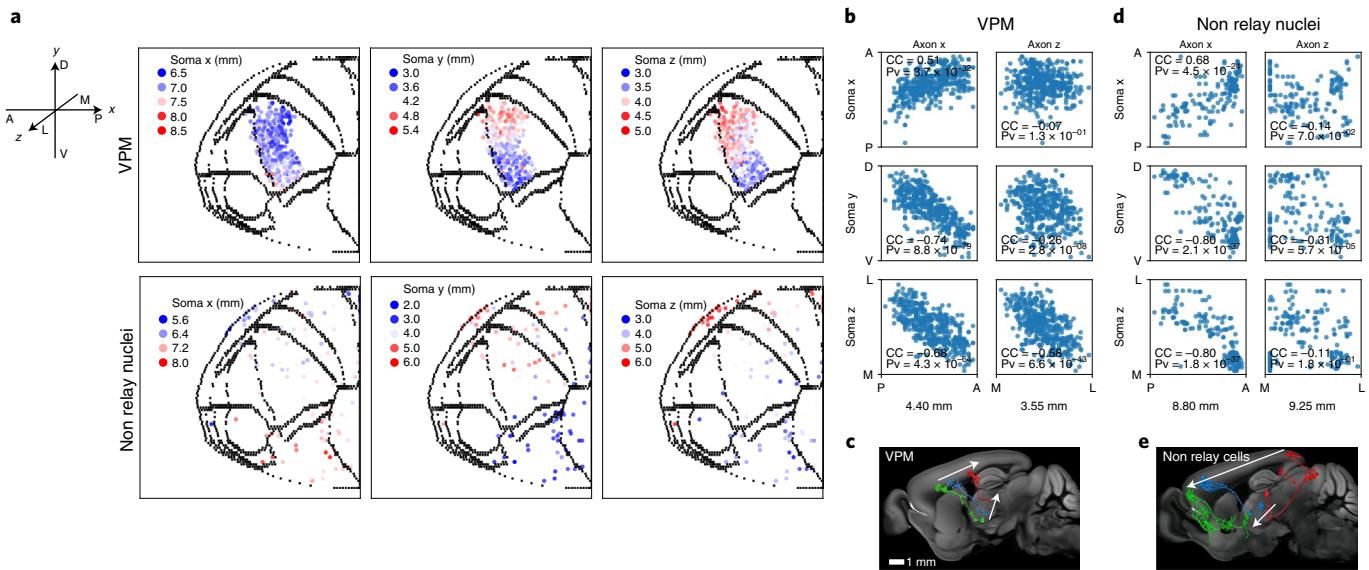
As an illustrative analysis, we created a two-dimensional (2D) top view of the Common Coordinate Framework (CCF) cortical structures by virtual flattening and quantified the axonal arbor locations within the 2D plane. We found that axonal arbors of VPM neurons exhibited a striped distribution, with the stripe approximately aligning with the  $x$  axis of the 2D map. For VPM neurons, the  $x$  coordinates (in the 2D cortex) of the axon centers are positively correlated with the  $x$  coordinates (in 3D CCF space) and anticorrelated with  $y$  and  $z$  coordinates of somata (Fig. 5b). The dominant pattern ( $P=6.5\times 10^{-7}$ , two-sided Fisher's exact test) of VPM topography is that neurons originating in the dorsal VPM project to the posterior part of the primary somatosensory cortex, namely the barrel field (SSp-bfd). This pathway is shorter than the pathway between ventral-VPM and SSp-m (Fig. 5c). The  $z$  coordinate of the axon is anticorrelated with the  $y$  and  $z$  coordinates of soma (Fig. 5b). We observed the same relationship between axon and soma locations of VPL neurons, except that their axon  $z$  axis is anticorrelated with soma  $x$  axis ( $P=0.025$ , two-sided Fisher's exact test, Extended Data Fig. 7). For intergeniculate leaflet of the lateral geniculate complex (IGL), axonal arbors follow an anterior–posterior distribution pattern (Fig. 5e).

(LGd) neurons, the  $x$  and  $z$  coordinates of the axon centers are anticorrelated with the  $x$  and  $y$  coordinates of somata (Extended Data Fig. 7). Among the four sensory relay nuclei, MG neurons show the lowest level of topography where only the  $x$  axes of soma and axon were correlated ( $P=0.015$ , two-sided Fisher's exact test, Extended Data Fig. 7). For the nonrelay thalamic nuclei, especially high-order nuclei including ventral medial nucleus of the thalamus (VM) and nucleus of reunions (RE), we integrated neurons with different nuclei origins and performed an overall topography test (Fig. 5a,d). The overall relationship of soma and axon locations is highly consistent with that of VPM and VPL, especially the anterior–posterior arrangement of axonal arbors. From anterior–ventral–medial nuclei, such as parataenial nucleus (PT) and rhomboid nucleus (RH), to posterior–dorsal–lateral nuclei, such as intergeniculate leaflet of the lateral geniculate complex (IGL), axonal arbors follow an anterior–posterior distribution pattern (Fig. 5e).

## Discussion

Motivated in part by the ongoing large-scale collaboration of the BRAIN Initiative's BICCN program to characterize brain cell types across multiple modalities, we developed mBrainAligner, a registration pipeline to tackle challenges of cross-modal registration. A whole-brain map of full neuronal morphology in three dimensions generated in this way helped test a number of hypotheses about neuronal projection diversity at multiple organizational levels<sup>32</sup>.

With mBrainAligner, we generated a mouse brain atlas in the fMOST domain. As a large amount of fMOST images are being generated, and ‘it is essential that brain atlases keep pace with large-scale data collection’<sup>30</sup>, an atlas in the fMOST domain is desirable. However, the atlas building process is traditionally computationally



**Fig. 5 | Topography of thalamic neuronal projections by integrating Common Coordinate Framework-registered brains.** **a**, Location of axonal arbors of cortical projection neurons grouped by thalamic nuclei. Dots represent the center of axonal arbors in the flattened two-dimensional cortical map. Locations of corresponding soma are color-coded in three separate columns: (left to right) the anterior-posterior (soma x), dorsal-ventral (soma y) and lateral-medial (soma z) axes. **b**, Left panels: scatter plots showing correlation between soma and axon arbor locations of VPM neurons. Soma axes are defined in the same way as **a**. Axon x and axon z represent anterior-posterior and lateral-ventral axes, separately. Right panel shows soma distribution of three types of VPM axon arbors. **c**, Examples of single neuron projections of VPM neurons. Colors represent individual neurons. **d**, The same layout as **b** but for neurons of nonrelay thalamic nuclei. **e**, Examples of single neuron projections of nonrelay neurons. Colors represent individual neurons.

demanding and labor intensive. mBrainAligner enabled us to build a high-quality ‘standard’ average fMOST-based template from a small sample set without manual intervention. By precisely registering the fMOST average template to the CCFv3 with mBrainAligner and then back-projecting the annotation of the CCFv3 onto fMOST, a smooth transformation of annotations and identical parcellation of brain regions between two atlases are guaranteed. This atlas building pipeline is generic and can be applied to other modalities.

Finally, for cross-modal brain registration, how to quantify the registration accuracy is often debated<sup>49</sup>. As soma locations, axon projection and neuronal morphology were used as defining features in the task of cell typing<sup>50</sup>, in a biologically meaningful way, we show that mBrainAligner produces improved results compared to alternative approaches. In addition to mBrainAligner’s use cases for intra and cross-modal registration of whole mouse brain, this tool can also be used to register partially imaged brains. We have open-sourced mBrainAligner, as well as some useful tools including a 2.5D corner detector, stripe artifacts removal, and image or metadata warping tools, to promote interdisciplinary and reproducible researches.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-021-01334-w>.

Received: 12 March 2021; Accepted: 28 October 2021;

Published online: 09 December 2021

## References

- Gong, H. et al. High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nat. Commun.* **7**, 1–12 (2016).
- Economou, M. N. et al. A platform for brain-wide imaging and reconstruction of individual neurons. *eLife* **5**, e10566 (2016).
- Osten, P. & Margrie, T. W. Mapping brain circuitry with a light microscope. *Nat. Methods* **10**, 515–523 (2013).
- Ragan, T. et al. Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat. Methods* **9**, 255–258 (2012).
- Wang, H. et al. Scalable volumetric imaging for ultrahigh-speed brain mapping at synaptic resolution. *Nat. Sci. Rev.* **6**, 982–992 (2019).
- Xu, F. et al. High-throughput mapping of a whole rhesus monkey brain at micrometer resolution. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-00986-5> (2021).
- Niedworok, C. J. et al. Charting monosynaptic connectivity maps by two-color light-sheet fluorescence microscopy. *Cell Rep.* **2**, 1375–1386 (2012).
- Dodt, H.-U. et al. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nat. Methods* **4**, 331–336 (2007).
- Chung, K., Wallace, J., Kim, S. Y., Kalyanasundaram, S. & Deisseroth, K. Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013).
- Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nat. Protoc.* **9**, 1682–1697 (2014).
- Murray, E. et al. Simple, scalable proteomic imaging for high-dimensional profiling of intact systems. *Cell* **163**, 1500–1514 (2015).
- Renier, N. et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).
- Pan, C. et al. Shrinkage-mediated imaging of entire organs and organisms using uDISCO. *Nat. Methods* **13**, 859–867 (2016).
- Susaki, E. A. et al. Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. *Nat. Protoc.* **10**, 1709–1727 (2015).
- Richardson, D. S. & Lichtman, J. W. Clarifying tissue clearing. *Cell* **162**, 246–257 (2015).
- Thomas, R., Smallwood, P. M., John, W., Jeremy, N. & Eshel, B. J. Genetically-directed, cell type-specific sparse labeling for the analysis of neuronal morphology. *PLoS One* **3**, e4099 (2008).
- Aransay, A., Rodríguez-López, C., García-Amado, M., Clascá, F. & Prensa, L. Long-range projection neurons of the mouse ventral tegmental area: a single-cell axon tracing analysis. *Front. Neuroanat.* **9**, 59 (2015).
- Ghosh, S. et al. Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* **472**, 217–220 (2011).
- Lin, R. et al. Cell-type-specific and projection-specific brain-wide reconstruction of single neurons. *Nat. Methods* **15**, 1033–1036 (2018).

