An interactive 3D atlas of gene expression in the mouse brain

Harry Carey[1]#*, Sébastien Piluso[2]#, Ingvild Bjerke[1], Xiaoyun Gui[3], Sharon Yates[1], Sophia Pieschnik[1], Signý Benediktsdóttir[1], Archana Golla[1], Maja Puchades[1], Timo Dickscheid[3], Trygve Leergaard[1], Simon McMullan[4], Henry Markram[2], Daniel Keller[2], Jan G. Bjaalie[1]*

* Correspondence:

Jan G. Bjaalie: j.g.bjaalie@medisin.uio.no

Harry Carey: harry.carey@medisin.uio.no

#Co-first authorship

Affiliations:

1. Neural Systems Laboratory, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway
2. Blue Brain Project, École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech, 1202 Geneva, Switzerland
3. Institute of Neuroscience and Medicine (INM-1, INM-4), Research Centre Jülich, D-52425 Jülich, Germany.
4. Macquarie Medical School, Macquarie University, Sydney, Australia

Abstract

The Allen Mouse Brain Atlas provides gene expression data for over 20,000 genes and has been extensively used in neuroscience. However, it's constrained by a resolution of 200 μm^3. We improved this to 25 μm^3 for 4,083 genes by re-registering 401,660 sections to atlas space. From this, we produced an atlas with regions defined by shared molecular signatures and built an interactive online platform allowing exploration of gene expression patterns.

Main

Gene expression atlases have substantially advanced our understanding of the organization of the mouse brain[1–3]. Of particular impact has been the Allen Mouse Brain Atlas (AMBA), offering brain-wide expression data for over 20,000 genes across the C57BL/6 (standard laboratory mouse) whole brain[1–4]. This resource is based on classical *in situ* hybridization (ISH) experiments, where the expression of each gene is measured in histological sections. Sections are registered to the three-dimensional (3D) Common Coordinate Framework version 3 (CCFv3) mouse brain atlas[5], and reconstructed into 3D maps of gene expression. The AMBA has supported discoveries across neurodegeneration, neuroplasticity, and genetic disorders[6]. However, the AMBA is constrained by relatively coarse spatial resolution[7] (200 μm^3) and contains registration inaccuracies[8] which blur anatomical detail.

Recent advances in registration methods now allow for more accurate alignment of histological sections to 3D reference atlases than what was possible when the AMBA was first developed[8–11]. Using the DeepSlice toolkit[8], along with the Advanced Normalisation Tools (ANTs) nonlinear alignment tool[9], we re-registered the coronal portion of the AMBA ISH data (4,083 genes) to an extended version of the CCFv3, called the CCFv3 Blue Brain Project ($CCFv3_{BBP}$)[12]. We then combined these registrations and segmentations to generate 3D gene expression volumes at 25 μm^3, a substantial increase over the 200 μm^3 of the AMBA. To support exploration and reuse, the data has been shared through the EBRAINS Knowledge Graph (RRID:SCR_017612) and volumes have been integrated into the siibra explorer atlas viewer[13].

The main advance of this work is that it provides far more detailed spatial information than was previously possible[9]. By reconstructing thousands of datasets at high resolution in a common coordinate framework, the resource enables more refined and intuitive exploration

of gene expression patterns across the whole mouse brain. It also enables the automated identification of areas based on shared expression signatures.

To create this resource, we began by downloading all coronal ISH datasets from the AMBA which were from postnatal day 56 C57BL/6 mice (totalling 401,660 sections from 1,078 brains). We used DeepSlice (RRID:SCR_023854) to place each section into the $CCFv3_{BBP}$. DeepSlice has previously been shown to achieve expert level section image to atlas registration accuracy[8]. For each brain, a subset of the alignments were reviewed and corrected with QuickNII (RRID:SCR_016854)[10], and any changes to angle or position were propagated to all sections from that brain. Nonlinear tissue distortions were corrected using the Advanced Normalisation Tools (ANTs) toolbox, which registered each section to the corresponding Nissl template slice from the $CCFv3_{BBP}$[12]. This procedure was repeated for all sections (Figure 1a).

To validate our registration pipeline, we compared its accuracy with the original AMBA registrations and with manual alignments (both from experts and novices) generated using QuickNII and VisuAlign (RRID:SCR_017978)[11] (figure 1b). For the manual alignments, each rater placed sections from nine AMBA brains into the $CCFv3_{BBP}$ using QuickNII, and performed non-linear corrections for distortions using VisuAlign. We then measured how far each rater's registrations within the $CCFv3_{BBP}$ deviated from the group average, providing an estimate of alignment error in microns[8]. We repeated this process, measuring the deviations of our alignments, and the AMBA's, from this group average.

Experts were most accurate (mean error: 193.6 µm), with novices being significantly less accurate (mean error: 257.9 µm; Tukey $P<0.001$). Our pipeline matched novice performance (mean error 252.8 µm; Tukey $P=0.995$), while remaining significantly less accurate than experts (Tukey $P<0.05$). Our pipeline outperformed the original AMBA registrations (mean

AMBA error 485.9 µm; Tukey $P<0.001$), as did novices (Tukey $P<0.001$) and experts (Tukey $P<0.001$). A major limitation of the AMBA registrations was poor estimation of cutting angles, which showed little correlation with those determined by neuroanatomists (r2 = 0.03, figure 1c). By comparison, the angles specified with our pipeline were highly correlated with the average neuroanatomist-determined angle (r2 = 0.823, Figure 1c).

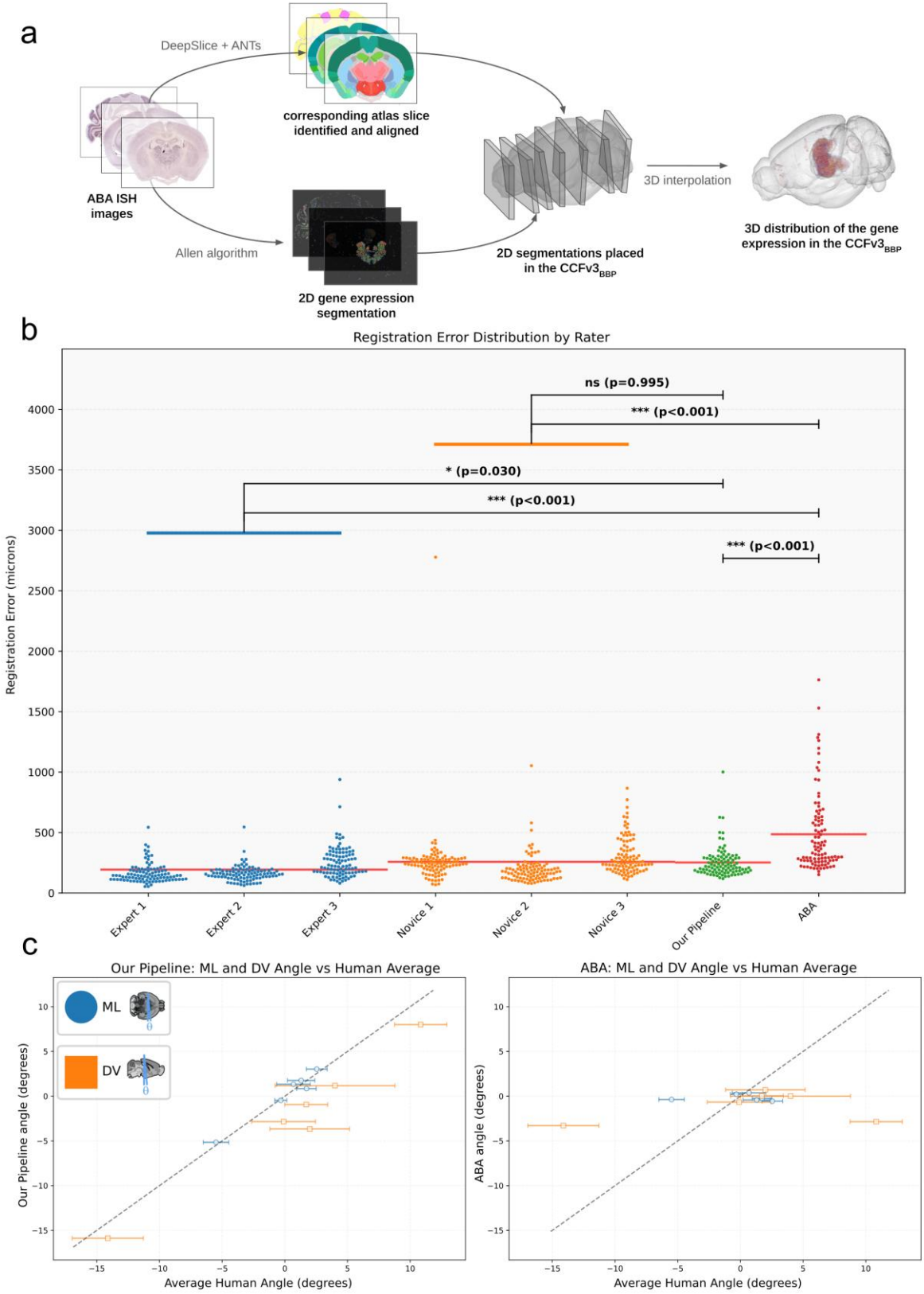


Figure 1

a, Pipeline overview. Coronal ISH sections from the Allen Brain Atlas are registered to the CCFv3bbp reference atlas using DeepSlice. Non-linear distortions are corrected using Advanced Normalisation Tool (ANTs). Empty space between sections is then interpolated creating high resolution 3D gene expression maps. **b**, Quantification of

registration error. The plot shows the distribution of alignment errors for novices (n=3), experts (n=3), our automated pipeline, and the original AMBA alignments. Our pipeline significantly outperforms the AMBA ($p<0.001$) and performs at a level comparable to novices ($p=0.995$). Red lines indicate the mean error for each group. Each data point represents one section **c**, Comparison of predicted sectioning angles. Scatter plots show the correlation between the average human-annotated angles (x-axis) and the angles predicted by our pipeline (left) and the original AMBA (right) for mediolateral (ML, blue) and dorsoventral (DV, orange) axes. Our pipeline's predictions show a strong correlation ($r2=0.823$) with human annotations, unlike the AMBA's ($r2=0.03$).

