Alignment of spatial transcriptomics data using diffeomorphic metric mapping

Kalen Clifton[1,2,*], Manjari Anant[1,3,*], Osagie K. Aimiuwu[4], Justus M. Kebschull[2], Michael I. Miller[2,5], Daniel Tward[6,7], Jean Fan[1,2]

[1] Center for Computational Biology, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21211

[2] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218,

[3] Department of Neuroscience, Johns Hopkins University, Baltimore, MD 21218

[4] University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

[5] Kavli Neuroscience Discovery Institute, The Johns Hopkins University, Baltimore, MD 21211

[6] Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA 90024

[7] Department of Neurology, University of California Los Angeles, Los Angeles, CA 90024

* These authors contributed equally

Correspondence should be addressed to:

Jean Fan (jeanfan@jhu.edu) and Daniel Tward (dtward@mednet.ucla.edu)

Key words: spatial transcriptomics, spatial alignment, common coordinate framework, LDDMM, diffeomorphic mapping

Abstract

Spatial transcriptomics (ST) technologies enable high throughput gene expression characterization within thin tissue sections. However, comparing spatial observations across sections, samples, and technologies remains challenging. To address this challenge, we developed STalign to align ST datasets in a manner that accounts for partially matched tissue sections and other local non-linear distortions using diffeomorphic metric mapping. We apply STalign to align ST datasets within and across technologies as well as to align ST datasets to a 3D common coordinate framework. We show that STalign achieves high gene expression and cell-type correspondence across matched spatial locations that is significantly improved over manual and landmark-based affine alignments. Applying STalign to align ST datasets of the mouse brain to the 3D common coordinate framework from the Allen Brain Atlas, we highlight how STalign can enable the interrogation of compositional heterogeneity across anatomical structures. STalign is available as an open-source Python toolkit at https://github.com/JEFworks-Lab/STalign and as supplementary software with additional documentation and tutorials available at https://jef.works/STalign.

Introduction

Spatial transcriptomics (ST) technologies have enabled high-throughput, quantitative profiling of gene expression within individual cells and small groups of cells in fixed, thin tissue sections. Comparative analysis of ST datasets at matched spatial locations across tissues, individuals, and samples provides the opportunity to interrogate spatial organizational variation in the context of health and disease. Such comparative analysis is complicated by technical challenges in sample collection such as missing tissue, tears, and distortions induced by sectioning. Other challenges include biological variation and technological variation in ST methodologies. While previous

computational methods have focused on spatial alignment of ST datasets from serial sections of the same tissue sample assayed using the same ST technology[1,2], spatial alignment of datasets across samples and technologies remains challenging. To address this, we present an approach called STalign that builds on recent developments in Large Deformation Diffeomorphic Metric Mapping[3,4] (LDDMM) to align ST datasets using image varifolds. STalign is amenable to data from single-cell resolution ST technologies as well as data from multi-cellular pixel-resolution ST technologies for which a corresponding registered single-cell resolution image such as a histology image is available. STalign is further able to accommodate alignment in both 2D and 3D coordinate systems. STalign is available as an open-source Python toolkit at https://github.com/JEFworks-Lab/STalign and as supplementary software with additional documentation and tutorials available at https://jef.works/STalign.

Results

Overview of Method

To align two ST datasets, STalign solves a mapping that minimizes the dissimilarity between a source and a target ST dataset subject to regularization penalties (Online Methods). Within single-cell resolution ST technologies, both the source and target ST datasets are represented as cellular positions (x^{ρ_S}, y^{ρ_S}) and (x^{ρ_T}, y^{ρ_T}) respectively (Fig 1a). Solving the mapping with respect to single cells has quadratic complexity and is computationally intractable, so STalign applies a rasterization approach to reduce computational time (Fig 1b). Briefly, STalign models the positions of single cells as a marginal space measure ρ within the varifold measure framework[5]. STalign then convolves the space measure ρ with Gaussian kernels k to obtain the smooth, rasterized function $I(x, y) = \left[k^{\frac{1}{2}} * \rho\right](x, y)$. Finally, STalign samples from the continuous

69 $I(x, y)$ to get a discrete image of a specified size with a specified pixel resolution. STalign focuses
70 on minimizing the dissimilarity between the source and target images I^S and I^T rather than
71 minimizing the dissimilarity between the source and target space measures because, while
72 approximately equivalent, the former can be calculated more efficiently (Online Methods). To
73 solve for a mapping that minimizes the dissimilarity between source and target images I^S and I^T,
74 STalign utilizes the LDDMM framework (Fig 1c). Using LDDMM to identify a diffeomorphic
75 solution allows us to have a smooth, continuous, invertible transformation which permits mapping
76 back and forth from the rasterized image and original cell positions while respecting the biological
77 constraints such that cell neighbor relationships stay relatively the same[6]. The mapping $\phi^{A,v}$ is
78 constructed from two transformations, an affine transformation A and a diffeomorphism φ_1^v such
79 that $\phi^{A,v}(x) = \mathrm{A}\varphi_1^v(x)$, where φ_1^v is generated by integrating a time varying velocity field v_t over
80 time and A acts on $\varphi_1^v(x)$ through matrix vector multiplication in homogeneous coordinates. The
81 optimal $\phi^{A,v}$ is computed by minimizing an objective function that is the sum of a regularization
82 term, $R(v)$ and a matching term, $M_\theta(\phi^{A,v} \cdot I^S, I^T)$. The regularization term controls spatial
83 smoothness. In this term, we optimize over $v_t,\ t \in [0,1]$ noting that if v_t is constricted to being
84 a smooth function, the φ_1^v constructed from v_t is guaranteed to be diffeomorphic. The matching
85 term incorporates a Gaussian mixture model $W(x)$ to estimate matching, background, and artifact
86 components of the image to account for missing tissue such as due to partial tissue matches or
87 tears. Additionally, the matching term contains an image contrast function f_θ to account for
88 differences due to variations in cell density and/or imaging modalities. To solve all parameters in
89 each term a steepest gradient descent is performed over a user-specified number of epochs. Once
90 $\phi^{A,v}$ is computed, STalign applies this computed transformation to the source's original cell
91 positions (x^{ρ_S}, y^{ρ_S}) to generate aligned source coordinates $(x^{\rho_{SA}}, y^{\rho_{SA}})$ (Fig 1d).

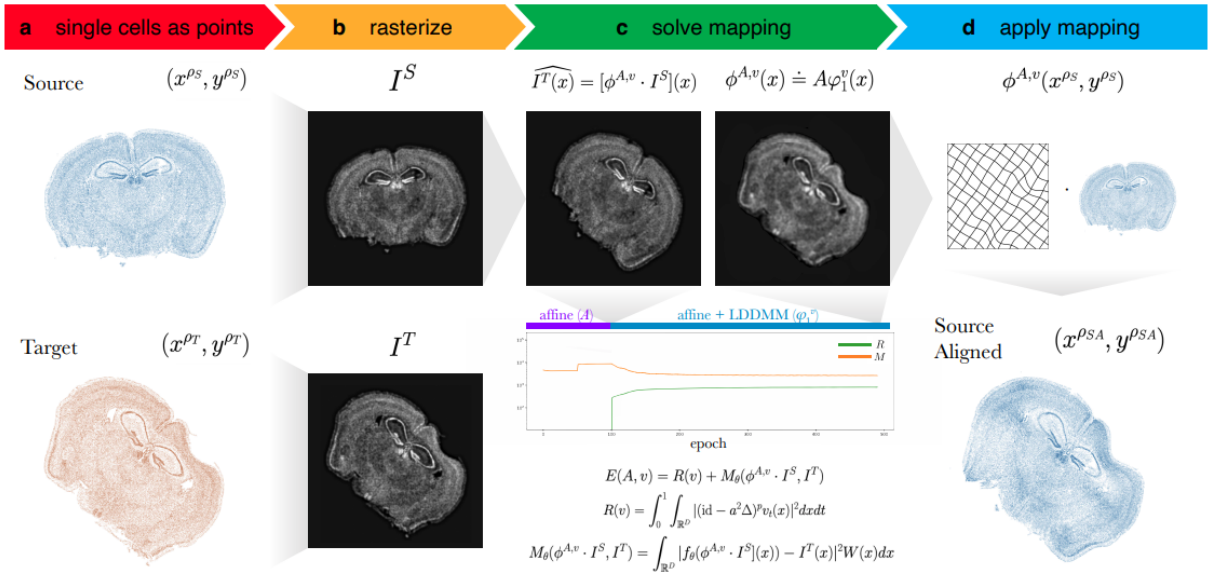


Figure 1. ***Overview of STalign on ST data from a single-cell resolution technology.*** **a.** STalign takes as input a source and target ST dataset as x- and y-coordinates of cellular positions. **b.** Source and target coordinates are then rasterized into images I^S and I^T. **c.** To align I^S and I^T, STalign solves for the mapping $\phi^{A,v}$ that when applied to I^S estimates I^T such that $I^T(x) = [\phi^{A,v} \cdot I^S](x)$. Gradient descent is used to solve affine transformation A and large deformation diffeomorphic metric mapping (LDDMM) φ_1^v that compose $\phi^{A,v}$ such that $\phi^{A,v}(x) = A\varphi_1^v(x)$. The objective function minimized includes a regularization term $R(v)$ to penalize non-smooth solutions and a matching term $M_\theta\left(\phi^{A,v} \cdot I^S, I^T\right)$ that minimizes the dissimilarity between the transformed source image and the target image while accounting for tissue and technical artifacts with $W(x)$ and f_θ, respectively. Components of the objective function decrease over epochs with transforms at different stages of the diffeomorphism. **d.** Once $\phi^{A,v}$ is solved, visualized as a deformation field, the mapping is applied to the coordinates of the source to obtain the coordinates for the aligned source.

<u>STalign enables alignment of single-cell resolution ST datasets within technologies</u>

107 As a proof of concept, we first applied STalign to align two single-cell resolution ST datasets from
108 the same technology. Specifically, we aligned, in a pairwise manner at matched locations, ST data
109 from 9 full coronal slices of the adult mouse brain representing 3 biological replicates spanning 3
110 different locations with respect to bregma assayed by MERFISH (Methods). Inherent local spatial
111 dissimilarities between slices, due to biological variability and further exacerbated by technical
112 variation as well as tears and distortions sustained in the data acquisition process, render affine
113 transformations such as rotations and translations often insufficient for alignment.

114 To evaluate the performance of STalign, we first evaluated the spatial proximity of
115 manually identified structural landmarks between the source and target ST datasets, expecting the
116 landmarks to be closer together after alignment. We manually placed 12 to 13 landmarks that could
117 be reproducibly identified (Supp Fig 1, Supp Table 1). We then compared the positions of the
118 corresponding landmarks after alignment using root-mean-square error (RMSE). When affine
119 transformations alone were used for alignment, RMSE was 202 +/- 17.1 μm, 170 +/- 3.47 μm, and
120 266 +/- 6.65 μm for each location respectively. When STalign based on an LDDMM
121 transformation model was used for alignment, RMSE was 113 +/- 10.5 μm, 169 +/- 4.53 μm, and
122 175 +/- 5.47 μm for each location respectively. STalign was thus able to consistently reduce the
123 RMSE between landmarks after alignment compared to an affine transformation, suggestive of
124 higher alignment accuracy [and consistent with previous performance characterizations of
125 LDDMM alignment versus affine transformations based on landmarks (recommendation of
126 citations from Daniel?)].

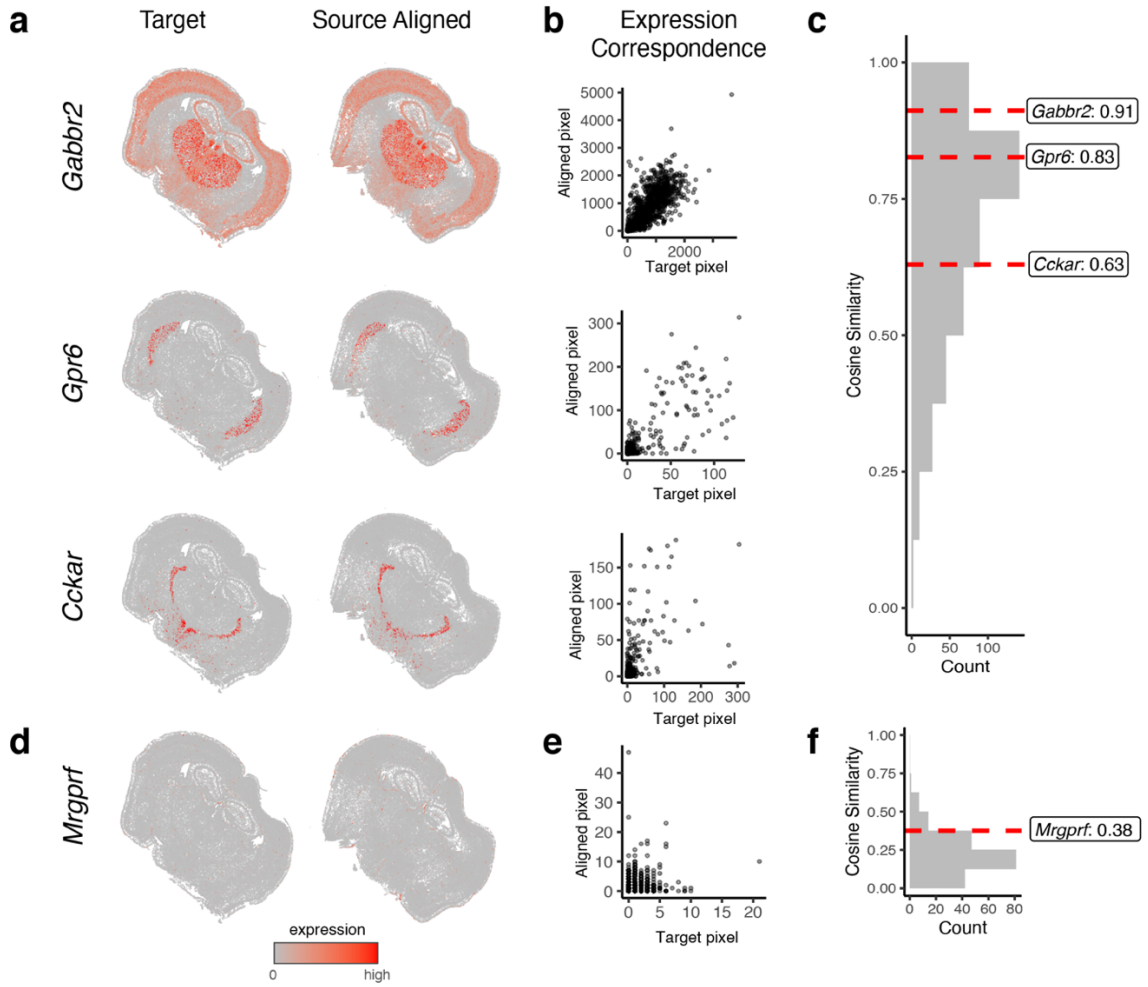
127 Given the ambiguity of where landmarks may be manually reproducibly placed and their
128 inability to evaluate alignment performance for the entire ST dataset, we next took advantage of
129 the available gene expression measurements to further evaluate the performance of STalign.