20. Wang, Y. et al. TeraVR empowers precise reconstruction of complete 3-D neuronal morphology in the whole brain. *Nat. Commun.* **10**, 3474 (2019).
21. Winnubst, J. et al. Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell* **179**, 268–281.e213 (2019).
22. Zhou, H. et al. GTTree: an Open-source tool for dense reconstruction of brain-wide neuronal population. *Neuroinformatics* **19**, 305–317 (2020).
23. Peng, H. et al. BigNeuron: large-scale 3D neuron reconstruction from optical microscopy images. *Neuron* **87**, 252–256 (2015).
24. Ecker, J. R. et al. The BRAIN initiative cell census consortium: Lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96**, 542–557 (2017).
25. Oh, S. W. et al. A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
26. Zingg, B., Hintiryan, H., Gou, L., Song, M. Y. & Dong, H.-W. Neural Networks of the Mouse Neocortex. *Cell* **156**, 1096–1111 (2014).
27. Hintiryan, H. et al. The mouse cortico-striatal projectome. *Nat. Neurosci.* **19**, 1100–1114 (2016).
28. Bienkowski, M. S. et al. Integration of gene expression and brain-wide connectivity reveals the multiscale organization of mouse hippocampal networks. *Nat. Neurosci.* **21**, 1628–1643 (2018).
29. Dong, H. W. *The Allen Reference Atlas: A Digital Color Brain Atlas of the C57BL/6J Male Mouse* (John Wiley & Sons Inc., 2008).
30. Wang, Q. et al. The Allen mouse brain common coordinate framework: a 3D reference atlas. *Cell* **181**, 936–953.e920 (2020).
31. 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 (2020).
32. Peng, H. et al. Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021).
33. Long, F., Peng, H., Liu, X., Kim, S. K. & Myers, E. A 3D digital atlas of *C. elegans* and its application to single-cell analyses. *Nat. Methods* **6**, 667–672 (2009).
34. Peng, H. et al. BrainAligner: 3D registration atlases of *Drosophila* brains. *Nat. Methods* **8**, 493–500 (2011).
35. Randlett, O. et al. Whole-brain activity mapping onto a zebrafish brain atlas. *Nat. Methods* **12**, 1039–1046 (2015).
36. Griffiths, V. A. et al. Real-time 3D movement correction for two-photon imaging in behaving animals. *Nat. Methods* **17**, 741–748 (2020).
37. Avants, B. B., Tustison, N. & Song, G. Advanced normalization tools (ANTS). *Insight J.* **2**, 1–35 (2009).
38. 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 (2009).
39. Modat, M. et al. Fast free-form deformation using graphics processing units. *Computer Methods Prog. Biomedicine* **98**, 278–284 (2010).
40. Niedworok, C. J. et al. aMAP is a validated pipeline for registration and segmentation of high-resolution mouse brain data. *Nat. Commun.* **7**, 1–9 (2016).
41. Renier, N. et al. Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* **165**, 1789–1802 (2016).
42. Goubran, M. et al. Multimodal image registration and connectivity analysis for integration of connectomic data from microscopy to MRI. *Nat. Commun.* **10**, 1–17 (2019).
43. Fan, J., Cao, X., Yap, P.-T. & Shen, D. BIRNet: brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* **54**, 193–206 (2019).
44. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q. & Shen, D. Deformable image registration based on similarity-steered CNN regression. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention* (eds. Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S.) 300–308 (Springer, Cham, 2017).
45. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
46. Harris, J. A. et al. Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
47. Muñoz-Castañeda, R. et al. Cellular anatomy of the mouse primary motor cortex. *Nature* **598**, 159–166 (2021).
48. Jones, E. G. *The Thalamus* (Springer Science & Business Media, 2012).
49. Tward, D. J. et al. Solving the where problem in neuroanatomy: a generative framework with learned mappings to register multimodal, incomplete data into a reference brain. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.03.22.002618> (2020).
50. Ramon y Cajal, S. *Histoologie du système nerveux de l'homme et des vertébrés*. Maloine, Paris **2**, 153–173 (1911).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**Animal care and use.** All experiments related to the use of mice followed National Institutes of Health guidelines, and received approval from the Institutional Animal Care and Use Committee of the Allen Institute for Brain Science, the University of Science and Technology of China, the Icahn School of Medicine at Mount Sinai and the Center for Genomic Regulation, respectively. All mice used in this study were group-housed with a 12-hour light/dark cycle, controlled temperature (20–22 °C) and humidity (50–70%), and free access to food and water.

**Data preparation.** The average and annotation template of the CCFv3 were downloaded from Allen Institute web portal (<http://atlas.brain-map.org/>). The detailed sample preparation, imaging and atlas generation methods have been described previously<sup>30</sup>. As the foreground brain area in the CCFv3 average template is close to image boundaries, to eliminate computational ambiguity, we padded the CCFv3 average template (25 μm) by 20 pixels at both ends of the anterior–posterior axes and used the padded image as the target.

Our fMOST dataset consisted of 57 brains of male and female transgenic mice (P56), with sparsely labeled cells in cortical, thalamic, claustral and striatal regions, and for cholinergic, noradrenergic and serotonergic neuronal types. The detailed sample preparation and imaging methods were described previously<sup>32</sup>. The auto-fluorescence channel was used for image registration. We used the TeraConverter module of Vaa3D<sup>51</sup> to perform image downampling. The neurons and somas of 31 fMOST brains were reconstructed via the TeraVR module<sup>20</sup> of Vaa3D.

The MRI dataset was downloaded from <https://www.nitrc.org/projects/infchfwmouse>, which contains image volumes representing the canonical Waxholm Space adult C57BL/6J mouse brain. This dataset provides five images, including T1, T2\* and T2-weighted magnetic resonance volumes, Nissl-stained optical histology and a volume of labels. All volumes are represented at 21.5 μm isotropic resolution. The labeled volume represents phase one delineations of the Waxholm Space mouse brain. The atlas contains 26 structures (+ inner ear) in accordance with the NeuroLex Brain Partonomy at: <https://scirunch.org/scirunch/interlex/dashboard>. We chose the T2\*-weighted MRI image dataset as test data, and removed brain tissues that did not match the CCFv3 based on the provided labels.

The VISO-R dataset was generated from an 8-week-old C57BL/6J mouse. The brain sample was sectioned into 300 μm slices, cleared and then stained with 4,6-diamidino-2-phenylindole to label the nuclei of all cells. All slices were imaged with a custom VISO-R microscope for high-throughput volumetric imaging at 1 × 1 × 2.5 μm<sup>3</sup> voxel resolution, and stitched together to reconstruct the whole mouse brain volume. The dataset was downsampled to 4 × 4 × 4 μm<sup>3</sup> resolution before registration.

The LSFM whole-brain dataset was generated from an 8-week-old adult male mouse brain. Perfused and para-formaldehyde-fixed brain was delipidated following a modified Adipo-Clear protocol<sup>32</sup>. The brain sample was washed with a methanol/B1n gradient (20%, 40%, 60%, 80%) for 1 h each step, then 100% methanol for 1 h × 2, then di-chloride-methane 1 h × 2 and overnight. Then the sample was washed with 100% methanol for 1 h, 2 h × 2, then bleached in 30% hydrogen peroxide/methanol (1:5) at 4°C overnight. After bleaching the sample was rehydrated back in reverse methanol/B1n gradient (80%, 60%, 40%, 20%) for 1 h each step, then in B1n for 1 h, then washed in 5% dimethylsulfoxide/0.3 M glycine/phosphate buffer solution/Triton X-100/Tween-20/Heparin (PTxwHPTxwH) for 1 h and 3 h, then washed and stored in PTxwH before clearing. The clearing was done following the standard iDISCO+ procedure<sup>41</sup>. The delipidated sample was washed in methanol/water gradient (20%, 40%, 60%, 80%) for 1 h each step, then in 100% methanol for 1 h × 3, then in 100% di-chloride-methane for 1 h, 2 h, and overnight, then cleared and stored in dibenzylether overnight. The dibenzylether-cleared sample was imaged horizontally with the LaVision LSFM at Imaging Core of Icahn School of Medicine at Mount Sinai, using 488 nm laser for tissue auto-fluorescence with the 1.3×/0.1 detection lens and Zoom v1.0, light sheet with NA 0.16. The X/Y sampling rate was 5 μm and Z step at 6 μm.

The LSFM partial brain hippocampus dataset was generated from a 12-week-old Thy1-yellow fluorescent protein mouse (strain B6.Cg-Tg(Thy1-YFP)2Jrs/J no. 003782; The Jackson Laboratories). The mice were killed and transcardially perfused with cold phosphate buffered saline (PBS), followed by cold CLARITY hydrogel solution. The brains were removed from the skull, immersed in 20 ml of hydrogel solution and stored at 4°C shielded from light. After 2–3 days at 4°C, hydrogel polymerization was induced by 3 h incubation at 37°C. Contact with oxygen was prevented with a vegetable oil layer. Thereafter, the already polymerized surrounding hydrogel was removed. Entire brains were immersed in 50 ml of clearing solution and stored in a water bath at 45°C. After 2 months of incubation in clearing solution for passive clearing, optically transparent brains were washed with 50 ml PBS (0.1% TritonX in 1× PBS) twice for 24 hours each, at 45°C. After that, the samples were immersed in a refractive index matching (75 g diatrizoic acid, 70 g D-sorbitol and 23 g N-methyl-D-glucamine in 100 ml of water) medium for 2 days before imaging. After incubation in refractive index matching the samples were imaged in a custom-built single plane illumination microscopy. We used a detection lens with

magnification 5×, a laser wavelength of 488 for fluorescence excitation and a 5 μm z step, obtaining a 1.29 × 1.29 × 5 μm<sup>3</sup> voxel size.