Since our pipeline produced significantly more accurate alignments than those in the AMBA, we were able to construct gene expression maps at much higher resolution than the original 200 μm^3 volumes provided by the Allen Institute. Using these improved alignments, we generated 25 μm^3 3D volumes for 4,083 genes. Gaps between sections were interpolated, and experiments targeting the same gene were combined into a single averaged volume. This increase in resolution greatly improved visualization of gene expression in fine subregions (figure 2A). Genes with many available experiments such as Cap1 and Cacna1g (top two panels of figure 2A) yielded the most detailed reconstructions. However, even genes with a low number of samples can be reconstructed into high-quality 3D maps using our approach (bottom two panels of figure 2A). Since many datasets also contained Nissl stained sections, these were also registered and combined into a population average Nissl volume, this volume is reported on in more depth in Piluso et al.[12].

To make the resource more accessible, we built an online portal, available at https://neural-systems-at-uio.github.io/spatial_brain_maps/, where users can search for genes of interest and explore their expression patterns. The portal also lets users select atlas regions and rank genes by specificity of gene expression in the chosen brain region, coverage, intensity, or a weighted mix of these metrics (figure 2b, supplementary figure 1). For interactive 3D viewing, we integrated all volumes into the siibra-explorer[13] on the EBRAINS platform, allowing each gene to be visualised alongside an atlas.

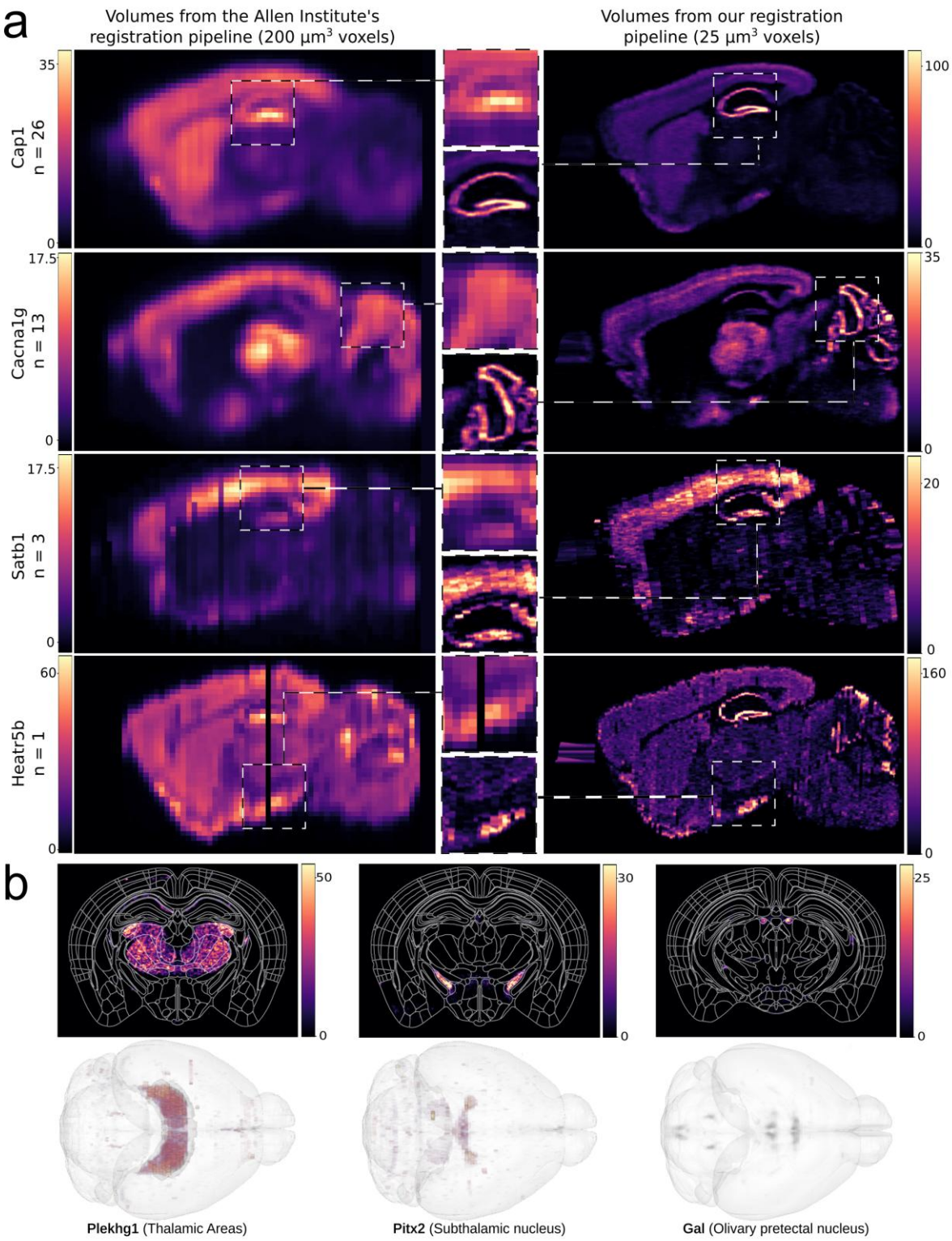


Figure 2

a, Comparison of gene expression volumes from the original Allen Brain Atlas (left, 200 μm^3 voxels) and our pipeline (right, 25 μm^3 voxels). Our high-resolution maps reveal finer anatomical details for genes with varying numbers of experimental samples (n). From top to bottom: Cap1 (n=26), Cacna1g (n=13), Satb1 (n=3), and Heatr5b (n=1). The values on the colour bar represent the range of shown pixel values. We set the max intensity of each plot at 70% of the max value in the volume so as to better display the variance present in each gene. The Allen institute has a smaller range of pixel values since it is lower resolution, greater down sampling has a smoothing effect which removes outlier values. **b**, Examples from the spatial search tool. The top-ranked gene is shown for queries targeting

thalamic areas (Plekhg1), the subthalamic nucleus (Pitx2), and the olivary pretectal nucleus (Gal). The 3D view (left) shows the expression of Plekhg1 within the thalamus (red) in the context of the whole brain. The coronal views show the precise expression patterns of Pitx2 and Gal within their respective target regions (outlined).

For more than a century, brain atlases have been shaped by the methods available for defining neuroanatomical boundaries. Early atlases relied on features visible under the microscope through histochemical staining, later supplemented by functional criteria. Today, high-resolution spatially registered datasets offer a way to automatically discover brain regions from patterns of gene expression. Earlier efforts have used clustering of AMBA data to create regions based on differential gene-expression[14–17], but these attempts were limited by the original 200 μm^3 AMBA resolution. Our volumes overcome this constraint, with 25 μm^3 resolution allowing clustering at a resolution comparable to contemporary 3D atlases[5]. Using our gene-expression volumes we first mirrored one hemisphere onto the other to create a symmetrical dataset. The dimensionality of the data was reduced with Principal Component Analysis (PCA) and K-means clustering was applied, grouping voxels with similar expression profiles. The resulting clusters were then projected back into 3D to generate a fully data-driven atlas.

To select the number of clusters, we tried various numbers of clusters on a subset of voxels and identified an "elbow point" where adding more clusters produced only small gains in performance (supplementary figure 2). This led us to choose 55 clusters for the final atlas. We refer to the final volume as the clustered area ("CArea") atlas and treat each cluster as a region. Many regions in the CArea atlas recapitulate established neuroanatomical boundaries, such as the cortical layers (figure 3b, slices 186, 225, 286, and 379), and the border between the caudate-putamen and accumbens (figure 3b, slice 379). Some of these regions have been difficult to define in traditional murine brain atlases[18], highlighting the value of a method which is able to identify these boundaries more objectively.