130 Because of the highly prototypic spatial organization of the brain, we expect high gene expression
131 correspondence across matched spatial locations after alignment. We focused our evaluation on
132 one pair of ST datasets of coronal slices from matched locations (Methods). We visually confirm
133 that alignment results in a high degree of spatial gene expression correspondence (Fig 2a, Supp
134 Fig 2a). To further quantify this spatial gene expression correspondence, we evaluated the gene
135 expression magnitudes at matched spatial locations across the aligned ST datasets. Specifically,
136 we aggregated cells into pixels in a 200μm grid to accommodate the differing numbers of cells
137 across slices and then quantified gene expression magnitude correspondence at spatially matched
138 200μm pixels using cosine similarity (Fig 2b-c, Supp Fig 2b). For a good alignment, we would
139 expect a high cosine similarity approaching 1, particularly for spatially patterned genes. To identify
140 such spatially patterned genes, we applied MERINGUE[7] to identify 457 genes with highly
141 significant spatial autocorrelation (Methods). For these genes, we observe a high spatial
142 correspondence after alignment as captured by the high median cosine similarity of 0.73. In
143 contrast, for the remaining 192 non-spatially patterned genes, we visually confirm as well as
144 quantify the general lack of spatial correspondence (Fig 2d-f, Supp Fig 3a-b). We note that these
145 non-spatially patterned genes are enriched in negative control blanks (57%), which do not encode
146 any specific gene but instead represent noise such that we would not expect spatial correspondence
147 even after alignment. Further, we observe a low median cosine similarity of 0.21 across non-
148 spatially patterned genes that is significantly lower than for spatially patterned genes (Wilcoxon
149 rank-sum test p-value < 2.2e-16).

150 We next compare the alignment achieved with STalign to manual affine transformations
151 (Supp Fig 4a, Methods). We again evaluate performance of the manual affine transformation using
152 a pixel-based cosine similarity quantification. We find that for spatially patterned genes, the cosine

153 similarity is consistently higher with a mean difference of 0.20 for the alignment by STalign
154 compared to manual affine (Supp Fig 4b). In contrast, for non-spatially patterned genes, the cosine
155 similarity is more comparable with a mean difference of 0.04 for the alignment by STalign
156 compared to manual affine (Supp Fig 4c). This greater improvement in spatial gene expression
157 correspondence for the alignment achieved with STalign compared to manual affine
158 transformation for spatially patterned genes suggests that STalign can achieve a higher alignment
159 accuracy.

160



161

162 **Figure 2.** ***Evaluation of STalign based on spatial gene expression correspondence.*** **a.** Correspondence of gene
163 expression spatial organization between the target and aligned source for select spatially patterned genes. **b.** Transcript

164 counts in the target compared to the aligned source at matched pixels for select genes. **c**. Distribution of cosine
165 similarities between transcript counts in target versus aligned source at matched pixels for 457 spatially patterned
166 genes with select genes marked. **d**. Spatial pattern of expression for a select non-spatially patterned gene in the target
167 and aligned source. **e**. Counts for the target versus aligned source at matched pixels for a select non-spatially patterned
168 gene. **f**. Distribution of cosine similarities between counts in target compared to the aligned source at matched pixels
169 for 192 non-spatially patterned genes.
170

171 STalign enables alignment of ST datasets across technologies

172 Many technologies for spatially resolved transcriptomic profiling are available, varying in
173 experimental throughput and spatial resolution[8]. We thus applied STalign to align two ST datasets
174 from two different ST technologies with differing experimental throughput and spatial resolution.
175 Specifically, we applied STalign to align the previously analyzed single-cell resolution ST dataset
176 of a full coronal slice of the adult mouse brain assayed by MERFISH to a multi-cellular pixel
177 resolution ST dataset of an analogous hemi-brain slice assayed by Visium (Fig 3a). As such, in
178 addition to being from different ST technologies, these two ST datasets further represent partially
179 matched tissue sections. For the Visium dataset, we leveraged a corresponding registered single-
180 cell resolution hematoxylin and eosin (H&E) staining image obtained from the same tissue section
181 for the alignment (Methods).

182 To evaluate the performance of this alignment, we again take advantage of the available
183 gene expression measurements. Due to partially matched tissue sections, we restricted downstream
184 comparisons to tissue regions STalign assessed with a matching probability > 0.85. We again
185 visually confirm that the spatial alignment results in a high spatial gene expression correspondence
186 albeit at differing resolutions across the two technologies (Fig 3b, Supp Fig 5a). To further quantify
187 this spatial gene expression correspondence, we evaluated the gene expression magnitudes at

188 matched spatial locations across the aligned tissue sections for the 415 genes with non-zero
189 expression in both ST datasets. We evaluated these genes for spatial autocorrelation on the Visium
190 data to identify 227 spatially patterned genes and 188 non-spatially patterned genes (Methods).
191 Due to the resolution differences between the two technologies, to ensure appropriate comparisons,
192 we used the positions of the Visium spots to aggregate MERFISH cells into matched resolution
193 pseudospots. Likewise, to control for detection efficiency differences between the two
194 technologies, we performed the same counts-per-million normalization on the Visium spot gene
195 expression measurements and the aggregated MERFISH pseudospots gene expression
196 measurements (Fig 3c, Supp Fig 5b). We again evaluated gene expression correspondence at
197 spatially matched spots using cosine similarity and observed a median cosine similarity of 0.55
198 across spatially patterned genes (Fig 3d) and a median cosine similarity of 0.06 across non-
199 spatially patterned genes (Supp Fig 6). We note that this gene expression correspondence after
200 spatial alignment is lower than what was previously observed within technologies most likely due
201 to variation in detection efficiency across technologies in addition to variation in tissue
202 preservation rather than poor spatial alignment. While MERFISH detects targeted genes at high
203 sensitivity, Visium enables untargeted transcriptome-wide profiling though sensitivity for
204 individual genes may be lower[8]. Likewise, while the MERFISH dataset was generated with fresh,
205 frozen tissue, the Visium dataset was generated with FFPE preserved tissue. Still, we anticipate
206 that while sensitivity to specific genes may vary across technologies and with different tissue
207 preservation techniques, the underlying cell-types should be consistent.
208

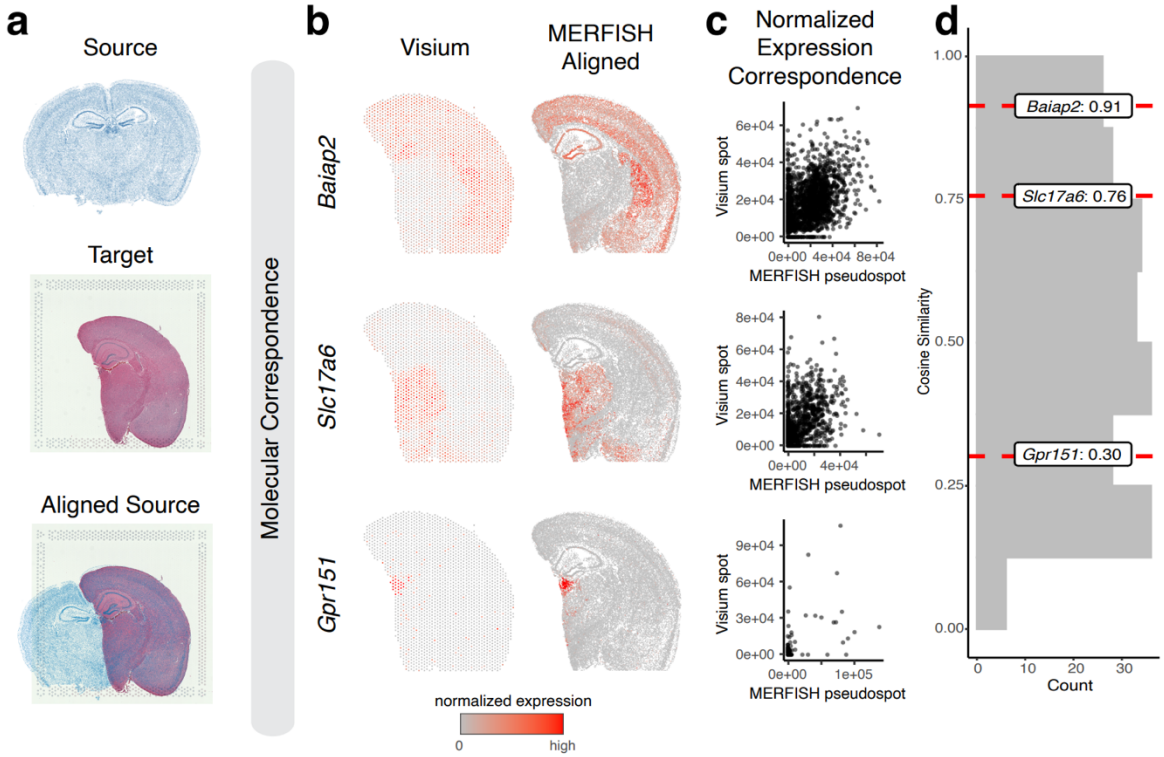
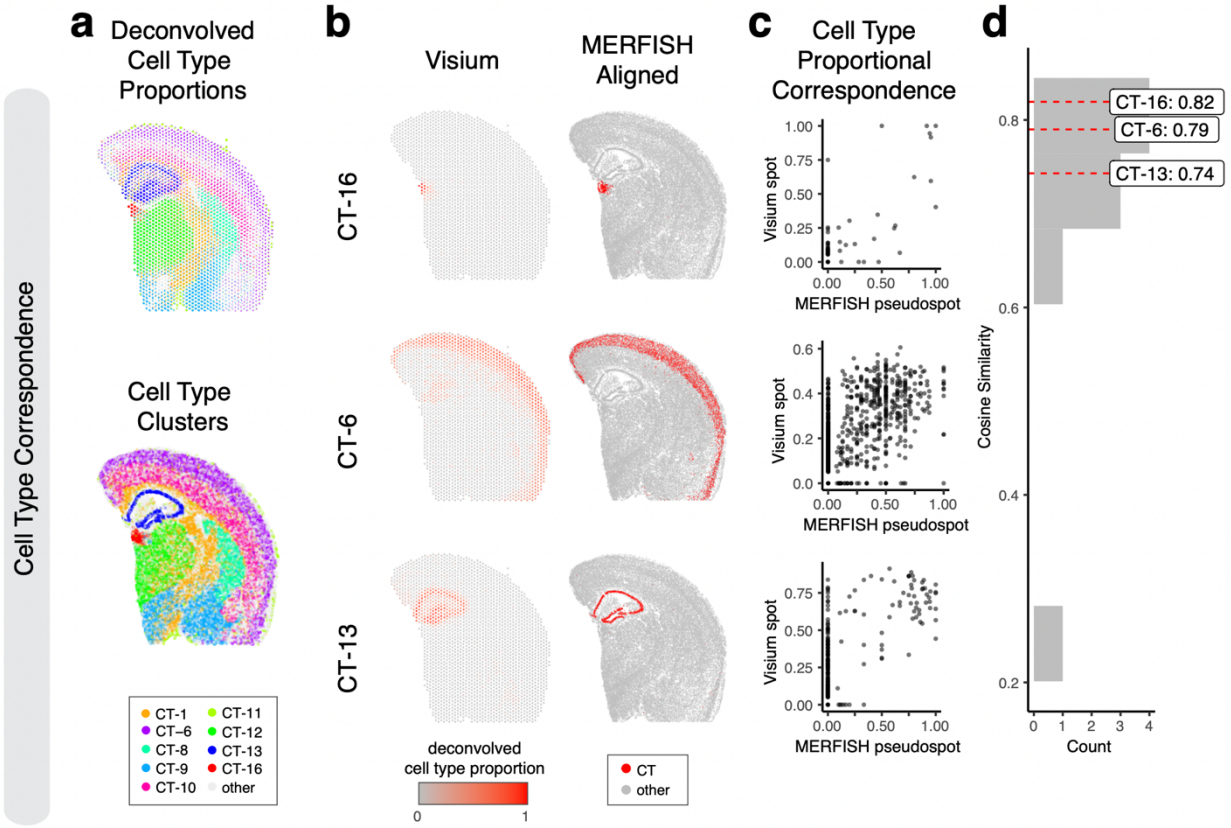


Figure 3. ***Application and evaluation of STalign on spatial transcriptomics data from different ST technologies based on normalized spatial gene expression correspondence.*** **a**. Overview of STalign on ST data from different ST technologies. Single-cell resolution ST is used as the source, with the initial image being produced from the x- and y-coordinates of each cell's position (top). For the multi-cellular resolution ST technologies, the corresponding single-cell resolution histological image is used as target (middle). STalign aligns the source to target (bottom). **b**. Correspondence of gene expression spatial organization between the Visium target and aligned MERFISH source for select spatially patterned genes. **c**. Normalized gene expression in the Visium target compared to the aligned MERFISH source at matched spots and pseudospots respectively for select spatially patterned genes. **d**. Distribution of cosine similarities between normalized gene expression in the Visium target versus aligned MERFISH source at matched spots and pseudospots for 227 spatially patterned genes detected by both ST technologies with select genes marked.