**Removal of stripe artifacts.** The stripe noise present in the raw fMOST images is mainly caused by fluorescent bleaching during the knife cutting and imaging process. We modeled this kind of stripe artifacts as multiplicative periodic noise and designed a log-space frequency notch filter to realize high-quality stripe artifacts removal (Supplementary Video 1). For each fMOST image, (1) one 2D coronal slice with the largest foreground brain area was extracted (Extended Data Fig. 8a) and converted to the frequency domain using fast Fourier transform (Extended Data Fig. 8b). (2) By examining the frequency spectrum, the bandwidth, orientation and cutoff frequency of the notch filter were identified and then used to construct a Gaussian notch filter (Extended Data Fig. 8c). (3) We imposed the log transform on the intensity of the fMOST image in the spatial domain to convert the multiplicative noise to additive one (Extended Data Fig. 8d). (4) The log-transformed fMOST image was filtered in the frequency domain in a slice-by-slice manner using the Gaussian notch filter generated in step 2 (Extended Data Fig. 8e). (5) Finally, we computed the inverse fast Fourier transform on the filtered spectrum and used inverse-log transform to bring the filtered image back from the log space (Extended Data Fig. 8f).

**Robust point matching-based automatic global registration.** To minimize the impact of intensity variation between different modalities, we implemented the automatic global alignment by affine aligning brains' outer contour points through point-set matching. First, the foreground brain region was extracted via Otsu thresholding<sup>53</sup>, followed by morphology filtering to remove the small and thin artifacts. Then we uniformly sampled (7 × 7 × 7 grid) the contour of the foreground region to obtain a point set (~1,500 points per brain) that represents the shape of the brain's outer contour. To tackle the large direction variation between brains, the principal axes of target and subject point sets were extracted using principal component analysis and then rigidly aligned. Finally, following the deterministic annealing framework of robust point matching<sup>54</sup>, we iteratively refined an affine transform to maximize the overall shape of two point sets (Supplementary Video 2). We set the initial annealing temperature to 0.25, and gradually decreased the temperature with an annealing rate of 0.96 during iteration.

**2.5D corner detection.** As CLM establishes the correspondence between two brains by deforming the landmarks defined in the target brain to fit the subject brain, the quantity and quality of target landmarks will have a big impact on the overall performance of registration. We designed a 2.5D corner detector automatically to extract points of high curvature (corners or junction of different brain regions) of a 3D image and applied it on the annotation template of the CCFv3 to generate anatomically meaningful target landmarks. We realized the 2.5D corner detector via a two-level detection and screening strategy.

In the first level, given a 3D image, we started by extracting all its coronal, sagittal and horizontal planes. For each plane, the Harris response<sup>55</sup> of each pixel was computed, thresholded (discarded if response is less than 0.1), and nonmaximum suppressed (with window size 7 × 7) to find the local maxima as candidate corners. We denoted these candidate corners as the first-level corners, and saved their Harris responses at coronal, sagittal and horizontal planes, respectively, for the second-level screening.

In the second level, we screened the first-level corners by only retaining ones that were detected as first-level corners in at least two orthogonal planes. We called the remained corners second-level ones and updated their Harris response by accumulating their responses at corresponding orthogonal directions. The final landmarks were obtained by applying 3D nonmaximum suppression (with windows size 31 × 31 × 31) on the second-level corners.

**CLM-based automatic local registration.** *Coherent landmark mapping.* CLM establishes a dense correspondence between two images by coherently deforming target landmarks to fit the subject brain (Supplementary Video 3). The landmarks before and after deforming naturally form matching pairs, and were used to calculate the displacement field for local image warping.

Suppose we have a subject image that consists of  $N$  voxels  $V = \{v_i, i=1, 2, \dots, N\}$ , and a target landmark-set of  $M$  points  $L = \{l_j, j=1, 2, \dots, M\}$ , where  $v_i$  is the  $i$ th voxel and  $l_j$  is the  $j$ th landmark in subject and target image, respectively. Let  $P = \{p_{ij}, p_{ij} \in [0, 1]\}$  be the classification probability of  $i$ th voxel to the  $j$ th landmark, we form the CLM as the following energy minimization problem:

$$E(P, \Theta) = \sum_{j=1}^M \sum_{i=1}^N p_{ij} (1 - s_{ij}) \frac{\|v_i - f(l_j, \Theta)\|^2}{r^2} + T \sum_{j=1}^M \sum_{i=1}^N p_{ij} \log p_{ij}, \text{s.t. } \sum_{i=1}^N p_{ij} = 1 \quad (1)$$

where  $f(l_j, \Theta)$  is a nonlinear spatial transform which maps the landmark  $l_j$  to a new location with parameters  $\Theta$ ,  $s_{ij} \in [0, 1]$  denotes the matching score between voxel  $v_i$  and currently deformed landmark  $f(l_j, \Theta)$ , and  $r$  is the distance influence factor that controls the impact of voxel-to-landmark distance to their classification

probability. Minimizing the first term encourages assigning larger classification probability to the voxel-to-landmark pair with higher matching score and smaller Euclidean distance. While minimizing the second term is equivalent to maximizing the entropy of  $P$ , which aims to enforce the convexity of the energy function and alleviate the risk of algorithm getting trapped in the local minimum during optimization.  $T$  acts as a balance factor between two terms.

We adopted a dual update strategy to solve the joint classification and spatial transform optimization problem posed in energy function (1). In each iteration, we first fix the transformation  $\Theta$  and calculate the voxel-to-landmark classification  $P$  with currently deformed landmarks, then update the transform and perform landmark deformation with classification probability fixed. We denote the above two steps as landmark-guided voxel classification and voxel classification-guided landmark deformation, respectively.

**Landmark-guided voxel classification.** Given spatial transform  $\Theta$ , the deformed target landmarks can be obtained as  $f(L, \Theta)$ . For any  $i, j$  value, minimizing  $E(P, \Theta)$  with respect to  $P$  will lead to:

$$\frac{\partial E}{\partial p_{ij}} = 0 \Rightarrow p_{ij} = \exp \left( -\frac{(1-s_{ij}) \|v_i - f(l_j, \Theta)\|^2}{r^2 T} - 1 \right) \quad (2)$$

we chose  $T=1$ , and normalize classification probability so that  $\sum_i p_{ij} = 1$  for any  $i \in [0, N]$  once they are updated using equation (2). To speed up the iterative voxel classification, we only concerned a subdomain ( $21 \times 21 \times 21$ ) around each landmark and chose  $r=10$ .

**Voxel classification-guided landmark deformation.** With fixed voxel classification probability  $P$ , we realized the coherent landmark deforming in three steps. First, we preliminarily updated each landmark to the weighted center of mass of all voxels  $\bar{p}_j^{\text{pre}} = \sum_{i=1}^N p_{ij} v_i$ . Then we updated transform parameters  $\Theta$  so that deformed landmarks  $f(L, \Theta)$  can best approximate their preliminary updated ones  $\{\bar{p}_j^{\text{pre}}, j = 1, 2, \dots, M\}$  under smoothness and brain shape prior constraints. We chose smooth-thin-plate-spline (STPS)<sup>36</sup> to model the spatial transform  $f$  with parameters  $\Theta$ , which can be decomposed into an affine transform and a smoothness parameter  $\lambda$  regularized nonlinear warping component. If  $\lambda=0$ , an exact interpolation was performed. We chose  $\lambda=0.001$  to balance the goodness of fit between two point sets and the smoothness of deformation. The spatial transform parameters  $\Theta$  were solved as a least-squares solution of STPS. Finally, we used LittleQuickWarp<sup>37</sup> (a fast and memory efficient implementation of STPS) with updated  $\Theta$  to transform the initial target landmarks and realize the coherent landmark deforming. As target landmarks are obtained on the CCFv3 annotation template, we ensured that all landmarks are on the boundaries of the brain region. They naturally capture the shape prior of these regions and form a simplified point-set representation of the CCFv3. The smooth (regularized by STPS) and coherent deformation of such a simplified atlas onto the subject brain implicitly utilizes the shape priors and spatial relationship of brain regions, while these priors are hard to incorporate in the commonly used B-spline based methods<sup>37-42</sup>.

If provided, the modality-specific shape priors can also be accommodated under the CLM framework to enhance further the robustness and accuracy of registration. Given the fMOST atlas, as the transformation between the CCFv3 and fMOST atlas was known and fixed, we utilized these modality-specific shape priors by first mapping the initial landmarks of different brain regions to fMOST space, and then deforming them in a region-wise manner following the above-mentioned STPS deforming strategy. In our implementation, this region-wise constraint was applied before the brain-wise constraint in each iteration.

**Matching score computation.** We based the calculation of matching score between a voxel-to-landmark pair upon three complementary features: (1) gradient, (2) histogram of oriented gradients (HOG)<sup>38</sup>, and (3) segmentation probability. As these features encode the edge, shape context and semantic information, respectively, their integration is expected to provide CLM with more reliable matching metric to cope better with the challenges of cross-modal registration.

We generated the gradient feature by applying the Sobel filter ( $3 \times 3 \times 3$ ) on Gaussian smoothed ( $3 \times 3 \times 3$ ) volume. At a given location, its HOG feature was calculated by concatenating the normalized gradient histograms of eight blocks within the patch ( $21 \times 21 \times 21$ ) that centered on that location. We divided each patch into 27 cells ( $7 \times 7 \times 7$ ) and combined its neighboring ( $2 \times 2 \times 2$ ) cells to create blocks. For each block, the gradient histogram (nine orientations) of each cell within it was computed, concatenated and normalized to form the block feature. This resulted in a 576-dimensional HOG feature for each location.

