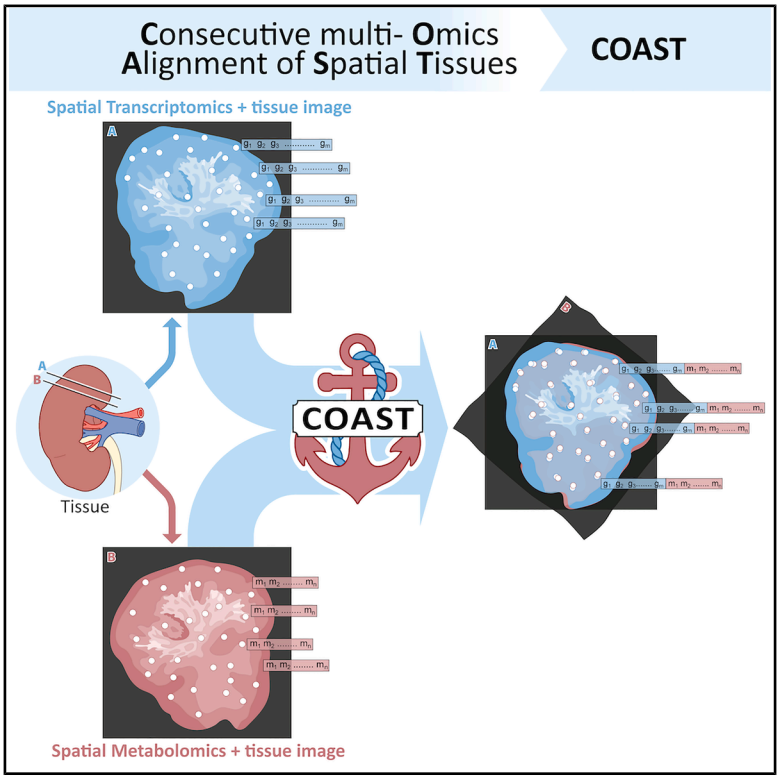
Cell Reports
Methods

Article

Image-guided alignment of consecutive multi-modal tissue slides

Graphical abstract



Highlights

- COAST aligns consecutive tissue sections across different spatial technologies
- The framework generalizes to different types of multi-modal data
- COAST accurately aligns uni-modal data based solely on image information
- Performance is robust to variations across imaging modalities and staining types

Authors

Benedetta Manzato, Claudio Novella Rausell, Gangqi Wang, ..., Sebastien J. Dumas, Ton J. Rabelink, Ahmed Mahfouz

Correspondence

a.mahfouz@lumc.nl

In brief

Manzato et al. develop COAST, a method to physically register images from consecutive tissue sections, enabling reliable downstream multi-modal analysis. Applying COAST to Stereo-seq and MALDI-MSI, the authors map the transcriptional and metabolic changes associated with kidney regeneration in an ischemia-reperfusion injury mouse model.

Manzato et al., 2026, Cell Reports Methods 6, 101575
November 16, 2026
https://doi.org/10.1016/j.crmeth.2026.101575

CellPress

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Image-guided alignment of consecutive multi-modal tissue slides

Benedetta Manzato,[1,2] Claudio Novella Rausell,[1,2] Gangqi Wang,[3,4] Nina Ogrinc,[5] Rosalie G.J. Rietjens,[2,3] Marleen E. Jacobs,[2,3] Christos Botos,[1] Sebastien J. Dumas,[2,3,6] Ton J. Rabelink,[2,3] and Ahmed Mahfouz[1,7,8,*]

[1]Department of Human Genetics, Leiden University Medical Center, Leiden 2333ZA, the Netherlands
[2]The Novo Nordisk Foundation Center for Stem Cell Medicine (reNEW), Leiden University Medical Center, Leiden 2333ZA, the Netherlands
[3]Department of Internal Medicine (Nephrology) & Einthoven Laboratory of Vascular and Regenerative Medicine, Leiden University Medical Center, Leiden 2333ZA, the Netherlands
[4]Children's Hospital of Fudan University, and the Shanghai Key Laboratory of Medical Epigenetics, Institutes of Biomedical Sciences, Fudan University, Shanghai 201102, China
[5]Center for Proteomics and Metabolomics (CPM), Leiden University Medical Center, Leiden 2333ZA, the Netherlands
[6]National Institute of Health and Medical Research (INSERM), UMR 1297, Institute of Metabolic and Cardiovascular Diseases, University of Toulouse, Toulouse, France
[7]Delft Bioinformatics Lab, Delft University of Technology, Delft 2628XE, the Netherlands
[8]Lead contact
*Correspondence: a.mahfouz@lumc.nl
https://doi.org/10.1016/j.crmeth.2026.101575

MOTIVATION With rapid advances in spatial technologies, researchers are using combinations of technologies to map different molecular features in tissues. However, performing multi-omics experiments on the same tissue section remains challenging, and experiments are often done on consecutive sections. To unlock the full potential of multi-modal measurements, precise alignment is essential. Current methods require matching molecular features or landmark detection, limiting their applicability.

SUMMARY

We present COAST (consecutive multi-omics alignment of spatial tissues), a method to reliably physically align consecutive tissue sections to produce a unified multi-modal molecular dataset suitable for downstream applications. COAST relies exclusively on the images associated with spatial data, eliminating the need for common molecular features or prior annotations. We demonstrate the effectiveness of COAST using spatial transcriptomics slides from different technologies, tissues, and resolutions, in which it achieves performance comparable to established uni-modal alignment tools. Applying COAST to spatial transcriptomics and metabolomics/lipidomics tissue sections from a mouse model of ischemia-reperfusion injury allowed the investigation of lipid/metabolite features of transcriptionally defined cell types. Overall, COAST offers a streamlined and integrative solution for multi-modal spatial data alignment.

INTRODUCTION

Spatial omics revolutionized many areas of biology as it allows molecular profiling of cells within their native spatial context. This represents a significant advancement over single-cell omics, which, while highly sensitive and capable of identifying novel cell types through the analysis of dissociated cells, lack the critical spatial context.[1] Among spatial omics technologies, spatially resolved transcriptomics (SRT) has emerged as the most widely adopted approach, similar to the prevalence of single-cell RNA sequencing (scRNA-seq) within the single-cell domain.[2] However, recent advancements have extended spatial omics capabilities to other molecular layers, such as metabolomics and proteomics.[3,4] Integrating these multiple omics enables the exploration of complementary molecular characteristics that contribute to cellular phenotypes and is essential to understand their collective roles in driving complex biological phenomena.[5]

Technologies which allow capturing multiple molecular layers simultaneously from the same tissue slide are rapidly developing.[6–8] These approaches remain technically demanding and constrained by challenges, such as differing slide preparation requirements between modalities and tissue degradation resulting in lower data quality.[9] As a result, researchers often rely on applying different spatial modalities to adjacent or consecutive tissue sections.[10,11] Aligning the consecutive sections is then essential to allow multi-modal downstream analysis. This task, however, is not trivial due to structural differences between

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

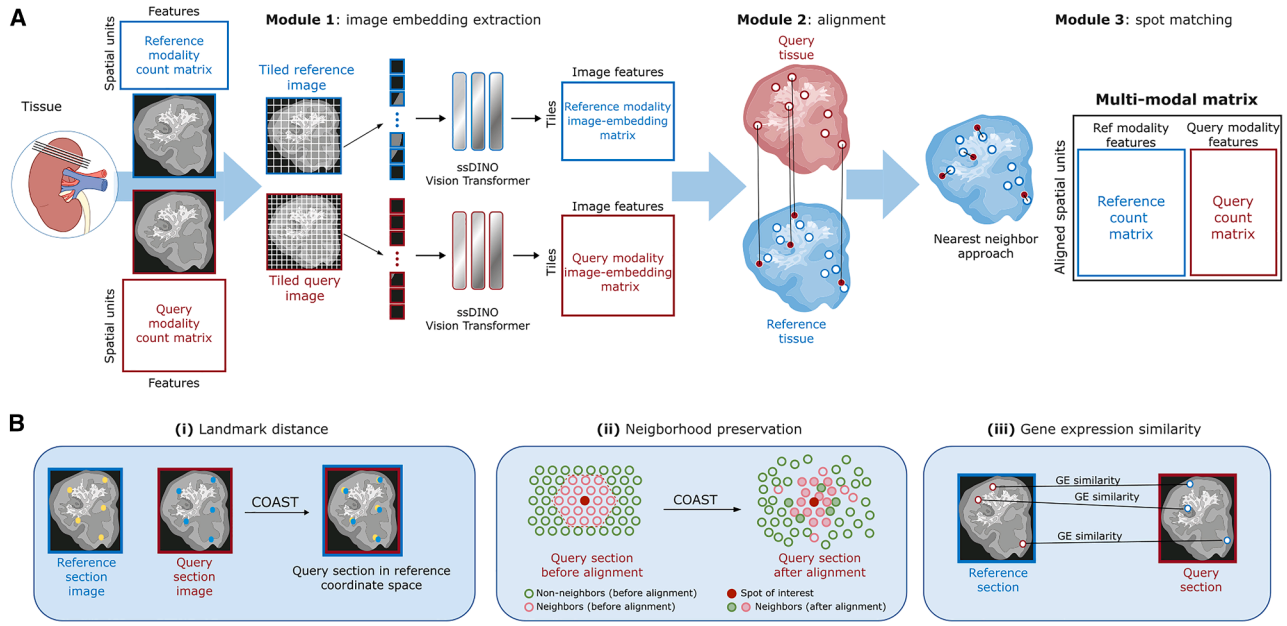


Figure 1. Schematic overview of consecutive multi-omics alignment of spatial tissues

(A) Consecutive multi-omics alignment of spatial tissues (COAST) takes as input two images that are tiled and fed separately to the ssDINO vision transformer to obtain an embedding of 384 image features for each image, where the size of the embedding is the size of the output layer of the vision transformer (module 1). Query spots are mapped to the reference coordinate space based on the similarity between their image embeddings (module 2). A unified uni- or multi-modal matrix is created by matching query and reference spots based on their proximity (module 3).

(B) The method is evaluated with three main quantitative metrics: (i) landmark distance, (ii) neighborhood preservation, and (iii) gene expression (GE) similarity.

sections as well as potential tissue deformation, distortion, and other artifacts introduced during the sectioning process.[12,13]

Several methods have been proposed to address this problem of tissue alignment, which can be roughly divided into two categories. The first category achieves alignment by maximizing the similarity of the measured molecular features (e.g., gene expression in case of SRT) using graph neural networks,[14] optimal transport,[15,16] or Gaussian processes[17] to align the feature distributions. Moreover, the reliance on shared molecular features allows these uni-modal alignment approaches to perform longitudinal alignment across temporal scales.[16,18,19] These methods, however, are largely limited to aligning consecutive sections of the same spatial modality to construct virtual 3D tissue blocks[20–22] but do not generalize to multi-modal sections without matching features across sections (commonly referred to as diagonal integration[23]). For example, CAST (cross-sample alignment of spatial omics) allows the integration of SRT and spatially resolved chromatin accessibility data by converting the accessibility data to gene activity scores, resulting in data loss. This approach is also not applicable to multimodal data in which molecular feature matching is more challenging (e.g., metabolomics and transcriptomics). Another class of methods identifies matching landmarks in tissue sections (i.e., locations with recognizable structures) that can be used to guide the alignment by applying affine transformation or geometric transformations.[24–26] Tissue landmarks can be detected manually or automatically, which can be laborious and/or impractical when tissues lack clear visually recognizable structures.

Here, we present COAST (consecutive multi-omics alignment of spatial tissues), a method to align consecutive sections profiled with any combination of multi-modal spatial omics technologies. COAST relies on the paired stained images generated alongside spatial omics data (i.e., nuclear staining or H&E) to construct a shared feature space in which spatial observations from separate sections can be aligned. Our approach reframes the alignment problem of multimodal spatial data from consecutive sections as a vertical integration task,[23] where correspondence is established based on the similarity of matched image-derived features rather than unmatched molecular profiles. We benchmark COAST against existing uni-modal alignment methods using SRT datasets, highlighting its applicability to H&E and nuclear staining. To illustrate COAST's true potential in multi-modal integration, we apply it to consecutive mouse kidney sections measured with Stereo-seq (spatial enhanced resolution omics-sequencing) for spatial transcriptomics and MALDI-MSI (matrix-assisted laser desorption/ionization mass spectrometry imaging) for spatial metabolomics.

