

SPRINT enables spatial proteomic profile reconstruction through multimodal fusion of histology and transcriptomics

Yadong Wang

ydwang@hit.edu.cn

Harbin Institute of Technology <https://orcid.org/0000-0001-6500-6217>

Yuzhi Sun

Harbin Institute of Technology

Xian Liu

Harbin Institute of Technology

Tianyi Zhao

Harbin Institute of Technology <https://orcid.org/0000-0003-1913-081X>

Article

Keywords:

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10783563/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

SPRINT enables spatial proteomic profile reconstruction through multimodal fusion of histology and transcriptomics

Yuzhi Sun¹, Xian Liu¹, Tianyi Zhao^{2,3,*}, Yadong Wang^{1,2,*}

Abstract

Spatially resolved protein expression is essential for characterizing tissue organization and immune microenvironments, yet existing computational inference methods predominantly rely on either histological images or gene expression alone, limiting their ability to recover proteins that are weakly associated with a single modality. Here we present SPRINT, a multimodal framework that reconstructs spatial proteomic profiles by jointly leveraging spatial gene expression and accompanying H&E-stained histological images. SPRINT combines resolution-adaptive image encoders with token-level self-attention to fuse morphological and transcriptional information. Across human tonsil, mouse spleen and human brain tissues, SPRINT reconstructed protein-specific spatial patterns across sections, donors and spatially disjoint regions. Its compatibility with different image resolutions, transcriptomic gene and assay configurations broadens its applicability across spatial transcriptomics datasets. Ablation analyses showed that multimodal fusion compensated for proteins with weak associations to either tissue morphology or transcript abundance, thereby expanding the range of immune-microenvironment proteins that could be reliably inferred. SPRINT further supported cross-species protein reconstruction, and the reconstructed profiles improved the recovery of spatial tissue domains. The framework was also extendable beyond proteomics, reconstructing spatial metabolite distributions in an independent tissue section when trained directly on metabolomic measurements. Together, these results establish SPRINT as a general framework for completing spatial molecular profiles across tissue sections, datasets, image-resolution and species, while enabling downstream characterization of tissue organization.

Main

Spatially resolved proteomic profiling provides a direct approach for characterizing the molecular organization of tissues by revealing the distribution of proteins within their native anatomical contexts[1]. Proteins represent functional executors of cellular states and often exhibit regulatory patterns that cannot be inferred solely from transcript abundance[2], [3], making spatial proteomics essential for dissecting cellular phenotypes, tissue architecture and disease-associated microenvironments[1]. Recent advances in imaging-

* Corresponding authors: Tianyi Zhao (zty2009@hit.edu.cn), Yadong Wang (ydwang@hit.edu.cn).

¹ Faculty of Computing, Harbin Institute of Technology, Harbin, China.

² Harbin Institute of Technology Zhengzhou Research Institute, Zhengzhou, China.

³ School of Medicine and Health, Harbin Institute of Technology, Harbin, China.

29 based and sequencing-based spatial technologies, including spatially resolved antibody-based
30 measurements[4], [5], [6], [7] and multimodal platforms such as spatial CITE-seq[8], [9], have enabled
31 simultaneous investigation of transcriptomic and protein landscapes at single-cell or near single-cell
32 resolution. However, despite their biological value, spatial proteomic technologies remain substantially less
33 accessible than spatial transcriptomic approaches because of the high cost, limited throughput and
34 experimental complexity associated with multiplex protein measurement[1], [10]. These limitations motivate
35 the development of computational approaches that can reconstruct spatial proteomic landscapes from more
36 readily available tissue images or spatial transcriptomic measurements.

37 Existing approaches can be broadly classified into image-driven and transcriptome-driven strategies. Image-
38 based methods, including GigaTIME[11], HEX[12], GHIST[13], HIPI[14] and ROSIE[15], leverage
39 morphological features extracted from haematoxylin and eosin (H&E)-stained tissue images to infer spatial
40 protein abundance. However, H&E images primarily capture macroscopic and cellular morphological
41 characteristics, whereas many proteins involved in signalling pathways, immune regulation and intracellular
42 processes exhibit limited morphological signatures. As a result, image-only models may preferentially
43 reconstruct H&E-informative proteins while failing to accurately recover proteins whose spatial variation is
44 weakly associated with tissue morphology. Conversely, transcriptome-based approaches, including
45 sciPENN[16], scTranslator[17] and DGAT[18], infer protein abundance from RNA expression profiles by
46 exploiting gene-protein associations. Although transcriptomic information provides complementary
47 molecular context, protein abundance is not determined exclusively by transcriptional regulation, and
48 substantial post-transcriptional regulation, protein degradation and translational control contribute to
49 transcript-protein discordance. Therefore, RNA-only models may also exhibit limited predictive capability
50 for proteins with weak transcriptomic coupling. Recent efforts have attempted to integrate additional spatial
51 information for multimodal molecular reconstruction. For example, SpatialEx[19] and NicheTrans[20]
52 introduced strategies for leveraging spatially resolved multimodal information to improve molecular
53 prediction. However, these approaches generally require additional experimentally matched modalities, such
54 as adjacent sections measured by complementary technologies, or rely on predefined spatial relationships and
55 auxiliary information. Such requirements limit their applicability to the rapidly expanding collections of
56 existing spatial transcriptomic datasets.

57 Beyond the availability of input modalities, another largely overlooked obstacle is the substantial variation
58 in spatial transcriptomic platforms[21], [22], [23], [24], [25]. Different technologies generate spatial images
59 with heterogeneous resolutions, ranging from near single-cell resolution to large tissue spots, resulting in
60 considerable differences in the amount and characteristics of morphological information captured by H&E
61 images. A computational framework capable of directly reconstructing spatial proteomes from existing H&E-
62 coupled spatial transcriptomic datasets must therefore simultaneously address two fundamental challenges:
63 effective integration of complementary molecular and morphological signals and adaptation to heterogeneous
64 image resolutions.

65 Here, we develop SPRINT (Spatial Proteomic Inference Network), a resolution-adaptive computational
66 framework that reconstructs spatial proteomic profiles from paired H&E images and transcriptomic data.
67 SPRINT addresses two key challenges: 1) It employs a dual-modality fusion architecture based on self-
68 attention mechanisms to jointly model image-derived morphological features and transcriptomic

representations, overcoming the limitations of existing approaches that rely solely on either imaging or RNA information. 2) It incorporates a technology-compatible multimodal processing strategy with complementary high-resolution and low-resolution image feature extractors, enabling histological information generated under diverse spatial transcriptomic technologies and imaging conditions to be effectively fused with the corresponding transcriptomic measurements. The model can therefore selectively capture fine-scale morphological patterns or global tissue contexts according to the input image resolution. After training on a dataset with matched histological, transcriptomic and proteomic measurements from a given tissue type, SPRINT can be directly applied to other spatial transcriptomic datasets from the same tissue context to predict spatial protein landscapes. Across nine spatial datasets covering four human and mouse tissues, SPRINT consistently outperformed existing state-of-the-art approaches and generalized across tissue sections, donors, spatially non-overlapping regions, imaging resolutions and species. This generalizability enabled SPRINT to reconstruct proteins that were weakly associated with either histological morphology or transcript abundance alone and to recover protein-level molecular information that improved downstream spatial domain identification. Moreover, when trained directly on spatial metabolomic measurements, SPRINT extended beyond proteomics to reconstruct metabolite distributions across independent tissue sections. These results establish SPRINT as a general computational framework for completing unmeasured spatial molecular profiles across diverse biological contexts, experimental settings and molecular modalities.