Therefore, we sought to evaluate the performance of our alignment based on cell-type spatial correspondence. To identify putative cell-types, we performed transcriptional clustering analysis on the single-cell resolution MERFISH data (Supp Fig 7a) and deconvolution analysis[9] on the multi-cellular pixel-resolution Visium data (Fig 4a, Methods). We matched cell-types based on transcriptional similarity between cell clusters and deconvolved cell-types (Supp Fig 7b). Indeed, we visually observe high spatial correspondence across matched cell-types (Fig 4a-b). We evaluated the proportional correspondence of cell-types at aligned spot and pseudospot spatial locations by cosine similarity and observed a high median cosine similarity of 0.75 across cell-types (Fig 4c-d). As such, STalign achieves high cell-type spatial correspondence across aligned ST datasets, suggestive of high alignment accuracy.



235 **Figure 4.** ***Evaluation of STalign on ST data from technologies at different resolutions based on cell-type***
236 ***correspondence.*** **a**. Transcriptionally matched cell-types from deconvolution analysis of spot-resolution Visium data
237 (top) and clustering analysis of spatially aligned single-cell-resolution MERFISH data (bottom). **b**. Cell type
238 correspondence between the Visium target and aligned MERFISH source with select cell-types shown. **c**.
239 Correspondence of cell-type proportion between the Visium target and aligned MERFISH source at matched spots
240 and pseudospots respectively for select cell-types. **d**. Distribution of cosine similarities between cell-type proportions
241 in the Visium target and aligned MERFISH source at matched spots and pseudospots respectively for all matched cell-
242 types with cell-types marked.
243
244
245 STalign enables alignment of ST datasets to a 3D common coordinate framework
246 Tissues are inherently 3-dimensional (3D), and tissue sections are subject to distortions in 3D as
247 well as 2D. As such, a more precise spatial alignment of 2D tissue sections must accommodate
248 this 3D distortion. The underlying mathematical framework for STalign is amenable to alignment
249 in 2D as well as 3D (Online Methods). We thus applied STalign to align ST datasets to a 3D
250 common coordinate framework (CCF). Specifically, we applied STalign to align 9 ST datasets of
251 the adult mouse brain assayed by MERFISH to a 50μm resolution 3D adult mouse brain CCF
252 established by the Allen Brain Atlas[10] (Methods, Fig 5a). We note that such a 3D alignment can
253 accommodate deformations in and out of 2D planes (Fig 5b). We separately perform unified
254 transcriptional clustering analysis on these 9 ST datasets to identify transcriptionally distinct cell
255 clusters and annotate them as cell-types based on known differentially expressed marker genes
256 (Methods, Fig 5c-d, Supp Fig 8). By aligning to a CCF like the Allen Brain Atlas that includes an
257 anatomic reference atlas of brain structures, we can lift over these annotations to each cell, enabling
258 further evaluation of variations of gene expression and cell-type composition within and across
259 these annotated brain regions (Fig 5e, Supp Fig 9a).

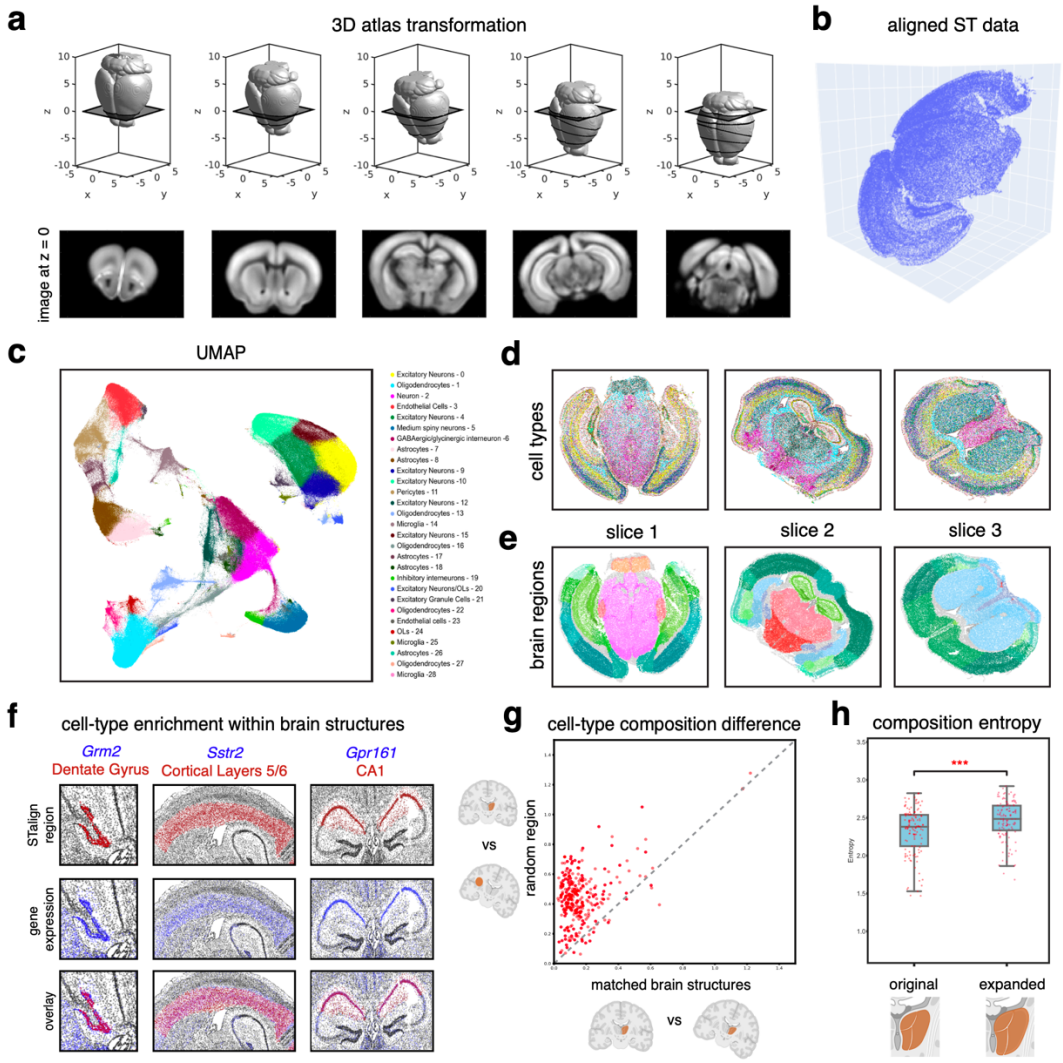


Figure 5. ***Evaluation of 3D-2D alignment using STalign.*** **A.** Transformation of 3D CCF atlas to align to ST data at z=0. **B.** Aligned ST data (MERFISH Slice 1 Replicate 1) plotted in 3D Allen Brain Atlas coordinates. **C.** UMAP embedding of different cell types defined by differential gene expression and Leiden clustering. **D.** Spatial location of cell types from (c) on MERFISH brain slices. **E.** Lift-over brain regions from aligning to the Allen Brain Atlas CCF with STalign. **F.** Brain regions (top) defined by STalign with expression of expected genes (middle) and overlay (bottom). **G.** Cell-type composition difference between paired brain regions from two MERFISH replicates. The x axis represents cell-type composition difference within matched brain structures defined by STalign across replicates and the y axis represents cell-type composition difference between STalign-defined regions and size-matched random brain regions. **h.** Significant difference between distribution of cell-type composition entropy for brain regions defined by STalign versus regions expanded by 100 nearest neighbors.

271

272 To assess the performance of our atlas alignment and lift-over annotations, we first
273 confirmed the enrichment of genes within certain brain regions. Consistent with previous studies,
274 we found *Grm2* to be visually primarily enriched in the dentate gyrus[11], *Sstr2* to be enriched in
275 cerebral cortical layers 5 and 6[12], and *Gpr161* to be enriched in the CA1[13] (Fig 5f), which was
276 consistent across replicates (Supp Fig 9b).

277 Next, we took a more agnostic approach to assess the performance of our atlas alignment
278 and lift-over annotations by evaluating the consistency of cell-type compositional heterogeneity
279 within brain structures across replicates. Consistent with previous studies[14], we observed cell-types
280 to be spatially and compositionally variable across brain regions. We visually confirmed that this
281 spatial and compositional variability is consistent across replicates (Supp Fig 8,9). To further
282 quantify this consistency, for each brain region, we evaluated whether its cell-type composition
283 was more similar between replicates than compared to a random brain region of matched size
284 (Methods). For an accurate atlas alignment, we would expect the lift-over brain region annotations
285 to be more similar in cell-type composition across replicates, particularly for brain structures with
286 distinct cell-type compositions, as compared to random brain regions of matched size. Indeed, we
287 found that in 93% of evaluated brain structures (131/141), the cell-type composition was
288 significantly more similar (Paired t-test p-value = 6.805e-121) between replicates than compared
289 to a random brain region of matched size. (Fig 5g). For the 7% (10/141) of brain regions that were
290 less similar across replicates, we found that the number of cells in these brain regions were
291 significantly fewer (Wilcoxon rank-sum test p-value = 0.002) than other brain regions (Supp Fig
292 10a). Notably, 6 of these brain regions had a minimum width of under 50µm, including both

293 compact and long, thin structures (Supp Fig 10b), highlighting potential limitations with respect
294 to alignment accuracy of structures at this given resolution of alignment.

295 Finally, we also sought to assess the performance of our atlas alignment and lift-over
296 annotations by evaluating cell-type compositions within and beyond annotated brain region
297 boundaries (Methods). Specifically, we calculated the entropy of each brain region based on the
298 region's cell-type composition compared to if the boundaries of these regions were expanded (Fig
299 5h). Because brain regions tend to have a distinct cell-type compositions, for an accurate atlas
300 alignment, we would expect the lift-over brain region annotations to exhibit entropies that are
301 comparatively lower than if the boundaries of these regions were expanded, as more cell-types
302 would be incorporated into the region and entropy would increase. We therefore expanded the
303 brain structures lifted over by STalign by 100 nearest neighbors (NN), or approximately 100µm,
304 and evaluated the change in entropy. We performed the same analysis on randomly defined brain
305 regions of matched size, which were expanded by 100 NN to account for increases in entropy due
306 to an incorporation of more cells. We found that the entropies for the expanded regions were
307 significantly higher (paired t-test p-value=8.6e-18) than for the original brain region annotations
308 lifted over by STalign. In contrast, the entropies for the expanded regions were not significantly
309 higher (paired t-test p-value = 0.12) for random brain regions (Supp Fig 11). Taken together, these
310 results demonstrate that STalign can align ST datasets to a 3D CCF to consistently lift over atlas
311 annotations that recapitulate the unique gene expression and cell-type composition within brain
312 regions.

313

314

315 **Discussion**

316 Alignment of ST datasets is a prerequisite step to enable comparisons across samples, subjects,
317 and technologies. Alignment can also enable pooling of measurements across biological replicates
318 to construct consensus ST profiles[1] as well as enable 3D reconstruction by serial registration[15].
319 Here, we presented STalign, which builds on advancements in LDDMM, to perform alignment of
320 ST datasets in a pairwise manner within ST technologies, across ST technologies, as well as to a
321 3D common coordinate system. We have shown that STalign achieves high accuracy based on the
322 spatial proximity of manually identified shared landmarks as well as gene expression and cell-type
323 correspondence at matched spatial locations after alignment. We note that based on these metrics,
324 STalign outperforms affine transformations alone, highlighting the utility of local, non-linear
325 transformations in alignment. STalign can further accommodate partially matched tissue sections,
326 where one tissue section may be a fraction of another. We further apply STalign to align ST
327 datasets to a 3D CCF to enable automated lift-over of CCF annotations such as brain regions in a
328 scalable manner. We confirm that lift-over brain region annotations identify cells that express
329 expected genes for a variety of brain regions. We also show that brain region annotations lifted
330 over by STalign exhibit consistent cell-type compositions across replicates and within boundaries
331 compared to random brain regions matched in size.

332 We anticipate that future applications of STalign to ST data particularly across ST
333 technologies will enable cross-technology comparisons as well as cross-technology integration
334 through spatial alignment. Likewise, with atlasing efforts like The Human BioMolecular Atlas
335 Program and others producing 3D CCFs[16], application of STalign to align ST data to such CCFs
336 to enable automated lift over of atlas annotations will facilitate downstream spatial comparisons.