RESULTS

Overview of COAST

COAST is a framework that takes consecutive spatial omics data and uses their paired staining images as an anchor to obtain a shared representation to map points of the two sections into a common coordinate framework (CCF), enabling the construction of unified uni- or multi-modal datasets of spatially aligned observations (Figure 1A). From the two stained images, we start by extracting structural information that can serve as a shared representation. Vision transformers have been shown to effectively capture

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

cellular heterogeneity,[27] and due to their ability to model both local and global context, they are well-suited for capturing long-range dependencies, thereby addressing morphological redundancy within tissue samples.[28] Once we establish the shared image features for both datasets, we build on CAST, a method designed for aligning uni-modal consecutive sections, to map the coordinates of the query image/spatial data into a CCF. Specifically, we employ the module CAST Stack that performs an initial rigid transformation followed by a B-spline transformation based on embedding similarity. This process maps the coordinates of the query image/spatial data into a CCF. Then, using the transformed coordinates of the query image, spatial units are matched based on their physical proximity to observations in the reference image. The result is a unified uni- or multi-modal matrix, where features represent the combined set of measurements from both modalities, and the spatial observations are the matched locations from both sections following alignment.

To quantify the quality of the resulting alignments, we measured the distance between landmark pairs post-alignment. These landmarks served as a form of ground truth and were identified as corresponding points within the tissue that should overlap following alignment, such as tissue edges, tears, or recognizable anatomical structures visible from the staining. We then calculated the Euclidean distance between the corresponding reference and query landmarks after applying COAST, which should be minimal when the alignment is accurate (Figure 1B[i]). To ensure that tissue integrity is maintained after alignment, we defined a neighborhood preservation score by comparing the third-order neighborhood of each query spot before and after alignment (Figure 1B[ii]). Moreover, when aligning uni-modal data (e.g., both consecutive sections measure gene expression), we can assess the quality of the alignment based on the similarity of the molecular profiles between the matched spots (Figure 1B[iii]). Importantly, we do not expect matched spots to have identical gene expression profiles, as a key characteristic of a good alignment is the preservation of (spatial) biological structure, not merely the optimization of gene expression or image feature similarity. This is especially important given that gene expression differences can naturally occur between tissue sections.

COAST aligns 10X Visium consecutive sections based on the paired H&E images

Although COAST is designed to align multi-modal tissue sections, we first evaluated its performance using consecutive spatial transcriptomics sections to allow comparisons against existing uni-modal alignment methods. Specifically, we aligned three pairs of consecutive 10X Visium sections of the mouse brain: sagittal posterior, coronal, and sagittal anterior (Figures 2A and S1A). First, we assessed how well the image features generated by the transformer captures tissue structures by clustering the joint vision transformer embeddings. The resulting clusters capture some of the known tissue architecture, such as the cerebellar fold in both sections (Figure S2A). Next, we compared the performance of COAST to CAST and MOSCOT (multi-omics single-cell optimal transport),[16] two uni-modal alignment methods, as well as manual rigid alignment and the original unaligned sections. We selected a total of 19, 13, and 32 landmarks for the mouse brain sagittal posterior, anterior, and coronal image sets, respectively (Figures S3A–S3F). The median distance between matching landmarks for COAST was 102.20, 12.39, and 135.29 pixels, which is relatively small considering the inter-spot distance for Visium is 69 (horizontally) and 138 (diagonally) pixels. For all the three datasets, COAST showed a significant improvement over the unaligned sections (Figures 2B and S1B) ($d = -6.39$, $p < 0.001$ for the sagittal posterior dataset; $d = -5.73$, $p < 0.001$ for the sagittal anterior dataset, $d = -5.74$, $p < 0.001$ for the coronal dataset). Similarly, COAST outperformed rigid alignment in the sagittal posterior dataset ($d = -0.60$, not significant) and was significantly better in the sagittal anterior and coronal dataset ($d = -1.525$, $p = 0.026$ and $d = -1.68$, $p = 0.015$). In all the datasets, COAST performed on par with CAST and with MOSCOT on the sagittal posterior dataset, while MOSCOT performed significantly worse than COAST ($d = -2.61$, $p < 0.001$) and CAST ($d = -2.92$, $p < 0.001$) in the coronal dataset. Moreover, the number of spots matched between the consecutive sections is largely consistent across all the alignment tools, indicating that none of the methods heavily mismatched border regions during the alignment (Figures S5A–S5C). These results indicate that COAST can accurately align consecutive 10X Visium sections based on H&E images only, and achieves similar or better performance than CAST and MOSCOT without gene expression data.

To further evaluate the accuracy of the alignment, we zoomed-in on the granular layer of the cerebellum. Since there are no anatomical annotations available, we first clustered the transcriptomic data of both sections from the sagittal posterior dataset and selected the cluster corresponding to the cerebellar granular layer (see STAR Methods). We then plotted the location of the reference and query coordinates after alignment, for COAST both the spots and tiles alignment are shown for clarity (Figure 2C). COAST, as well as CAST and MOSCOT, aligned the query and reference spots belonging to the cerebellar granular layer seamlessly, whereas there is a discrepancy when the spots are rigidly aligned. To assess neighborhood preservation we compared the third-order neighborhood of each query spot before and after alignment (Figure 1B[ii]). For Visium data, the third order neighborhood corresponds to 24 spots. Across the three datasets, at least 75% and 83% of the original neighborhood (corresponding to 18, 18, and 20 out of 24 spots, respectively, for mouse brain sagittal posterior, anterior, and coronal) were preserved after alignment for every spot in the query tissue, demonstrating strong preservation of local tissue structure (Figure 2D).

Since we are aligning uni-modal data, we can assess the quality of the alignment based on gene expression similarity between the matched spots (Figure 1B[iii]). Using this measure, in the mouse brain sagittal posterior dataset, COAST and the uni-modal methods CAST and MOSCOT outperform rigid alignment and unaligned baseline (Figure 2E, left). For the mouse brain coronal, COAST performs on par with CAST and rigid alignment while MOSCOT achieves lower performance as observed earlier through landmark evaluation (Figure 2E, right). For the mouse brain sagittal anterior, COAST achieves better gene expression similarity than unaligned baseline and comparable similarity as the rigid alignment, CAST and MOSCOT (Figure S1E). Of note, unlike CAST and MOSCOT which utilize the expression data

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

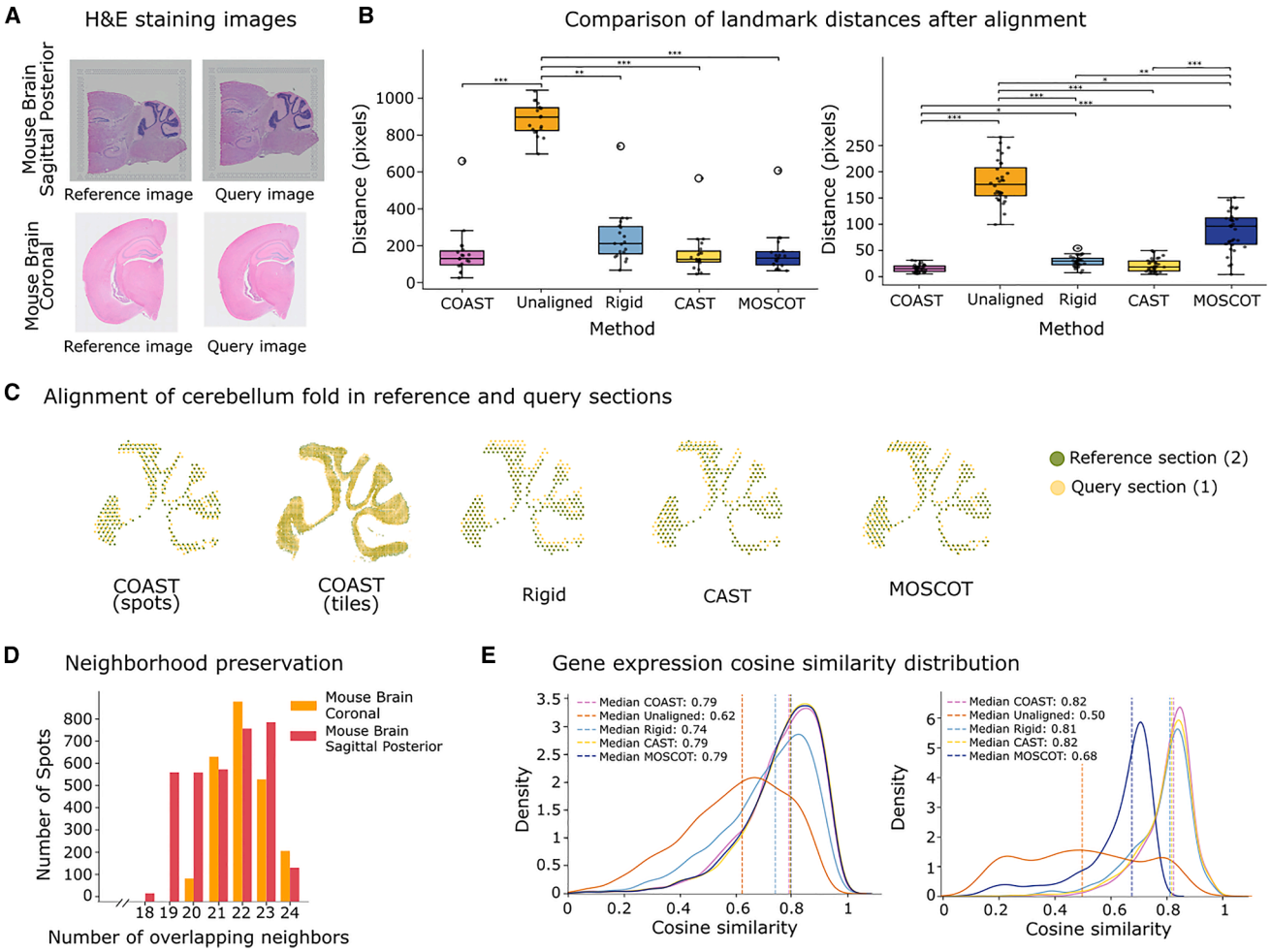


Figure 2. Aligning consecutive Visium slides using COAST

(A) H&E images of the two consecutive mouse brain sections. Top: 10X mouse brain sagittal posterior, bottom: 10X mouse brain coronal.

(B) Boxplot representing the distances between landmarks after registration via COAST, rigid alignment, unaligned, CAST, and MOSCOT. Left: mouse brain sagittal posterior, right: mouse brain coronal. The center line indicates the median, box bounds represent the interquantile range, and whiskers extend to 1.5 x interquantile range. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(C) The cerebellar fold of mouse brain sagittal posterior dataset after alignment.

(D) Neighborhood preservation of the third order neighborhood (24 spots) in the query section before and after alignment.

(E) Cosine similarity of gene expression of matched spots with COAST, rigid alignment, CAST, and MOSCOT. Left: mouse brain sagittal posterior, right: mouse brain coronal. See also Figures S1 and S4.

for alignment, and as such have an advantage when evaluating expression similarity of the aligned sections, COAST only uses the associated H&E images.

While the three Visium datasets demonstrate good performance, these are relatively easy scenarios, characterized by well-preserved tissue morphology, minimal distortion, and consistent orientation across slides. To evaluate COAST under more challenging conditions, we artificially distorted the query images by applying a 90° rotation and a horizontal flip (around the *x* axis), respectively. We then ran only the feature extraction and alignment modules on these altered images. COAST was still able to recover a meaningful alignment in all three cases, demonstrating its capacity to handle substantial spatial transformations (Figure S4). This suggests that the model captures underlying tissue structure in a way that is robust to such perturbations.

All together, these results show that COAST effectively aligns consecutive sections using the H&E stainings associated with SRT data, performing comparably to, and in some cases surpassing, existing unimodal methods incorporating molecular features.