While spatial transcriptomic methods now allow the simultaneous imaging of thousands of genes, they are not yet scalable to the large cohorts needed for population-level atlasing[19–21]. Our resource integrates data from 1,078 animals, making the CArea atlas regions broadly representative rather than driven by a few mice. Since the atlas is in 3D, it is compatible with existing volumetric atlas analysis software[22–24]. Given that the number of clusters output from our pipeline is adjustable, it is possible for users to create use-case specific atlases if a particular number of parcellations is required. We consider the CArea atlas to be complementary to existing human-delineated atlases such as the CCFv3.

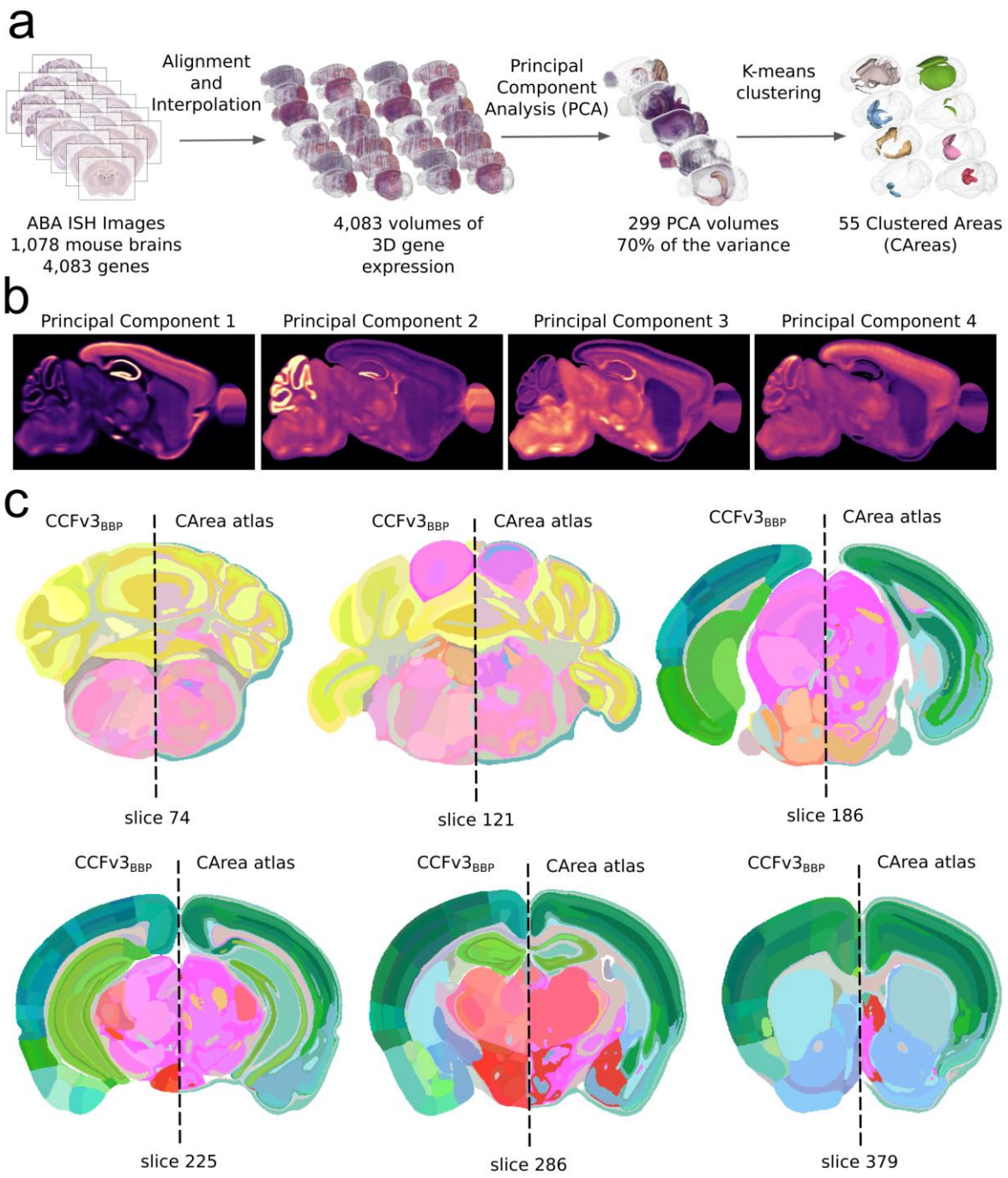


Figure 3
a, Sagittal views of the top 4 principal components after running a voxel-wise iterative PCA on all 4,083 gene volumes. The colour scale represents the principal component (PC) value with bright yellow indicating high positive values and dark purple indicating low or negative values. These components capture the dominant patterns of co-expression across the mouse brain **b,** Side by side comparison of the regions from an established anatomical atlas ($CCFv3_{BBP}$, left of dashed line) and those from our data-driven Clustered Area atlas (CArea atlas, right of dashed line) on representative coronal slices. The $CCFv3_{BBP}$ colours were modified such that each region was uniquely coloured. Each CArea was assigned the colour of the $CCFv3_{BBP}$ region it most overlapped with, if two CAreas were assigned the same colour, one was adjusted such that each CArea was uniquely coloured. Slice numbers refer to the coronal plane of the atlas.

In summary, we present high-resolution 3D gene expression volumes for the mouse brain. The volumes enable a population-scale, data-driven anatomical atlas that captures shared molecular signatures while faithfully recapitulating many known anatomical structures. Beyond the core dataset, we provide an intuitive search and visualization interface, ensuring accessibility. The full analysis pipeline is reusable and can be adapted to a wide range of future applications, bridging high-resolution molecular mapping with practical tools for exploration and investigation.

Data sharing statement

- Our online search portal can be accessed at https://neural-systems-at-uio.github.io/spatial_brain_maps/.
- All code for replicating our analysis is available from https://github.com/Neural-Systems-at-UIO/interactive_gene_expression.
- All registrations, gene expression volumes, and clustered atlases are openly accessible via the EBRAINS Knowledge Graph from https://search.kg.ebrains.eu/instances/7f8ef0e2-121a-4892-8a5e-1c7a8b693503.
- The CArea atlas will be released via the brainglobe atlas API (RRID:SCR_023848)[24].

Acknowledgements

We thank Heidi Kleven for her invaluable insights into atlasing terminology and methods, Lydia Ng for early discussions that helped shape the project's direction, and Michael Hawrylycz for valuable comments on the manuscript. We are also grateful to the Allen Institute for Brain Science for sharing their data in a curated and usable form. Finally, we acknowledge Sergio Rivas-Gomez from the Blue Brain Project for his high-performance computing support. This study was supported by the European Union's Research and

Innovation Program Horizon Europe under Grant Agreement no. 101147319 (EBRAINS 2.0), The Research Council of Norway under Grant Agreement no. 333157 (Norwegian INCF Node), the Blue Brain Project, a research centre of the École polytechnique fédérale de Lausanne (EPFL), and the Swiss government's ETH Board of the Swiss Federal Institutes of Technology.

Online Methods

Data acquisition

All ISH and corresponding segmentation data were downloaded via the Allen Institute's API (RRID:SCR_005984) and filtered to include only coronally cut mouse brains which passed quality control and were from postnatal day 56, C57BL/6 mice. We used the $CCFv3_{BBP}$ atlas[9] as the anatomical reference. This atlas is an extended and refined version of the Allen Institute's CCFv3 atlas, incorporating several important improvements. It includes a non-truncated main olfactory bulb, cerebellum, and medulla, thereby providing coverage of the entire mouse brain. It also offers more detailed anatomical annotations, including of the laminar organization within the cerebellum. In addition, the $CCFv3_{BBP}$ provides a 3D Nissl-stained reference volume that is more accurately co-registered with the atlas labels than the Nissl volume provided with the CCFv3.

Data registration

To register the data, we first ran each whole brain dataset through DeepSlice (version 1.2.4), using the angle integration options and providing the section thickness reported by the Allen so as to improve accuracy. From each dataset, 5 sections covering the length of the brain were selected for quality control (most brains contained between 100 and 500 sections) and reviewed to correct errors in angle estimation or anteroposterior positioning. The cross-

section from the $CCFv3_{BBP}$ Nissl reference template corresponding to each recalibrated ISH section was used for non-linear refinements. Using this Nissl section as reference image and the corresponding ISH histological tissue as the moving image, we applied the ANTs registration package and calculated both affine and non-linear transformations.