We chose to generate the segmentation probability feature using a semantic segmentation network. As a proof of principle, we adopted a slightly modified 3D U-Net<sup>39</sup> to generate the segmentation probability (with probability values range from 0 to 1) of each voxel to six main brain regions (HY, HPF, CTX, CBX, BS, CP) and background. Indeed, under the CLM framework, 3D U-Net can be readily substituted with other more sophisticated semantic segmentation networks to improve the registration performance further. The network architecture used in this study is shown in Extended Data Fig. 2a, which contains a

classic encoder-decoder structure. We used four layers at the encoder stage and four layers at the decoder stage, and they have 32, 64, 128 and 256 channels, respectively. The seven-dimensional score maps generated at the final softmax layer were used as the segmentation probability feature. The network was trained from scratch on 57 brains with ground truth segmentation labels obtained via mBrainAligner's semiautomatic registration module. We used 23 brains as the training set, three brains as the validation set and the remainder for testing. The segmentation results of six brain regions are visualized in Extended Data Fig. 2b. Potential segmentation errors can be mitigated as the segmentation information is only one of the three discriminative features that guide the deformation of landmarks, for which the deformation is further constrained by the overall shape priors.

Finally, we computed the matching score of each landmark-to-voxel pair by accumulating the similarity score of their centered patches ( $21 \times 21 \times 21$ ) on three feature maps. In our implementation, (1) inverse mean squared error, (2) mutual information and (3) normalized cross-correlation were used as similarity metrics.

**Semiautomatic refinement.** The semiautomatic registration module provides an interactive interface for investigators to visualize, examine and fine-tune the registration results (Fig. 1d, Extended Data Fig. 1 and Supplementary Video 4). In this interface, the boundary of brain regions defined in the CCFv3 was vectorized as Bezier curves and visualized as a layer overlaid on the registered brain. The control nodes of Bezier curves were evenly sampled from brain region boundaries defined in the CCFv3 annotation template. The registration quality could be examined by checking how well the registered brain would fit the CCFv3 boundaries in coronal, sagittal and horizontal views at different scales. Different to the reported method that performed slice-by-slice adjustment in coronal view<sup>40</sup>, our method could fine-tune three-view panels synchronously by simply dragging the curves, thus distortions in Z direction could also be rectified. In addition, we also provided a magnetic lasso<sup>41</sup> mode to assist in the alignment of curves in a semiautomatic manner. Three levels of subregion granularity (coarse, medium, fine) were provided for each main brain region to accommodate different local granularity requirements (Extended Data Fig. 1b). We realized the final 3D image deforming by using STPS transformation that computed on the offsets of fine-tuned control nodes. We demonstrated that the semiautomatic registration module could register partially imaged brain (LSFM partially imaged brain of hippocampus in Extended Data Fig. 9), a known challenging registration problem for fully automatic methods.

**3D fMOST atlas building of mouse brain.** We built the fMOST average template from 36 mouse brains in an iterative way. Before entering the iteration, all fMOST brains were preprocessed and globally aligned to the CCFv3 following the methods described above. To encourage a symmetric atlas, the globally aligned brains were mirrored across the midline, which resulted in a total of 72 brains. We iterated the following two steps until convergence: (1) each brain was locally registered to the template (we chose a high-quality globally aligned fMOST brain as the template in the first iteration) by using the automatic local registration module of mBrainAligner and averaged together; (2) the average deformation field over all brains was computed, inverted, and used to perform shape normalization on the average image obtained in (1) to compensate the potential shape bias towards CCF. This shape-normalized average brain was then used as the anatomical template in step (1) for the next iteration. We stopped the iteration until the sum of the magnitude of the average deformation field was below the threshold of 1.0.

We generated the fMOST annotation template by first aligning the fMOST average template to the CCFv3 and then back-projecting the annotation template of the CCFv3 using the inverse of the deformation field generated. To ensure a highly accurate registration, we registered the fMOST average template to the CCFv3 by concatenating the automatic and semiautomatic registration module of mBrainAligner.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The full resolution fMOST image datasets (<https://download.brainimagelibrary.org/biccn/zeng/luo/fMOST/>) of all mouse brains used in this study, as well as the original and CCFv3 registered single neuron reconstructions (<https://doi.org/10.35077/g.25>), are available at BICCN's Brain Image Library (BIL) at Pittsburgh Supercomputing Center ([www.brainimagelibrary.org](http://www.brainimagelibrary.org)). The single-neuron reconstructions, the CCFv3 registered version of these reconstructions, as well as 3D navigation movie-gallery of these data are also available at SEU-ALLEN Joint Center, Institute for Brain and Intelligence (<https://braintell.org/projects/fullmorpho/>). The Allen CCFv3 is available at (<http://atlas.brain-map.org/>), and MRI mouse brain used in this study can be obtained through (<https://www.nitrc.org/projects/infchwsmouse>). We also provided downsampled fMOST, LSFM, VISO and MRI whole mouse brain data and LSFM partially imaged brain data along with code and testing scripts at Github page ([https://github.com/Vaa3D/vaa3d\\_tools/tree/master/hackathon/mBrainAligner](https://github.com/Vaa3D/vaa3d_tools/tree/master/hackathon/mBrainAligner)). The full resolution LSFM and VISO brain images are available on request. Source data are provided with this paper.

## Code availability

The source code of all mBrainAligner modules, including stripe artifacts removal, automatic global registration, automatic local registration, semiautomatic refinement along with the binary executable files and sample data can be found at mBrainAligner's GitHub page ([https://github.com/Vaa3D/vaa3d\\_tools/tree/master/hackathon/mBrainAligner](https://github.com/Vaa3D/vaa3d_tools/tree/master/hackathon/mBrainAligner)). The automatic global, local registration and semiautomatic registration modules were written in C++ (v.11) with Qt (v.5.6), and the stripe artifacts removal module was implemented using Matlab 2016.

## References

51. Peng, H., Ruan, Z., Long, F., Simpson, J. H. & Myers, E. W. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat. Biotechnol.* **28**, 348–353 (2010).
52. Chi, J. et al. Three-dimensional adipose tissue imaging reveals regional variation in beige fat biogenesis and PRDM16-dependent sympathetic neurite density. *Cell Metab.* **27**, 226–236.e223 (2018).
53. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (2007).
54. HL, C. & Rangarajan, A. A new point matching algorithm for non-rigid registration. *Computer Vis. Image Underst.* **89**, 114–141 (2003).
55. Harris, C. G. & Stephens, M. A combined corner and edge detector. In *Proc. Alvey Vision Conference* Vol. 15, 10–5244 (Citeseer, 1988).
56. Wahba, G. *Spline Models for Observational Data* (SIAM, 1990).
57. Qu, L. & Peng, H. Littlequickwarp: an ultrafast image warping tool. *Methods* **73**, 38–42 (2015).
58. Dalal, N. & Triggs, B. Histograms of oriented gradients for human detection. In *Proc. Computer Society Conference on Computer Vision and Pattern Recognition* Vol. 1 886–893 (IEEE, 2005).
59. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Proc. International Conference on Medical Image Computing and Computer-assisted Intervention* (eds. Ourselin S., Joskowicz L., Sabuncu M., Unal G., Wells W.) 424–432 (Springer, 2016).
60. Fürth, D. et al. An interactive framework for whole-brain maps at cellular resolution. *Nat. Neurosci.* **21**, 139–149 (2018).
61. Mortensen, E. N. & Barrett, W. A. Intelligent scissors for image composition. In *Proc. 22nd Annual Conference on Computer Graphics and Interactive Techniques* 191–198 (ACM TOG, 1995).

## Acknowledgements

This work was initialized and funded by Southeast University to support Open Science collaboration through the SEU-ALLEN Joint Center, Institute of Brain and Intelligence. We thank Y. Wang and Q. Wang (Allen Institute for Brain Science) for assistance with manual annotation of some somas, Y. Yu (Allen Institute for Brain Science) for downsampling several fMOST datasets, S. Jiang (Southeast University) for assistance in data storage and transmission, D.H. Lu (Tencent) for building mBrainAligner's web

portal interface, M. Dierssen (Center for Genomic Regulation) and L. Manubens-Gil (Southeast University) for providing partial brain LSFM data, W. Luo, H. Wang, J. Yang, M. Wang, Y. Tang and R. Zhang (Anhui University) for help with semiautomatic registration, the team at SEU-ALLEN Joint Center for neuron reconstruction, and the supercomputing platform of Anhui University for computing support. This work was also partially funded by the National Science Foundation of China (NSFC) grant 61871411 and the University Synergy Innovation Program of Anhui Province GXXT-2019-008 and GXXT-2021-001 to L.Q., the NSFC grant 32071367 to Y.W., the NSFC-Guangdong Joint Fund U20A6005 to G.B. and the Fundamental Research Funds for the Central Universities 2242021k30046.

## Author contributions

H.P. conceptualized the study and provided guidance for this study. L.Q. led the development of mBrainAligner. H.P., L.Q., P.X., Li L. and Y.L. designed all experiments. L.Q. and Y.L. designed and implemented the CLM. Y.L. conducted validation analyses of mBrainAligner, generated the fMOST atlas and created the figures. Y.L. designed and Long L. developed the semiautomatic registration module. Long L. designed and implemented the 2.5D corner detection algorithm. L.Q. and W.W. designed and implemented the stripe artifacts removal tool. L.Q. and T.W. codeveloped tools for image and various neuronal structure warping, and T.W. assisted in building the fMOST atlas. K.G. and Y.Z. collaborated on deep-learning models. L.O. developed the global registration module. F.X. developed registration scripts and helped in aligning neurons to CCF. J.W. and Y.L. collaborated with L.Q. to set up and manage the software-developing team and the data-analyzing team. P.X. conducted the cell type analysis and provided tools for soma localization and axon projection analyses. Li L. provided and validated soma localization and axon projection data. Y.W. contributed to neuronal reconstruction. H.G. and Q.L. provided the fMOST imaging data. A.K. and Z.W. provided the LSFM imaging data. F.X. and G.B. provided the VISO-R imaging data. H.D. assisted in anatomy analysis. H.Z. provided the fMOST dataset. H.P., L.Q. and Y.L. wrote the manuscript with input from all coauthors.

## Competing interests

The authors declare no competing interests.