Alignment of high-resolution Xenium datasets using COAST

The spatial resolution of spatial transcriptomics measurements varies between technologies. Here, we evaluate the applicability of COAST to a more challenging dataset where the observations are single cells, compared to Visium spots (55 μm) and Stereo-seq

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

CellPress
OPEN ACCESS

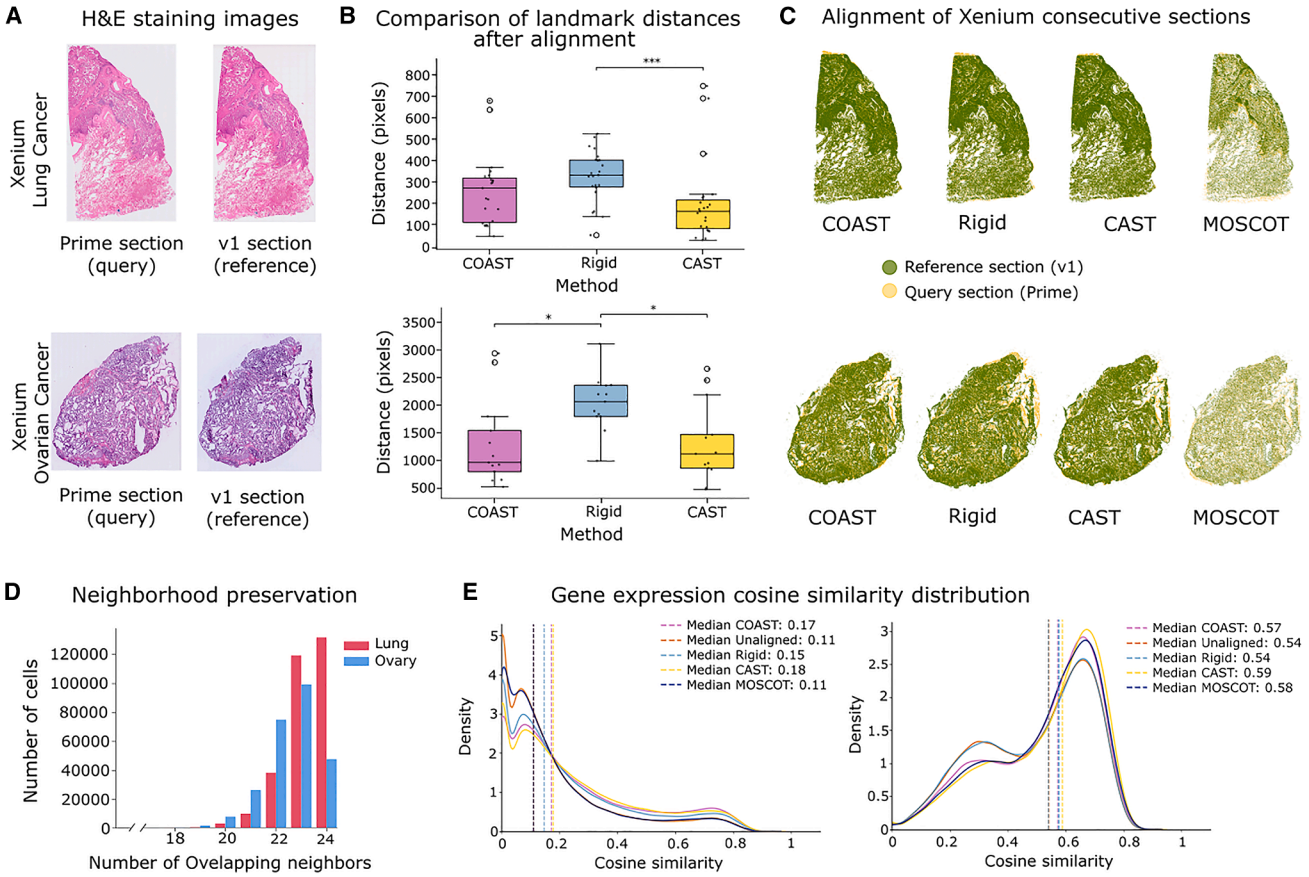


Figure 3. Aligning consecutive Xenium slides using COAST

(A) H&E images of the two consecutive Xenium human sections. Top: 10X human lung cancer (196 genes), bottom: 10X human ovarian cancer (278 genes).

(B) Boxplot representing the distances between landmarks after registration via COAST, rigid alignment, and CAST. Top: human lung cancer, bottom: human ovarian cancer. The center line indicates the median, box bounds represent the interquantile range, and whiskers extend to 1.5 x interquantile range. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(C) Spatial distribution of reference and query cells after alignment with COAST and benchmarking tools.

(D) Neighborhood preservation of nearest 24 cells in the query section before and after alignment.

(E) Cosine similarity of gene expression of matched cells with COAST, rigid alignment, CAST, and MOSCOT. Left: human lung cancer, right: human ovarian cancer.

bins (20 μm). Compared to spatial units and spots, single-cell resolution data leave less margin to mismatch errors during the alignment. We used two pairs of consecutive sections from Xenium datasets: human lung cancer (Figure 3A, top) and human ovarian cancer (Figure 3A, bottom), with around 250,000 measured cells per section (see STAR Methods). We manually defined 23 and 12 landmarks for the lung and ovarian datasets, respectively (Figures S3G and S3L), and benchmarked the performance of COAST against CAST and standard rigid alignment. Our results indicate that both CAST and COAST significantly improve alignment accuracy over rigid methods (Figures 3B and 3C). The performance of MOSCOT in terms of landmark distances was not tested because we could only apply MOSCOT to a subset of the dataset due to memory constraints. For both datasets, spatial neighborhoods were effectively preserved when aligning the sections using COAST, with all query cells maintaining at least 75% of their original neighbors post-alignment (Figure 3D). Finally, we evaluated the gene expression matching on the overlapping gene sets between the pair of consecutive slides (196 for lung and 278 for ovary). Generally, the gene expression similarities for the Xenium datasets are lower than other datasets investigated in this study (Figure 3E). This is partially due to the difference in gene detection and sensitivity between the different gene panels (Figure S6), and for the ovarian dataset partly to the visual differences between the sections (Figure 3A, bottom). Despite these challenges, COAST performed comparably to established uni-modal methods in terms of expression profile similarity between matched cells. Collectively, these results demonstrate that COAST effectively aligns consecutive sections at single-cell resolution and is highly adaptable across different spatial transcriptomics technologies.

COAST generalizes to uni-modal data with nuclear staining images

The type of staining images accompanying spatial omics data varies across technologies. To further evaluate the generalizability of COAST to different stainings, we used the mouse

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

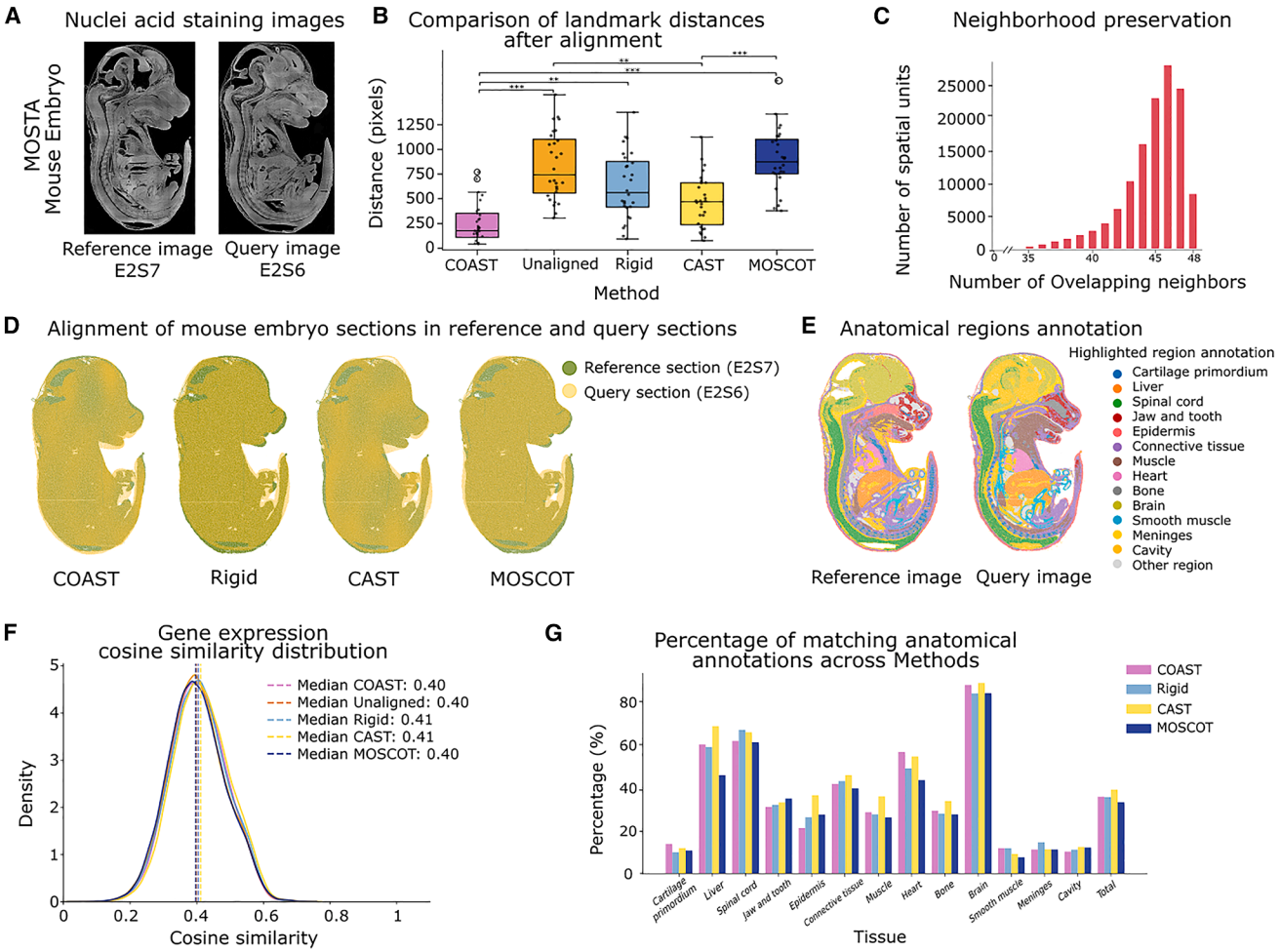


Figure 4. Aligning serial Stereo-seq slides using COAST

(A) Nucleic acid staining images of two serial mouse embryo sections.

(B) Boxplot representing the distances between landmarks after registration via COAST, rigid alignment, unaligned, CAST, and MOSCOT. The center line indicates the median, box bounds represent the interquantile range, and whiskers extend to 1.5 x interquantile range. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(C) Neighborhood preservation of the third order neighborhood (48 spots) in the query section before and after alignment.

(D) Spatial distribution of reference and query spots after alignment with COAST and benchmarking tools.

(E) Mouse embryo sections colored by anatomical region (regions selected for downstream analysis only).

(F) Cosine similarity of gene expression of matched spots with COAST, rigid alignment, CAST, and MOSCOT (raw counts, 3,000 highly variable genes).

(G) Percentage of correctly matched anatomical regions after alignment with COAST and benchmarking tools.

embryo MOSTA spatial transcriptomics dataset. This dataset was generated using Stereo-seq,[18] which acquires a nucleic acid staining images from each tissue section. We applied COAST to two serial sections—E16.5_E2S6 and E16.5_E2S7—sliced several hundred micrometers apart (Figure 4A).