Interpolation

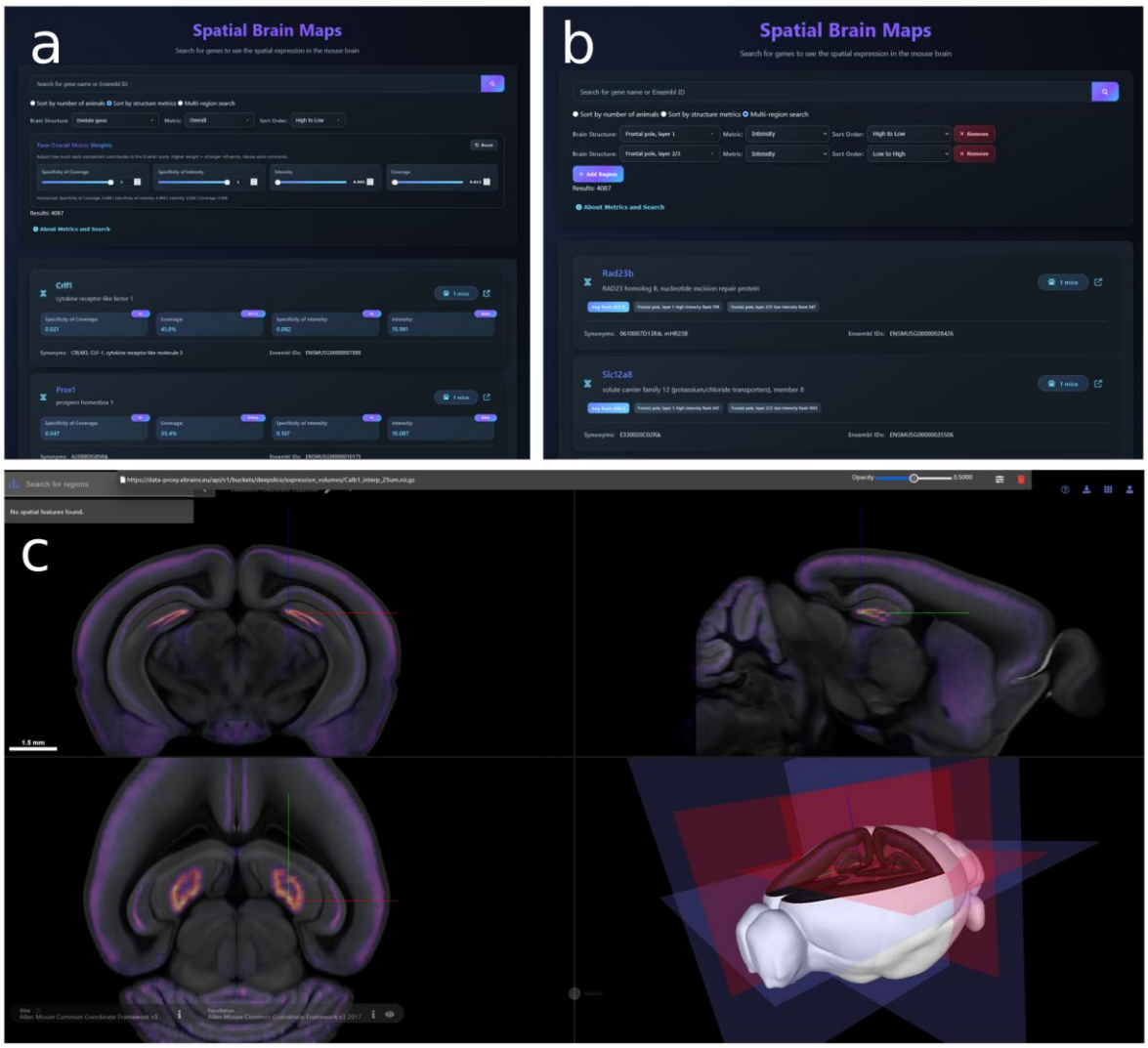
Segmentation images were first downsampled via average pooling to match the resolution of the $CCFv3_{BBP}$ (25 μm^3). We then placed the segmentation data into $CCFv3_{BBP\ space}$ using the position as determined by our registration files. To correct for empty space between the sections we used a K nearest neighbour (KNN) interpolation method, where each voxel was assigned the average value of the five nearest segmentation pixels. To ensure that all voxels in each volume were equivalent and comparable we applied KNN approach to all voxels not just those for which segmentation data was missing, avoiding a scenario where some voxels are an average of 5 and others just single datapoints. This process produced smooth 3D volumes of gene expression at 25 μm^3 resolution.

Validation

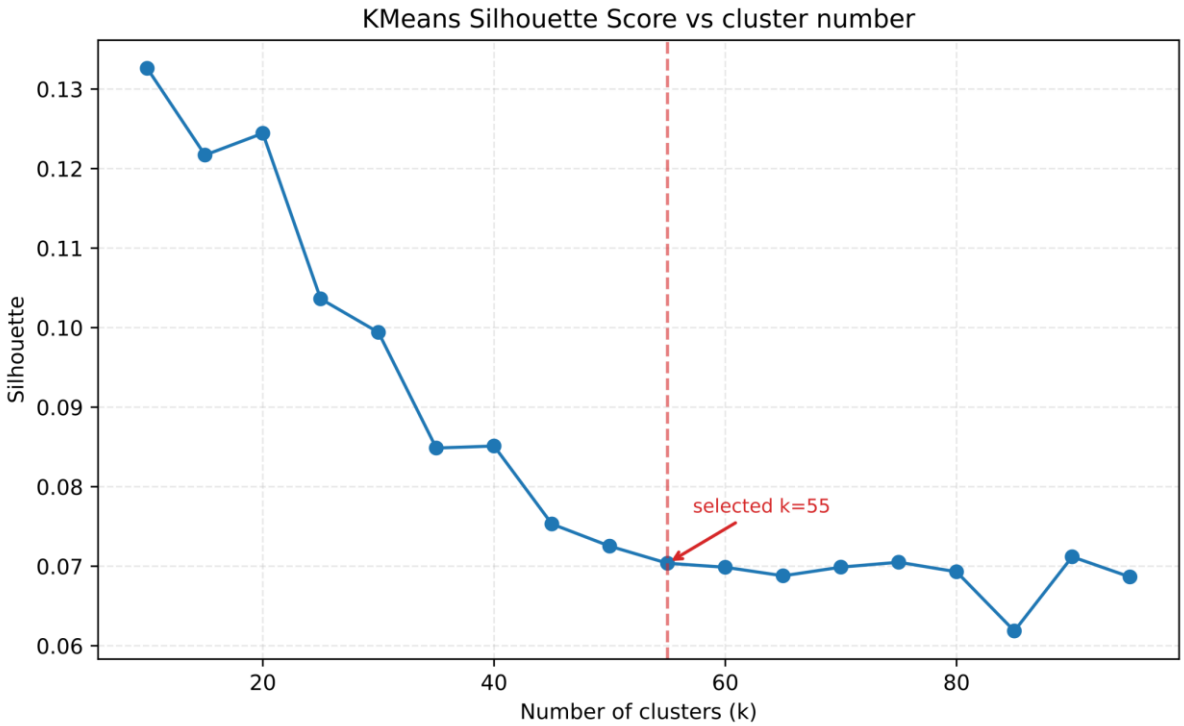
To quantify alignment error for each section, raters positioned sections in atlas space using QuickNII and VisuAlign. The registration generated by each rater was then compared to the mean position derived from all other raters. To obtain an error measure that reflected the full extent of each section rather than only its edges, a 2D grid was projected into the 3D atlas space. Every grid point was transformed according to the rater's placement, and an identical grid was generated using the mean placement of the remaining raters. The Euclidean distance between corresponding points in these two grids (supplementary figure 3), expressed in microns, was used as the section-wise error metric (Figure 1b).

Clustering

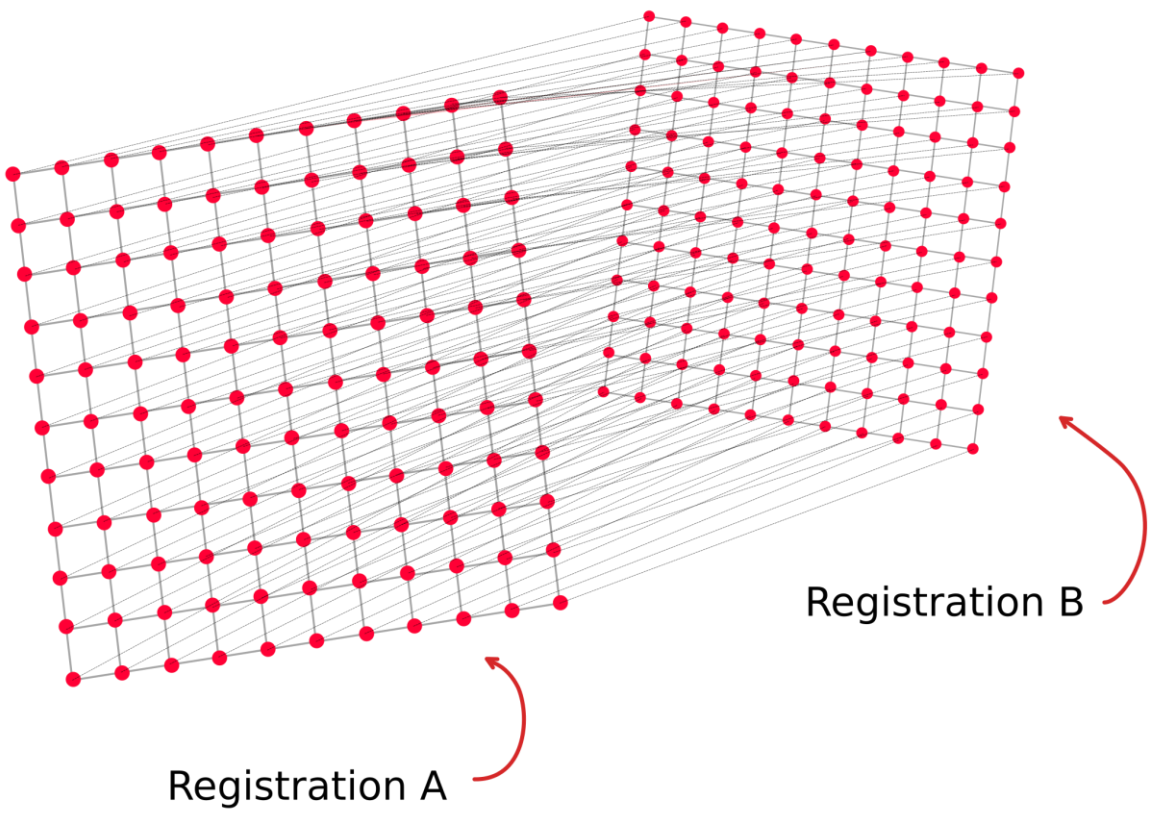
In order to make clustering across all 4,083 genes computationally feasible, we first ran a voxel-wise iterative Principal Component Analysis (PCA) on the gene volumes in order to reduce the dimensionality down from 4,083 genes per voxel to 299 principal components, retaining 70% of the original variance. K-means clustering was then applied to this reduced data, clustering together voxels with similar molecular signatures. We then projected these clusters back into 3D producing atlas like volumes identifying regions of the brain with similar patterns of gene expression. The K-means clustering algorithm allowed us to choose the number of clusters, we therefore tested cluster numbers from 10 to 100 (incrementing by 5) on a 1,000,000 voxel subset (supplementary figure 2). By analysing the change in the silhouette score, we were able to manually identify an elbow point in the graph at 55 clusters. For this reason, we chose the 55 regions for our “CArea" (Clustered-Area) atlas (each cluster functions as a region of the CArea atlas).