## Additional information

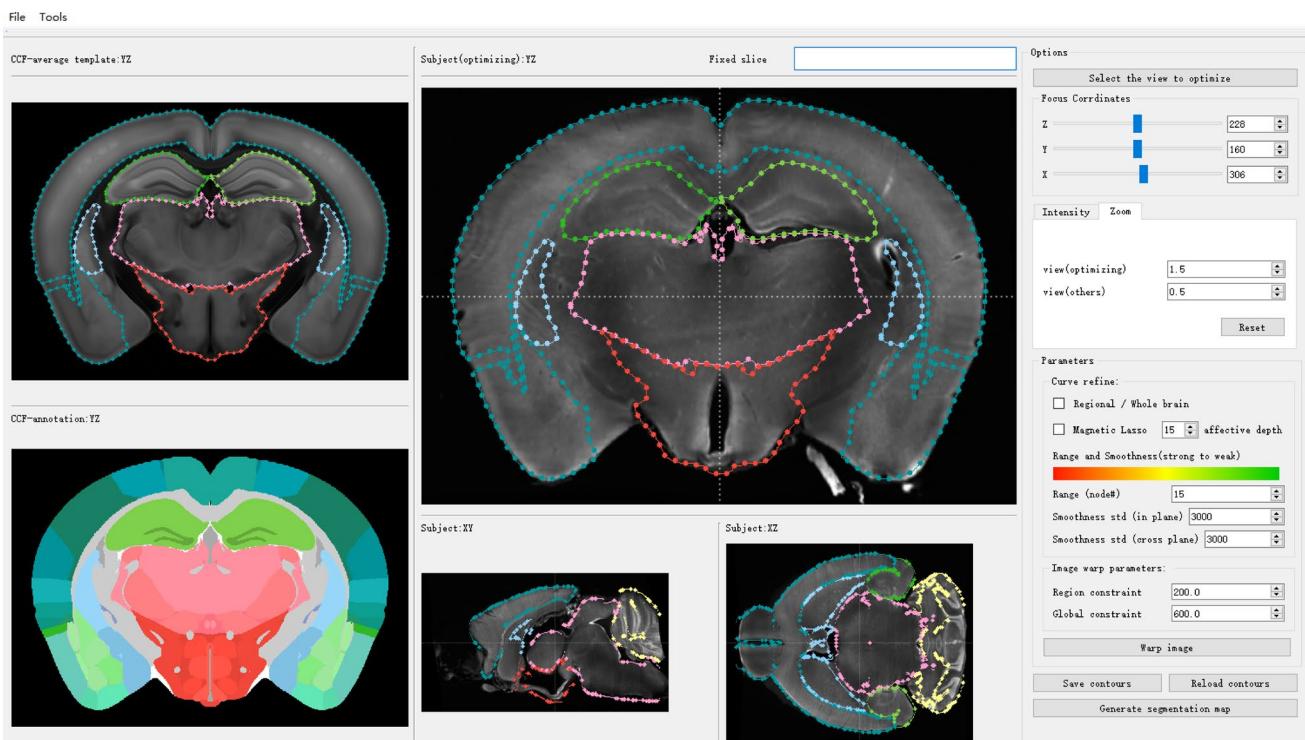
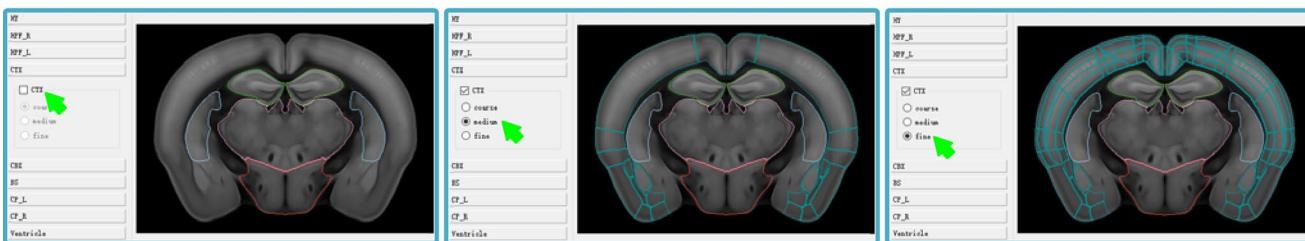
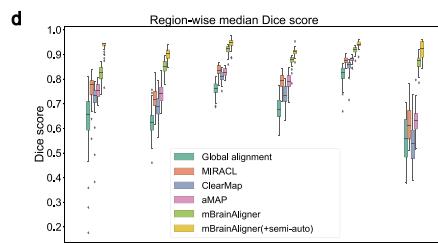
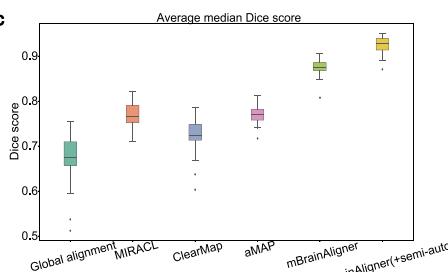
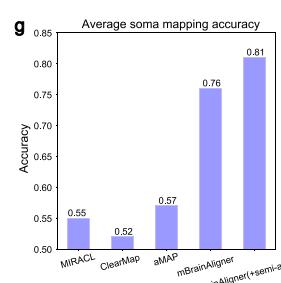
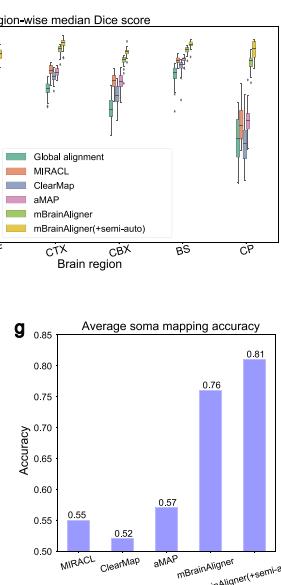
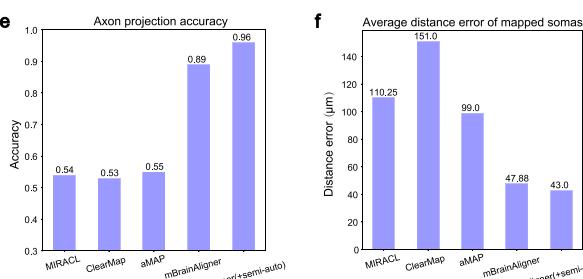
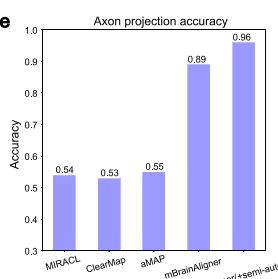
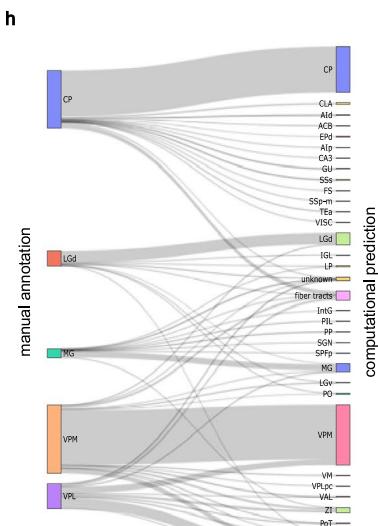
**Extended data** are available for this paper at <https://doi.org/10.1038/s41592-021-01334-w>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-021-01334-w>.

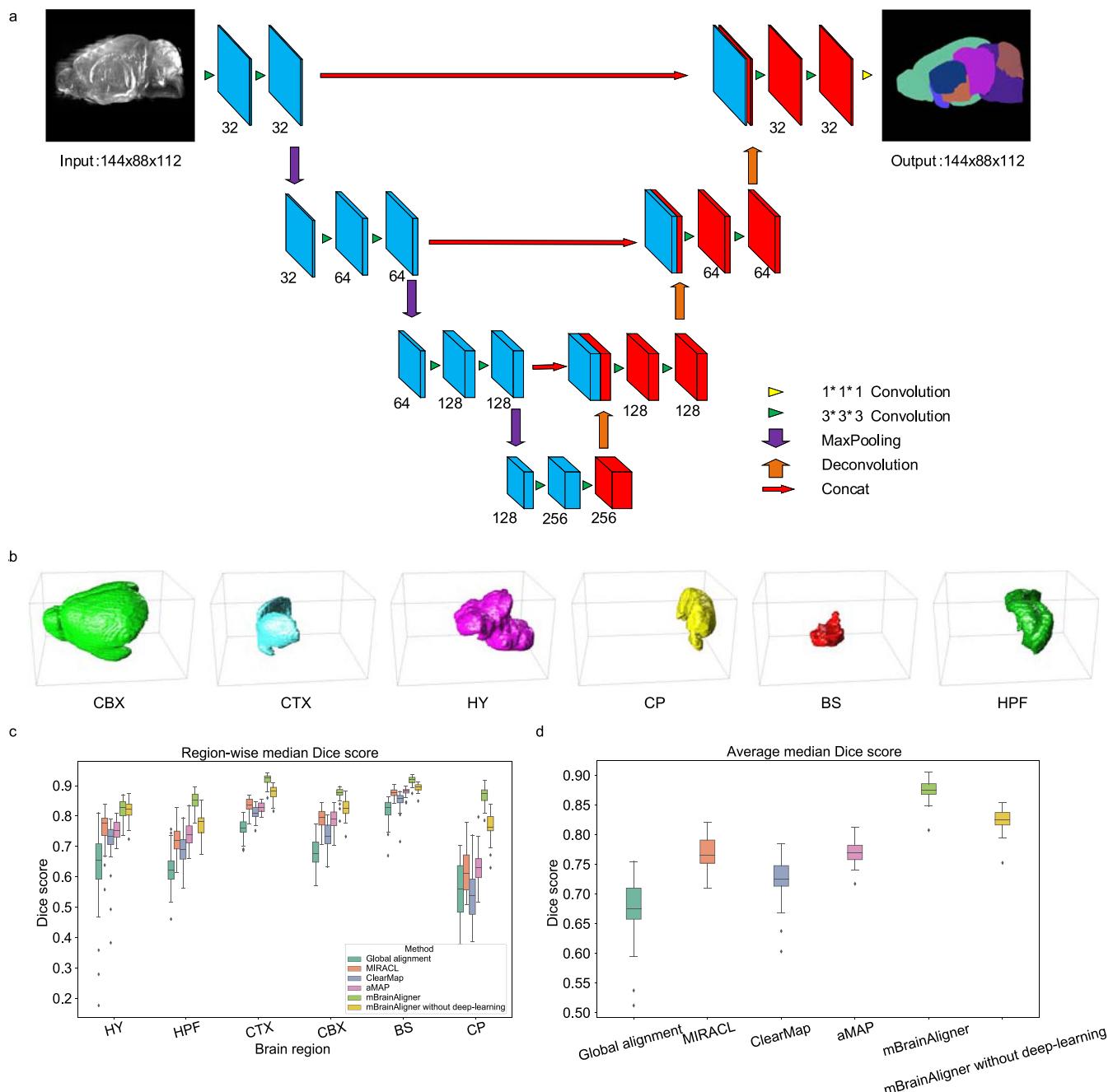
**Correspondence and requests for materials** should be addressed to Hanchuan Peng.