We manually placed 28 landmarks and measured the pairwise distances between matching landmarks across sections after alignment (Figures S3M and S3N). COAST and CAST significantly outperformed MOSCOT with COAST also significantly outperforming rigid alignment ($p < 0.001$, d > 1.2) (Figure 4B). Although the distances between matching landmarks were smaller with COAST compared to CAST, this difference was not significant. We also assessed neighborhood preservation to quantify tissue distortion. The spatial observations in Stereo-seq are organized in a grid-structure and therefore the third-order neighborhood includes 48 spots. We found that 100% of the spots retained at least 35 matching neighbors before and after alignment using COAST, indicating the method's ability to uniformly reposition spots belonging to the query tissue (Figure 4C). We calculated gene expression similarity between matched cells after alignment. All three methods as well as rigid alignment achieved a similar performance (Figure 4F). The overall low similarity in gene expression compared to the 10X Visium data (Figure 4E) is to be expected, given that the two sections are separated by several hundred micrometers. Additionally, the number of spots retained after alignment was comparable

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

across methods, with the highest number retained by COAST (93.9%; 117,474 out of 125,090 cells) (Figure S5D).

Since the MOSTA dataset includes tissue annotations, we evaluated the quality of alignment by comparing the annotations of the matched cells. Since the distance between the sections is relatively large, there are significant changes in anatomical structures. We selected a subset of highly abundant annotations (more than 1,000 cells) that were shared across both sections (Figure 4E). This ensured that comparisons were based on robust and consistently represented populations. We then assessed the proportion of matched spots that retained the same region annotation post-alignment, presenting the results as a percentage of the total number of spots within each category (Figure 4G). Overall, the percentage of matching annotations is not very high, similar to the overall low gene expression similarity noted above, due to the large differences between the sections. Importantly, there are no significant differences across the different alignment methods, with COAST consistently performing comparably to the other alignment methods.

These results demonstrate the generalizability of COAST to nucleic acid staining sections as well as its robustness to the distance between the consecutive sections, highlighting its applicability across spatial technologies and experimental settings.

COAST aligns spatial transcriptomics, metabolomic, and lipidomic data from the mouse kidney

To demonstrate COAST's utility for studying complex biological processes through multimodal spatial data, we focused on kidney regeneration. Using an ischemia-reperfusion injury (IRI) mouse model,[29] we collected lipid and metabolite profile spectra via MALDI-MSI, and from a consecutive tissue section, 10 μm apart, we obtained whole-transcriptome gene expression data using Stereo-seq (STAR Methods). The dataset includes three biological replicates. Both modalities are accompanied by corresponding DAPI and ssDNA-stained images (Figure 5A).

For each pair of images (DAPI for MALDI-MSI and segmented-ssDNA for Stereo-seq, see STAR Methods, Figures 5B and S7), we applied COAST to extract the image embeddings as described earlier (Figure 1A). We clustered the resulting embeddings for each image separately to assess whether they capture the same tissue structures. We observed that not all clusters were shared between the two modalities, likely due to the fact that, unlike previously analyzed image pairs, these images were acquired using different stainings, introducing a platform effect (Figure S2H). Nevertheless, when clustering was performed separately on the two consecutive section images, internal kidney structures remained clearly distinguishable (Figure S2G). Despite this platform effect, the alignment module of COAST was able to accurately match the two images across the three IRI kidney samples (Figures 5C and 5D).

To evaluate alignment accuracy, we manually annotated corresponding anatomical landmarks across the three tissue pairs, identifying 21, 31, and 19 landmark pairs for IRI1, IRI2, and IRI3, respectively (Figures S3O and S3P). COAST consistently achieved significantly lower landmark distances compared to both unaligned and rigidly aligned sections, indicating more accurate spatial alignment (Figure 5D).

To assess local structural preservation after alignment, we evaluated the overlap in spatial neighborhoods for each spot in the MALDI-MSI sections before and after alignment. For IRI1, 99.4% of spots retained at least 40 out of 48 of their original neighbors, and 95.6% retained all 48 neighbors. Similarly, in IRI2, 98.4% of spots retained at least 40 neighbors, with 93.5% preserving all 48. In IRI3, 98.7% of spots retained at least 40 neighbors, and 94.2% preserved all 48 (Figure 5E). These results indicate that the alignment introduces minimal local distortion and preserves the original tissue architecture at a fine-grained spatial level across all samples.

These results demonstrate that COAST achieves high alignment accuracy while preserving tissue architecture, outperforming both unaligned baselines and, in some cases, manual rigid alignment. Importantly, these results demonstrate that COAST can align images even when input images originate from different nuclear stainings.

COAST alignment identifies of multi-modal cell-type markers

Both transcriptomics and metabolomics have been used to investigate progression and repair mechanisms after IRI. These modalities were previously studied independently,[29–32] but more recently efforts have shifted toward their integration, often across adjacent sections or without spatial context.[33–35] Using COAST, we can now analyze these spatial multi-modal data simultaneously. Based on the resulting integrated data, we compared lipid and metabolite features and transcriptomic markers from matched spatial bins. This comparison was designed to highlight consistencies between the transcriptomic and metabolomic modalities, which should emerge if the alignment is sufficiently accurate at the bin level. For this task, lipid features were selected from the literature.[29,36] Lipid markers showed negative enrichment in regions annotated as tissue gaps (Figure 6A), confirming the successful matching of gap regions across modalities. Among proximal tubule (PT) markers, the PT-S1/S2-specific lipid feature *m/z* 782.6 was enriched in PT-S1/S2, as well as in distal convoluted tubule (DCT) and connecting tubule (CNT) cells, consistent with their shared epithelial identity. Out of the three PT-S3 markers we included, two (*m/z* 772.6, 809.6) were enriched in PT-S3 cells, whereas *m/z* 462.4 was not. Injury-associated lipid features showed cell-type-specific enrichment. *m/z* 844.6 and 822.7 showed higher expression in injured tubules, even though no enrichment in the failed repair proximal tubules (FR-PT). Interestingly, *m/z* 822.7 and 838.7 are more expressed in PT-S3 compared to PT-S1/S2. This is consistent with the fact that the S3 segment operates under lower oxygen availability than S1/S2, rendering it more vulnerable to metabolic stress and injury.[37,38] The thick ascending limb (TAL) shared multiple features (*m/z* 878.6, 744.6, 788.6, and 750.6) with the collecting duct, in accordance with previous findings.[36] Regarding the glomerular lipid features, *m/z* 556.4 (shared with injured tubule and FR-PT) and 647.4 (shared with PT-S1/S2) show expression in the glomerular population, as previously shown in literature.

Next, we performed unbiased metabolite and lipid differential expression analysis between the transcriptionally defined cell types (Figure S8). CNT and DCT exhibited nearly complete

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

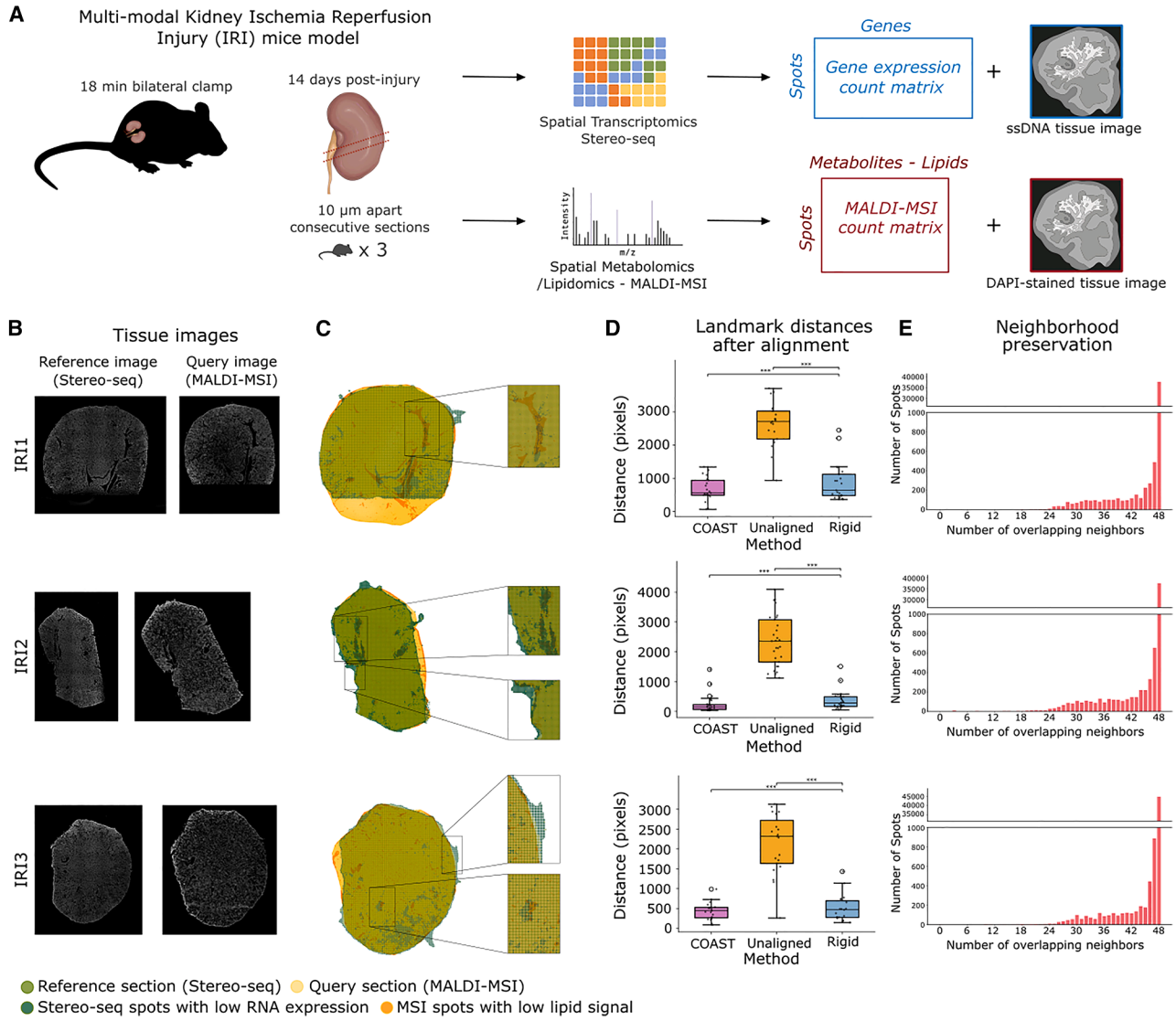


Figure 5. Aligning consecutive multi-modal slides using COAST

(A) Multi-omics ischemia-reperfusion injury (IRI) experimental design.

(B) Tissue images. Left: ssDNA image of the reference tissues (Stereo-seq), right: DAPI-stained images of query tissues (MALDI-MSI).

(C) Alignment of the two consecutive sections with COAST (left) zoom-in of the alignment in selected regions of the tissues (right).

(D) Boxplot representing the distances between landmarks after registration via COAST, unaligned, and rigid alignment. The center line indicates the median, box bounds represent the interquantile range, and whiskers extend to 1.5 x interquantile range. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(E) Neighborhood preservation of the third order neighborhood (48 spots) in the query section before and after alignment. See also Figure S7.

overlap in their identified markers, sharing 9 of the top 10 markers (m/z 185.0, 193.1, 195.1, 225.1, 239.2, 241.1, 267.1, 303.1, and 539.2), supporting the close metabolic continuity of connecting and distal nephron segments. PT-S3 population displayed the most divergent lipidomic and metabolomic profile compared to other cell types, sharing none of the top 10 markers with any other cell type. Moreover, PT-S3 showed exclusive enrichment for multiple high-mass lipid species (m/z 809.6, 837.6, 838.7, 778.6, and 888.8), highlighting a distinct lipidomic characterization. Injured tubules and FR-PT presented a mixed metabolic and lipidomic phenotype, with two overlapping markers (m/z 247.2 and 435.2) and sharing several markers with CD, TAL, CNT, DCT, and glomeruli (Figure S8). Although only two features were unique to the glomerular population (m/z 554.355 and 275.1), other non-unique features still showed enrichment (*m/z* 634.4, 556.4, and 647.6). Importantly, also in this analysis, the gaps cluster showed no discriminant features, supporting its status as a region in the tissue with no lipid or metabolite expression.