Results

The SPRINT framework

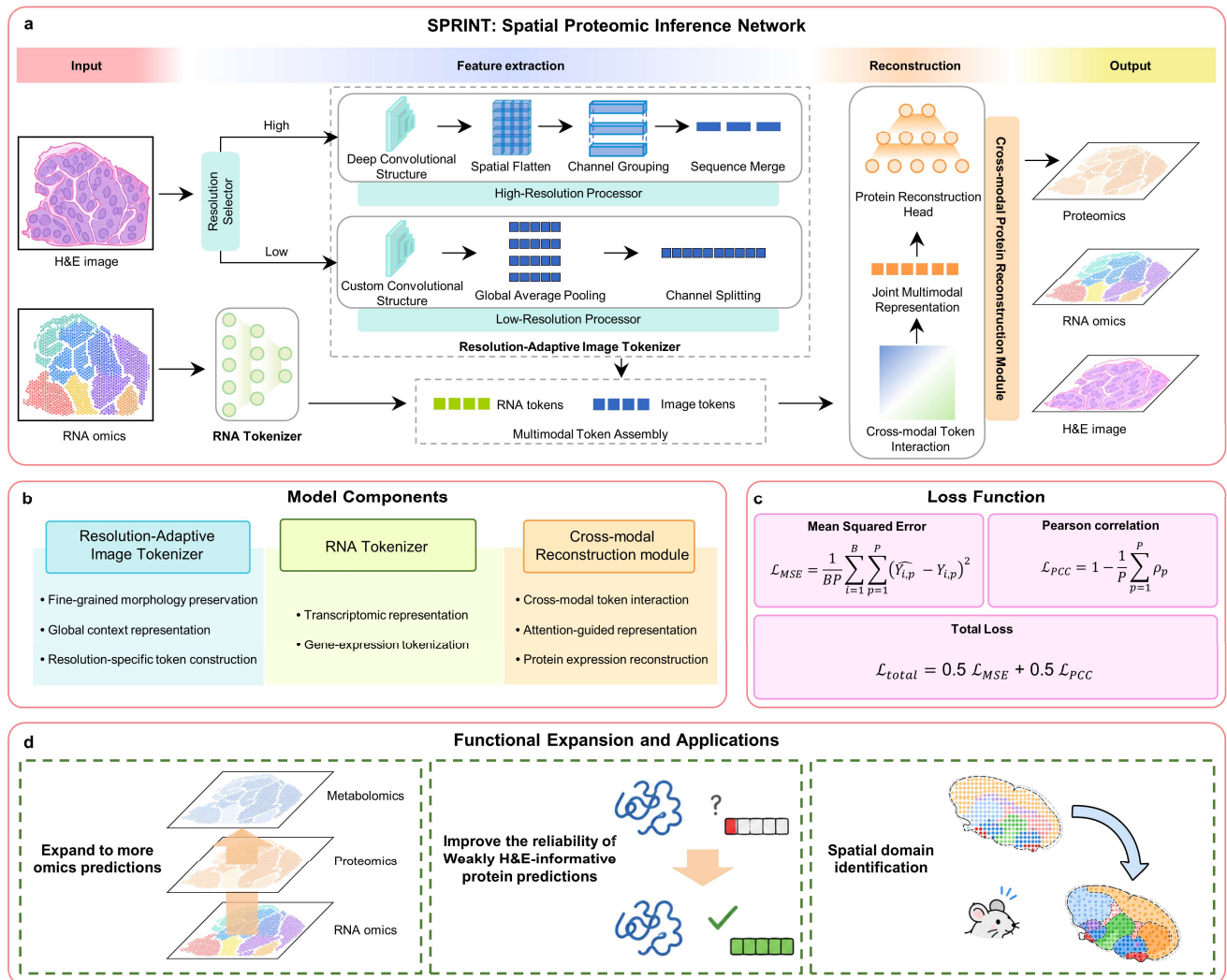


Fig. 1. Overview of SPRINT. **a.** Architecture of the SPRINT. SPRINT takes paired H&E-stained histological images and spatial transcriptomic expression profiles as input. For the image modality, a resolution selector first determines the image resolution and routes the input to a resolution-adaptive image tokenizer. High-resolution images are processed using a pretrained deep convolution structure feature extractor, followed by spatial flattening, channel grouping and sequence merging to preserve fine-grained spatial morphology. Low-resolution images are processed using a custom convolution structure trained from scratch, followed by global average pooling and channel splitting to generate compact representations of global tissue context. For the transcriptomic modality, an RNA tokenizer transforms transcriptome-wide expression profiles into RNA tokens. RNA and image tokens are assembled into a joint multimodal token sequence and processed by the cross-modal protein reconstruction module. Within this module, self-attention enables cross-modal token interaction to derive a joint multimodal representation, which is subsequently mapped to protein abundance through a protein reconstruction head. SPRINT ultimately generates reconstructed spatial proteomic profiles aligned with the original transcriptomic profiles and H&E images. **b.** Core components of SPRINT. The resolution-adaptive image tokenizer contains complementary high- and low-resolution processing branches that support fine-grained morphology preservation, global

context representation and resolution-specific token construction. The RNA tokenizer encodes transcriptome-wide molecular information into tokenized transcriptomic representations. The cross-modal protein reconstruction module performs cross-modal token interaction, learns integrated multimodal representations and reconstructs spatial protein expression. **c.** Loss function of SPRINT. The training objective consists of two complementary components: mean squared error (MSE) loss and Pearson correlation coefficient (PCC)-based loss. The two losses are equally weighted, with each contributing 50% to the total optimization objective. **d.** Functional expansion and applications of SPRINT. SPRINT provides a general framework for spatial omics reconstruction and can be extended to other molecular modalities, such as spatial metabolomics prediction. Compared with image-driven approaches, SPRINT maintains robust reconstruction performance for weakly H&E-informative proteins by integrating transcriptomic information. The reconstructed spatial proteomic profiles, combined with transcriptomic data, improve spatial domain identification.

Existing computational approaches for spatial protein inference generally rely on either histological morphology or transcriptomic information, leaving each strategy constrained by the information content of a single modality. Morphology-based prediction is most effective for proteins whose spatial variation is reflected in tissue appearance, whereas transcript-based prediction can be limited by the incomplete correspondence between RNA and protein abundance. We therefore developed SPRINT, a spatial proteomic inference framework that jointly exploits morphological and transcriptomic information through two coordinated design strategies: resolution-adaptive image tokenization and cross-modal protein reconstruction (Fig. 1a). During training, paired H&E images and spatial transcriptomic profiles are used as inputs, while experimentally measured protein abundances provide supervision. Once trained, SPRINT requires only the matched image and transcriptomic modalities to reconstruct protein expression at the corresponding spatial locations.