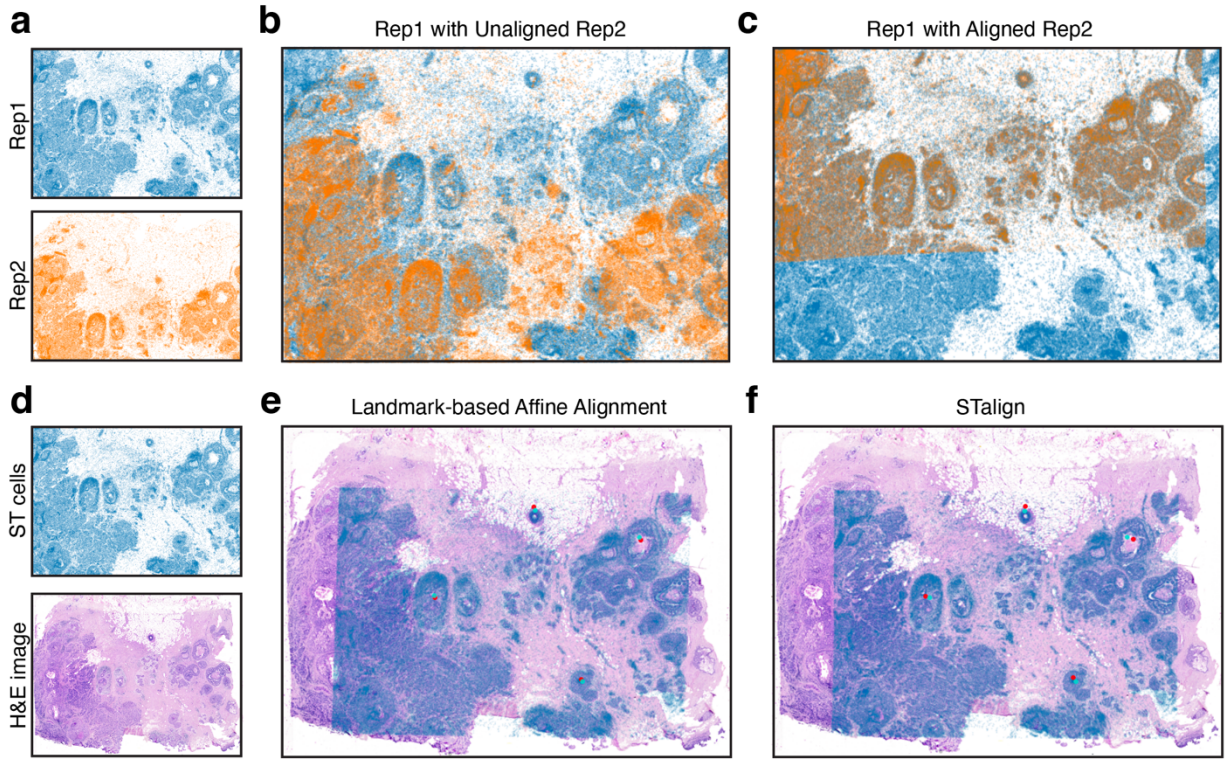
337 Still, among the limitations of STalign with respect to ST data, it is currently applicable to
338 only ST datasets with single-cell resolution or those accompanied with a registered single-cell

339 resolution histology image from same assayed tissue section, which may not be available to all
340 non-single-cell resolution ST technologies. We further assume that the cell position information
341 available in ST data is representative of the tissue's cell density.

342 Further, as STalign is based on an LDDMM transformation model for alignment, it inherits
343 the same limitations. As LDDMM relies on optimization using gradient descent, the resulting
344 alignment solution may converge on local minima. Strategies to guide the optimization away from
345 potential local minima may be applied in the future. Likewise, the more different the source and
346 targets for alignment, particularly for partially matching sections, the more important the
347 initialization will be for this optimization. As such, landmark points may be used to guide the
348 initialization of an orientation and scaling for alignment. In addition, LDDMM enforces an inverse
349 consistency constraint such that every observation in the target must have some correspondence in
350 the source in a manner that cannot accommodate holes or other topological differences in the tissue
351 through the deformation only[6]. As such, when performing alignments, we advise choosing the
352 more complete tissue section as the source because our Gaussian mixture modeling for
353 accommodating partially matched tissues and other artifacts applies to the target image intensity
354 only.

355 Currently, STalign relies on variation in cell densities that generally form visible structures
356 that can be used for alignment. As we have shown, alignment across samples and animals is
357 possible for tissues with highly prototypic structures such as the brain. For other tissues with
358 substantially more inter-sample and inter-animal variation, we anticipate that such alignment will
359 prove more difficult. Although cancer tissues are highly non-prototypic in structure, there is still
360 often sufficient structural consistency across serial sections to enable alignment. As such, we have
361 applied STalign to align single-cell resolution ST datasets arising from partially matched serial

362 sections of the same breast cancer sample assayed by Xenium (Methods, Fig 6a-c). Likewise, we
363 have applied STalign to align a single-cell ST dataset assayed by Xenium to a corresponding H&E
364 image of the same tissue section (Methods, Fig 6d-f). We visually observe a high degree of spatial
365 correspondence and overlap of structural features after alignment. For alignment of other tissues
366 with non-prototypic structures, we anticipate that restricting to structured regions prior to
367 alignment will be needed for meaningful spatial comparisons.
368



369

370 **Figure 6.** ***Application of STalign to ST data of breast cancer.*** **A.** Two single-cell resolution ST datasets from partially
371 matched, serial breast cancer sections visualized as x- and y-coordinates of cellular positions. Overlay of cellular
372 positions before (b) and after (c) alignment with STalign. **D.** A single-cell resolution ST dataset with a corresponding
373 H&E image from the same tissue section. **E.** Overlay of cellular positions and H&E image based on affine
374 transformation by minimizing distances between manually placed landmarks, shown as points in red and turquoise.
375 **F.** Overlay of cellular positions and H&E image after alignment with STalign.

Still, alignment accuracy at the resolution of single cells is limited by the fact that there is generally no one-to-one correspondence between cells across samples, particular for complex tissues. As such, accuracy can typically only be expected to be achieved up to a "mesoscopic scale" at which it is reasonable to define cell density[17]. As we have shown, this presents challenges particularly in aligning thin structures. While STalign currently uses an isotropic (Gaussian) kernel to estimate cell densities, future work considering non-isotropic kernels may improve accuracy for these thin structures. However, generally, our choice of kernel will inherently bias our alignment towards accuracy at a certain structural scale. Likewise, although we focused here on aligning based on cell densities, STalign and the underlying LDDMM framework can also be applied to align based on cellular features such as gene expression magnitude and cell-type annotations, which may improve the accuracy of alignment for regions with homogenous cell density but heterogeneous gene expression and cell-type composition. Integration of such features in the alignment process necessitates orthogonal means of performance evaluation beyond the correspondences in gene expression magnitude and cell-type proportions that we have used here. By aligning based on cell densities, we do not require shared gene expression quantifications or unified cell-type annotations, potentially enhancing flexibility and providing opportunities for integrating across other data modalities for which spatially resolved single cell resolution information is available such as other spatial omics data in the future.

Overall, we anticipate that moving forward STalign will help provide a unified mathematical framework for ST data alignment to enable integration and downstream analyses requiring spatial alignment to reveal new insights regarding transcriptomic differences between different tissue structures and across various physiological axes.

Acknowledgments

K.C., M.A., and J.F. are supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35-GM142889 and the National Science Foundation under Grant No. 2047611. M.A. and J.M.K are supported by the National Institute of Drug Abuse under Award Numbers DP1-DA056668. Fig. 5 was created with component graphics licensed from Biorender.com.

Author Contributions

DT led the development of the STalign software and mathematical modeling with input from JF, KC, MA, and MIM. JF, KC, and MA led the application of STalign to various ST datasets with input from DT. KC evaluated the performance of STalign for 2D alignment under the guidance of JF. MA evaluated the performance of STalign for 3D alignment under the guidance of JF and JMK. OKA evaluated the performance of STalign using landmark-based approaches under the guidance of DT. All authors contributed to the writing of the manuscript. All authors approved the final manuscript.

Competing financial interests

MIM is a founder of AnatomyWorks. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies.

The other authors declare that they have no competing financial interests.

References

1. Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nature Methods 2022 19:5* **19**, 567–575 (2022).

2. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 1–7 (2020).

3. Glaunès, J., Qiu, A., Miller, M. I. & Younes, L. Large deformation diffeomorphic metric curve mapping. *Int J Comput Vis* **80**, 317–336 (2008).

4. Beg, M. F., Miller, M. I., Trouvé, A. & Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vis* **61**, 139–157 (2005).

5. Miller, M., Tward, D. & Trouvé, A. Molecular Computational Anatomy: Unifying the Particle to Tissue Continuum via Measure Representations of the Brain. *BME Front* **2022**, (2022).

6. Dupuis, P., Grenander, U. & Miller, M. I. Variational Problems on Flows of Diffeomorphisms for Image Matching. *Appl Math (Irvine)* **56**, 587–600 (1998).

7. Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X. & Fan, J. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res* **31**, 1843–1855 (2021).

8. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nature Methods 2022 19:5* **19**, 534–546 (2022).

9. Miller, B. F., Huang, F., Atta, L., Sahoo, A. & Fan, J. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nature Communications 2022 13:1* **13**, 1–13 (2022).

10. Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936-953.e20 (2020).

11. Karimi, S. A., Komaki, A., Salehi, I., Sarihi, A. & Shahidi, S. Role of group II metabotropic glutamate receptors (mGluR2/3) blockade on long-term potentiation in the dentate gyrus region of hippocampus in rats fed with high-fat diet. *Neurochem Res* **40**, 811–817 (2015).

12. Stumm, R. K. *et al.* Somatostatin Receptor 2 Is Activated in Cortical Neurons and Contributes to Neurodegeneration after Focal Ischemia. *The Journal of Neuroscience* **24**, 11404 (2004).

13. Stubbs, T., Bingman, J. I., Besse, J. & Mykytyn, K. Ciliary signaling proteins are mislocalized in the brains of Bardet-Biedl syndrome 1-null mice. *Front Cell Dev Biol* **10**, 2491 (2023).

14. Song, L. *et al.* STAB: a spatio-temporal cell atlas of the human brain. *Nucleic Acids Res* **49**, D1029 (2021).

15. Kiemen, A. L. *et al.* CODA: quantitative 3D reconstruction of large tissues at cellular resolution. *Nat Methods* **19**, 1490–1499 (2022).

16. Börner, K. *et al.* Construction and Usage of a Human Body Common Coordinate Framework Comprising Clinical, Semantic, and Spatial Ontologies. (2020).

17. Mitra, P. P. The Circuit Architecture of Whole Brains at the Mesoscopic Scale. *Neuron* **83**, 1273–1283 (2014).

18. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 1–5 (2018).

19. Tward, D. J. An Optical Flow Based Left-Invariant Metric for Natural Gradient Descent in Affine Image Registration. *Front Appl Math Stat* **7**, 61 (2021).

20. Amari, S. & Douglas, S. C. Why natural gradient? *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* **2**, 1213–1216 (1998).

21. Miller, M. I., Trouvé, A. & Younes, L. Geodesic shooting for computational anatomy. *J Math Imaging Vis* **24**, 209–228 (2006).

Methods

Datasets

Nine MERFISH datasets consisting of 734,696 cells and 483 total genes, across 9 brain slices (3 replicates of 3 coronal sections from matched locations with respect to bregma) were obtained from the Vizgen website for *MERFISH Mouse Brain Receptor Map data release* (https://info.vizgen.com/mouse-brain-map).

A Visium dataset of an FFPE preserved adult mouse brain were obtained from the 10X Datasets website for *Spatial Gene Expression Dataset by Space Ranger 1.3.0* (https://www.10xgenomics.com/resources/datasets/adult-mouse-brain-ffpe-1-standard-1-3-0)

Two Xenium datasets (In Situ Replicate 1 and In Situ Replicate 2) of a single breast cancer FFPE tissue block were obtained from the 10X Datasets website for *High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue* (https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast)

The 50μm resolution 3D adult mouse brain CCF was obtained from the Allen Brain Atlas website (https://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/annotation/ccf_2017/annotation_50.nrrd).

Application of STalign

498 To align MERFISH datasets, we applied STalign in a pairwise manner across replicates for
499 sections from matched locations with respect to bregma, rasterized at a 50μm resolution, and
500 iterated over 1000 epochs, with the following changes to default parameters (sigmaM: 0.2).
501
502 To align a MERFISH dataset to a Visium dataset, we applied STalign with MERFISH Slice 2
503 Replicate 3, rasterized at a 50μm resolution, as the source and the high resolution Visium
504 hematoxylin and eosin (H&E) staining image as the target.
505
506 To align MERFISH to the Allen CCF, we applied STalign using the 3D reconstructed Nissl image
507 from the Allen CCF atlas as a source, and each of our 9 MERFISH images as a target.
508
509 To align Xenium datasets, we applied STalign with Xenium Breast Cancer Replicate 1 as the
510 source and with Xenium Breast Cancer Replicate 2 as the target, rasterized at 30μm resolution.
511 We placed a set of 3 manually chosen landmark points to computer an initial affine transformation.
512 We iterated for 200 epochs with the following changes to default parameters (sigmaM:1.5,
513 sigmaB:1.0, sigmaA:1.5, epV: 100).
514
515 To align Xenium to H&E, we applied STalign with Xenium Breast Cancer Replicate 1, rasterized
516 at 30μm resolution, as the source and the corresponding H&E image from the same tissue as the
517 target. We placed a set of 3 manually chosen landmark points to computer an initial affine
518 transformation. We iterated for 2000 epochs with the following changes to default parameters
519 (sigmaM:0.15, sigmaB:0.10, sigmaA:0.11, epV: 10, muB: black, muA: white) where muB and

520 muA initializes the mixture model for the background and artifact components as corresponding
521 to black and white colors respectively in the target image.
522
523
524 Expression based performance evaluation for STalign-based alignment of single-cell resolution
525 ST datasets within technologies
526 To evaluate the performance of STalign on aligning datasets from the same technologies based on
527 expression correspondence, we focused on the alignment of Slice 2 Replicate 3 and Slice 2
528 Replicate 2 from the MERFISH datasets, with the former as the source and the latter as the target.
529 A grid was created to partition all cells into 200µm square pixels. For each 200µm pixel, the gene
530 expression of cells in the pixel was summed for the aligned source and for the target to get gene
531 expression at 200µm resolution.
532
533 MERINGUE (v1.0) was applied to calculate Moran's I on the 200µm resolution summed gene
534 expression of the target. Genes with an adjusted p-value < 0.05 were identified as significantly
535 spatially patterned genes and genes with an adjusted p-value >= 0.05 were identified as non-
536 significantly spatially patterned genes.
537
538 For each gene, the cosine similarity was calculated between the 200µm resolution summed gene
539 expression counts in the aligned source and the 200µm resolution summed gene expression counts
540 in the target across pixels. A Wilcoxon rank sum test was used to compare the distributions of
541 cosine similarities for spatially patterned and non-significantly spatially patterned genes.
542

543 Comparison to manual alignment of single-cell resolution ST datasets within technologies

544 In addition to alignment by STalign, we performed manual alignment of Slice 2 Replicate 3 and
545 Slice 2 Replicate 2 from the MERFISH datasets, with the former as the source and the latter as the
546 target. The cell positions of source were rotated to the best alignment with the target based on
547 visual inspection, resulting in a rotation of pi/5 (or approximately 64 degrees). The cell positions
548 of the source were then translated to align the lowest x and y cell position values to the target. With
549 the manually aligned source and target, we repeated the expression-based performance evaluation
550 described in section "Expression based performance evaluation for STalign-based alignment of
551 single-cell resolution ST datasets within technologies."