Supplementary Figure 1
a, The region based search interface for our dataset. The input allows you to choose a region and then weight the ranking according to an array of metrics. **b,** with a more advanced multi region search functionality users can find genes which are highly expressed in one region but lowly expressed in another. Users can combine as many regions as they wish into their query. **c,** All volumes are viewable in 3D in the browser via the Siibra atlas explorer.



Supplementary Figure 2
Running the k-means algorithm on a subset of the PCA voxels (1,000,000 voxels) helped us identify an elbow point at 55 regions. Due to this we chose 55 as the number of regions in the CArea atlas.



Supplementary Figure 3
To measure the distance between two registrations we project a grid into atlas space, where the corners of the grid

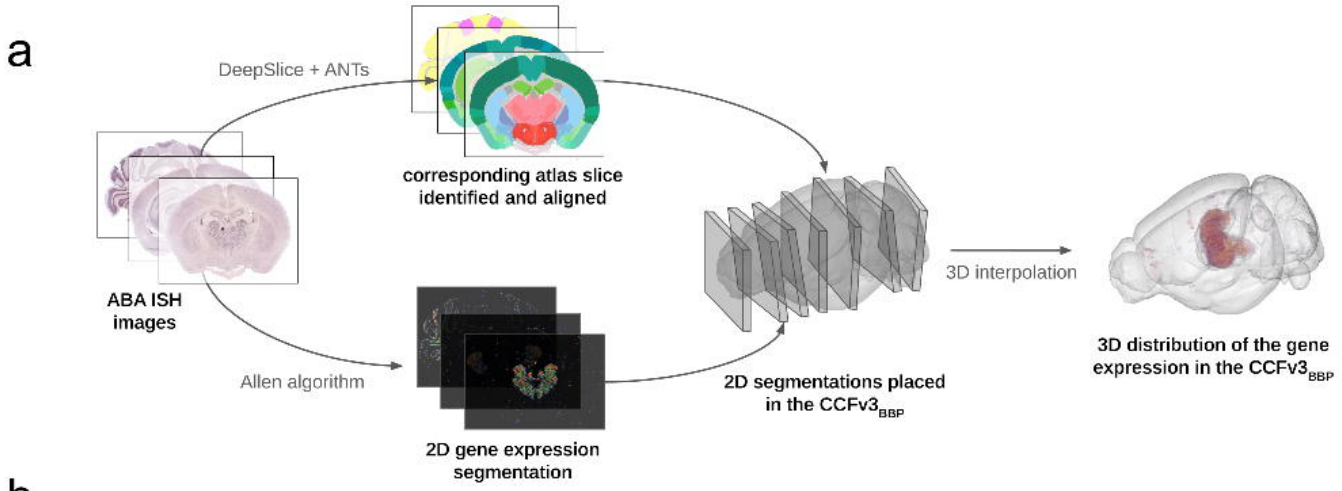
align with the corners of the registration. We then measure the pairwise distance between corresponding points in each grid.