**Peer review Information** *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Nina Vogt was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

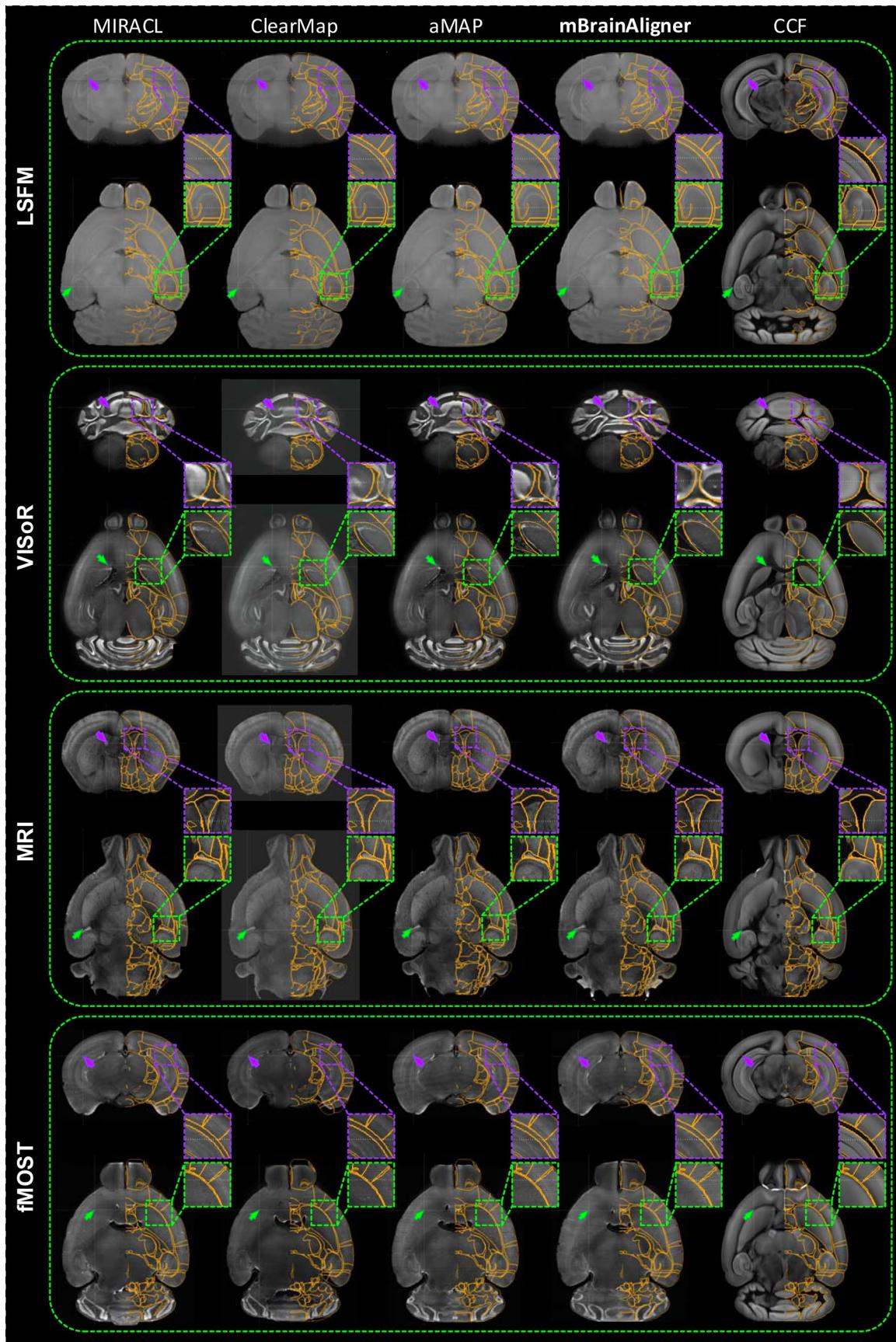
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**a****b****c****d****e****f**

**Extended Data Fig. 1 | Semi-automatic registration.** **a**, Interface of semi-automatic registration module. **b**, Three levels of sub-region granularity (coarse, medium, and fine) are provided for each main brain region. **c-h**, Comparison of registration accuracy on 31 fMOST brains when semi-automatic registration is applied after automatic registration. Box plot: center line, median; box limits, upper and lower quartiles; whiskers, 1.5 x interquartile range; points, outliers.

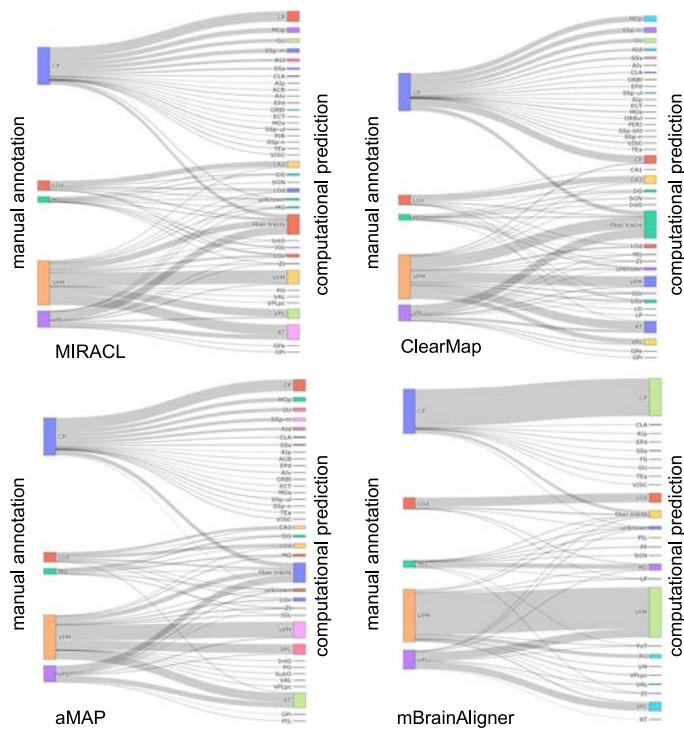


**Extended Data Fig. 2 | Architecture and performance of deep learning-based segmentation network.** **a**, Architecture of segmentation network. **b**, 3D visualization of segmentation results of six brain regions. **c-d**, Comparison of registration accuracy of mBrainAligner on 31 fMOST brains with deep learning features toggled on and off. Box plot: center line, median; box limits, upper and lower quartiles; whiskers, 1.5 x interquartile range; points, outliers.

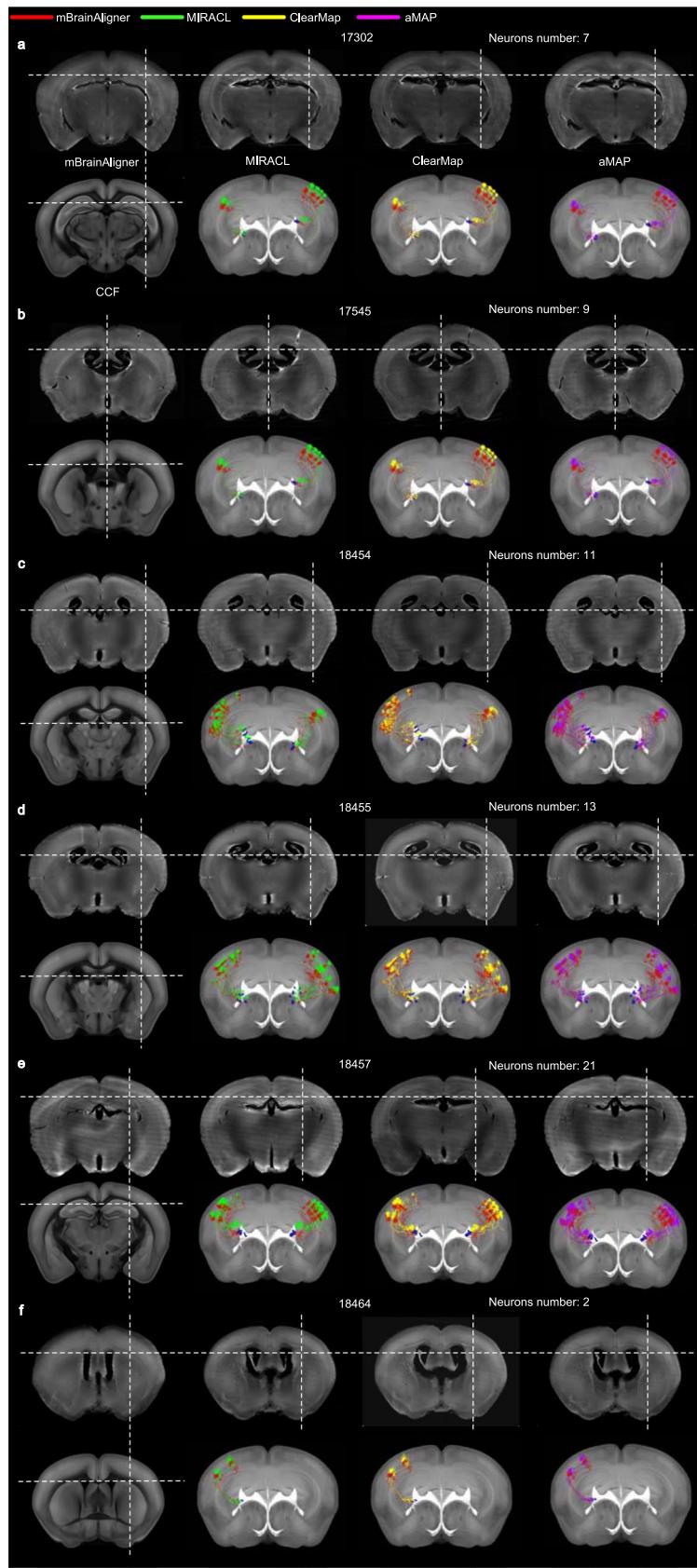


Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Comparison of different methods in registering images of different modalities to CCFv3.** Four green dashed boxes show the brain registration results of four different modalities (LSFM, VISO-R, MRI, fMOST) generated by different methods (MIRACL, ClearMap, aMAP and mBrainAligner.). The first and second rows of each box show the coronal and sagittal view of registered brains respectively with zoom-in view of two subregions. Registered brains of different modalities with brain region boundaries (orange) defined in CCFv3 overlaid. In each box, arrows of the same color point to the same spatial coordinates.