552

553 Evaluation alignment across technologies

554

555 *Expression based performance*

556 Given that the MERFISH tissue section is larger than the Visium, we considered the aligned region
557 to be limited to the MERFISH tissue that had a matching probability > 0.85 and restricted the list
558 of cells in the aligned MERFISH dataset to only those in this region.

559

560 To aggregate the cells in the aligned MERFISH dataset into pseudospots that match with the
561 Visium spots, we calculated the distances between the positions of the MERFISH cells and the
562 positions of the Visium spot centroids. Cells were classified as within the pseudospot that
563 corresponds to the Visium spot if the distance of the cell to the Visium centroid was less than the
564 Visium spot radius. The Visium spot radius information was obtained by multiplying the
565 'spot_diameter_fullres' by the "tissue_hires_scalef" in the Visium scalefactors_json.json file and

566 dividing by 2. For each pseudospot, the gene expression of all cells within the pseudospot was
567 summed.
568
569 For gene expression correspondence analysis, we restricted to the 415 genes that had at least one
570 copy in both the MERFISH and Visium datasets and that were detected in more than one spot in
571 the Visium dataset.
572
573 MERINGUE (v1.0) was applied to calculate Moran's I on the Visium counts-per-million (CPM)
574 normalized counts. Genes with an adjusted p-value < 0.05 were identified as significantly spatially
575 patterned genes and genes with an adjusted p-value >= 0.05 were identified as non-significantly
576 spatially patterned genes.
577
578 CPM normalization and log10 transformation with a pseudocount of 1 were applied on the gene
579 expression of the MERFISH pseudospots and Visium spots. For each gene, the cosine similarity
580 was calculated between the normalized and log-transformed gene expression magnitudes across
581 matched MERFISH pseudospots and Visium spots.
582
583 *Cell-type correspondence performance*
584 To identify cell-types in the Visium data, we applied STdeconvolve on a corpus of 838 genes after
585 filtering out lowly expressed genes (<100 copies), genes present in < 5% of spots and genes present
586 in > 95% of spots and restricting to significantly over-dispersed with alpha =1e-16 in order to
587 obtain a corpus < 1000 genes, resulting in 16 deconvolved cell-types.
588

589 To identify cell-types in the aligned MERFISH data, PCA was performed on the CPM normalized
590 cell by gene matrix. Louvain clustering was performed on a neighborhood graph of cells using the
591 top 30 PCs and 90 nearest neighbors to identify 16 transcriptionally distinct clusters of cells.
592 To match deconvolved cell-types and single-cell clusters, we used the deconvolved cell-type-
593 specific transcriptomic profiles from STdeconvolve and averaged the transcriptional profiles per
594 cluster from single-cell clustering. We restricted to the 257 shared genes, CPM normalized, and
595 correlated the resulting normalized transcriptional profiles using Spearman correlation. We
596 considered a Visium deconvolved cell-type and MERFISH single-cell cluster as a match if they
597 had transcriptional similarity > 0.5.
598
599 For each matched cell-type, we evaluated spatial compositional correspondence using cosine
600 similarity of the cell-types proportional representation across matched MERFISH pseudospots and
601 Visium spots.
602
603 Evaluation of 2D to 3D CCF alignment
604
605 *Unified transcriptional clustering analysis and cell-type annotation*
606 All MERFISH datasets were combined. Transcriptional clustering analysis and cell type
607 annotation was performed using the SCANPY package[18] [version 1.9.1]. Data were normalized to
608 counts per million (scanpy: normalize_total) and log transformed (scanpy: log1p). PCA (scanpy:
609 pca) was computed on the cell by gene matrix. A neighborhood graph of cells using the top 10 PCs
610 and 10 nearest neighbors was created (scanpy: neighbors), and Leiden clustering was performed
611 on this graph (scanpy: leiden) to identify 29 clusters. Differentially expressed genes were extracted

612 from each cluster (scanpy: rank_genes_groups), and cell-types were annotated based on marker
613 genes in each cluster.
614
615 *Annotated brain region composition analysis*
616 To generate random brain regions, a random number generator (random.randint) defined the x, y
617 coordinate of the center of the random region, and the random region was composed of the N
618 closest points to the center, where N is the number of cells in the brain region. A slice/replicate
619 with random regions was constructed for all slice/replicates with STalign defined regions, and the
620 number of cells (N) were the same for STalign and randomly defined regions.
621
622 To compare cell-type compositions, each region was represented by a cell-type vector, which was
623 composed by the proportion of each cell type in the region (29x1 vector). We calculate the
624 Euclidean distance between cell type vectors of the same region across replicates in Slice 2 using
625 the regions defined by STalign. The Euclidean distance was also found across replicates in Slice
626 2 using randomly defined brain regions and STalign brain regions. The Euclidean distances of both
627 groups were compared using a paired t-test. 431 data points for each group were used, comparing
628 replicate 1 to replicate 2, replicate 2 to replicate 3, and replicate 3 to replicate 1.
629
630 To evaluate annotated brain region boundaries, brain regions were expanded using k-nearest
631 neighbors (k=50, 100, 200) using the 'ball tree' algorithm for each region and each replicate in
632 Slice 2 (sklearn.neighbors.NearestNeighbors). The procedure was conducted for STalign defined
633 brain regions and randomly defined brain regions. Shannon's entropy was evaluated for STalign
634 defined and randomly defined brain regions that were expanded by 0, 50, 100 and 200 nearest

635 neighbors. Paired t-tests were used to compute p-values between original and expanded brain
636 regions for each STalign and random groups. Effect size was computed as a difference in the means
637 of the compared distributions. PP plots were used to visualize normality, and we used a Gaussian
638 fit with R>0.8 and a variance ratio less than 4 to confirm normality and equal variances. 431 data
639 points for each group were used.
640
641 To evaluate regions that had a greater Euclidean distance between two STalign regions compared
642 to random versus STalign regions, we calculated the number of cells and Shannon's entropy of
643 each region and tested for significance using a Wilcoxon Rank Sum test due to the small sample
644 size. Shannon's entropy was calculated using the formula [–summation(p(x)*logp(x))], where p(x)
645 is the probability of picking cell-type x from the given brain region (scipy.special.entr).
646
647 To analyze the lowest and highest entropic regions, we calculated the Shannon's entropy of all
648 replicates in Slice 2 and plotted regions that had the bottom 10% and top 10% of entropies and
649 plotted these regions on Slice 2 Replicate 3.
650
651 Implementation and software availability
652 STalign is available as an open-source Python toolkit at https://github.com/JEFworks-Lab/STalign
653 and as supplementary software with additional documentation and tutorials available at
654 https://jef.works/STalign.
655
656 The implementation of STalign uses the following parameters and default values.

Symbol	Explanation	Default
σ	Width of rasterization kernel	30 μm

σM	Weight on image matching functional	0.1
σR	Weight on regularization matching functional	5.00E-06
σP	Weight on landmark matching functional	2.00E+01
σA	Variance of artifact component for Gaussian Mixture Modeling	5
σB	Variance of background component for Gaussian Mixture Modeling	2
a	Smoothness scale of diffeomorphism	500.0 μm
p	Power of Laplacian for regularization	2
niter	Number of iterations of gradient descent	5000
diffeo_start	Iteration to start optimizing vt for coarse-to-fine	0
nt	Number of timesteps for integration of vt	3
epL	Gradient descent step size: linear part of A	2.00E-08
epT	Gradient descent step size: translation part of A	2.00E-01
epv	Gradient descent step size: vt	2.00E+03

The PyTorch framework was used for automatic gradient calculations. Based on the PyTorch backend, STalign supports parallelization across multiple cores or on GPUs. Derivatives (covectors) are converted to gradient vectors[4,19] for natural gradient descent[20].

For improved robustness, STalign allows users to input pairs of corresponding points in the source and target images. These points can be used to initialize the affine transformation A through least squares to steer our gradient based solution toward an appropriate local minimum in this challenging nonconvex optimization problem as well as be added to the objective function to drive the optimization problem itself. Landmark based optimization in the LDDMM framework has been studied extensively[21]. A script curve_annotator.py is provided to assist with interactive placement of these points.

670 **Supplementary Tables**

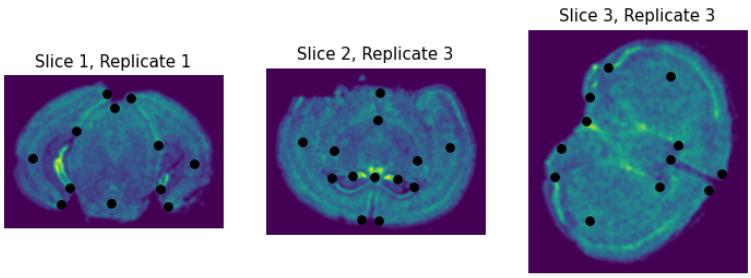
671 **Supplemental Table 1.** Description of manually placed landmark locations on ST datasets of the

672 mouse brain.

673

674

675 **Supplementary Figures**

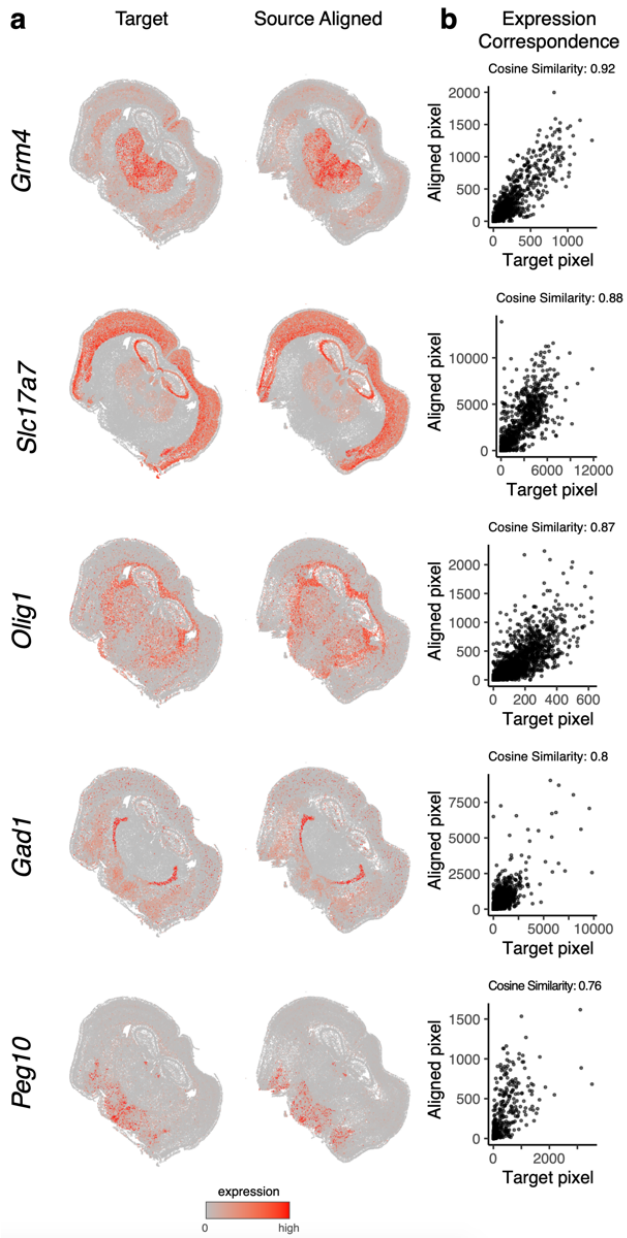


676

677 **Supplemental Figure 1.** Manually placed landmark locations on ST datasets for one

678 representative biological replicate spanning 3 different locations with respect to bregma.