A central design of SPRINT is its resolution-adaptive image tokenizer, which accommodates differences in image resolution across spatial profiling platforms while preserving the morphological information most relevant at each scale (Fig. 1a, b). Rather than imposing a single image-processing strategy on all inputs, SPRINT first determines the imaging resolution and routes the image through one of two complementary processing paths. For high-resolution images, the model retains spatially resolved morphological patterns and reorganizes them into a compact sequence of image tokens that preserves local tissue structure. For lower-resolution inputs, SPRINT emphasizes global tissue context and generates a compact token representation that is less dependent on fine-scale image detail. Although the two branches operate differently, both produce image tokens in a common latent representation space, allowing the subsequent reconstruction procedure to remain unchanged across imaging resolutions. This design separates resolution-specific morphological encoding from downstream multimodal inference and enables SPRINT to accommodate heterogeneous image characteristics without requiring a common native image resolution.

The second core design is the cross-modal protein reconstruction module, which learns protein-associated representations from interactions between image and transcriptomic tokens rather than treating the two modalities as independent predictors (Fig. 1a, b). Transcriptomic profiles are first transformed into RNA tokens with the same latent dimensionality as the image tokens. The two token sets are then assembled into a joint multimodal sequence, enabling information exchange between transcriptomic states and spatial morphological patterns. Within this shared representation space, cross-modal interactions allow the contribution of morphological and molecular information to vary across the learned representation instead of being fixed by a predefined modality weighting. The resulting joint multimodal representation is subsequently

mapped to the abundance of the target proteins through a protein reconstruction head. SPRINT is optimized using complementary quantitative and correlation-based objectives: mean squared error constrains absolute reconstruction accuracy, whereas a Pearson correlation-based loss promotes preservation of protein-specific spatial variation (Fig. 1c). Together, these objectives encourage reconstructed profiles to match both the magnitude and spatial pattern of experimentally measured protein expression.

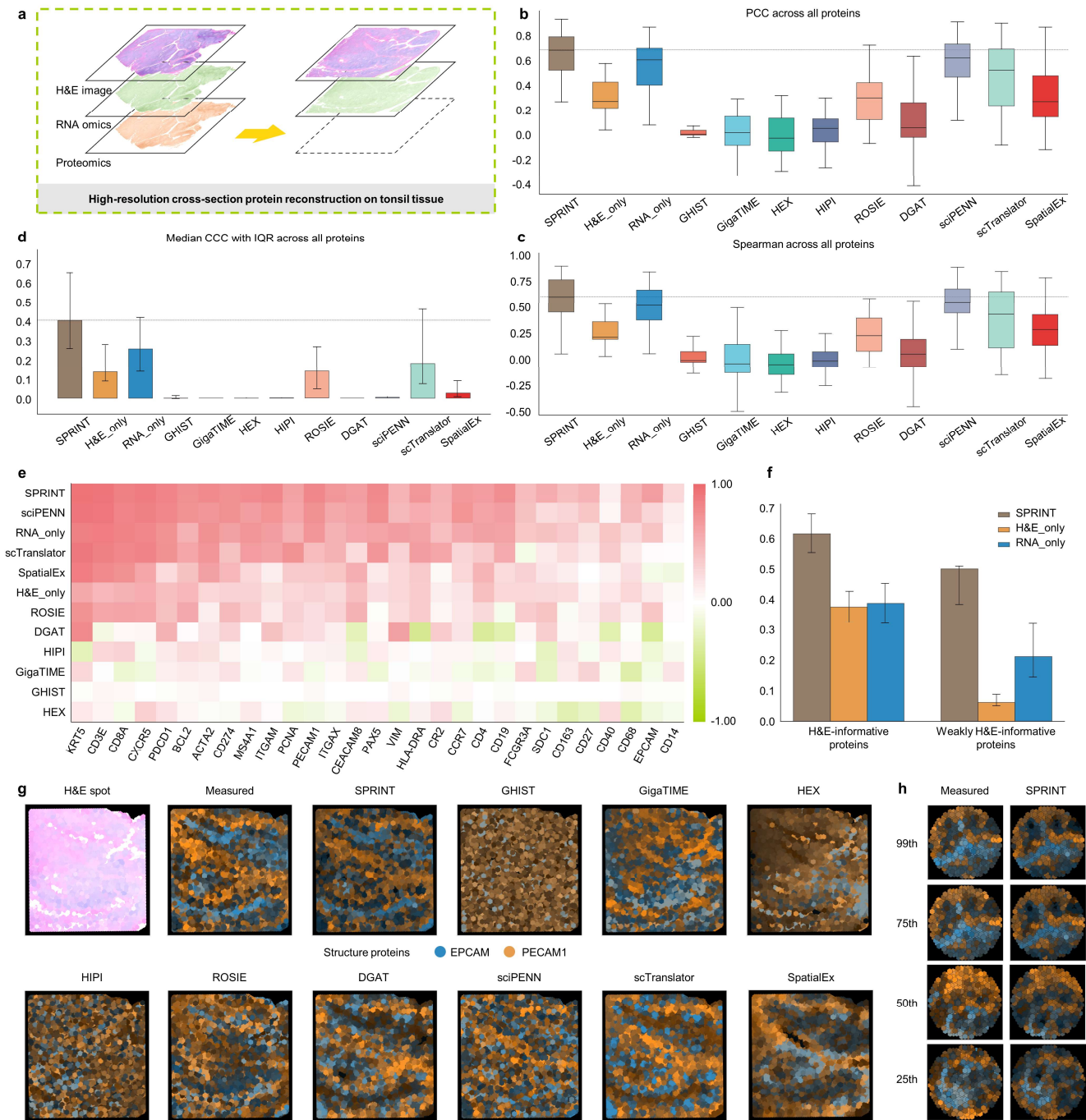
The reconstructed protein profiles remain spatially aligned with the original transcriptomic measurements and H&E images, yielding an integrated spatial multi-omics representation of the same tissue section (Fig. 1a). By integrating complementary morphological and transcriptomic information through resolution-adaptive encoding and token-level cross-modal interaction, SPRINT broadens the range of proteins that can be reliably reconstructed beyond those preferentially captured by either modality alone. In particular, transcriptomic information provides an additional molecular context for proteins with weak H&E-associated morphological signals, while image-derived features preserve spatial structural information that is not directly represented by gene expression. The reconstructed molecular profiles can subsequently support downstream analyses such as spatial-domain identification (Fig. 1d). Although SPRINT was developed primarily for spatial proteomic reconstruction, its reconstruction framework is not intrinsically restricted to proteins and can be trained against other spatial molecular targets, providing a route toward extension to additional modalities such as spatial metabolomics.