Moreover, comparing the spatial distributions of lipid and transcriptomic markers showed an overall high concordance with cell type annotations obtained from transcriptomics. The

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

CellPress
OPEN ACCESS

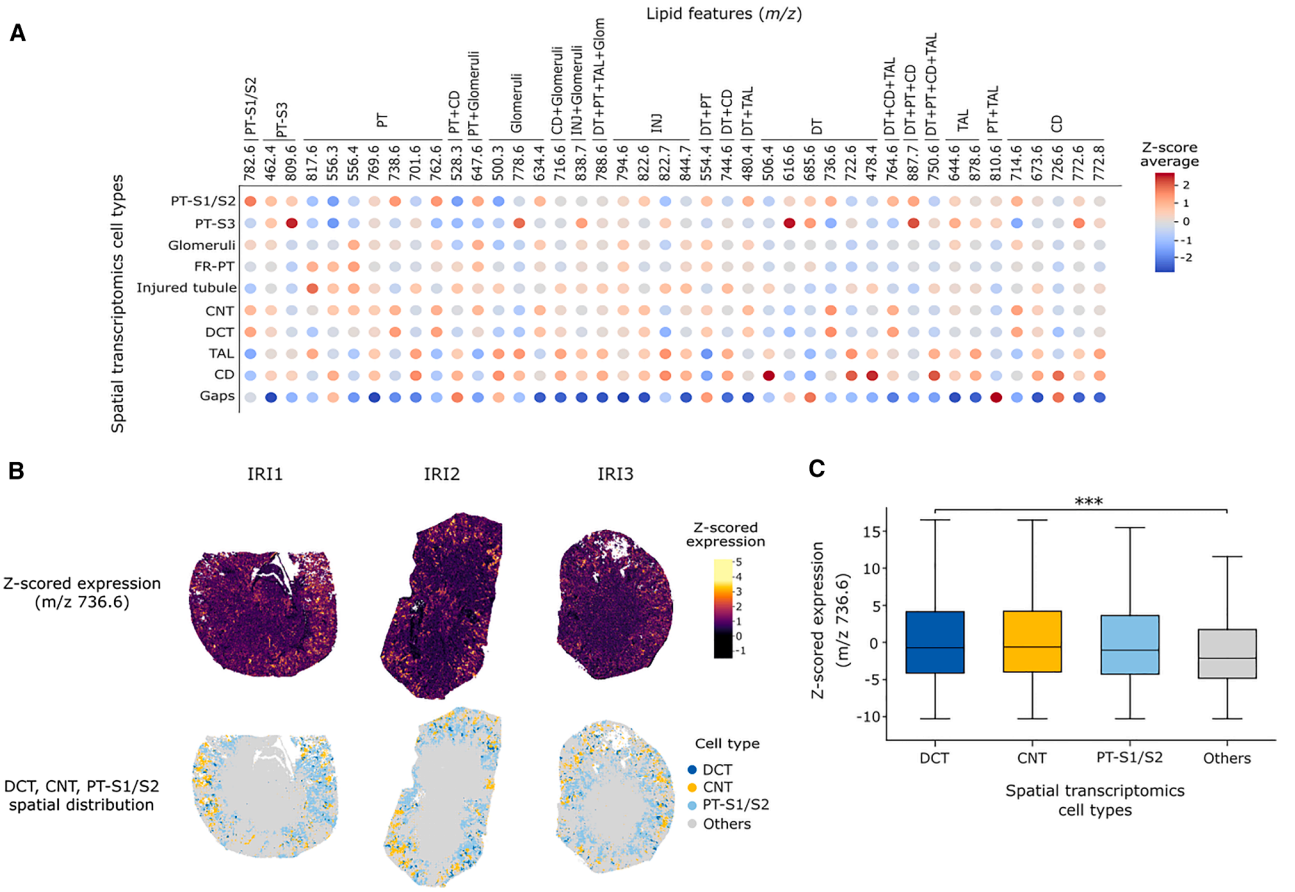


Figure 6. Downstream analysis on multi-modal dataset aligned with COAST

(A) Correspondence between lipid features (columns) and spatial transcriptomics-defined cell types (rows). Color indicates *Z* score enrichment. The top column annotation indicates the cell type for which each lipid feature has been reported as a marker in the literature.

(B) Representative spatial maps showing the localization of lipid feature 736.6 (left) alongside corresponding transcriptomic cell-type annotations (right).

(C) Statistical analysis of lipid feature 736.6 in cell types of interest.

Abbreviations: CD, collecting duct; PT-S1/S2, proximal tubule segment 1/2; PT-S3, proximal tubule segment 3; FR-PT, failed-repair proximal tubule; CNT, connecting tubule; DCT, distal convoluted tubule; TAL, thick ascending limb; INJ, injury.

PT-S3-specific lipid feature *m/z* 736.6 and transcriptomic DCT/CNT and PT-S1/S2 bins showed overlapping spatial distributions (Figure 6B). Furthermore, the expression of m/z 736.6 was significantly higher in DCT compared to the other cell types (Figure 6C). These results support the accuracy of alignment and highlight consistencies between lipid and transcriptomic measurements, while also revealing differences not detectable without spatial integration.

Overall, the downstream analyses carried out in this section using the COAST multi-modal output identified potential metabolite and lipid features that align known anatomical and functional structure of the kidney by reflecting biologically coherent patterns of feature sharing across related nephron sections.

DISCUSSION

We presented COAST, a method to align consecutive tissue sections using only image data. In contrast to existing approaches, COAST does not rely on matched molecular profiles or cell-type annotations, avoiding potential biases and expanding applicability to diverse multi-modal datasets. Despite using only image features for alignment, COAST effectively aligns spatial spots with similar visual characteristics and, notably, achieves high similarity in gene expression between matched spots in the case of uni-modal datasets. It also preserves tissue architecture and minimizes unnecessary spot displacement, as demonstrated by the neighborhood preservation metric. COAST is applicable to different staining types, making it applicable to diverse spatial technologies. This generalizability indicates that image data and the use of pre-trained ssDINO vision transformer can reflect meaningful internal tissue structure for guiding the alignment. Across three quantitative metrics—landmark distance, gene expression similarity, and neighborhood preservation—as well as qualitative visualizations, we showed that COAST performs on par with established unimodal tools and surpasses manual rigid alignment, which is often time consuming and less precise. Finally, we illustrated how COAST output can be used for downstream multimodal analyses,

revealing lipids and metabolites differentially expressed that reflect the nephron structure, such as a strong overlap between CNT and DCT, and recapitulate known injury-associated features, including the enrichment of the injury markers *m/z* 822.7 and 838.7 in PT-S3 compared to PT-S1/S2, consistent with the IRI phenotype reported in literature.

COAST offers a time- and cost-efficient solution that, while requiring GPU resources, runs rapidly. Its modular architecture, comprising three main components, enables flexibility: individual modules (such as the vision transformer or the spot-matching strategy) can be adapted or replaced depending on the technology or study objective.

Finally, establishing a reliable physical alignment between consecutive tissue sections is a critical prerequisite for multi-modal downstream analyses and for drawing biologically meaningful conclusions. COAST fulfills this need by providing accurate physical alignment that preserves tissue architecture and spatial context across modalities, constituting a robust, flexible, and streamlined pipeline that performs strongly without relying on prior knowledge or molecular annotations.

Limitations of the study

One current limitation is that the method benefits greatly from well-defined tissue borders; as a result, performance is reduced when aligning smaller tissue regions or immunofluorescence samples where staining is stronger in the tissue core than at the edges. Moreover, while a satisfactory alignment was ultimately achieved for the multi-modal dataset, it required image preprocessing steps to mitigate platform effects arising from differences between imaging systems. Furthermore, because the framework depends on morphological features for cross-modal bridging, COAST is inherently constrained by the availability of histological images; alignment cannot be performed in the absence of high-resolution imaging data. Crucially, while COAST leverages morphological similarities to bridge modalities, this assumption for alignment is lost if sections are separated by significant anatomical distances, such as those representing distinct developmental time points or distant spatial sections within the same sample.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ahmed Mahfouz (a.mahfouz@lumc.nl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Dataset availability:
  - 10X Mouse Brain Sagittal Posterior (Visium). Section 1: https://www.10xgenomics.com/datasets/mouse-brain-serial-section-1-sagittal-posterior-1-standard-1-0-0. Section 2: https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard.
  - 10X Mouse Brain Coronal (Visium). Section 1: https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-1-ffpe-2-standard. Section 2: https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-2-ffpe-2-standard.
  - 10X Mouse Brain Sagittal Anterior (Visium). Section 1: https://www.10xgenomics.com/datasets/mouse-brain-serial-section-1-sagittal-anterior-1-standard-1-1-0. Section 2: https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-anterior-1-standard.
  - MOSTA Mouse Embryo: https://db.cngb.org/stomics/mosta/download/.
  - Stereo-seq dataset: GSE298899.
  - Stereo-seq images: https://doi.org/10.6084/m9.figshare.30720194.v1. doi: https://doi.org/10.6084/m9.figshare.30720194.
  - MALDI-MSI dataset: https://zenodo.org/records/17722895. doi: https://doi.org/10.5281/zenodo.17722894.
- Code availability:
  - Pretrained ssDINO vision transformer: https://github.com/facebookresearch/dino (Architecture: ViT-S/16).
  - Code for COAST and figure reproducibility is available at https://doi.org/10.5281/zenodo.21454307 and https://github.com/mahfouzlab/COAST.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We acknowledge Manon Zuurmond (Department of Internal Medicine, LUMC, Leiden, the Netherlands) for making the illustrations for this article. The Novo Nordisk Foundation Center for Stem Cell Medicine (reNEW) is supported by Novo Nordisk Foundation grants (NNF21CC0073729). S.J.D. is supported by funding from Agence Nationale de la Recherche (ANR24-CPJ1-0181-01). T.J.R. is supported by the European Union through ERC grant (SPARK 101140863). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

AUTHOR CONTRIBUTIONS

Conceptualization, B.M. and A.M.; methodology, B.M.; software, B.M.; formal analysis, B.M., C.N.R., G.W., N.O., R.G.J.R., M.E.J., C.B., and S.J.D.; investigation, B.M. and A.M.; data curation, B.M.; writing – original draft, B.M.; writing – review and editing, B.M., C.N.R., N.O., A.M., and T.J.R.; visualization, B.M.; supervision, A.M. and T.J.R.; project administration, A.M. and T.J.R.; funding acquisition, T.J.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used Claude in order to edit the text. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ischemic-reperfusion injury (IRI) study
- METHODS DETAILS
  - COAST algorithm
  - Datasets

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2026.101575.