679



680

681 **Supplemental Figure 2.** ***Additional examples of MERFISH to MERFISH alignment for***

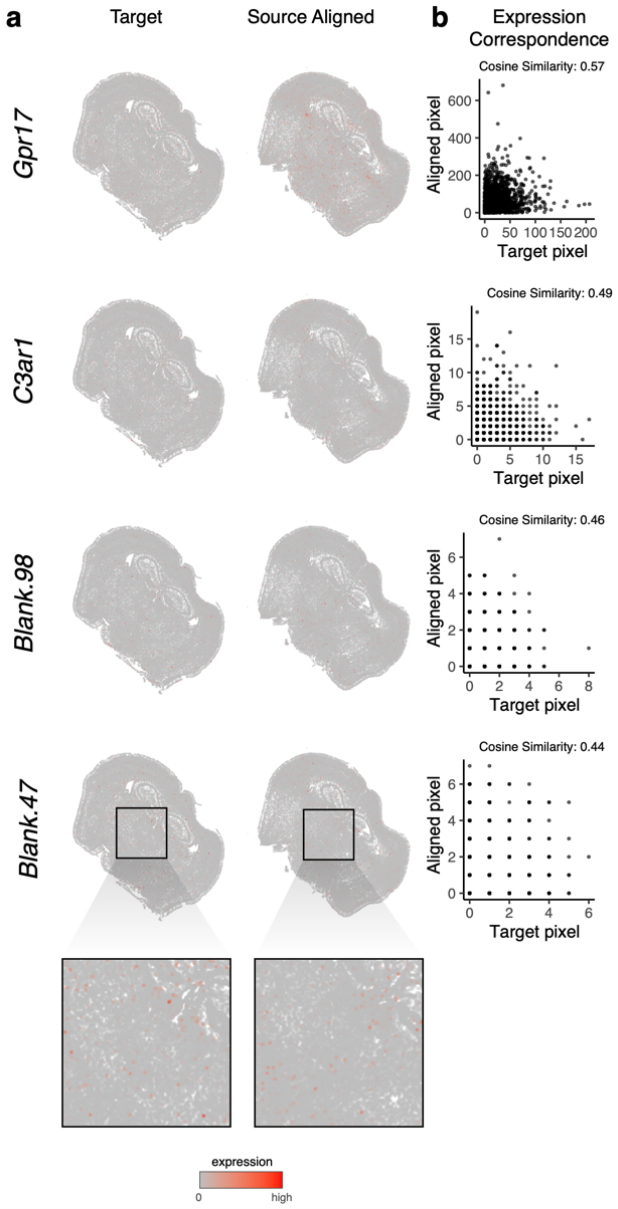
682 ***spatially patterned genes.*** **a.** Correspondence of gene expression spatial organization between the

683 target and aligned source for select spatially patterned genes. **b.** Transcript counts in the target

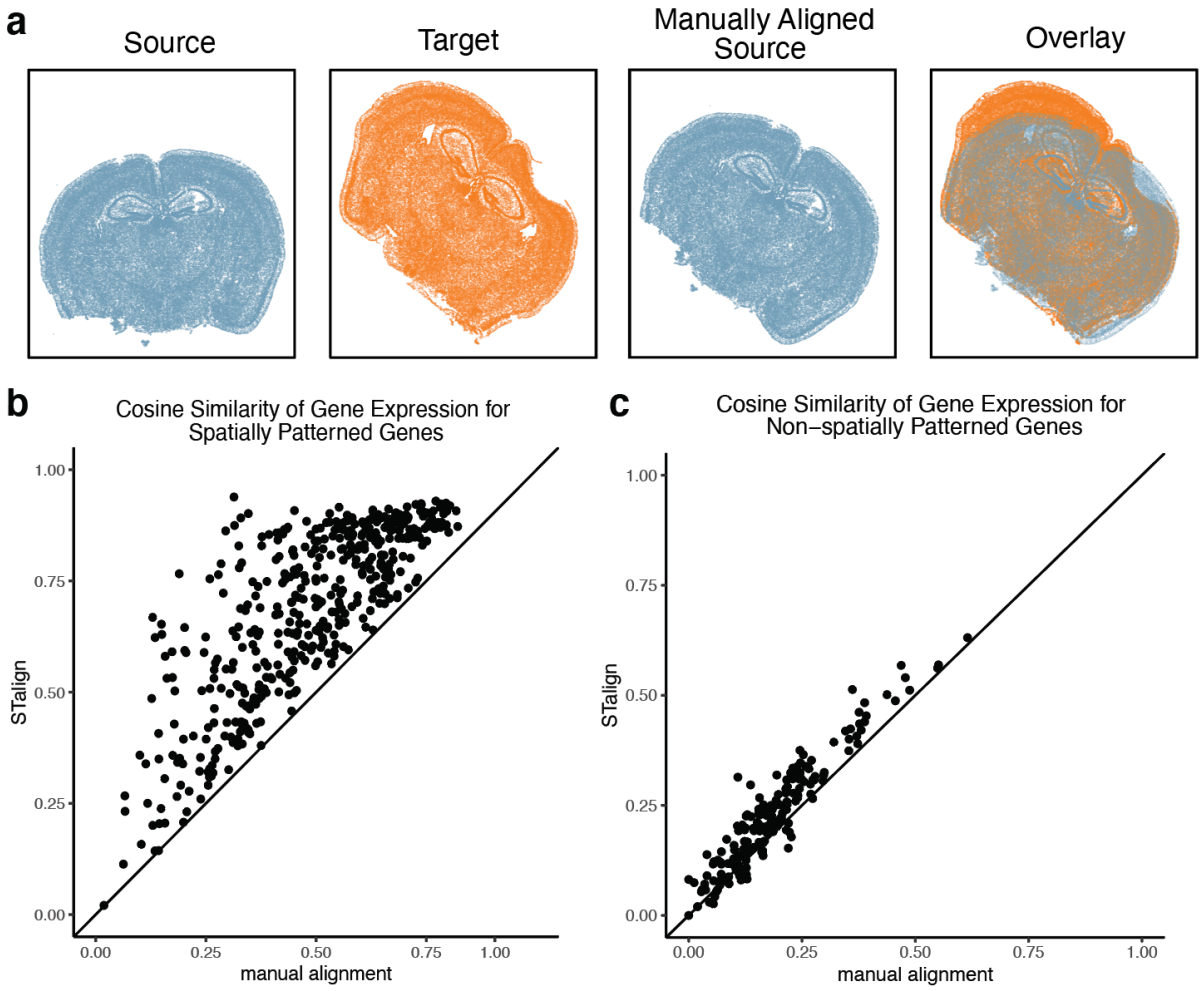
684 compared to the aligned source at matched pixels for select genes with cosine similarities between

685 transcript counts in target versus aligned source marked.

686

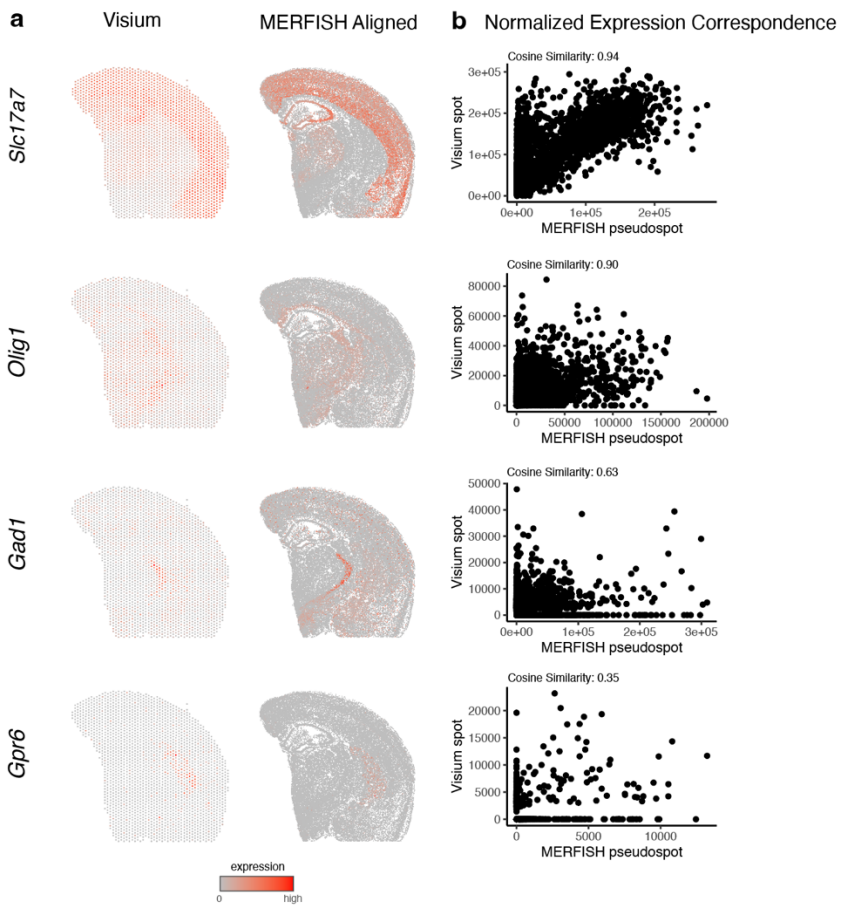


Supplemental Figure 3. ***Additional examples of MERFISH to MERFISH alignment for non-spatially patterned genes.*** **a.** Correspondence of gene expression spatial organization between the target and aligned source for select non-spatially patterned genes (inset displays cells at higher magnification). **b.** Transcript counts in the target compared to the aligned source at matched pixels for select genes with cosine similarities between transcript counts in target versus aligned source marked.



694

695 **Supplemental Figure 4.** ***Evaluation of STalign against manual alignment.*** **a.** Spatial agreement
696 of target and source that has been manually aligned based on simple affine transformations. **b.**
697 Cosine similarities between transcript counts in target versus aligned source for STalign compared
698 to manual alignment for 457 spatially patterned genes. (mean difference = 0.20) **c.** Cosine
699 similarities between transcript counts in target versus aligned source for STalign compared to
700 manual alignment for 192 non-spatially patterned genes. (mean difference = 0.04)



701

702 **Supplemental Figure 5.** ***Additional examples of MERFISH to Visium alignment for spatially***

703 ***patterned genes.*** **a.** Correspondence of gene expression spatial organization between the target

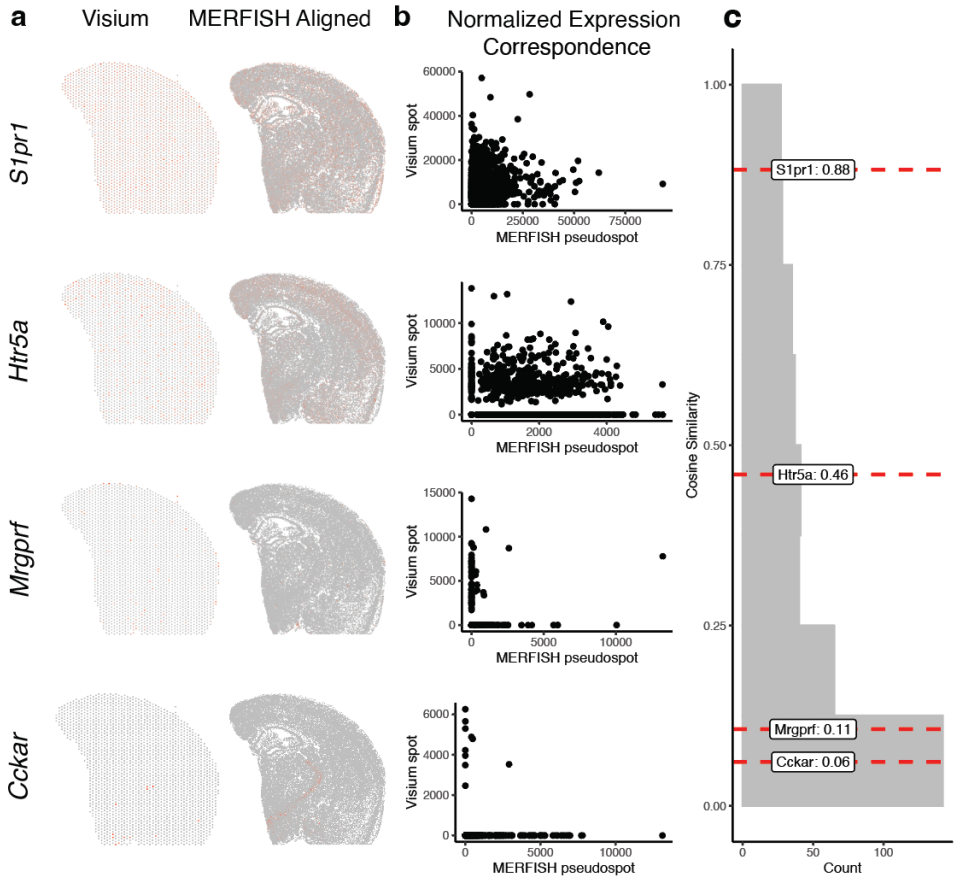
704 and aligned source for select spatially patterned genes. **b.** Normalized gene expression in the

705 Visium target compared to the aligned MERFISH source at matched spots and pseudospots for

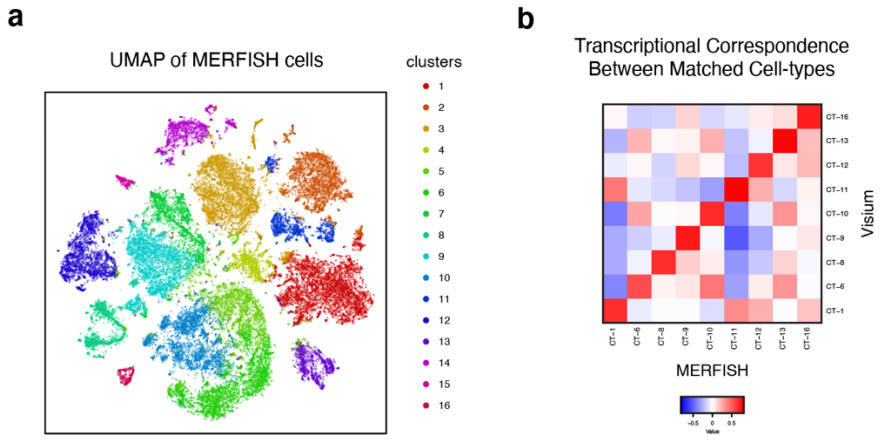
706 select genes with cosine similarities between transcript counts in target versus aligned source

707 marked.