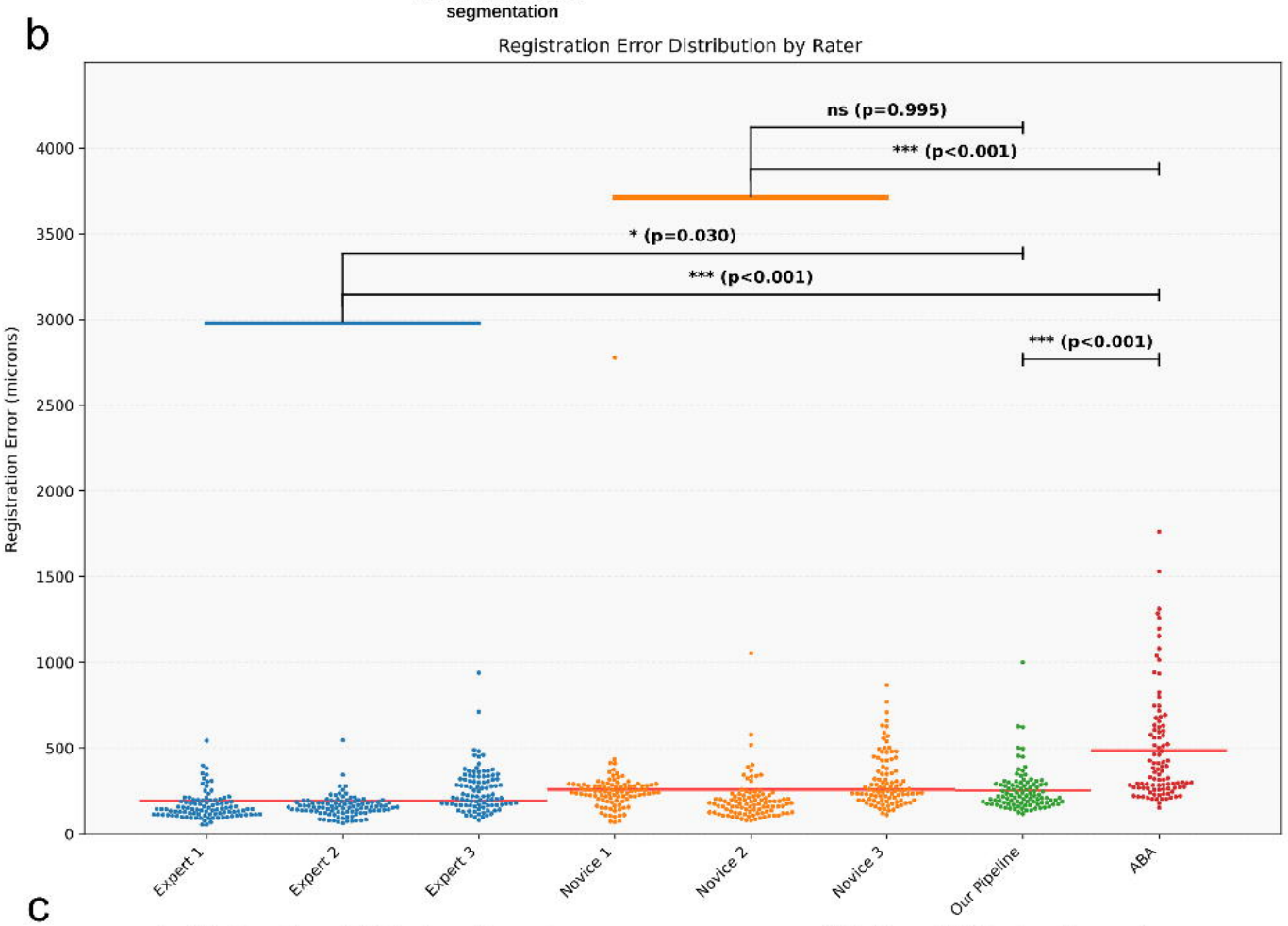
References

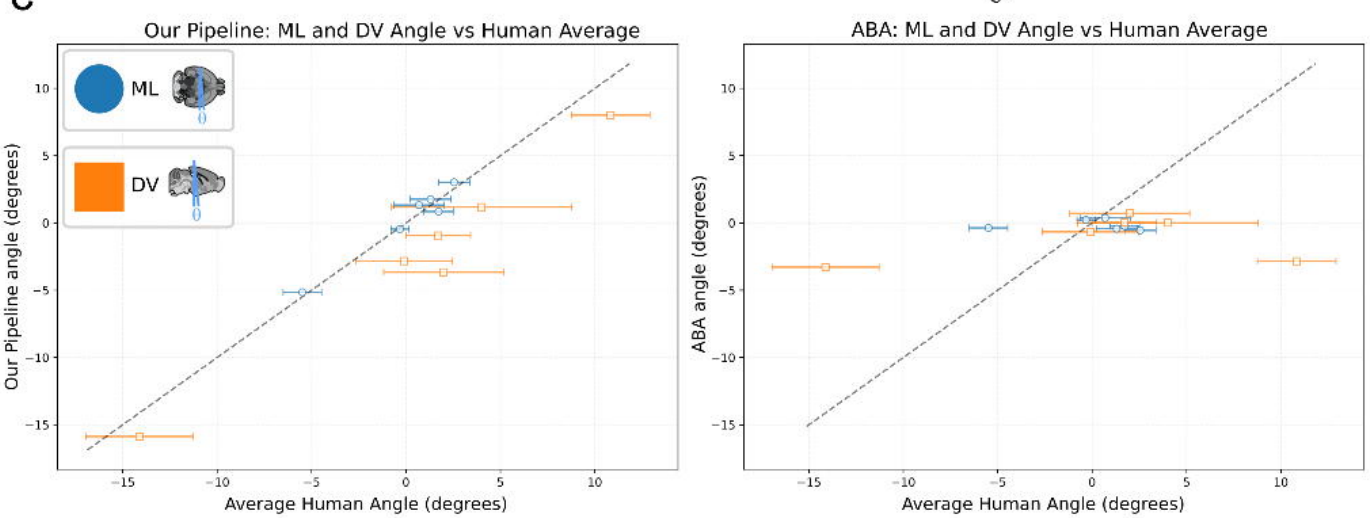
1. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
2. Mahfouz, A., Huisman, S. M. H., Lelieveldt, B. P. F. & Reinders, M. J. T. Brain transcriptome atlases: a computational perspective. *Brain Struct. Funct.* **222**, 1557–1580 (2017).
3. Sunkin, S. M. & Hohmann, J. G. Insights from spatially mapped gene expression in the mouse brain. *Hum. Mol. Genet.* **16**, R209–R219 (2007).
4. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).
5. Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936-953.e20 (2020).
6. Jones, A. R., Overly, C. C. & Sunkin, S. M. The Allen Brain Atlas: 5 years and beyond. *Nat. Rev. Neurosci.* **10**, 821–828 (2009).
7. Ng, L. *et al.* Neuroinformatics for Genome-Wide 3-D Gene Expression Mapping in the Mouse Brain. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **4**, 382–393 (2007).
8. Carey, H. *et al.* DeepSlice: rapid fully automatic registration of mouse brain imaging to a volumetric atlas. *Nat. Commun.* **14**, 5884 (2023).
9. Tustison, N. J. *et al.* The ANTsX Ecosystem for Mapping the Mouse Brain. 2024.05.01.592056 Preprint at https://doi.org/10.1101/2024.05.01.592056 (2024).
10. 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).

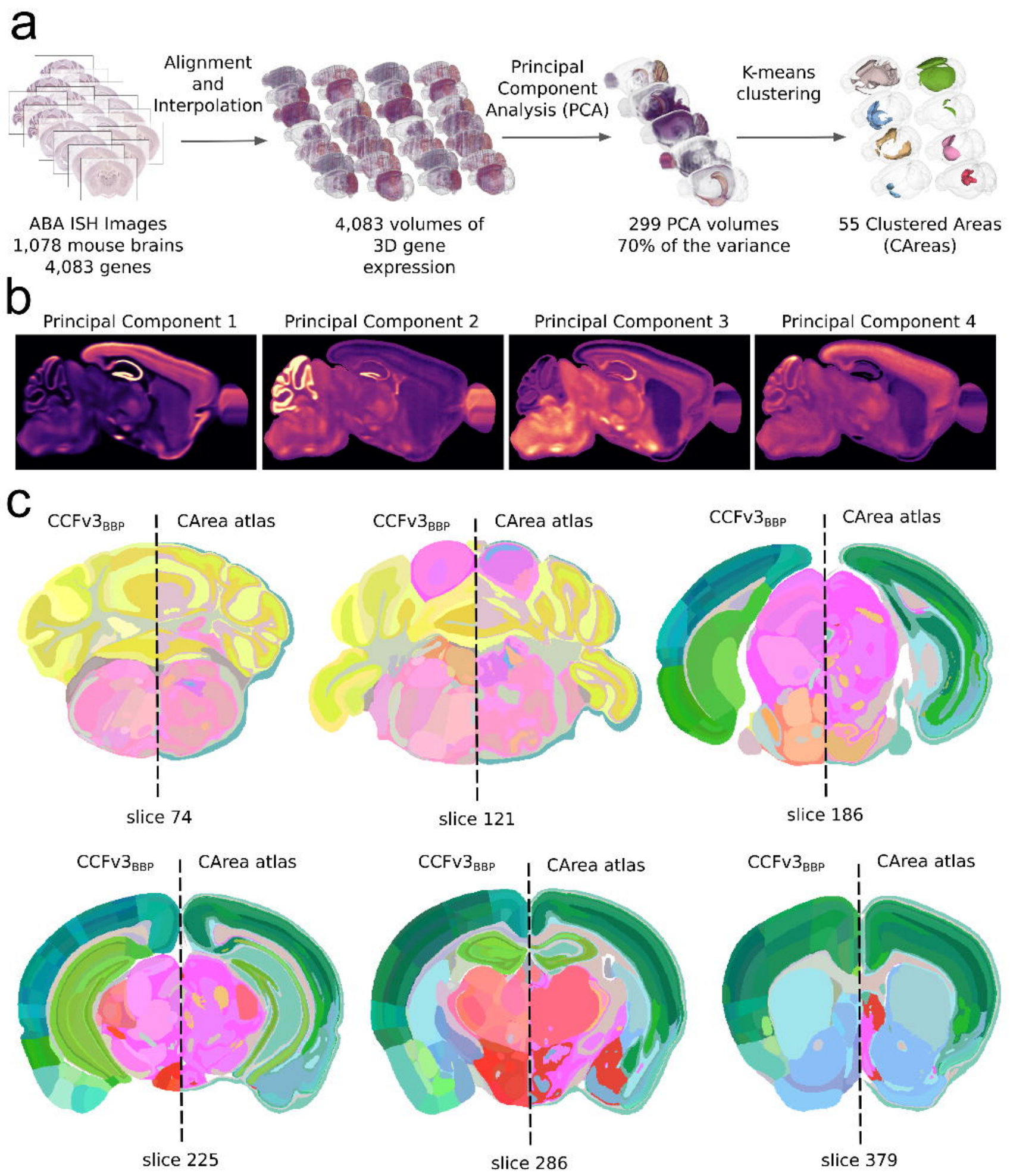
11. Puchades, M. A. *et al.* Software and pipelines for registration and analyses of rodent brain image data in reference atlas space. *Front. Neuroinformatics* **19**, (2025).
12. Piluso, S. *et al.* An extended and improved CCFv3 annotation and Nissl atlas of the entire mouse brain. *Imaging Neurosci.* **3**, imag_a_00565 (2025).
13. Dickscheid, T. *et al.* Siibra: A software tool suite for realizing a Multilevel Human Brain Atlas from complex data resources. Preprint at https://doi.org/10.1101/2025.05.20.655042 (2025).
14. Hawrylycz, M. *et al.* Multi-scale correlation structure of gene expression in the brain. *Neural Netw.* **24**, 933–942 (2011).
15. Bohland, J. W. *et al.* Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy. *Methods* **50**, 105–112 (2010).
16. Cahill, R. *et al.* Unsupervised pattern identification in spatial gene expression atlas reveals mouse brain regions beyond established ontology. *Proc. Natl. Acad. Sci.* **121**, e2319804121 (2024).
17. Ng, L. *et al.* An anatomic gene expression atlas of the adult mouse brain. *Nat. Neurosci.* **12**, 356–362 (2009).
18. Kleven, H., Schlegel, U., Groenewegen, H. J., Leergaard, T. B. & Bjerke, I. E. Comparison of basal ganglia regions across murine brain atlases using metadata models and the Waxholm Space. *Sci. Data* **11**, 1036 (2024).
19. Han, L. *et al.* Single-cell spatial transcriptomic atlas of the whole mouse brain. *Neuron* S0896627325001333 (2025) doi:10.1016/j.neuron.2025.02.015.
20. Zhang, M. *et al.* Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
21. Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).