Received: December 19, 2025
Revised: June 9, 2026
Accepted: August 7, 2026

REFERENCES

1. Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. Nat. Rev. Genet. *24*, 494–515. https://doi.org/10.1038/s41576-023-00580-2.
2. Cao, J., Li, C., Cui, Z., Deng, S., Lei, T., Liu, W., Yang, H., and Chen, P. (2024). Spatial Transcriptomics: A Powerful Tool in Disease Understanding and Drug Discovery. Theranostics *14*, 2946–2968. https://doi.org/10.7150/thno.95908.
3. (2024). Method of the Year 2024: spatial proteomics. Nat. Methods *21*, 2195–2196. https://doi.org/10.1038/s41592-024-02565-3.
4. Wang, G., Van Den Berg, B.M., Kostidis, S., Pinkham, K., Jacobs, M.E., Liesz, A., Giera, M., and Rabelink, T.J. (2025). Spatial quantitative metabolomics enables identification of remote and sustained ipsilateral cortical metabolic reprogramming after stroke. Nat. Metab. *7*, 1791–1800. https://doi.org/10.1038/s42255-025-01340-8.
5. Wörheide, M.A., Krumsiek, J., Kastenmüller, G., and Arnold, M. (2021). Multi-omics integration in biomedical research – A metabolomics-centric review. Anal. Chim. Acta *1141*, 144–162. https://doi.org/10.1016/j.aca.2020.10.038.
6. Nunes, J.B., Ijsselsteijn, M.E., Abdelaal, T., Ursem, R., Van Der Ploeg, M., Giera, M., Everts, B., Mahfouz, A., Heijs, B., and de Miranda, N.F.C.C. (2024). Integration of mass cytometry and mass spectrometry imaging for spatially resolved single-cell metabolic profiling. Nat. Methods *21*, 1796–1800. https://doi.org/10.1038/s41592-024-02392-6.
7. Vicari, M., Mirzazadeh, R., Nilsson, A., Shariatgorji, R., Bjärterot, P., Larsson, L., Lee, H., Nilsson, M., Foyer, J., Ekvall, M., et al. (2024). Spatial multimodal analysis of transcriptomes and metabolomes in tissues. Nat. Biotechnol. *42*, 1046–1050. https://doi.org/10.1038/s41587-023-01937-y.
8. Yeo, Y.Y., Chang, Y., Qiu, H., Yiu, S.P.T., Michel, H.A., Wu, W., Jin, X., Kure, S., Parmelee, L., Luo, S., et al. (2024). Same-Slide Spatial Multi-Omics Integration Reveals Tumor Virus-Linked Spatial Reorganization of the Tumor Microenvironment. Preprint at bioRxiv. https://doi.org/10.1101/2024.12.20.629650.
9. Hendriks, T.F.E., Eijkel, G.B., Visvikis, T., Balluff, B., Heeren, R.M.A., and Cuypers, E. (2025). One Section, Two Worlds: Single-Cell Integration of MALDI-MSI and Spatial Transcriptomics on the Same Single Tissue Section. Preprint at bioRxiv. https://doi.org/10.1101/2025.07.16.665110.
10. Ekvall M. Developing Heart Multi Modal 2023 doi:https://doi.org/10.6084/M9.FIGSHARE.22905374.V1
11. Wang, N., Wang, R., Zhang, X., Li, X., Liang, Y., and Ding, Z. (2021). Spatially-resolved proteomics and transcriptomics: An emerging digital spatial profiling approach for tumor microenvironment. Vis. Cancer Med. *2*, 1. https://doi.org/10.1051/vcm/2020002.
12. Casamitjana, A., Lorenzi, M., Ferraris, S., Peter, L., Modat, M., Stevens, A., Fischl, B., Vercauteren, T., and Iglesias, J.E. (2022). Robust joint registration of multiple stains and MRI for multimodal 3D histology reconstruction: Application to the Allen human brain atlas. Med. Image Anal. *75*, 102265. https://doi.org/10.1016/j.media.2021.102265.
13. Paknezhad, M., Loh, S.Y.M., Choudhury, Y., Koh, V.K.C., Yong, T.T.K., Tan, H.S., Kanesvaran, R., Tan, P.H., Peng, J.Y.S., Yu, W., et al. (2020). Regional registration of whole slide image stacks containing major histological artifacts. BMC Bioinf. *21*, 558. https://doi.org/10.1186/s12859-020-03907-6.
14. Tang, Z., Luo, S., Zeng, H., Huang, J., Sui, X., Wu, M., and Wang, X. (2024). Search and match across spatial omics samples at single-cell resolution. Nat. Methods *21*, 1818–1829. https://doi.org/10.1038/s41592-024-02410-7.
15. Liu, X., Zeira, R., and Raphael, B.J. (2023). PASTE2: Partial Alignment of Multi-slice Spatially Resolved Transcriptomics Data. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.08.523162.
16. Klein, D., Palla, G., Lange, M., Klein, M., Piran, Z., Gander, M., Meng-Papaxanthos, L., Sterr, M., Saber, L., Jing, C., et al. (2025). Mapping cells through time and space with moscot. Nature *638*, 1065–1075. https://doi.org/10.1038/s41586-024-08453-2.
17. Jones, A., Townes, F.W., Li, D., and Engelhardt, B.E. (2023). Alignment of spatial genomics data using deep Gaussian processes. Nat. Methods *20*, 1379–1387. https://doi.org/10.1038/s41592-023-01972-2.
18. Huizing, G.J., Samaran, J., Capocefalo, D., Audit, A., Peyré, G., and Cantini, L. (2026). STORIES: learning cell fate landscapes from spatial transcriptomics using optimal transport. Nat. Methods *23*, 522–531. https://doi.org/10.1038/s41592-025-02855-4.
19. Peng, Q., Zhou, P., and Li, T. (2024). stVCR: Spatiotemporal dynamics of single cells. Preprint at bioRxiv. https://doi.org/10.1101/2024.06.02.596937.
20. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.
21. Zhang, M., Pan, X., Jung, W., Halpern, A.R., Eichhorn, S.W., Lei, Z., Cohen, L., Smith, K.A., Tasic, B., Yao, Z., et al. (2023). Molecularly defined and spatially resolved cell atlas of the whole mouse brain. Nature *624*, 343–354. https://doi.org/10.1038/s41586-023-06808-9.
22. Pentimalli, T.M., Schallenberg, S., León-Periñán, D., Legnini, I., Theurillat, I., Thomas, G., Boltengagen, A., Fritzsche, S., Nimo, J., Ruff, L., et al. (2025). Combining spatial transcriptomics and ECM imaging in 3D for mapping cellular interactions in the tumor microenvironment. Cell Syst. *16*, 101261. https://doi.org/10.1016/j.cels.2025.101261.
23. Argelaguet, R., Cuomo, A.S.E., Stegle, O., and Marioni, J.C. (2021). Computational principles and challenges in single-cell data integration. Nat. Biotechnol. *39*, 1202–1215. https://doi.org/10.1038/s41587-021-00895-7.
24. Bogovic, J.A., Hanslovsky, P., Wong, A., and Saalfeld, S. (2016). Robust registration of calcium images by learned contrast synthesis. In 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) (IEEE), pp. 1123–1126. https://doi.org/10.1109/ISBI.2016.7493463.
25. Ekvall, M., Bergenstråhle, L., Andersson, A., Czarnewski, P., Olegård, J., Käll, L., and Lundeberg, J. (2024). Spatial landmark detection and tissue registration with deep learning. Nat. Methods *21*, 673–679. https://doi.org/10.1038/s41592-024-02199-5.
26. Manukyan, A., Bahry, E., Wyler, E., Becher, E., Pascual-Reguant, A., Plumbom, I., Dikmen, H.O., Elezkurtaj, S., Conrad, T., Altmüller, J., et al. (2023). VoltRon: A Spatial Omics Analysis Platform for Multi-Resolution and Multi-omics Integration using Image Registration. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.15.571667.
27. Pfaendler, R., Hanimann, J., Lee, S., and Snijder, B. (2023). Self-supervised vision transformers accurately decode cellular state heterogeneity. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.16.524226.
28. Song, A.H., Chen, R.J., Ding, T., Williamson, D.F.K., Jaume, G., and Mahmood, F. (2024). Morphological Prototyping for Unsupervised Slide

Representation Learning in Computational Pathology. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Seattle, WA, USA: IEEE), pp. 11566–11578. https://doi.org/10.1109/CVPR52733.2024.01099.

29. Wang, G., Heijs, B., Kostidis, S., Mahfouz, A., Rietjens, R.G.J., Bijkerk, R., Koudijs, A., van der Pluijm, L.A.K., van den Berg, C.W., Dumas, S.J., et al. (2022). Analyzing cell-type-specific dynamics of metabolism in kidney repair. Nat. Metab. *4*, 1109–1118. https://doi.org/10.1038/s42255-022-00615-8.
30. Miguel, V., Shaw, I.W., and Kramann, R. (2025). Metabolism at the crossroads of inflammation and fibrosis in chronic kidney disease. Nat. Rev. Nephrol. *21*, 39–56. https://doi.org/10.1038/s41581-024-00889-z.
31. Tang, W., and Wei, Q. (2023). The metabolic pathway regulation in kidney injury and repair. Front. Physiol. *14*, 1344271. https://doi.org/10.3389/fphys.2023.1344271.
32. Polonsky, M., Gerhardt, L.M.S., Yun, J., Koppitch, K., Colón, K.L., Amrhein, H., Wold, B., Zheng, S., Yuan, G.C., Thomson, M., et al. (2024). Spatial transcriptomics defines injury specific microenvironments and cellular interactions in kidney regeneration and disease. Nat. Commun. *15*, 7010. https://doi.org/10.1038/s41467-024-51186-z.
33. Li, W., Liu, X., Li, H., Zeng, J., Chen, Y., and Xu, B. (2024). Metabolomic and transcriptomic insights into the mechanisms of renal ischemia-reperfusion injury progression. Sci. Rep. *14*, 30101. https://doi.org/10.1038/s41598-024-81600-x.
34. Zhang, D., Xing, Y., Li, W., Yang, F., Lang, Y., Yang, J., and Liu, Z. (2018). Renal tubules transcriptome reveals metabolic maladaption during the progression of ischemia-induced acute kidney injury. Biochem. Biophys. Res. Commun. *505*, 432–438. https://doi.org/10.1016/j.bbrc.2018.08.111.
35. Rietjens, R.G.J., Manzato, B., Liao, Y., Sol, W.M.P.J., Mahfouz, A., and Giera, M., van den Berg B.M., Dumas S.J., Wang G., Rabelink T.J.. (2026). Impaired Metabolic Remodeling Characterizes Injured Niches during Post-ischemic Kidney Repair. J Am. Soc. Nephrol. https://doi.org/10.1681/ASN.0000001216.
36. Farrow, M.A., Tideman, L.E.M., Neumann, E.K., Migas, L.G., Patterson, N.H., Colley, M.E., Allen, J.L., Pingry, E.L., Dufresne, M., Yang, H., et al. (2025). A lipid atlas of the human kidney. Sci. Adv. *11*, eadu3730. https://doi.org/10.1126/sciadv.adu3730.
37. Kirita, Y., Wu, H., Uchimura, K., Wilson, P.C., and Humphreys, B.D. (2020). Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. Proc. Natl. Acad. Sci. *117*, 15874–15883. https://doi.org/10.1073/pnas.2005477117.
38. Li, H., Dixon, E.E., Wu, H., and Humphreys, B.D. (2022). Comprehensive single-cell transcriptional profiling defines shared and unique epithelial injury responses during kidney fibrosis. Cell Metab. *34*, 1977–1998.e9. https://doi.org/10.1016/j.cmet.2022.09.026.
39. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. (2021). Emerging Properties in Self-Supervised Vision Transformers. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (Montreal, QC, Canada: IEEE), pp. 9630–9640. https://doi.org/10.1109/ICCV48922.2021.00951.
40. Pippin, J.W., Sparks, M.A., Glenn, S.T., Buitrago, S., Coffman, T.M., Duffield, J.S., Gross, K.W., and Shankland, S.J. (2013). Cells of Renin Lineage Are Progenitors of Podocytes and Parietal Epithelial Cells in Experimental Glomerular Disease. Am. J. Pathol. *183*, 542–557. https://doi.org/10.1016/j.ajpath.2013.04.024.
41. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition [Internet] (IEEE), pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
42. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., and Chanan, G. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Preprint at arXiv. https://doi.org/10.48550/ARXIV.1912.01703.
43. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.
44. Stringer, C., and Pachitariu, M. (2025). Cellpose3: one-click image restoration for improved cellular segmentation. Nat. Methods *22*, 592–599. https://doi.org/10.1038/s41592-025-02595-5.
45. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.
46. Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., and Muller, K.E. (1990). Contrast-limited adaptive histogram equalization: speed and effectiveness. In [1990] Proceedings of the First Conference on Visualization in Biomedical Computing (IEEE Comput. Soc. Press), pp. 337–345. https://doi.org/10.1109/VBC.1990.109340.
47. Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom. Bioinform. *2*, lqaa078. https://doi.org/10.1093/nargab/lqaa078.