157

SPRINT reconstructs spatial protein landscapes across datasets by fusing histological

158

and transcriptomic information



160

Fig. 2 High-resolution cross-section protein reconstruction. **a**. Experimental workflow. A tonsil tissue section containing H&E imaging,

161

transcriptomic and proteomic data was used as the training set to infer protein expression in another tissue section. The measured proteomic data

162

from the predicted section were then used to evaluate model performance. **b-d**. Evaluation of SPRINT and the comparison methods using the

163

Pearson correlation coefficient (PCC), Spearman correlation coefficient and median concordance correlation coefficient (CCC). Higher values

164

indicate better performance for all three metrics. H&E_only denotes an ablation model retaining only the H&E feature-processing branch of

SPRINT, whereas RNA_only denotes an ablation model retaining only the transcriptomic feature-processing branch. **e.** Heatmap showing the PCC obtained for each protein across all methods. Methods positioned higher on the y axis have larger summed PCC values across all proteins, whereas proteins positioned further to the left on the x axis have larger summed PCC values across all models. **f.** Performance of the SPRINT ablation models on H&E-informative and weakly H&E-informative proteins. Representative groups of H&E-informative and weakly H&E-informative proteins were selected to evaluate the effectiveness of the image and RNA feature-extraction modules of SPRINT. **g.** In silico visualization of spatial protein expression in human tonsil tissue. The expression patterns of the structural markers EPCAM and PECAM1 are presented as an additive colour overlay. The results from all methods were quantile-matched to the measured protein values to ensure fair visual comparison. The leftmost image in the first row shows the corresponding H&E image, and “Measured” denotes the visualization of the measured protein values. **h.** Examples of SPRINT spatial protein predictions at different prediction-accuracy percentiles. All spots predicted by SPRINT were ranked according to the spot-wise Pearson correlation coefficient between the predicted and measured protein abundances. Representative spots from the 99th, 75th, 50th and 25th percentiles were selected, and their local neighbourhoods are shown as enlarged views. Each row displays the measured protein expression (ground truth) and the corresponding SPRINT prediction for the same local region.

To assess whether SPRINT could reconstruct spatial proteomic landscapes across tissue sections, we used two publicly available human tonsil datasets generated with the 10x Genomics Visium CytAssist Spatial Gene and Protein Expression platform[26]. The datasets were derived from independent FFPE tonsil specimens and contained matched high-resolution H&E images, spatial transcriptomic profiles and spatial protein measurements. SPRINT was trained on the dataset generated using the standard antibody panel and evaluated on the independent dataset generated with add-on antibodies. The analysis was restricted to the 29 proteins shared between the training and test antibody panels. During evaluation, SPRINT inferred protein abundance in the test specimen from its H&E image and spatial transcriptomic profiles, whereas the corresponding measured protein abundances were withheld during inference and used exclusively as ground truth (Fig. 2a). This design therefore provided a stringent assessment of protein reconstruction across both tissue sections and donors.

SPRINT achieved the strongest overall performance across all three complementary evaluation metrics and across individual proteins (Fig. 2b-e). It obtained the highest median Pearson correlation coefficient (PCC), Spearman correlation coefficient and concordance correlation coefficient (CCC), indicating accurate recovery of both spatial variation and quantitative protein abundance. Image-driven methods performed comparatively better for proteins associated with visible tissue structures but showed lower and more variable performance across the complete protein panel, whereas RNA-driven methods generally captured a larger proportion of protein variation, particularly for immune-associated markers, but showed weaker agreement with the measured abundance scale. In contrast, SPRINT maintained high PCC values across proteins with diverse spatial and biological characteristics. These included histologically discernible epithelial and vascular markers, such as KRT5, EPCAM and PECAM1[27], [28], [29], and H&E-informative proteins such as BCL2 and CD68[30], [31], [32], demonstrating that SPRINT can map morphological structures to protein-specific expression patterns rather than merely transferring coarse histological features. SPRINT also accurately reconstructed proteins associated with molecularly defined immune and tissue compartments, including the B-cell markers PAX5, CD19, MS4A1 and CR2[33]; the immune activation and localization markers PDCD1, CCR7 and CD27[34]; and the myeloid or vascular markers ITGAM, ITGAX and PECAM1[35]. Because these protein distributions reflect cell-type composition, immune state and local tissue organization that cannot be fully resolved from either histology or transcript abundance alone, their accurate reconstruction supports the ability of SPRINT to fuse complementary morphological and transcriptomic information.

206 Together, these results show that bimodal fusion enables robust reconstruction of protein-specific spatial
207 landscapes across independent datasets.

208 We next used modality-specific ablations to determine whether both input branches contributed meaningful
209 information to protein reconstruction. We selected several representative proteins from the weakly H&E-
210 informed and H&E-informed groups, respectively (Methods). For H&E-informative proteins, the H&E-only
211 and RNA-only models achieved comparable performance, indicating that proteins associated with visible
212 tissue morphology were also predictable from the corresponding transcriptomic context (Fig. 2f). For weakly
213 H&E-informative proteins, however, the RNA-only model substantially outperformed the H&E-only model,
214 demonstrating the greater contribution of molecular information when protein expression was not readily
215 reflected in tissue morphology. The complete SPRINT model achieved the highest performance in both
216 groups, showing that histological and transcriptomic information remained complementary even when one
217 modality was individually informative. These results validate both modality-specific feature-extraction
218 branches and demonstrate that the multimodal fusion strategy effectively combines their complementary
219 information to improve protein reconstruction.

220 Visual analysis further demonstrated that SPRINT faithfully reconstructed both the global organization and
221 local heterogeneity of protein spatial distributions within the tissue. In the measured test section, EPCAM
222 and PECAM1 formed distinct, spatially interwoven patterns delineating epithelial and vascular structures
223 within the tonsil. After quantile matching the predictions from all methods to the measured protein
224 distributions to ensure fair visual comparison, SPRINT closely reproduced both the overall organization and
225 local heterogeneity of these complementary structural markers (Fig. 2g). By contrast, several image-driven
226 methods produced weakly differentiated or morphology-dominated patterns, whereas other methods
227 introduced excessive mixing, fragmentation or smoothing of the two protein signals. These differences
228 indicate that SPRINT reconstructs protein-specific spatial architecture rather than merely transferring coarse
229 histological patterns to predicted protein maps, with consistent recovery of local protein composition across
230 representative spot-level accuracy percentiles and immune-associated protein groups (Fig. 2h and
231 Supplementary Fig. 1).

SPRINT reconstructs proteomic landscapes from low-resolution tissue images and enhances spatial domain identification