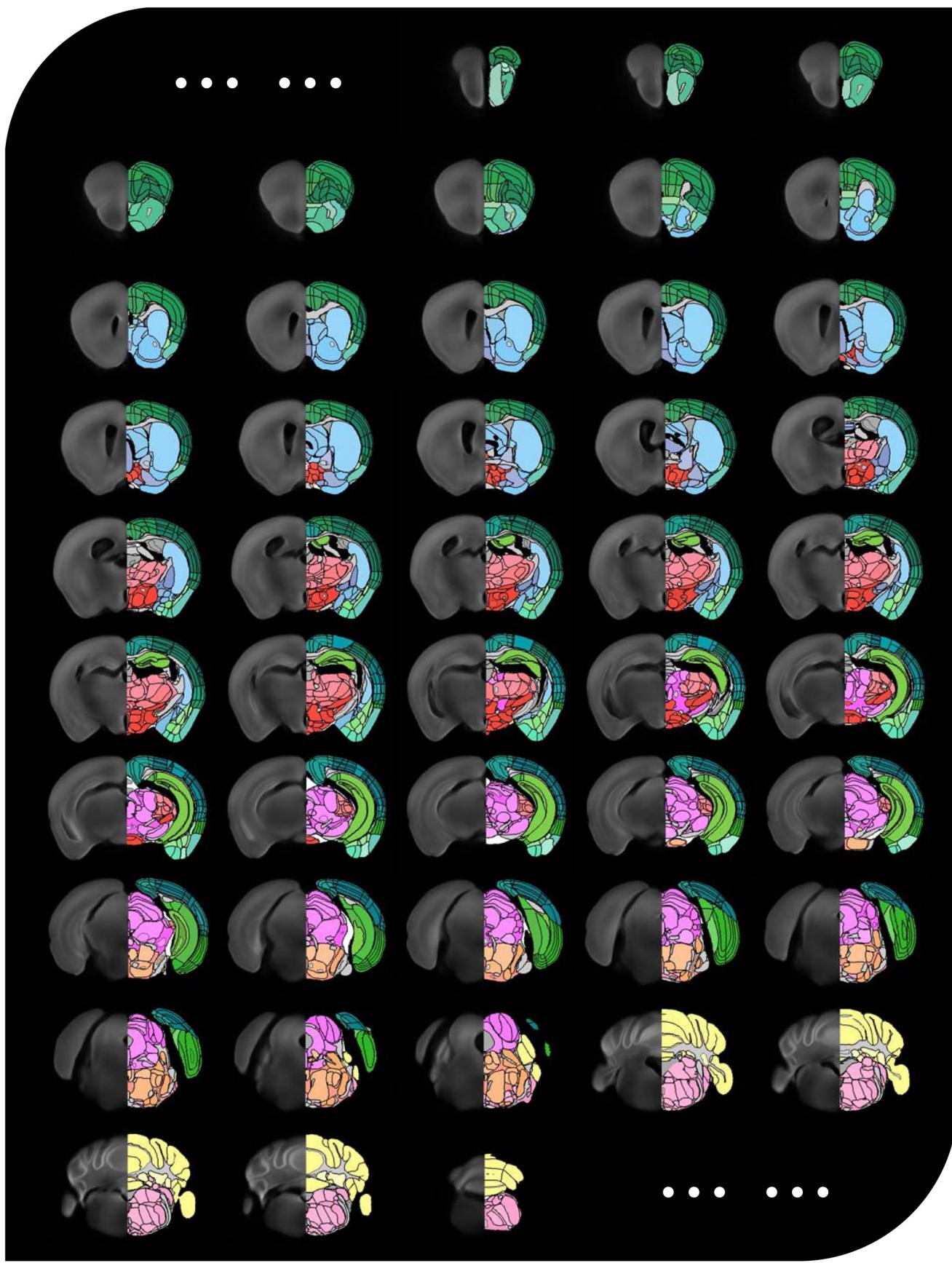


**Extended Data Fig. 4 | Comparison of soma localization between manual annotation and computational prediction results generated by different methods.** The left side of diagram shows manually annotated results, and the right side shows their computationally predicted results.

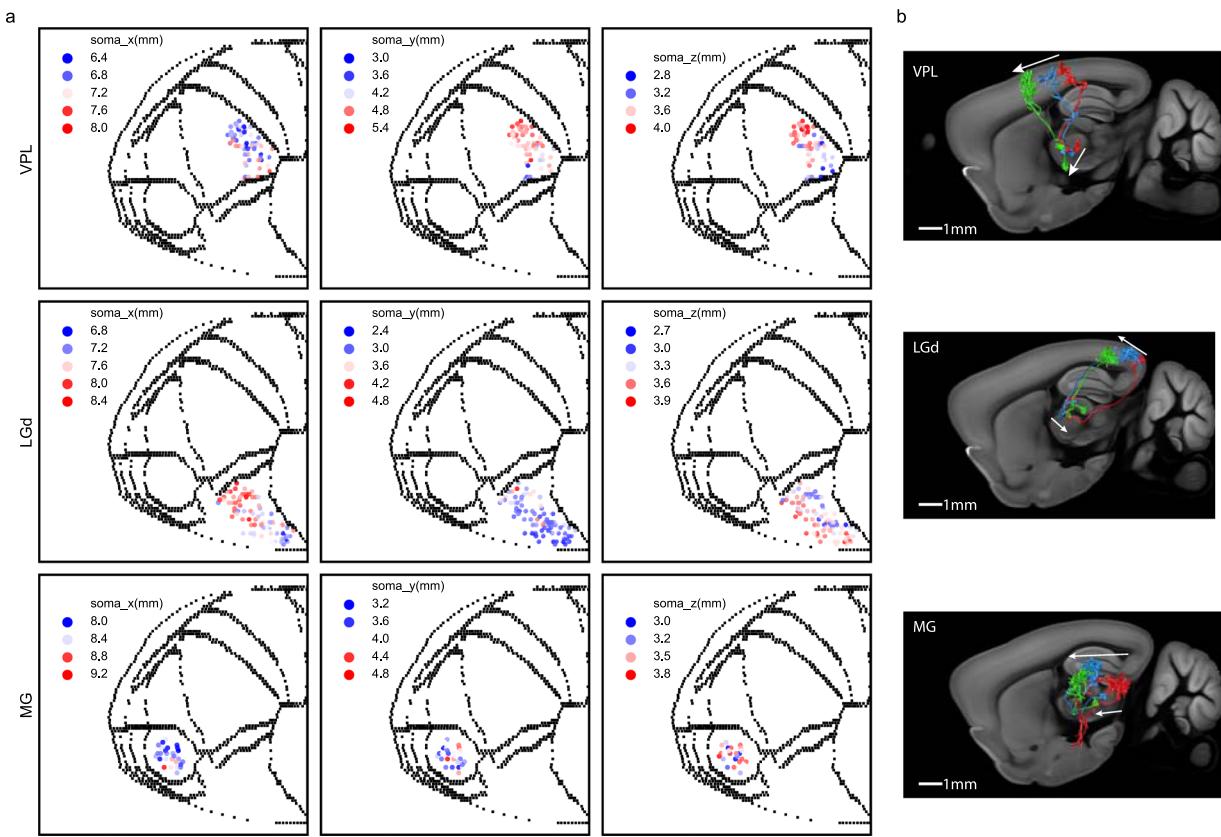


Extended Data Fig. 5 | See next page for caption.

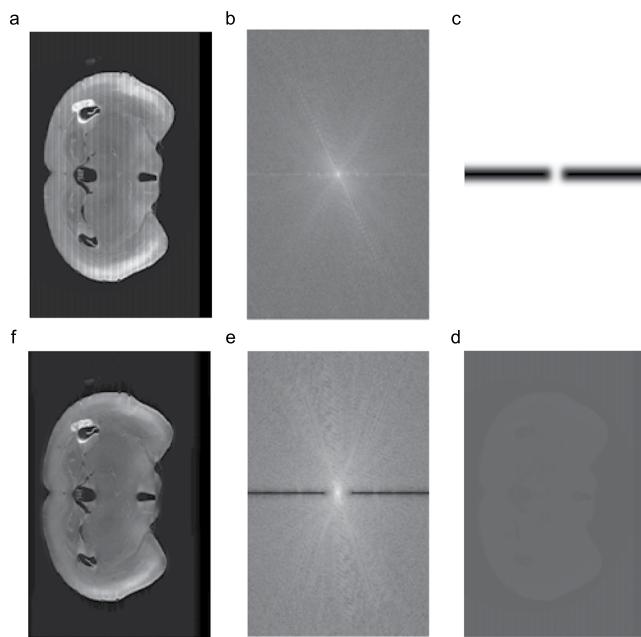
**Extended Data Fig. 5 | Comparison of neuron mapping results generated by different methods.** a-f, 63 CCFv3 mapped neurons of six brains (ID 17302, 17545, 18454, 18455, 18457 and 18464) with somas residing in VPM or VPL. The first row of a-f shows CCFv3 aligned brains generated by four different methods (mBrainAligner, MIRACL, ClearMap and aMAP). The cross-lines overlaid on the aligned brains point to the same coordinate on the inner boundary of cortex. The left-most image of second row of a-f shows corresponding slice of CCFv3 average template, and the right three images show the maximum intensity projection view of CCFv3 with mapped neurons overlaid, with red for mBrainAligner, green for MIRACL, yellow for ClearMap and purple for aMAP.