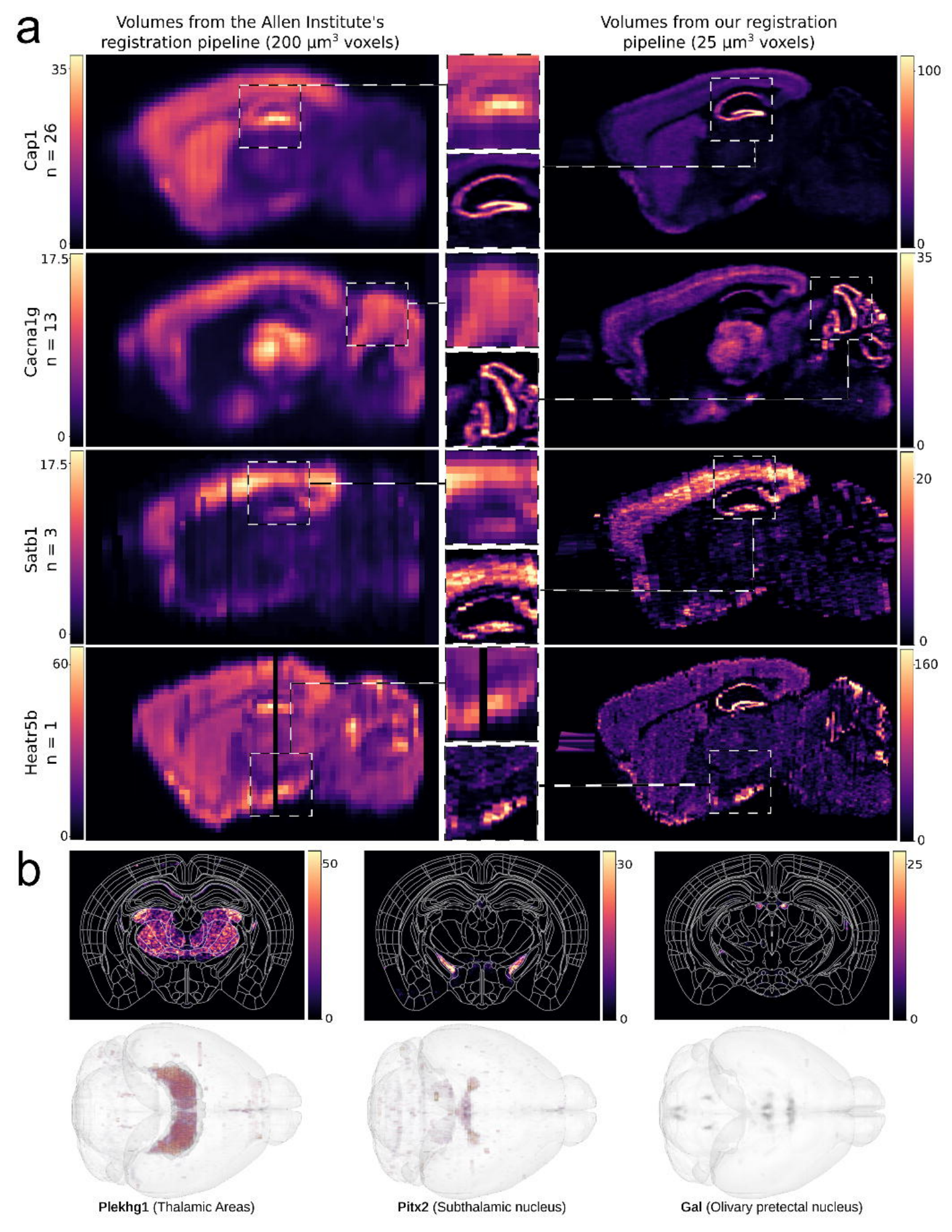
22. Yates, S. C. *et al.* QUINT: Workflow for Quantification and Spatial Analysis of Features in Histological Images From Rodent Brain. *Front. Neuroinformatics* **13**, (2019).
23. Chiaruttini, N. *et al.* ABBA, a novel tool for whole-brain mapping, reveals brain-wide differences in immediate early genes induction following learning. 2024.09.06.611625 Preprint at https://doi.org/10.1101/2024.09.06.611625 (2024).
24. Claudi, F. *et al.* BrainGlobe Atlas API: a common interface for neuroanatomical atlases. *J. Open Source Softw.* **5**, 2668–2668 (2020).

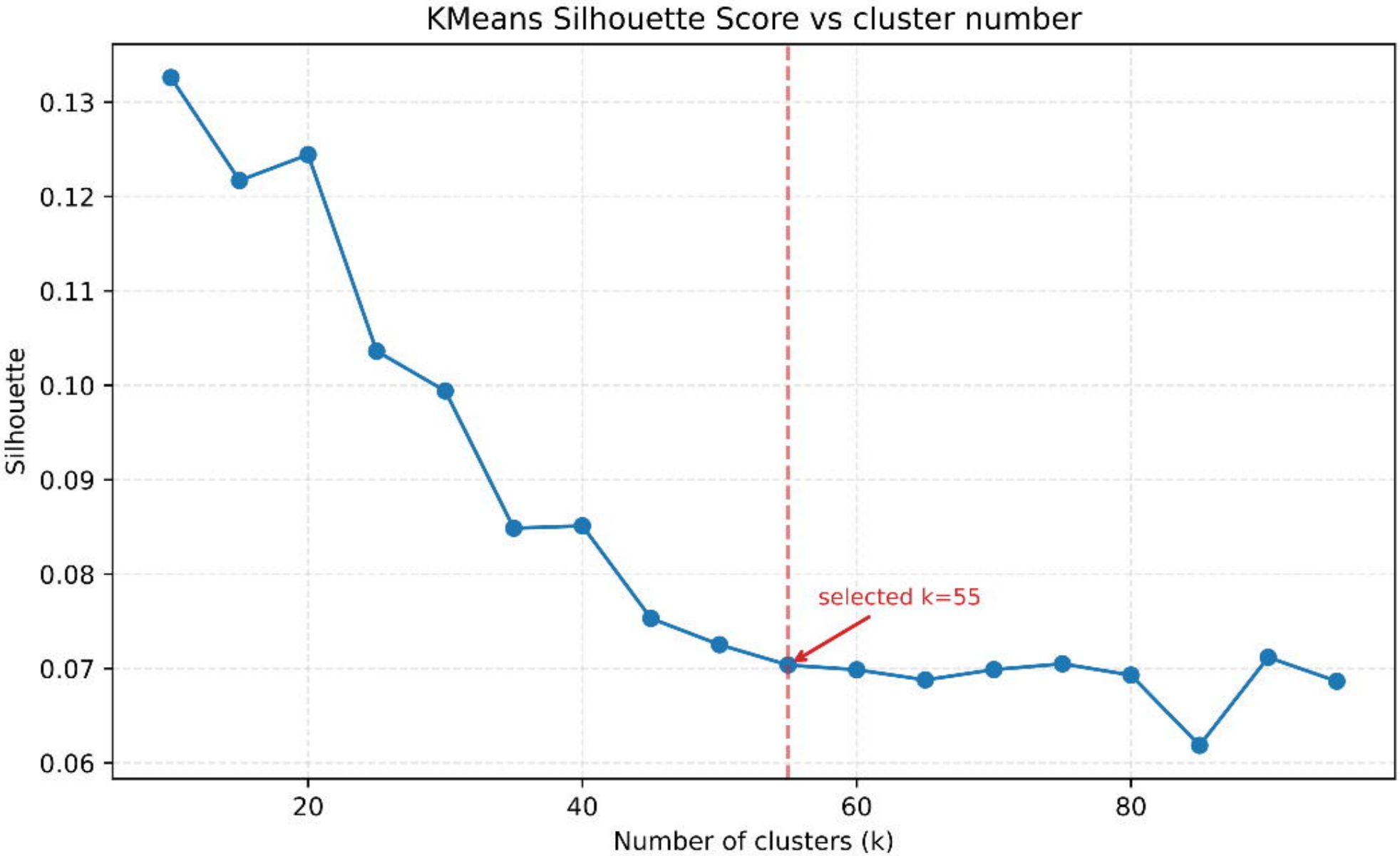

a
DeepSlice + ANTs
corresponding atlas slice identified and aligned
ABA ISH images
Allen algorithm
2D gene expression segmentation
2D segmentations placed in the CCFv3BBP
3D interpolation
3D distribution of the gene expression in the CCFv3BBP