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Mouse kidney samples (14 days post-IRI)	N/A	N/A
Chemicals, peptides, and recombinant proteins		
N-(1-naphthyl) ethylenediamine dihydrochloride (NEDC)	Sigma Aldrich	N/A
Deposited data		
MALDI-MSI kidney sections - Ischemia reperfusion injury mice model (day 14)	This paper	https://zenodo.org/records/17722895
High-resolution kidney sections images - Ischemia reperfusion injury mice model (day 14)	This paper	https://doi.org/10.6084/m9.figshare.30720194.v1
Stereo-seq Ischemia reperfusion injury mice model (14 days)	Rietjens et al.[35]	GEO dataset: GSE298899
10X Mouse Brain Sagittal Posterior (Visium) - Section 1	10X Genomics	https://www.10xgenomics.com/datasets/mouse-brain-serial-section-1-sagittal-posterior-1-standard-1-0-0
10X Mouse Brain Sagittal Posterior (Visium) - Section 2	10X Genomics	https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard
10X Mouse Brain Coronal (Visium) - Section 1	10X Genomics	https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-1-ffpe-2-standard
10X Mouse Brain Coronal (Visium) - Section 2	10X Genomics	https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-2-ffpe-2-standard
10X Mouse Brain Sagittal Anterior (Visium) - Section 1	10X Genomics	https://www.10xgenomics.com/datasets/mouse-brain-serial-section-1-sagittal-anterior-1-standard-1-1-0
10X Mouse Brain Sagittal Anterior (Visium) - Section 2	10X Genomics	https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-anterior-1-standard
10X Human Lung Cancer (Xenium)	10X Genomics	https://www.10xgenomics.com/datasets/xenium-human-lung-cancer-post-xenium-technote
10X Human Ovarian Cancer (Xenium)	10X Genomics	https://www.10xgenomics.com/datasets/xenium-comparison-fresh-frozen-human-ovarian-cancer
MOSTA Mouse Embryo	Chen et al.[20]	https://db.cngb.org/stomics/mosta/download/
Experimental models: Organisms/strains		
Model organism: B6.Ren1cCre/TdTomato/J male mouse	N/A	N/A
Software and algorithms		
SCiLS Lab (2024b Pro)	SCiLS, Bruker Daltonics	https://www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/scils-lab.html
Pretrained ssDINO Vision Transformer	Caron et al.[39]	https://github.com/facebookresearch/dino
COAST algorithm	This paper	https://github.com/mahfouzlab/COAST https://doi.org/10.5281/zenodo.21454307

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CAST algorithm	Tang et al.[14]	https://github.com/wanglab-broad/CAST
MOSCOT algorithm	Klein et al.[16]	https://github.com/theislab/moscot
Fiji/ImageJ Big Warp plug-in	Bogovic et al.[24]	https://imagej.net/plugins/bigwarp
Other		
Indium-tin-oxide (ITO)-coated glass slides	VisionTek Systems Ltd	ITO coated glass (ITOGLASS 12), ITO resistance is 12–15 ohms/sq, Glass size 25 × 75mm, Glass thickness 1.1mm https://www.visionteksystems.co.uk/ito-glass.html
HTX M3+ Sprayer	HTX Technologies, USA	https://www.htximaging.com/htx-m3-sprayer
RapifleX MALDI-TOF/TOF system	Bruker Daltonics GmbH	https://www.bruker.com/en/products-and-solutions/mass-spectrometry/maldi-tof/rapiflex.html
STOmics Stereo-seq V1.2	STOmics	Catalog number: 211ST114 https://en.stomics.tech/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ischemic-reperfusion injury (IRI) study

The experiment was carried out as in Rietjens et al.[35] and Wang et al.[4] Experiments consisted of three 12-week-old male constitutional renin reporter (B6.Ren1cCre/TdTomato/J) mice.[40] The renal vasculature was clamped for approximately 18 min. Clamp removal allowed blood flow to resume (reperfusion) and the abdomen was closed. At day 14 post-surgery, mice were sacrificed by perfusion with cold PBS-heparin (5 UI/mL) via the left ventricle for 6 min at a controlled pressure of 150 mmHg before removing the kidney. MALDI-MSI and Stereo-seq sections were taken 10 μm apart. Animal experiments were approved by the Ethical Committee on Animal Care and Experimentation of the Leiden University Medical Center (permit no. AVD1160020171145). The Stereo-seq data generation is described in detail in Rietjens et al.[35]

MALDI-MSI tissue preparation

Tissue slices were embedded in 10% gelatin and cryosectioned into 10 μm thick sections using a cryostat at −20°C. The sections were thaw-mounted onto indium-tin-oxide (ITO)-coated glass slides (VisionTek Systems Ltd., Chester, UK). Mounted sections were placed in a vacuum freeze-dryer for 15 min prior to matrix application. After drying, N-(1-naphthyl) ethylenediamine dihydrochloride (NEDC) (Sigma-Aldrich, UK) MALDI-matrix solution of 7 mg/mL in methanol/acetonitrile/deionized water (70, 25, 5 %v/v/v) was applied using an HTX M3+ Sprayer (HTX Technologies, USA). The setting was as below: temperature, 60°C; number of passes, 20 layers; flow rate, 80 μL/min; velocity, 2000 mm/min; track spacing, 3 mm; gas flow rate 10 psi, and drying time in between passes, 30 s.

MALDI-MSI measurement and data analysis

MALDI-TOF/TOF-MSI was performed using a RapifleX MALDI-TOF/TOF system (Bruker Daltonics GmbH, Bremen, Germany). Negative ion-mode mass spectra were acquired at a pixel size of 10 × 10 μm^2 over a mass range from m/z 80–1000. Prior to analysis the instrument was externally calibrated using red phosphorus. Spectra were acquired with 50 laser shots per pixel at a laser repetition rate of 10 kHz. Data acquisition was performed using flexControl (Version 4.0, Bruker Daltonics) and visualizations were obtained from flexImaging 5.0 (Bruker Daltonics).

MSI data were exported and processed in SCiLS Lab 2024b pro (SCiLS, Bruker Daltonics) with baseline correction using convolution algorithm. All MALDI-TOF-MSI data were normalized to the Root-Mean-Square (RMS). High-molecular-weight features (m/z > 400, predominantly phospholipids) were picked at signal-to-noise-ratio >3 on the average spectrum, and matrix peaks were excluded from the m/z feature list. The m/z values from MALDI-TOF were imported into the Human Metabolome Database (https://hmdb.ca/) after recalibration in mMass and annotated for metabolites with an error < ±20 ppm. Peak intensities of the selected features were exported for all the measured pixels from SCiLS Lab, which were used for the following analysis.

METHODS DETAILS

COAST algorithm

COAST takes as an input the high-resolution images associated with the spatial data and the spatial count matrices, where the x and y coordinates are present in the metadata and match the coordinates of the image. The method includes three main modules.

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

Feature extraction

Initially, both images are divided into tiles, that are obtained by taking a square of pixels around each spatial unit centroid, taken from the x and y columns in the metadata, as well as a grid-like manner across the whole image, including the background (“in-tissue” and “grid” tiles, respectively). *Tile_size* parameter controls the size of the tile in pixels and should reflect the capture area (in pixels) of the spatial units in the dataset (i.e., the pixel gap between their centroids). *Step_size* parameter controls how much the sliding window moves between consecutive tiles in the grid-search; it should be chosen to take a number of tiles from the background proportional to the overall image resolution and should be large enough to avoid generating unnecessary tiles and slow down computations. The tile and step sizes were selected to align with the varying spatial resolutions of each platform, ensuring the captured regions matched the underlying distribution of the biological data. For high-resolution datasets, tile sizes were calibrated to encapsulate the local cellular neighborhood while maintaining a manageable computational load. The parameters chosen for each dataset are specified in Table S1. Tiles are resized to 224 $\times$ 224 to be compatible with the input of the vision transformer and image features are subsequently extracted using an ImageNet-pre-trained Vision Transformer (ssDINO, self-supervised self-distillation with no labels)[39,41] resulting in a dataset with 384 image features for each tile. The method has an optional argument to add a 300 pixels padding to one side of the image to help the alignment. Padding has been added on the left side to the three multi-modal IRI datasets.

Alignment

Building on CAST, a method for alignment of uni-modal consecutive sections, we use the module CAST Stack to map the coordinates of the query section in the reference coordinate system based on the similarity of the 384-dimensional transformer embeddings. First, a gradient-descent-based rigid alignment (translation, rotation, scaling and reflection) is applied to have both sections in the same scale and coordinate system, followed by a free-form deformation (FFD), a robust non-linear warping approach, to handle local morphological differences among tissue samples. Instead of implementing the alignment by satisfying every cell at its optimum, CAST Stack prioritizes preserving biologically meaningful tissue structure, through the optimization of molecular data. In our case, we replace the molecular data in this step with the image embeddings obtained from the vision transformer.

Due to the large number of tiles in high-resolution images we use the function CAST.sub_data_extract to select a random subset of tiles (20,000) and CAST Stack is applied on the dataset at a reduced size. After generating the transformation parameters, we then apply the calculated alignment transformation to all cells in the original sample at the full resolution.

Spot matching

After alignment, only the tiles centered around molecular data (“in-tissue” tiles) are retained by filtering out tiles originally obtained through grid-like search (“grid” tiles). Each spot in the reference section was matched to its closest spatial unit in the query section by querying the nearest neighbor in the *transformed* spatial coordinates, using a k-d tree data structure for efficient Euclidean distance search. A match was retained only if the closest spot fell within a defined maximum distance threshold; spots without a valid match within this threshold were excluded from further multimodal integration. This maximum distance (max_distance) serves as a key hyperparameter and should be chosen to reflect the physical distance, measured in image pixels, between the centroids of adjacent spatial units, based on the resolution and layout of the specific spatial transcriptomics platform (Table S1). If a query spot maps to more than one reference spot or if the query section has a higher spatial resolution than the reference, this causes the reference spots to be duplicated, resulting in an “artificial upscaling” of the reference spatial units. The final output is a set of matched spot pairs across modalities, which enables the creation of a multimodal (or combined uni-modal) feature matrix combining molecular information across aligned spatial locations (Figure 1C).

Image features clustering and visualization

Following the feature extraction module, we performed unsupervised clustering on the image-derived features from each pair of tissue sections. The feature vectors from both sections were concatenated into a single matrix and standardized using Z score normalization. K-means clustering was then applied to the combined dataset. Using the associated spatial coordinates, we visualized the resulting clusters by assigning each a unique color and displaying them as scatterplots.

Implementation

COAST was implemented in Python 3.10.14 and developed to run on a CUDA-enabled GPU. Key libraries used included PyTorch for model loading,[42] Hugging Face for the DINO Vision Transformer backbone, CAST for the alignment step,[14] torchvision and PIL for image handling and numpy and pandas for data manipulation.

Method scalability

COAST is scalable and applicable to a wide range of dataset resolutions. Table S2 reports the runtime for all the COAST modules (feature extraction, CAST alignment, spot matching and aligned dataset saving and results visualization) on all datasets used in this study. All experiments were run on GPUs on the Leiden University Medical Center Cluster (SHARK) (specifically Tesla T4 (16 GB), TITAN Xp (12 GB), GRID V100D-16Q (16 GB) and Quadro RTX 6000 (24 GB)).

Datasets

Visium Mouse Brain Sagittal Posterior dataset

Fresh frozen adult mouse brain tissue sections were processed using the 10X Genomics Visium Gene Expression protocol (Spatial 3′ v1). The two consecutive slides were sectioned in the sagittal orientation (posterior region). Gene expression was quantified for 32,285 genes.

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

The dataset was preprocessed using Scanpy.[43] Cells with fewer than 5,000 total counts or more than 35,000 total counts were removed, and genes expressed in fewer than 10 cells were excluded from downstream analysis. Following filtering, normalization and log-transformation, we performed PCA, neighbor graph construction, UMAP embedding, and unsupervised clustering separately on the two sections using the Leiden algorithm (default resolution 1). Spatial visualization of the resulting clusters was conducted using high-resolution histology images overlaid with spatial coordinates. The cluster overlapping the cerebellum fold was manually identified and retained for further analysis. This cluster contained 197 spots in section 1 and 201 spots in section 2. To generate the plot in Figure 2C, we identified the tiles corresponding to the cerebellar fold cluster, as determined from clustering the image-derived features (Figure S2A, cluster 1).

Visium Mouse Brain Coronal dataset

Fresh frozen adult mouse brain tissue sections were processed using the 10X Genomics Visium CytAssist Gene Expression protocol (Visium V4 Slide - FFPE v2). The consecutive slides were sectioned in the coronal orientation. Gene expression was quantified for 32,285 genes. Preprocessing was performed in the same way as 10X Genomics Mouse Brain - Sagittal Posterior.