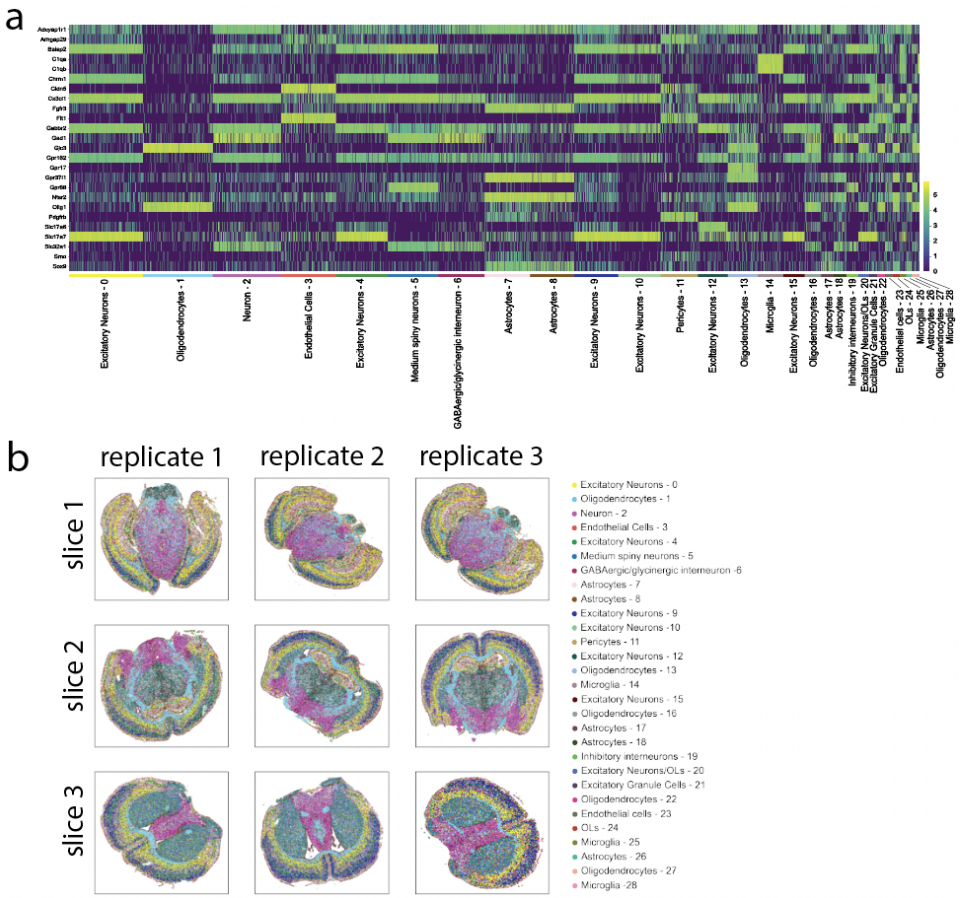
708



Supplemental Figure 6. ***Examples of MERFISH to Visium alignment for spatially non-patterned genes.*** **a.** Correspondence of gene expression spatial organization between the target and aligned source for select non-spatially patterned genes. **b.** Normalized gene expression in the Visium target compared to the aligned MERFISH source at matched spots and pseudospots for select non spatially patterned genes. **c**. Distribution of cosine similarities between normalized gene expression in the Visium target versus aligned MERFISH source at matched spots and pseudospots for 188 non-spatially patterned genes detected by both ST technologies with select genes marked.



Supplemental Figure 7. ***Cell-type correspondence between clustering of MERFISH data and deconvolution of Visium data***. **a.** UMAP embedding of MERFISH cells colored by cluster. **b.** Heatmap of transcriptional correlation between the average expression for MERFISH clusters and deconvolved expression for Visium cell-types from STdeconvolve.



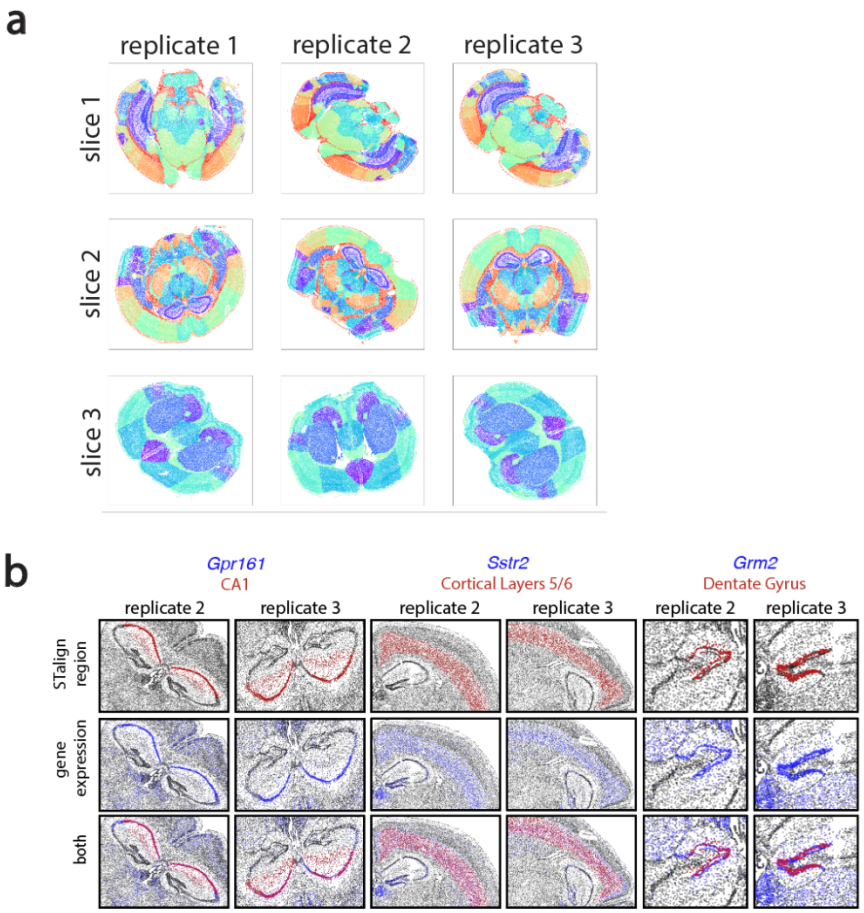
725

726 **Supplemental Figure 8.** C***ell types in MERFISH dataset.*** **a.** Heat map of differentially expressed

727 genes in cell types defined by Leiden clustering for all cells across 9 MERFISH datasets. **b.** Cell

728 types defined by differential gene expression and Leiden clustering, represented by different

729 colors, for three biological replicates of three brain slices.



730

731 **Supplemental Figure 9.** ***STalign-defined brain regions.*** **a.** Brain regions defined by STalign,

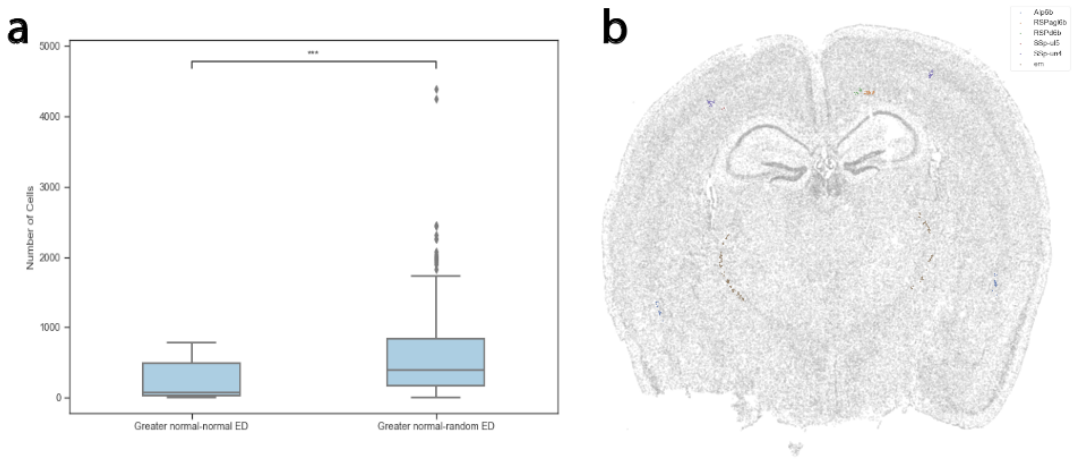
732 represented by different colors, for three biological replicates of three brain slices. **b.** Examples of

733 genes (blue) expressed in brain regions (red) obtained through 3D alignment of MERFISH slices

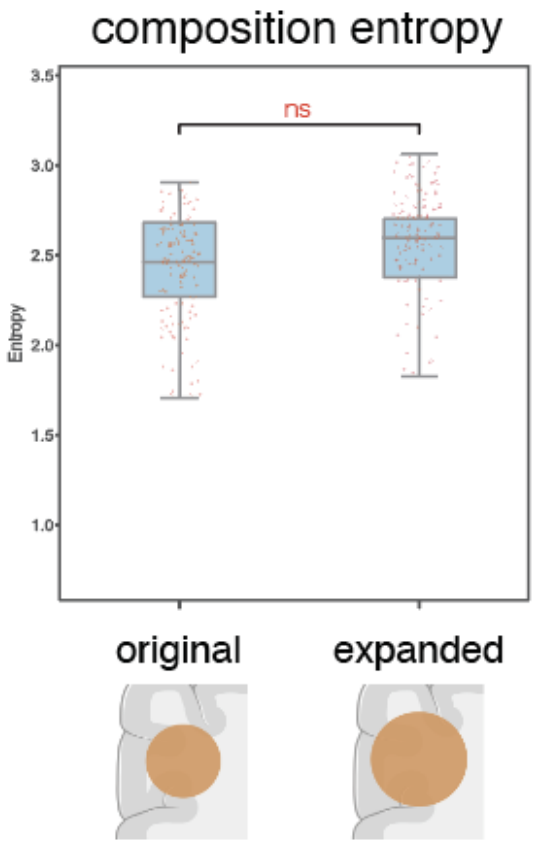
734 using STalign. Based on gene expression, brain region annotations show consistency and accuracy

735 across replicates of brain slices.

736



Supplemental Figure 10. ***Analyzing brain regions with lower cell-type compositional similarity between replicates compared to size-matched random regions.*** **a.** Distribution of number of cells in brain regions for which Euclidean-distance (ED) was greater (left) or smaller (right) between replicates compared to matched randomly defined brain regions **b.** Representative MERFISH dataset (Slice 2 Replicate 3) of brain regions in the 'Greater normal-normal ED' suggestive of lower cell-type compositional similarity between replicates compared to size-matched random regions from (a) which were under 50µm in at least one dimension.



747

748 **Supplemental Figure 11.** ***Entropy for size-matched and expanded random brain regions.*** Non-
749 significant (ns) difference between distribution of cell-type composition entropy for randomly
750 defined brain regions that were matched in size for STalign-defined brain regions (left) versus
751 regions expanded by 100 nearest neighbors (~100µm).

752

Online Methods for Alignment of spatial transcriptomics data using diffeomorphic metric mapping

Kalen Clifton[1,2,*] Manjari Anant[1,3,*]
Osagie K. Aimiuwu[4] Justus M. Kebschull[2]
Michael I. Miller[2,5] Daniel Tward[6,7] Jean Fan[1,2]

[1]Center for Computational Biology, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, 21211
[2]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21218
[3]Department of Neuroscience, Johns Hopkins University, Baltimore, MD, 21218
[4]University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
[5]Kavli Neuroscience Discovery Institute, The Johns Hopkins University, Baltimore, MD 21211
[6]Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA 90024
[7]Department of Neurology, University of California Los Angeles, Los Angeles, CA 90024

April 11, 2023

1 Methods

1.1 Cell density data model

For single-cell resolution spatially resolved transcriptomics data, we model the point sets of detected cells in the framework of mathematical varifold measures [1]. While the theory extends to more complex spaces of features, here we focus on the locations of cells only, termed the *marginal space measure* (after marginalizing out features other than spatial location) as defined in [2].

Briefly, these space measures are weighted sums of Dirac δ distributions $\rho \doteq \sum_i^{N_q} w_i^\rho \delta_{x_i^\rho}$, where $x_i^\rho \in \mathbb{R}^D$ stores the spatial coordinate of the ith out of N_q cells, and $w_i^\rho \in \mathbb{R}$ stores its weight. In this work, cell positional data is two

dimensional, so $D = 2$, and with some abuse of notation we sometimes write (x, y) instead of x.

We aim to evaluate the similarity between two single-cell resolution spatially resolved transcriptomics datasets, which we call a source and a target. Note other commonly used terms for source are: template, atlas, or moving image, while another commonly used terms for target is: fixed image. To compute distances between datasets, we embed their corresponding space measures ρ_S and ρ_T respectively in the dual of a Reproducing Kernel Hilbert Space V^* and use the standard operator norm (see for example [3]). For some choice of kernel function k, the norm squared is

$$\begin{aligned}\|\rho_S - \rho_T\|_{V^*}^2 = &\sum_{i_S, j_S}^{N_{\rho_S}} w_{i_S}^{\rho_S} w_{j_S}^{\rho_S} k(x_{i_S}^{\rho_S}, x_{j_S}^{\rho_S}) \\ &- 2 \sum_{i_S}^{N_{\rho_S}} \sum_{j_T}^{N_{\rho_T}} w_{i_S}^{\rho_S} w_{j_T}^{\rho_T} k(x_{i_S}^{\rho_S}, x_{j_T}^{\rho_T}) \\ &+ \sum_{i_T, j_T}^{N_{\rho_T}} w_{i_T}^{\rho_T} w_{j_T}^{\rho_T} k(x_{i_T}^{\rho_T}, x_{j_T}^{\rho_T}) \ . \end{aligned} \tag{1}$$

Here we chose k as a Gaussian with $k(x_i, x_j) = \exp\left(-\frac{1}{2}|x_i - x_j|^2/2\sigma^2\right)$, where $|\cdot|$ denotes the standard Euclidean norm, and σ is a user specified kernel width parameter.