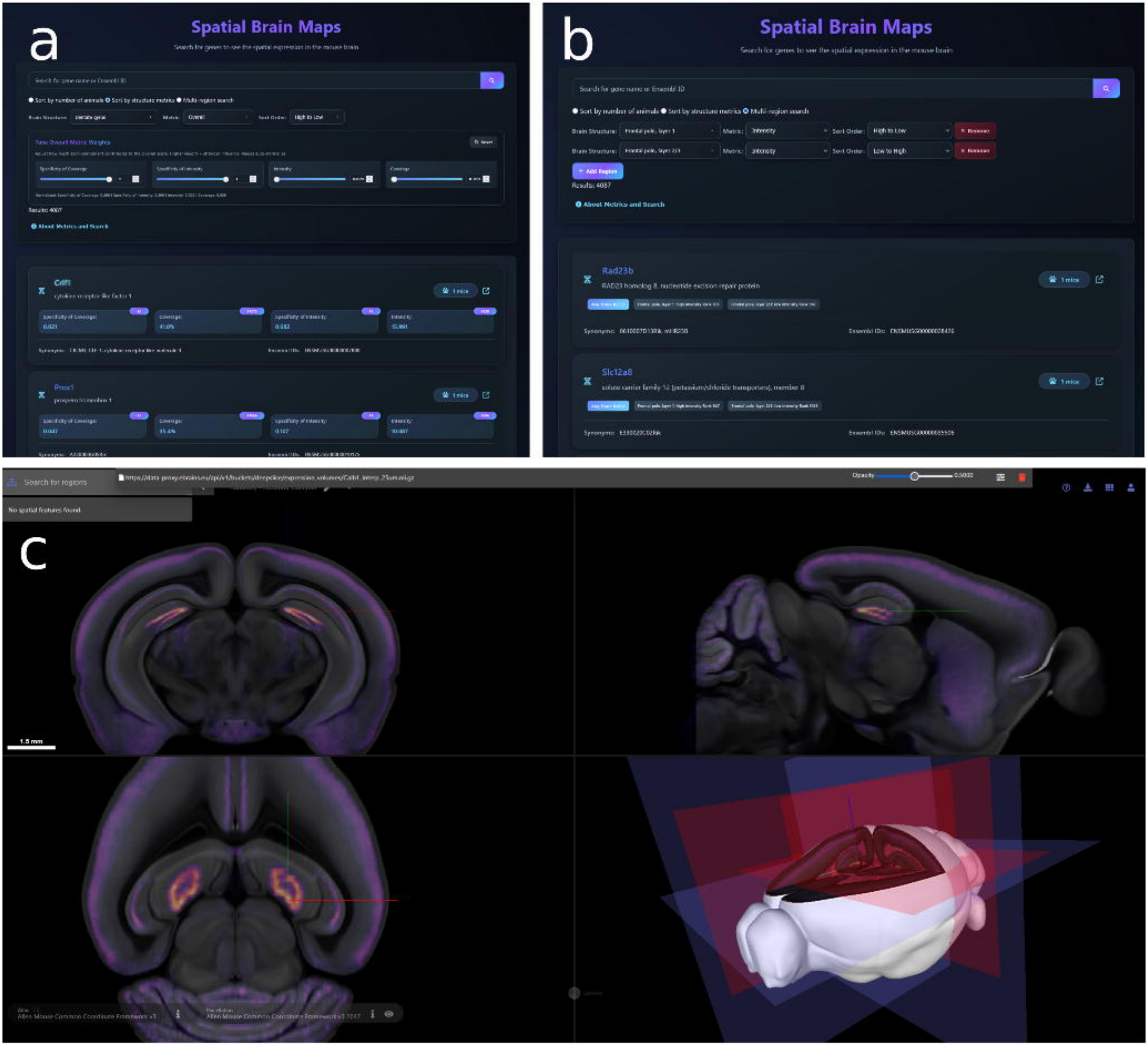

b
Registration Error Distribution by Rater
Registration Error (microns)
ns (p=0.995)
*** (p<0.001)
* (p=0.030)
*** (p<0.001)
*** (p<0.001)
Expert 1
Expert 2
Expert 3
Novice 1
Novice 2
Novice 3
Our Pipeline
ABA

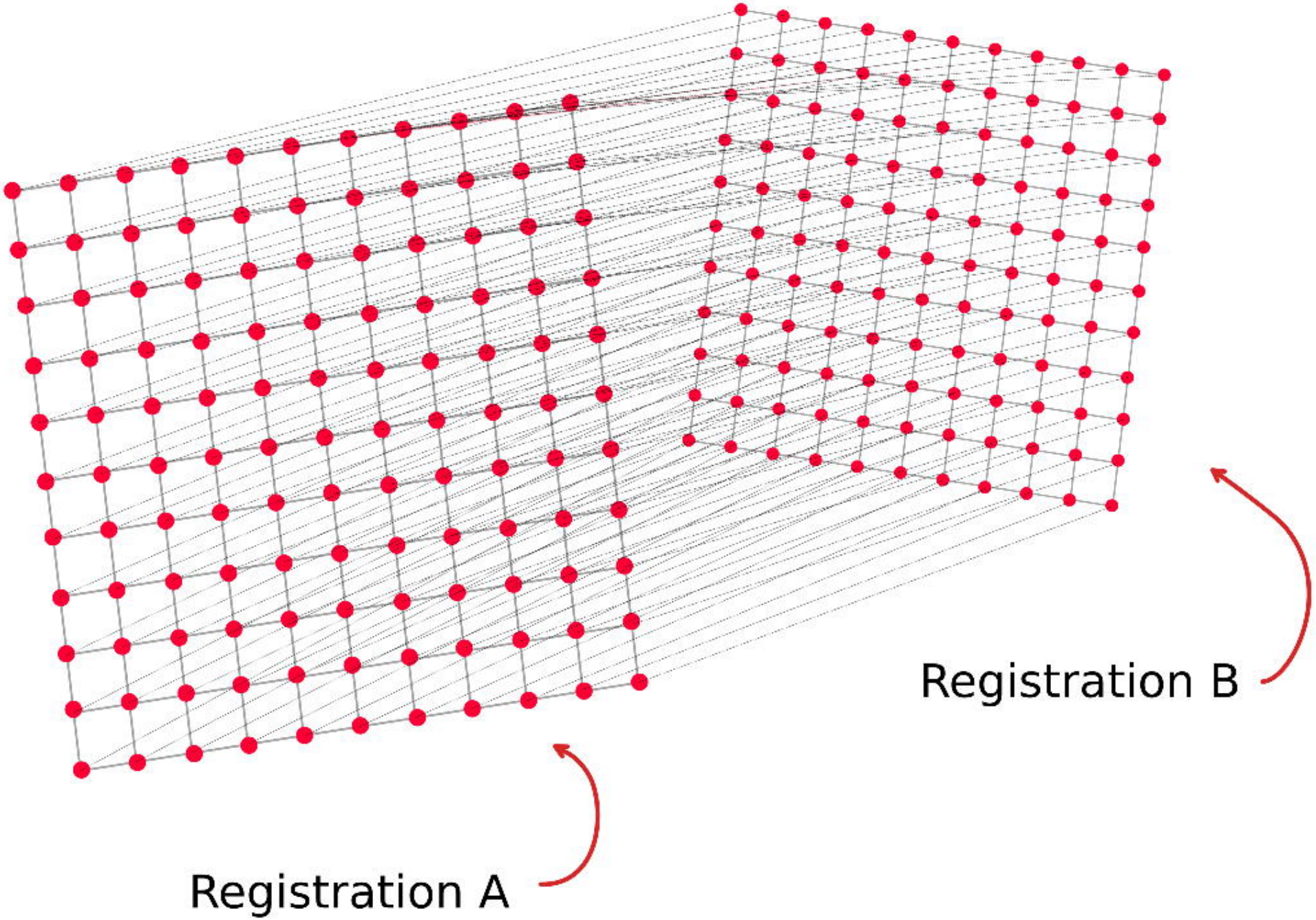

c
Our Pipeline: ML and DV Angle vs Human Average
ML
DV
Our Pipeline angle (degrees)
Average Human Angle (degrees)
ABA: ML and DV Angle vs Human Average
ABA angle (degrees)
Average Human Angle (degrees)


a
Alignment and Interpolation
Principal Component Analysis (PCA)
K-means clustering
ABA ISH Images
1,078 mouse brains
4,083 genes
4,083 volumes of 3D gene expression
299 PCA volumes
70% of the variance
55 Clustered Areas (CAreas)
b
Principal Component 1
Principal Component 2
Principal Component 3
Principal Component 4
c
CCFv3BBP
CArea atlas
slice 74
slice 121
slice 186
slice 225
slice 286
slice 379


a
Volumes from the Allen Institute's registration pipeline (200 μm³ voxels)
Volumes from our registration pipeline (25 μm³ voxels)
Cap1 n = 26
Cacna1g n = 13
Satb1 n = 3
Heatr5b n = 1
35
0
100
0
17.5
0
35
0
17.5
0
20
0
60
0
160
0
b
50
0
30
0
25
0
Plekhg1 (Thalamic Areas)
Pitx2 (Subthalamic nucleus)
Gal (Olivary pretectal nucleus)


KMeans Silhouette Score vs cluster number
Silhouette
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
20
40
60
80
Number of clusters (k)
selected k=55


a
Spatial Brain Maps
b
Spatial Brain Maps
Search for genes to see the spatial expression in the mouse brain
Search for gene name or Ensembl ID
Sort by number of animals
Sort by structure metrics
Multi-region search
Brain Structure:
Metric:
Sort Order:
High to Low
Intensity
Low to High
+ Add Region
Results: 4087
About Metrics and Search
Rad23b
RAD23 homolog B, nucleotide excision repair protein
Slc12a8
solute carrier family 12 (potassium/chloride transporters), member 8
Search for regions
No spatial features found
c
1.5 mm
Allen Mouse Common Coordinate Framework v3


Registration B
Registration A