Visium Mouse Brain Sagittal Anterior dataset

Fresh frozen adult mouse brain tissue sections were processed using the 10X Genomics Visium Gene Expression protocol. The two consecutive slides were sectioned in the sagittal orientation (anterior region). Gene expression was quantified for 32,285 genes. Preprocessing was performed in the same way as 10X Genomics Mouse Brain - Sagittal Posterior.

Xenium Human Lung Cancer

Two consecutive sections of human lung adenocarcinoma tissue (FFPE) were used: slide 1 (Xenium v1) using the Xenium Human Lung Gene Expression Panel with nuclear expansion and consecutive slide 2 using Xenium Prime 5K *In Situ* Gene Expression with Cell Segmentation and the Xenium Prime 5K Human Pan Tissue and Pathways Panel. The size of the overlapping feature set is 196 genes. We applied quality control filtering to both sections by removing low-quality cells with fewer than 5 detected genes and excluding genes expressed in fewer than 5 cells. Both sections were normalized and log-transformed.

Xenium Human Ovarian Cancer

Two consecutive sections of human ovarian cancer (FF) were used: slide 1 (Xenium v1) using Xenium v1 *In Situ* Gene Expression with Cell Segmentation and the Xenium Human Multi-Tissue and Cancer Panel plus 100 custom genes and consecutive slide 2 (Xenium Prime 5K) using *In Situ* Gene Expression with Cell Segmentation data and the Xenium Prime 5K Human Pan Tissue and Pathways Panel.

The size of the overlapping feature set is 278 genes. Preprocessing was performed in the same way as 10X Genomics Human Lung Cancer.

MOSTA Mouse Embryo (Stereo-seq platform)

We analyzed the E16.5 embryonic mouse dataset, previously published by Chen et al.[20] The data were acquired using the Stereo-seq platform on frozen consecutive sections spaced several hundred micrometers apart, each 10 μm thick. We used sections E16.5_E2S6 (125,090 cells) and E16.5_E2S7 (128,143 cells), as they are the only consecutive sections with associated nuclei acid staining images available. Both sections measured gene expression across 27,574 genes. The authors of the paper normalized and log-transformed the raw count matrix using Scanpy[43] (sc.pp.normalize_total and sc.pp.log1p). Moreover, anatomical region annotations were available and were used for Figure 3G (Table of markers available on MOSTA dataset website). Since the Stereo-seq data were provided in bin-50 resolution, we reprocessed the spot coordinates to align them with pixel-based image coordinates, using the Bin1 matrix files available on the dataset website (E16.5_E2S6_GEM_bin1.tsv.gz, E16.5_E2S7_GEM_bin1.tsv.gz).

IRI Stereo-seq dataset

The design, tissue segmentation, and preprocessing of the Stereo-seq experiments, including cell-type annotation, were performed as described in Rietjens et al.[35] Importantly, the spatial coordinates, and the corresponding count matrix, were binned using 40 as bin size (20 μm × 20 μm resolution), resulting in new spot identifiers. The final Stereo-seq datasets contained 48,347, 40,918, and 40,494 spatial spots for IRI1, IRI2, and IRI3, respectively. The image associated with the Stereo-seq experiment comprised all three tissue sections and was segmented into individual regions for downstream analysis, yielding image sizes of 12,659 × 11,336 pixels for IRI1, 9,710 × 15,063 pixels for IRI2, and 10,444 × 12,561 pixels for IRI3. The three replicates were corrected for batch effect using rPCA.

In Figure 5B, we observe that the ssDNA tissue images from Stereo-seq exhibit artifacts (a grid-like pattern of squares resulting from the imaging process). To address this, we performed cell segmentation on the Stereo-seq image and applied a binary mask, where detected cells are marked in white and all other areas in black (Figure S7). More in detail, whole-slide mouse kidney tissue DAPI-stained images were processed using a custom nuclei segmentation pipeline implemented in Python 3. The pipeline employs the nuclei model of Cellpose3,[44] combined with a tile-based strategy, in which the image is split into overlapping tiles, for scalable and adaptive mask extraction. In the preprocessing stage non-tissue background pixels were manually zeroed using Fiji.[45] The images are enhanced using contrast-limited adaptive histogram equalization (CLAHE)[46] to improve nucleus-background separability and then split into overlapping square tiles (size = 512 × 512 pixels, overlap = 0.2). Tile size and overlap are determined by the parameters "tile_side_length" and "tile_overlap," set to 512 and 0.2, respectively. Tiling enables memory-efficient processing of arbitrarily large images and supports adaptive parameterization. WIn this study, we utilized this adaptability for Cellpose3's diameter parameter, which is estimated from the median equivalent diameter of objects initially segmented via flow fields and conveys to the model what average nuclei diameters to expect. To ensure that nuclei are not split across tile borders, the overlap must be large enough to fully capture even the largest nuclei but not so large as to cause the borders of non-adjacent tiles to approach each other. Consequently, we chose a tile size of 512 pixels to balance accuracy, memory load, and estimation stability in our systems and images. Tiles

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

are processed in batches through the Cellpose3 "nuclei" model. To increase sensitivity, the cell probability threshold was lowered from the default 0 to −14, CLAHE contrast was intensified (clip limit = 5.0, grid size = 32 × 32), and the flow threshold was raised from the default 0.4 to 0.9, requiring more coherent flow fields for a nucleus to be accepted. This parameter combination implements a sensitivity-specificity tradeoff, with the flow threshold acting as the principal specificity regulator. Parameter settings were determined by evaluating performance on annotated image subsets, using Intersection-over-Union (IoU) as the criterion for segmentation quality. For adaptive diameter estimation to function, diameter was set to None and resample to True. All other Cellpose3 parameters remained at their default values.

To reconstruct the full mask set for the original image, a custom tile-merging algorithm was applied. For each overlapping tile pair, the following four steps were taken sequentially.

1. Priority Tile Selection: The tile with the greater number of detected nuclei is designated as the priority tile to minimize inclusion of edge-adjacent masks, which are more likely to be affected by border artifacts.
2. Priority Tile Border Mask Removal: The location of the priority tile's edge within the non-priority tile is identified. All priority tile masks intersecting this boundary are removed to avoid including potentially distorted nuclei.
3. Preservation of Boundary Non-Priority Masks: Non-priority tile masks intersecting the corresponding boundary are marked down to be preserved.
4. Removal of Redundant Non-Priority Masks: All non-priority tile masks within the overlapping region that were not marked for preservation are discarded.

This process ensures that masks in the overlapping area originate only from the priority tile, except along its border, where non-priority masks are used to avoid edge artifacts. The method effectively prevents merging into oversized masks, duplication, and artifacts from the tile creation while being computationally efficient, even on CPU systems.

Finally, we used an optional filtering script to evaluate each mask based on its circularity, eccentricity, size, solidity, hole fraction, and aspect ratio. Masks with values outside user-defined thresholds were removed from the final output. We used the following threshold: min_pixels = 20, max_pixels = 900, min_circ = 0.56, max_circ = 1.00, min_sol = 0.765, max_sol = 1.00, min_ecc = 0.00, max_ecc = 0.975, min_ar = 0.50, max_ar = 3.20, min_hole = 0.00, max_hole = 0.001.

IRI MALDI-MSI dataset

The final datasets included 240,255 spots for IRI1, 161,596 for IRI2, and 165,870 for IRI3, at 10 × 10 μm resolution. Deisotoping was performed manually in SCiLS software. After deisotoping, a total of 301 lipid and 285 metabolite features were retained for all three tissues. The corresponding image dimensions were 4,352 × 5,120 pixels for IRI1, 4,352 × 4,864 pixels for IRI2, and 3,840 × 4,352 pixels for IRI3. The three replicates were corrected for batch effect with combat (implemented in scanpy.pp.combat).[47]

Benchmark methods

Rigid alignment

We performed rigid alignment in Python by manually translating, rotating and/or scaling the x and y coordinates of the spatial units of the query section to obtain an optimal visual alignment to the coordinates of the reference section (Table S3).

MOSCOT

We ran MOSCOT with default parameters. For all spatial transcriptomics datasets, alignment was performed using both affine and warp transformations separately and the alignment yielding the most accurate visual concordance was selected for downstream analysis. Final coordinates for the Mouse Brain Coronal dataset were obtained using the warp transformation, while affine transformation was used for all the other datasets in the study. For the 10X Xenium datasets, cells were subsampled to keep 30% of the original data due to memory constraints. For this reason the landmark distance metrics could not be computed on the MOSCOT alignment on the Xenium datasets.

CAST

CAST Mark and Stack were run on the uni-modal benchmarking datasets with the default parameters. The MOSTA Mouse Embryo dataset has been downsized with the function CAST.sub_data_extract (20,000 spots).

For all benchmarked methods, spatial units were matched between the query and reference in the same manner as the spot matching procedure described in the COAST algorithm section (see Methods). For the comparison to unaligned sections, given the uni-modal consecutive sections are in the same coordinate scale (i.e., same tissue coordinate magnitude), no transformation was applied. The two consecutive IRI kidneys are in completely different coordinate spaces therefore a center of mass transformation was applied to the coordinate of the MSI tissues in order to overlap the Stereo-seq tissues to some extent.

QUANTIFICATION AND STATISTICAL ANALYSIS

COAST evaluation metrics

Landmark alignment accuracy

To evaluate the accuracy of spatial alignment across different methods, we measured the Euclidean distances between manually annotated anatomical landmarks and their corresponding points after alignment. For each alignment method, we first assessed

Please cite this article in press as: Manzato et al., Image-guided alignment of consecutive multi-modal tissue slides, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101575

the normality of the distance distributions using the Shapiro-Wilk test. Depending on the outcome, we selected either a one-way ANOVA (for normally distributed data) or a non-parametric Kruskal-Wallis test (for non-normal data) to compare alignment performance across groups. A significance level of 0.05 was used throughout.

Following a significant result in the group-level test, we performed pairwise comparisons using Dunn's test with Bonferroni correction to adjust for multiple testing. We computed Cohen's d for all pairwise comparisons. To visualize the results, we generated boxplots for each alignment method, overlaying individual data points with jitter to illustrate the spread. Significant differences between groups were annotated using asterisk notation to indicate the level of statistical significance ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***).

Gene expression similarity

For each matched pair of spots, bins or cells, we extracted gene expression vectors from the molecular data for both sections and evaluated their similarity using cosine similarity. We first identified highly variable genes of the reference section using log-transformed expression values in Scanpy (sc.pp.highly_variable_genes).[43] Cosine similarity was computed between the highly variable gene expression vectors for each spot pair. This analysis was performed across five different alignment methods (COAST, Unaligned, Rigid alignment, MOSCOT, CAST). Results were visualized using a kernel density estimation plot (*seaborn.kdeplot*).

Neighborhood preservation score

We quantified neighborhood preservation by comparing spatial neighborhoods before and after alignment. For the query section (pre-alignment and post-alignment), we identified the nearest neighbors of each spot using a k-nearest neighbors approach with k set to 24 for Visium and Xenium data and k set to 48 for Stereo-seq. Specifically, for each spot, the k closest neighboring spots were determined based on Euclidean distance in the two-dimensional coordinate space, using the KD-tree algorithm. The distribution of overlapping neighbors across all spots was visualized using a bar plot showing the frequency of spots by the count of overlapping neighbors.

IRI dataset downstream analysis

Stereo-seq cell type - m/z marker analysis

Lipid cell-type markers were found in literature. More in detail, from Farrow et al. 2025[36] we selected the top 10 markers per cell type (glomeruli, TAL, PT, Distal, Collecting duct) that differed no more than ±0.1 m/z from our detected features. From Wang et al. 2022[29] we retained all markers within a 0.1 m/z tolerance. Average lipid marker expression per cell type was visualized using custom heatmaps. For each feature, expression values were extracted from the MSI data, averaged across cell types (cell-type annotation obtained as described in Rietjens et al. 2025[35]), and standardized by Z score transformation.

MALDI-MSI differential expression analysis

Differential gene expression analysis was performed using the Wilcoxon rank-sum test implemented in Scanpy (sc.tl.rank_genes_groups), grouping cells by transcriptomics-derived cell type annotations. The top 10 marker genes per cell type were identified based on ranked test statistics, using root mean squared-normalized expression values.