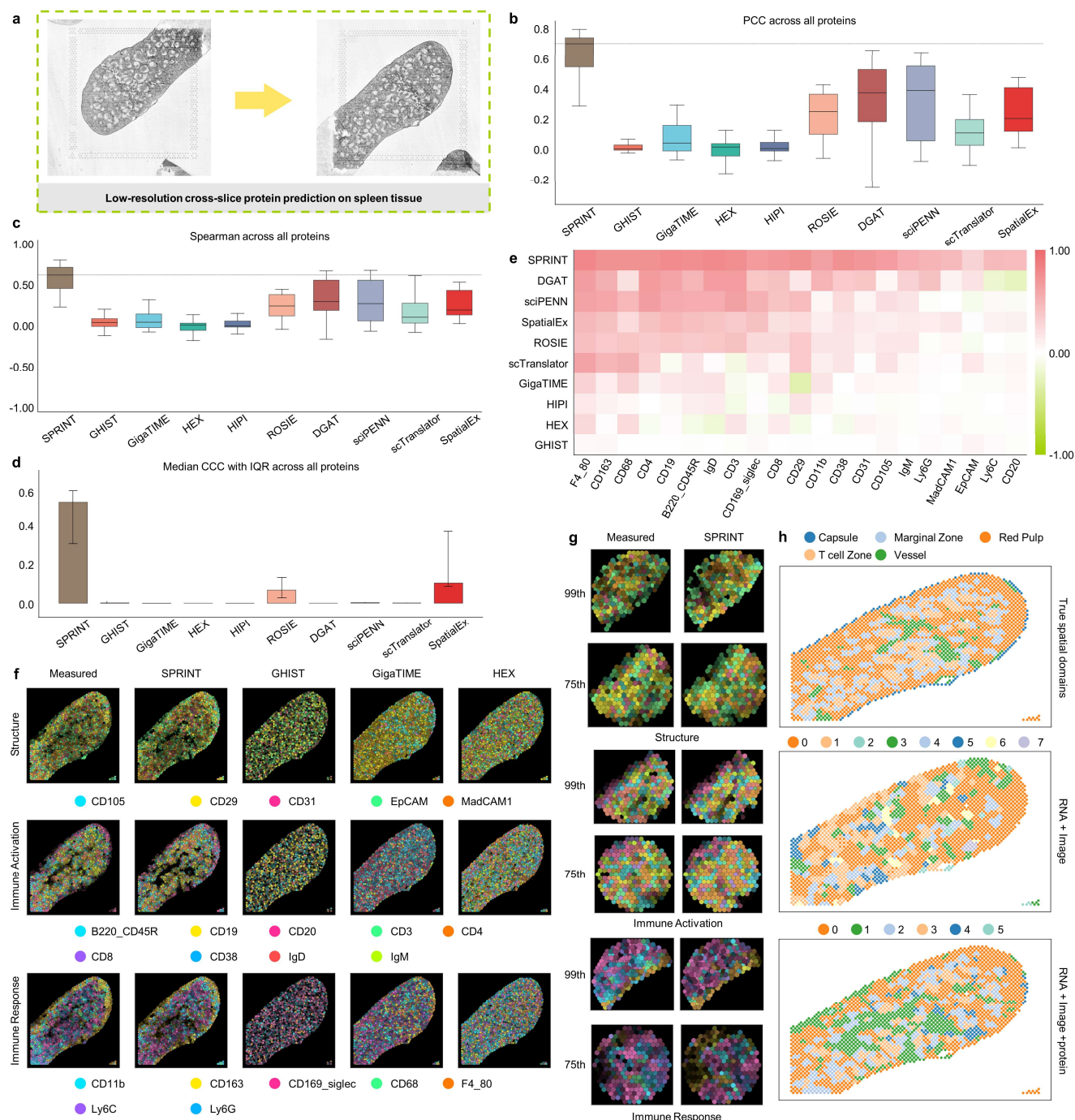


Fig. 3 Cross-section protein reconstruction from low-resolution histology in mouse spleen. **a.** Experimental workflow. SPRINT was trained on one mouse spleen section containing spatial transcriptomic and proteomic measurements together with a low-resolution histological image and evaluated on a held-out replicate. Protein measurements from the test section were excluded during inference and used only as ground truth. **b-d.** Performance of SPRINT and comparison methods across the 21 measured proteins, evaluated using the Pearson correlation coefficient (PCC), Spearman correlation coefficient and concordance correlation coefficient (CCC), respectively. Higher values indicate greater agreement between

predicted and measured protein abundances across tissue spots. **e.** Heatmap showing the protein-wise PCC obtained by each method. Methods are ordered by their summed PCC across proteins, and proteins are ordered by their summed PCC across methods. **f.** In silico visualization of spatial protein expression in the test spleen section. Proteins were organized into three functional groups representing tissue structure, immune activation and immune response, and their expression patterns were displayed as additive colour overlays. Predictions from all methods were quantile-matched to the measured protein values to enable consistent visual comparison. **g.** Representative local neighbourhoods of measured and SPRINT-predicted protein expression for the three functional groups. Spots were ranked according to the spot-wise Pearson correlation between predicted and measured protein-abundance profiles within each group, and representative neighbourhoods from the 99th and 75th percentiles are shown. **h.** Spatial domain identification using RNA and histological features without or with SPRINT-reconstructed protein expression. Inferred domains are compared with the annotated splenic regions, including the capsule, marginal zone, red pulp, T-cell zone and vessels.

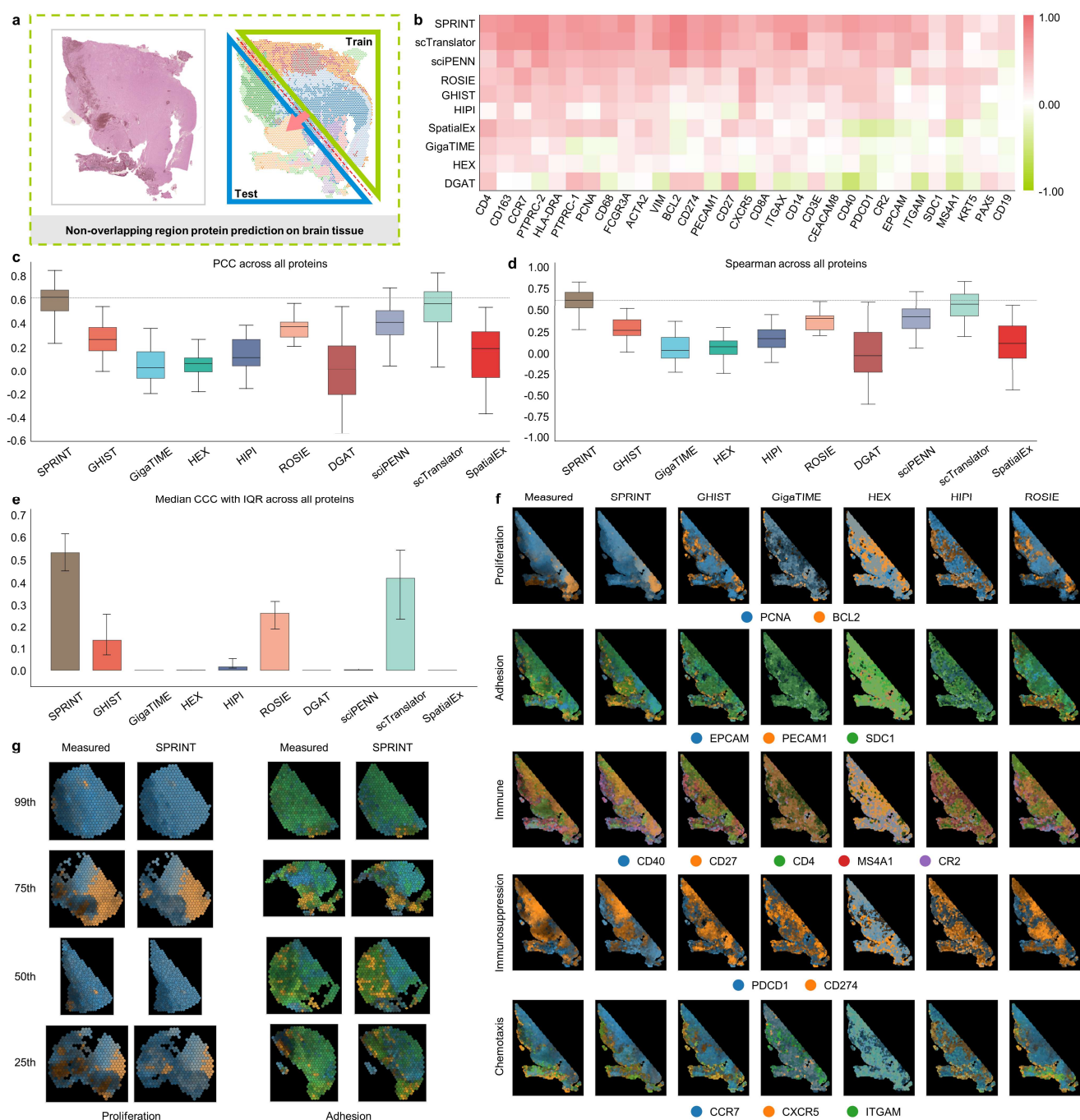
To evaluate SPRINT under low-resolution imaging conditions, we used two mouse spleen sections from the SPOTS dataset generated by Ben-Chetrit et al.[36] The dataset provides spatial transcriptomic measurements together with 21 antibody-derived protein markers. SPRINT was trained on replicate 1 and evaluated on the held-out replicate 2, whose measured protein abundances were excluded during inference and used only as ground truth (Fig. 3a). This setting allowed us to assess the cross-section reconstruction of splenic proteomic landscapes from low-resolution histology and spatial transcriptomics.