The variables $w_{i_S}^{\rho_S}$, $w_{j_S}^{\rho_S}$, and $x_{i_S}^{\rho_S}, x_{j_S}^{\rho_S}$ corresponds to the weights and spatial coordinates of the ith and jth cells in the source, while $w_{i_T}^{\rho_T}$, $w_{j_T}^{\rho_T}$, $x_{i_T}^{\rho_T}, x_{j_T}^{\rho_T}$ correspond to the weights and spatial coordinates of the cells in the target. For simplicity, in the main paper we write x_S and x_T for source and target points respectively.

In STalign, we initialize weights to 1, though applying nonlinear deformations will modify these weights as discussed below in section 1.4.

1.2 Rasterization

Since computation of this norm is of quadratic complexity in the number of points, we turn to a more efficient representation for computing optimal transformations through rasterization. We can reduce the complexity of our calculations significantly by approximating our space measures through sampling a density signal on a regular grid (known as rasterization), rather than keeping a list of points and weights.

By defining $k^{\frac{1}{2}}$ such that $k^{\frac{1}{2}} * k^{\frac{1}{2}} = k$ (where $*$ refers to convolution), the

above expression for norm squared (1) can be written as

$$\|\rho_S - \rho_T\|_{V^*}^2 = \int \left| \sum_{i_S=1}^{N_{\rho_S}} w_{i_S}^{\rho_S} k^{\frac{1}{2}}(x - x_{i_S}^{\rho_S}) - \sum_{i_T=1}^{N_{\rho_T}} w_{i_T}^{\rho_T} k^{\frac{1}{2}}(x - x_{i_T}^{\rho_T}) \right|^2 dx$$
$$= \|I^S - I^T\|_2^2 \ . \tag{2}$$

Note that when k is a radially symmetric Gaussian, $k^{\frac{1}{2}}$ is also a radially symmetric Gaussian but with half the variance. Here we have defined the smooth density function

$$I(x) \doteq [k^{\frac{1}{2}} * \rho](x) = \sum_{i=1}^{N_\rho} w_i^\rho k(x - x_i^\rho) \ , \tag{3}$$

and $\|\cdot\|_2$ is the L_2 norm on functions.

Due to the smoothness introduced by $k^{\frac{1}{2}}$, these functions can be accurately discretized by sampling them on a uniform pixel grid at a resolution rate defined by the user and comparable in size to σ.

Without rasterization, evaluating the function I at a given point would involve a sum over every x_i, an order N complexity operation. After rasterization the function I can be evaluated at any point in order 1 complexity using bilinear interpolation. This allows the norm to be evaluated by summing over a pixel grid in order P complexity (where P is the number of pixels), rather than a double sum over the points x_i in order N^2 complexity.

For example, MERFISH Slice 2 Replicate 3 has 85958 cells, and the rasterized dataset has $336 \times 256 = 86,016$ pixels. The Naive approach would involve 7,388,777,764 terms in the sum (pairs of cells), whereas in the rasterization approach there are only 86,016 terms in the sum (pixels). This is an approximately 86,000 times increase in efficiency which occurs for each iteration of optimization, ignoring the negligible time cost of rasterization itself, which occurs only once at the start of registration.

1.3 Diffeomorphic transformation model

We estimate alignments between two rasterized datasets by applying a transformation $\phi : \mathbb{R}^D \to \mathbb{R}^D$. $\phi(x) \doteq A\varphi_1(x)$, the composition of two transformations: a diffeomorphism (a smooth differentiable transformation with a smooth differentiable inverse) $\varphi_1 : \mathbb{R}^D \to \mathbb{R}^D$ generated in the Large Deformation Metric Mapping (LDDMM) framework [4], and an affine transformation A (i.e. a 3x3 matrix in homogeneous coordinates whose upper left 2x2 block is a linear transform and upper right 2x1 block is a translation vector). In this notation $A\varphi_1(x)$ denotes matrix multiplication of the matrix A and the vector φ_1 in homogeneous coordinates.

In the LDDMM framework a diffeomormphism is generated by integrating a time varying velocity field v_t over time $t \in [0, 1]$, by solving the ordinary

differential equation

$$\frac{d}{dt}\varphi_t = v_t(\varphi_t) \tag{4}$$

with initial condition $\varphi_0(x) = 0$. For identifying alignments, we optimize over v_t rather than φ_1 directly, and to emphasize this dependence we use the superscript φ^v in the main text. Similarly, we use $\phi^{A,v}$ to emphasize the dependence of ϕ on both A and v_t. As long as v_t a smooth function of space, φ_1^v is guaranteed to be diffeomorphic. We enforce this through regularization as described below in section 1.5.

While this section described how we parameterize our transformations, next we need to describe how they act to deform our datasets, in order to use them in an optimization problem.

1.4 Action of transformations on datasets

The action of a transformation ϕ on a space measure dataset ρ moves the spatial coordinate of each cell, and adjusts the weight of each cell based on the transformation's Jacobian determinant.

$$\phi \cdot \rho = \phi \cdot \left(\sum_{i=1}^{N_\rho} w_i^\rho \delta_{x_i^\rho} \right) \tag{5}$$

$$= \sum_{i=1}^{N_\rho} w_i^\rho |d\phi(x_i^\rho)| \delta_{\phi(x_i^\rho)} \tag{6}$$

where $d\phi(x)$ denotes the matrix of partial derivatives of the map ϕ at the point x, and $|\cdot|$ represents the determinant of a matrix.

We note that the standard image action $[\phi \cdot I](x) = I(\varphi^{-1}x)$ has been well studied theoretically (as a left group action), computationally (in terms of its efficient implementation through interpolation), and application-wise (in terms of its use in a variety of image registration platforms e.g. [4]). This image action does not match exactly with the action of a transformation on a space measure dataset (5), although it is a close approximation for datasets of interest. We wish to leverage the existing computational machinery for image actions, and therefore describe the nature of this approximation below:

$$[\phi \cdot I](x) = [\phi \cdot (k^{\frac{1}{2}} * \rho)](x) \tag{7}$$

$$= I(\phi^{-1}(x)) \tag{8}$$

$$= \sum_{i=1}^{N_\rho} w_i^\rho k^{\frac{1}{2}}(\phi^{-1}(x) - x_i^\rho) \tag{9}$$

$$\simeq \sum_{i=1}^{N_\rho} w_i^\rho |d\phi(x_i^\rho)| k^{\frac{1}{2}}(x - \phi(x_i^\rho)) \tag{10}$$

$$= [k^{\frac{1}{2}} * (\phi \cdot \rho)](x) \tag{11}$$

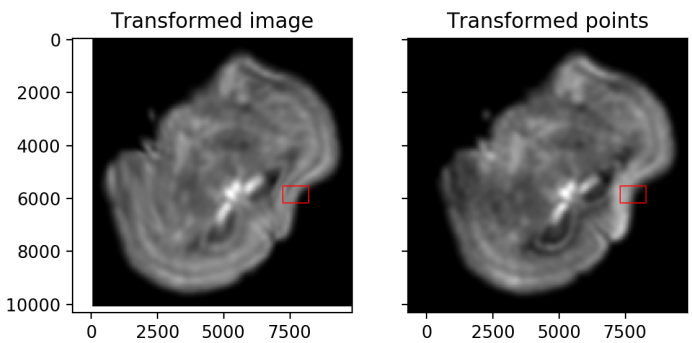


Figure 1: Rasterization followed by deformation with the image action (left), versus naively deforming the positions of cells followed by rasterization (right). Note the intensity changes that occur in regions of high deformation.

The approximate equality is accurate in our examples when $k^{\frac{1}{2}}$ is narrow relative to the smoothness of v_t and wide relative to the spacing between cells. The approximate equality would be exact if $k^{\frac{1}{2}}$ were a Dirac delta function. If cells are too far apart, a larger value of σ could be chosen. For a particularly sparse set of cells, a different method that does not include rasterization would be more appropriate, for example measure matching [5].

With this formulation, deformations can be applied to smooth density images I using interpolation in order P (number of pixels) complexity. Figure 1 illustrates the importance of the Jacobian factor. Without including this factor, transforming a density image alters its brightness, which is typically undesirable: a bigger organ tends to have more cells with the same cell density, rather than the same number of cells with a lower density.

1.5 Image registration

We compute a spatial alignment between two ST datasets by minimizing the sum of two objective functions: a regularization term R, and a matching term M_θ,

$$E(A, v) = R(v) + M_\theta(\phi^{A,v} \cdot I^S, I^T) \tag{12}$$

which we define below. Note that the computation of I is described in section 1.2, the parameterization of $\phi^{A,v}$ is described in 1.3, and the action $\phi^{A,v} \cdot I^S$ is described in the section 1.4. Recall that ϕ depends on both the velocity field v and the affine transform A.

Following the LDDMM framework, we regularize our diffeomorphism via

$$R(v) = \int_0^1 \int_{\mathbb{R}^D} |(\mathrm{id} - a^2\Delta)^p v_t(x)|^2 dx dt \tag{13}$$

where id is an identity matrix, Δ is the Laplacian, a is a constant with units of length that controls spatial smoothness, and $p = 2$ is a power that must be large enough to guarantee that results are diffeomorphisms [6].

Our matching takes the form of [7]

$$M_\theta(\phi^{A,v} \cdot I^S, I^T) = \int_{\mathbb{R}^D} |f_\theta(\phi^{A,v} \cdot I^S](x)) - I^T(x)|^2 W(x) dx \ . \tag{14}$$

Note that I^T need not correspond to a smooth density image as defined in section 1.2. For example, we include the case where it is a red green blue image corresponding to an H&E stain.

The function f_θ is a transformation of image contrast with unknown parameters θ. We use a polynomial for f_θ, in which case the minimizing parameters θ can be found exactly by solving a weighted least squares problem. The purpose of this transformation is to model differences in contrast between images from the same modality due to calibration issues; and contrast/color differences between different modalities.

The term W is a positive weight that represents the probability that a given pixel in the target image can be matched accurately to one in the source image. For example, if tissue is missing in the target image but not the source image, pixels in the region of missing tissue would get a small weight. Similarly, if the target image included a signal not present in the source (e.g. a bright fluorescence signal).

To optimize E, we alternate between updating W with Gaussian mixture modeling, and jointly updating $(\theta, \phi^{A,v})$ with gradient based methods, using expectation maximization algorithm as discussed in [7]. Recall that updating ϕ corresponds to updating the affine transformation matrix A, and the velocity field v_t which generates the deformation φ_1 from (4).

After solving for the optimal transformation parameters A and v_t, a transformation and its inverse are constructed by solving (4) sampled on a regular grid, using Semi-Lagrangian techniques [8]. With $\phi(x) = A\varphi_1^v(x)$ and $\phi^{-1,A,v}(x) = \varphi_1^{-1,v}(A^{-1}x)$ computed, cell locations $x_{i_S}^{\rho_S}$ in the source image can be mapped into the target by calculating $\phi(x_{i_S}^{\rho_S})$ through linear interpolation. Similarly, a point $x_{i_T}^{\rho_T}$ in the target image can be mapped to the atlas by calculating $\phi^{-1}(x_{i_T}^{\rho_T})$ through linear interpolation.

1.6 3D to 2D alignment

In addition to aligning spatially resolved transcriptomics datasets in which the cell positional information is 2D, we registered the 3D reconstructed Allen common coordinate framework (CCF) atlas (source) to each of the 9 MERFISH datasets (target). The image transformation is similar to the alignment discussed in the section 1.5 with a few exceptions:

It is important to note that all transformations are performed on the source 3D atlas. Since the 50 μm Nissl-stained Allen Brain Atlas CCF v3 was used as the source image, rasterization is not applied to the atlas. The affine transformation A for the 3D-2D alignment is a 4×4 3D matrix in homogeneous

coordinates. The space of dense 3D images in the orbit of the atlas, are defined via diffeomorphisms

$$\phi : (x_1, x_2, x_3) \in \mathbb{R}^3 \to \phi(x) = (\phi_1(x), \phi_2(x), \phi_3(x)) \in \mathbb{R}^3 \tag{15}$$

The diffeomorphism $\phi \in D$ acts on the atlas to generate the orbit of imagery $\mathcal{I}$,

$$I \in \mathcal{I}, I = \phi \cdot I_{atlas} \ . \tag{16}$$

The velocity field v_t is still defined by (4), but $v_t \in \mathbb{R}^3$.

The image I^S in the matching term M represents the transformed source atlas evaluated at $z = 0$, to enable comparison in the same dimension between the source and the target images.

References

[1] W. K. Allard, "On the first variation of a varifold," *Annals of mathematics*, vol. 95, no. 3, pp. 417–491, 1972.

[2] M. Miller, D. Tward, and A. Trouvé, "Hierarchical computational anatomy: unifying the molecular to tissue continuum via measure representations of the brain," *bioRxiv*, pp. 2021–04, 2022.

[3] J. Glaunès, A. Qiu, M. I. Miller, and L. Younes, "Large deformation diffeomorphic metric curve mapping," *International journal of computer vision*, vol. 80, no. 3, pp. 317–336, 2008.

[4] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International journal of computer vision*, vol. 61, no. 2, pp. 139–157, 2005.

[5] J. Glaunes, A. Trouvé, and L. Younes, "Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, pp. II–II, IEEE, 2004.

[6] P. Dupuis, U. Grenander, and M. I. Miller, "Variational problems on flows of diffeomorphisms for image matching," *Quarterly of applied mathematics*, pp. 587–600, 1998.

[7] D. Tward, T. Brown, Y. Kageyama, J. Patel, Z. Hou, S. Mori, M. Albert, J. Troncoso, and M. Miller, "Diffeomorphic registration with intensity transformation and missing data: Application to 3d digital pathology of alzheimer's disease," *Frontiers in neuroscience*, vol. 14, p. 52, 2020.

[8] A. Staniforth and J. Côté, "Semi-lagrangian integration schemes for atmospheric models—a review," *Monthly weather review*, vol. 119, no. 9, pp. 2206–2223, 1991.