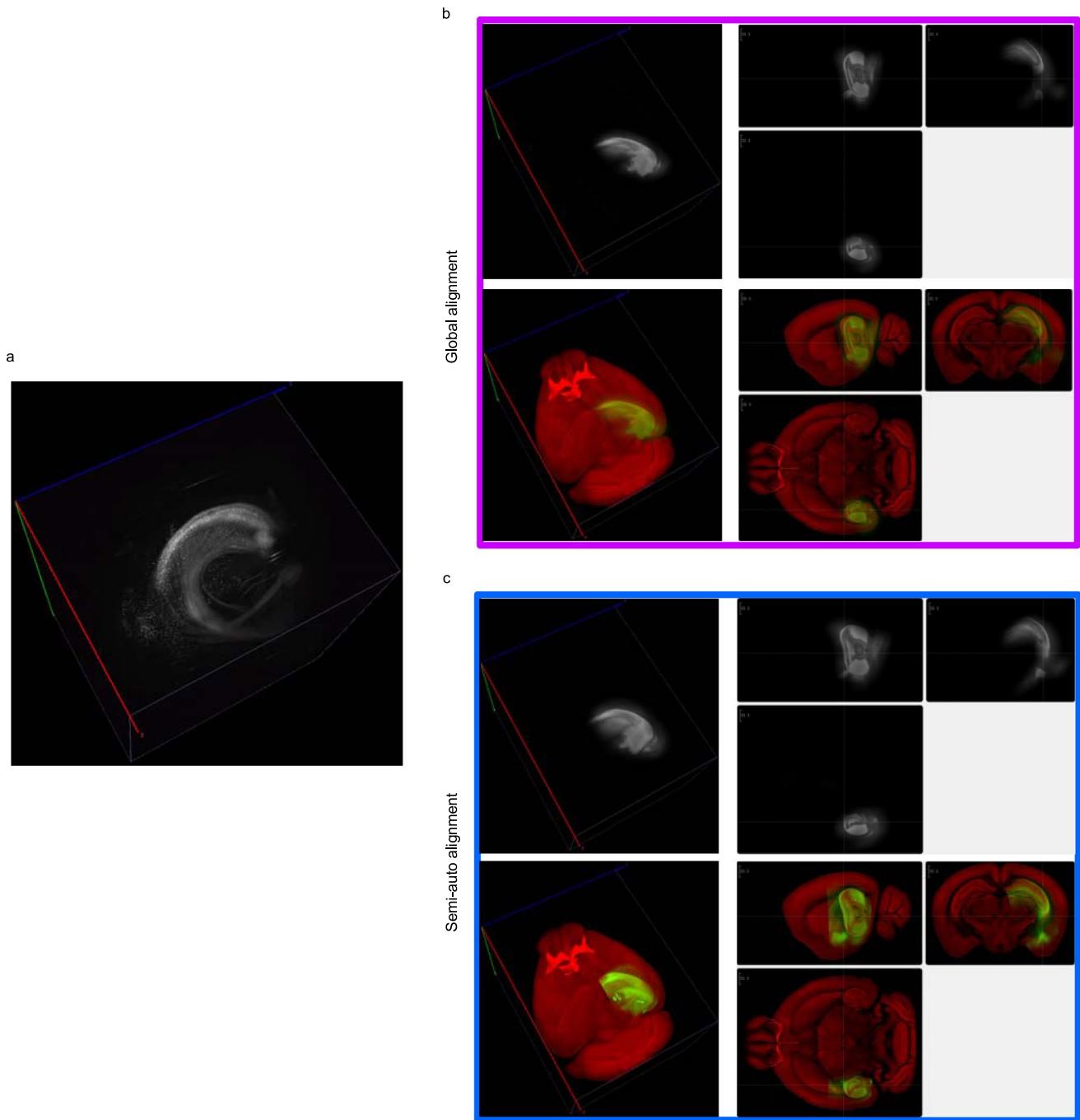
**Extended Data Fig. 6 | fMOST-domain atlas of mouse brain (2D views).** Coronal sections of fMOST-domain atlas of mouse brain, with average template shown on left and annotation template on right.



**Extended Data Fig. 7 | Topography of thalamic neuron projections by integrating CCF-registered brains.** **a**, Location of axonal arbors of cortical projection neurons grouped by thalamic nuclei. Dots represent center of axonal arbors in the flattened 2D cortical map. Locations of corresponding somas are color-coded in three separate columns: (left to right) the anterior-posterior (soma-x), dorsal-ventral (soma-y) and lateral-medial (soma-z) axes. **b**, Examples of single-neuron projections (VPL, LGd, MG). Colors represent individual neurons.



**Extended Data Fig. 8 | Stripe artifacts removal.** **a**, One coronal slice of fMOST mouse brain image. **b**, The Fourier spectrum of image shown in a. **c**, Constructed Gaussian notch filter. **d**, Log transformed image of a. **e**, Filtered spectrum of log-transformed image using filter as shown in c. **f**, Image after stripe removal.



**Extended Data Fig. 9 | Partial image alignment.** **a**, Partially imaged brain region of hippocampus (in LSFM modality). **b**, Global alignment result of partial image to CCFv3. The left column shows the maximum intensity projection of globally aligned partial brain (top) and its blended view with CCFv3 average template (bottom), with their three-view shown on the right. **c**, Result after semi-automatic refinement.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

We use TeraConverter module (version bundled with Vaa3D) of Vaa3D (version 3.604) to downsample the raw data, use the TeraVR (version bundled with Vaa3D) to reconstruct the single neuron full morphology of fMOST brains, and use LittleQuickWarp (version bundled with Vaa3D) to warp brain images. The Vaa3D, TeraConverter and TeraVR software is available through the GitHub release page of vaa3d.org (<https://github.com/Vaa3D/>).

Data analysis

Custom softwares for automatic global, local registration and semi-automatic registration were written in C++ (v11) with Qt (5.6) development kit. Custom software for stripe artifacts removal module was implemented using Matlab 2016. Custom codes, executables and the user guides of the custom software can be accessed at [https://github.com/Vaa3D/vaa3d\\_tools/tree/master/hackathon/mBrainAligner](https://github.com/Vaa3D/vaa3d_tools/tree/master/hackathon/mBrainAligner).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The full resolution fMOST image datasets (<https://download.brainimagelibrary.org/biccn/zeng/luo/fMOST/>) of all mouse brains used in this study, as well as the original and CCFv3 registered single neuron reconstructions (DOI: <https://doi.org/10.35077/g.25>), are available at BICCN's Brain Image Library (BIL) at Pittsburgh Supercomputing Center ([www.brainimagelibrary.org](http://www.brainimagelibrary.org)). The single neuron reconstructions, the CCFv3 registered version of these reconstructions, as well as 3D

navigation movie-gallery of these data are also available at SEU-ALLEN Joint Center, Institute for Brain and Intelligence (<https://braintell.org/projects/fullmorpho/>). The Allen CCFv3 atlas is available at (<http://atlas.brain-map.org/>), and MRI mouse brain used in this study can be obtained through (<https://www.nitrc.org/projects/incfwhsmouse>). We also provided downsampled fMOST, LSF, VIsor and MRI whole mouse brain data and LSF partially imaged brain data along with code and testing scripts at Github page ([https://github.com/Vaa3D/vaa3d\\_tools/tree/master/hackathon/mBrainAligner](https://github.com/Vaa3D/vaa3d_tools/tree/master/hackathon/mBrainAligner)). The full resolution LSF and VIsor brain images are available upon request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	57 fMOST mouse brains with 31 brains contain digitally reconstructed neurons. This is so far all the data collected in the US BRAIN's BICCN project. It's sufficient to lead to a determination that systematic errors surpassed statistical errors.
Data exclusions	We used all 1708 neurons reconstructed so far from 31 fMOST brains to conduct the soma localization analysis. All 63 neurons with somas in VPM and VPL, and all 162 neurons with apical dendrites distributed in the L1 layer of cortex were selected for axon projection analysis. There is no further exclusion criteria was applied. 31 fMOST brains containing reconstructed neurons were excluded from the network training dataset to ensure that there is no overlap between the training and testing datasets.
Replication	We successfully tested against 31 fMOST, 1 LSF, 1 VIsor and 1 MRI mouse brains.
Randomization	Randomization was not performed as multiple experimental groups across biological samples were not used in this study.
Blinding	All annotators worked in isolation and were blinded to group allocation.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

### Laboratory animals

57 mice used for fMOST whole brain imaging were 8-week-old and contained both male and female.  
 One mouse used for VIsor whole brain imaging was 8-week-old male.  
 One mouse used for LSF whole brain imaging was 8-week-old male.  
 One mouse used for LSF partial brain imaging was 12-week-old male.  
 All mice used in this study were group-housed with a 12-hour light/dark cycle, controlled temperature (20–22°C) and humidity (50%–70%), and free access to food and water.

### Wild animals

This study did not involve wild animals.

### Field-collected samples

This study did not involve samples collected from the field.

### Ethics oversight

All experiments related to the use of mice followed NIH guidelines, and received approval from the Institutional Animal Care and Use Committee of the Allen Institute for Brain Science, the University of Science and Technology of China, the Icahn School of Medicine at

Mount Sinai, and the Center for Genomic Regulation, respectively.

Note that full information on the approval of the study protocol must also be provided in the manuscript.