Across tissue spots, SPRINT consistently achieved the strongest agreement between predicted and measured protein abundances (Fig. 3b-d). SPRINT achieved the highest PCC, Spearman correlation and CCC, consistently outperforming all comparison methods. These results indicate that SPRINT recovered not only the relative variation across spots but also protein-abundance distributions that were quantitatively consistent with the measurements. Protein-level analysis further showed that the performance of SPRINT extended broadly across the antibody panel rather than being driven by a small number of readily predictable proteins (Fig. 3e). SPRINT maintained consistently high PCC values for structural and vascular proteins, including CD105, CD29 and CD31[37]; lymphoid markers, including CD3, CD4 and CD8[38]; and myeloid-response markers, including CD11b, CD163 and CD169[35]. Thus, even in the absence of high-resolution histological texture, SPRINT could integrate coarse morphological context with transcriptomic information to reconstruct spatial protein expression.

We next organized the 21 measured proteins into three biologically interpretable groups representing tissue structure, immune activation and immune response, and visualized their joint spatial distributions as additive colour overlays (Fig. 3f). The structure group comprised CD105, CD29, CD31, EpCAM and MadCAM1, capturing vascular, stromal and epithelial organization[39], [40]. The immune-activation group combined major B- and T-cell markers[41], whereas the immune-response group represented myeloid and inflammatory compartments[42], [43]. Across all three groups, SPRINT closely reproduced the measured tissue-wide protein organization while preserving distinct local combinations of markers. By contrast, the comparison methods produced excessively mixed or spatially diffuse signals that obscured the organization of individual protein programmes. The recovery of three functionally distinct protein landscapes indicates that SPRINT reconstructed coordinated biological organization rather than isolated protein signals, with the principal group-specific patterns also preserved across representative local neighbourhoods spanning different prediction-accuracy percentiles (Fig. 3g and Supplementary Fig. 2).

Finally, we investigated whether the reconstructed proteomic landscape could support spatial domain

279 identification. We compared unsupervised spatial domains inferred from RNA and image features alone with
280 those obtained after incorporating SPRINT-predicted protein abundances (Fig. 3h). Adding the reconstructed
281 proteomic modality produced more coherent and clearly delineated domains. Most notably, the marginal zone
282 was recovered as a largely continuous domain after the predicted proteins were incorporated, despite its
283 connected and structurally complex organization. In contrast, when only RNA and image features were used,
284 this domain was identified only in scattered regions and differed markedly from the annotated spatial structure.
285 Quantitatively, the adjusted Rand index increased from 0.045 to 0.256, and the normalized mutual
286 information increased from 0.067 to 0.248. Thus, the protein profiles generated by SPRINT provided
287 biological information complementary to transcriptomic and morphological features, improving downstream
288 tissue partitioning. Collectively, these results establish the robustness of SPRINT to low-resolution histology
289 and demonstrate that its reconstructed proteomic landscapes can support spatial analyses beyond protein
290 reconstruction itself.



292

Fig. 4. Protein reconstruction in spatially disjoint regions. **a.** Experimental design. A human brain section was divided diagonally into mutually exclusive training and test regions. SPRING was trained using the image, spatial transcriptomic profiles and measured protein abundances from the training region and subsequently used to infer protein expression in the held-out region. Protein measurements from the test region were excluded during inference and used only as ground truth. **b.** Heatmap showing the protein-wise Pearson correlation coefficient (PCC) obtained by SPRING and the comparison methods in the held-out region. **c-e.** Performance of SPRING and the comparison methods across the 31 measured proteins, evaluated using PCC, Spearman correlation coefficient and concordance correlation coefficient (CCC), respectively. Higher values indicate greater agreement between predicted and measured protein abundances across tissue spots. **f.** In silico visualization of spatial protein

expression in the test region. Proteins were organized into five functional groups representing proliferation, adhesion, immune activation, immunosuppression and chemotaxis, and their joint spatial distributions were displayed as additive colour overlays. **g.** Representative local neighbourhoods of measured and SPRINT-predicted protein expression for the proliferation and adhesion groups.

To determine whether SPRINT could generalize beyond spatial locations represented during training, we analysed a human brain FFPE section profiled using the 10x Genomics Visium CytAssist Spatial Gene and Protein Expression platform[44]. We divided the section diagonally into two mutually exclusive regions, using 3,283 spots for training and holding out the remaining 2,473 spots for testing (Fig. 4a). The measured protein abundances in the test region were excluded during inference and used only as ground truth. Unlike the preceding cross-section experiments, in which analogous anatomical locations may occur across different sections, this design explicitly excluded overlap between training and test coordinates and directly assessed spatial extrapolation within the same tissue. The diagonal partition also resulted in marked differences in the distributions and relative proportions of cell types between the two regions (Fig. 4a). SPRINT was therefore required to generalize not only to unseen spatial locations but also across a shift in cellular composition, representing a more stringent setting than a random spot-level split. At the same time, partitioning a single section minimized section-level experimental variation, allowing the capacity for spatial extrapolation to be evaluated directly.

SPRINT maintained strong predictive performance across the 31 measured proteins under this spatially disjoint setting (Fig. 4b-e). It achieved the highest median PCC, Spearman correlation and CCC, preserving both spot-to-spot variation and quantitative agreement with measured protein abundances. Protein-level analysis showed that this performance extended broadly across the panel rather than being driven by a small subset of readily predictable markers (Fig. 4b). High predictive agreement was observed for proteins associated with immune identity, trafficking and activation, including CD4, CD163, CCR7 and HLA-DRA[45], [46], as well as proteins representing proliferation, adhesion and tissue organization. These results demonstrate that SPRINT transferred multimodal relationships learned in one part of the tissue to a spatially non-overlapping region with a different cellular composition.

We next examined whether this spatial generalization extended to coordinated biological programmes. The measured proteins were organized into five functional groups representing proliferation, adhesion, immune activation, immunosuppression and chemotaxis, and their joint spatial distributions were visualized as additive colour overlays (Fig. 4f). SPRINT recovered the complementary distributions of the proliferation markers PCNA and BCL2 and preserved the coordinated adhesion patterns defined by EPCAM, PECAM1 and SDC1[47]. It also reconstructed spatially heterogeneous immune programmes represented by CD40, CD27, CD4, MS4A1 and CR2[48], the immunoregulatory proteins PDCD1 and CD274, and the chemotactic markers CCR7, CXCR5 and ITGAM[49]. The recovery of these functionally distinct patterns indicates that SPRINT reconstructed multivariate protein programmes within the unseen region rather than merely propagating a dominant tissue-wide signal from the training region. Results from the remaining comparison methods are provided in Supplementary Fig. 3. Local-neighbourhood analysis further demonstrated that these reconstructed programmes were preserved at finer spatial scales (Fig. 4g and Supplementary Fig. 3). Thus, the tissue-wide agreement achieved by SPRINT was supported by faithful reconstruction of local multivariate protein environments.

Collectively, this spatially disjoint evaluation complements the cross-section analyses by explicitly excluding overlap between training and test locations while introducing a substantial shift in cellular composition. SPRINT reconstructed diverse protein programmes throughout the held-out brain region, demonstrating that it learns transferable relationships among imaging, transcriptional state and protein abundance rather than memorizing location-specific protein patterns.

SPRINT can be transfer to cross-species and spatial metabolite reconstruction

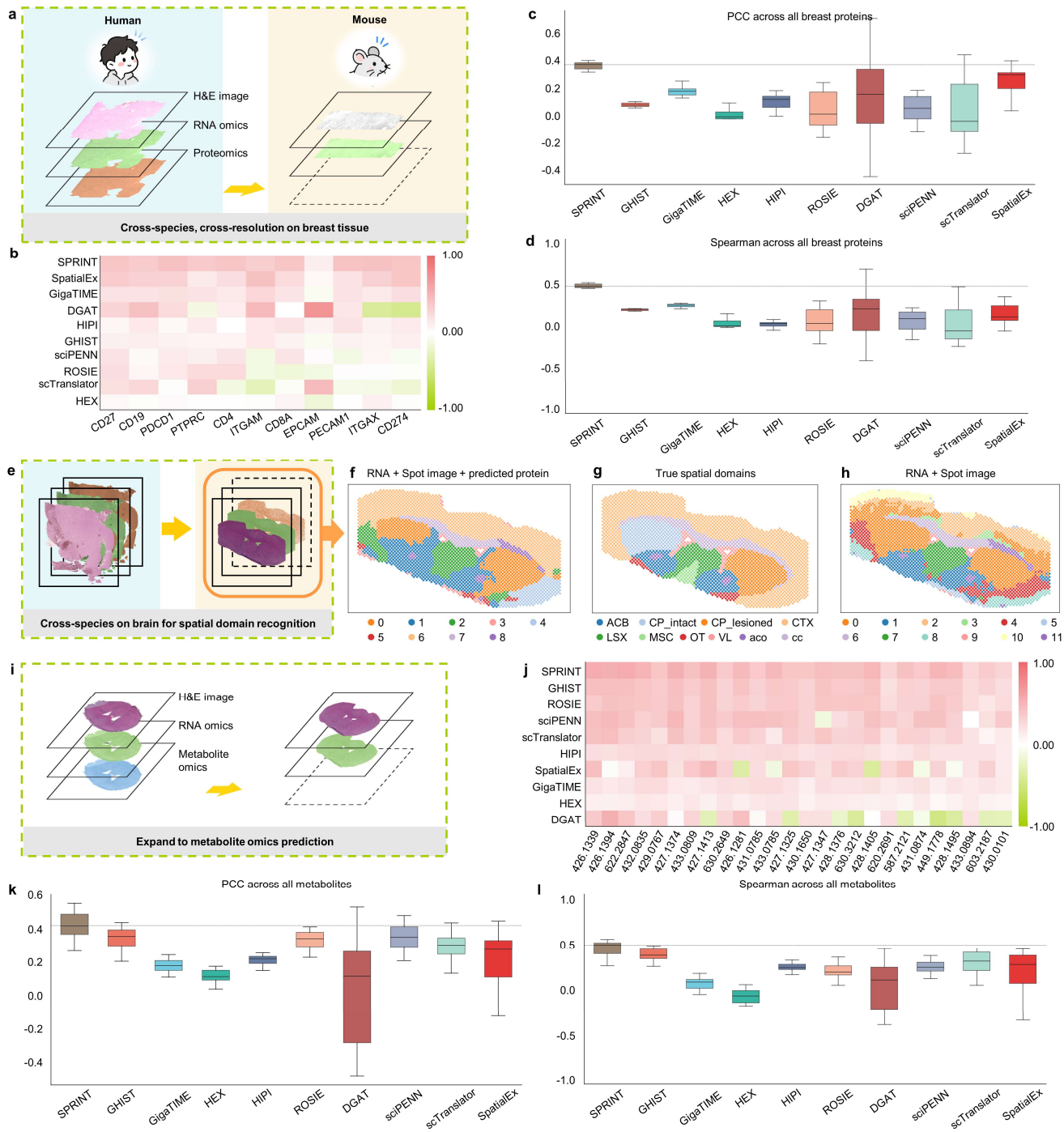


Fig. 5 Cross-species transfer and spatial metabolite reconstruction with SPRINT. a. Experimental design for cross-species and cross-image-

resolution protein reconstruction. SPRINT was trained using matched imaging, spatial transcriptomic and proteomic measurements from human breast tissue and applied to an independent mouse breast tissue section. Protein measurements from the mouse section were excluded during inference and used only for evaluation. **b.** Heatmap showing the protein-wise PCC values obtained by SPRINT and the comparison methods for the 11 proteins measured in mouse breast tissue. **c-d.** Distributions of PCC (c) and Spearman correlation coefficients (d) across the evaluated proteins. **e.** Workflow for cross-species reconstruction and spatial-domain analysis in mouse brain tissue. A protein reconstruction model trained on human brain tissue was applied to mouse brain tissue without target-sample protein supervision. The reconstructed protein profiles were subsequently integrated with RNA and spot-image features for spatial-domain identification. **f.** Spatial domains identified using RNA, spot-image and SPRINT-predicted protein features. **g.** Reference anatomical annotations of the mouse brain tissue, including the nucleus accumbens (ACB), intact and lesioned caudoputamen (CP_intact and CP_lesioned), cortex (CTX), lateral septal complex (LSX), medial septal complex (MSC), olfactory tubercle (OT), ventrolateral nucleus (VL), anterior commissure (aco) and corpus callosum (cc). **h.** Spatial domains identified using RNA and spot-image features without reconstructed proteins. **i.** Experimental design for spatial metabolite reconstruction. SPRINT was trained directly using matched imaging, spatial transcriptomic and metabolomic measurements from one mouse brain section and evaluated in an independent section. **j.** Heatmap showing the metabolite-wise PCC values obtained by SPRINT and the comparison methods for 26 reliably detected mass-to-charge features. **k-l.** Distributions of PCC (k) and Spearman correlation coefficients (l) across the evaluated metabolite features.

To evaluate the broader applicability of SPRINT, we examined two distinct forms of generalization. First, we tested whether a protein reconstruction model trained in one species could be transferred to another species. Second, we applied the SPRINT framework to spatial metabolomics by training it directly on matched metabolomic data, thereby assessing whether its multimodal architecture could reconstruct molecular outputs beyond proteins.

We first evaluated cross-species and cross-resolution protein reconstruction by training SPRINT on a human breast tissue dataset[50] and applying the trained model to an independent mouse breast tissue section [36] (Fig. 5a). Transcriptomic inputs were restricted to genes shared between the two species, and performance was evaluated using the 11 proteins measured in the mouse tissue. Despite differences in species, tissue scale and imaging resolution, SPRINT achieved the highest overall correlations between predicted and measured protein abundances, with median PCC and Spearman correlation coefficients of 0.379 and 0.471, respectively (Fig. 5b-d). Protein-level analysis showed that this performance extended across immune markers, including CD19, CD27, CD4 and CD8A, as well as epithelial and vascular markers such as EPCAM and PECAM1 (Fig. 5b). These results indicate that the relationships learned by SPRINT between morphology, transcription and protein expression remain informative across species and imaging conditions.

We next used human [44] and mouse brain datasets[20] to further evaluate the cross-species generalizability of SPRINT (Fig. 5e). The model was trained using human protein measurements and subsequently applied to mouse brain tissue without protein supervision from the target sample, producing spatial predictions for 31 proteins. We then compared spatial domains identified from RNA and spot-image features with those obtained after incorporating the predicted protein profiles. Under the tuned spatial-domain workflow, inclusion of the reconstructed proteins produced domains that more closely reflected the annotated anatomical organization of the mouse brain (Fig. 5f-h). The adjusted Rand index increased from 0.204 to 0.515, while normalized mutual information increased from 0.494 to 0.573. In particular, the protein-augmented representation more clearly resolved several anatomically distinct regions that were fragmented or insufficiently separated using RNA and image information alone. Thus, SPRINT predictions transferred across species retain biologically informative variation that can support downstream tissue organization analyses.

Finally, we investigated whether the SPRINT framework could be applied to a molecular modality other than proteins. We trained SPRINT directly on matched imaging, spatial transcriptomic and metabolomic data from one mouse brain section and evaluated metabolite reconstruction in an independent section [20] (Fig. 5i). Analysis was restricted to 26 reliably detected metabolite features (Methods). SPRINT achieved the highest median PCC and Spearman correlation coefficients (Fig. 5j-l). Its performance was broadly distributed across the evaluated mass-to-charge features rather than being driven by a small subset of metabolites (Fig. 5j). These findings show that the multimodal integration strategy underlying SPRINT can be adapted to reconstruct spatial molecular measurements beyond proteins when trained with the corresponding target modality.

Together, these experiments establish two complementary aspects of the generalizability of SPRINT. The breast and brain analyses demonstrate that a protein reconstruction model can transfer across species and remain informative for downstream spatial analysis without protein supervision in the target species. Separately, the metabolomics analysis shows that the SPRINT architecture can be trained to reconstruct molecular outputs beyond proteins, supporting its broader applicability as a framework for spatial molecular reconstruction.

Discussion

Spatial transcriptomic assays inherently provide two complementary sources of information: spatially resolved gene expression and matched tissue imaging. Jointly exploiting these modalities offers substantial potential for recovering molecular features that were not measured experimentally and for extending the analytical value of existing spatial datasets. Here we introduced SPRINT, a multimodal framework that fuses histological and transcriptomic information to reconstruct spatial molecular profiles. Across human and mouse tissues, SPRINT recovered protein-specific spatial landscapes across independent sections, donors, non-overlapping tissue regions, imaging resolutions, gene panels and species. The reconstructed protein profiles captured information that was not fully recoverable from either modality alone and improved downstream characterization of tissue domains. Moreover, the framework extended beyond proteomics to spatial metabolite reconstruction when trained directly using metabolomic measurements. Together, these findings establish multimodal fusion as an effective strategy for completing spatial molecular profiles and position SPRINT as a general framework for extracting additional biological information from routinely generated spatial transcriptomic datasets.

A principal advantage of SPRINT is its explicit integration of complementary morphological and molecular information. Tissue images contain information about cellular organization, tissue architecture and local microenvironments, but the extent to which individual proteins are reflected in visible morphology varies substantially. Conversely, transcript abundance provides protein-related molecular information but does not fully determine protein abundance owing to post-transcriptional regulation, differences in translation and degradation, and technical variation between molecular assays. The protein-level and ablation analyses in human tonsil tissue illustrate this complementarity. The H&E-only branch recovered meaningful signals for proteins whose distributions were associated with recognizable tissue structures, whereas its performance declined markedly for weakly H&E-informative proteins. The RNA-only branch retained information for

425 these less morphologically apparent targets, but the complete model performed consistently across both
426 groups. SPRINT thus does not treat tissue morphology as a direct proxy for molecular expression; instead, it
427 uses morphological context to condition and refine the relationship between transcriptomic state and protein
428 abundance.

429 The use of resolution-specific image encoders further distinguishes SPRINT from approaches that apply a
430 single visual representation across heterogeneous imaging settings. High-resolution histological images
431 contain fine-grained textures and local cellular boundaries that can be preserved through spatially organized
432 image tokens. Low-resolution images provide less reliable local detail but still encode broader tissue context,
433 for which compact global representations may be more appropriate. SPRINT remained effective in the low-
434 resolution mouse spleen experiment, indicating that its performance does not depend exclusively on high-
435 quality microscopic texture. This flexibility is relevant to spatial studies in which image resolution, staining
436 quality and acquisition protocols vary across platforms or archived specimens. Nevertheless, the current high-
437 and low-resolution branches represent two predefined processing regimes. Future implementations could
438 replace this discrete selection with scale-aware or multiscale encoders that adapt continuously to image
439 resolution and local information content.

440 SPRINT generalized beyond interpolation between nearby measured spatial locations. In the tonsil and spleen
441 analyses, models were trained and evaluated on independent tissue sections, thereby capturing variation
442 between specimens and acquisition runs. Because corresponding regions may still be present across sections
443 from the same tissue, we additionally divided a human brain section diagonally into spatially disjoint training
444 and test regions. This produced non-overlapping regions with appreciably different cellular compositions and
445 required the model to extrapolate protein expression beyond the spatial support of its training observations.
446 SPRINT retained strong reconstruction performance under this setting and recovered coordinated protein
447 programs associated with proliferation, adhesion, immune regulation and chemotaxis. These results suggest
448 that the model learns transferable relationships among morphology, transcription and protein expression
449 rather than relying primarily on spatial proximity or memorization of local protein measurements. More
450 extensive evaluation across anatomical regions, individuals and pathological states will nevertheless be
451 required to establish the limits of such spatial extrapolation.

452 SPRINT-reconstructed protein profiles improved the identification of biologically coherent spatial domains.
453 In both spleen and mouse brain tissue, incorporating SPRINT-predicted proteins into multimodal
454 representations produced spatial domains that more closely reflected reference tissue organization than
455 domains derived from RNA and image features alone. Predicted proteins can provide a compact
456 representation of cellular identity and functional state that complements the high dimensionality and sparsity
457 of spatial transcriptomic measurements. This may be particularly useful for resolving tissue compartments
458 defined by coordinated surface-marker or signalling programs rather than by large transcriptional differences.
459 However, the spatial-domain workflows were tuned separately for the different modality combinations. The
460 observed improvements therefore demonstrate the utility of the complete protein-augmented analysis
461 workflow, rather than isolating the causal contribution of predicted proteins under identical clustering
462 parameters. Future benchmarks based on prospectively fixed analysis settings and independent annotations
463 will enable a more controlled assessment of their contribution to downstream tasks.

SPRINT also showed that multimodal relationships learned in one species can remain informative in another. Models trained on human breast or brain tissue generated biologically structured protein predictions in mouse tissues without protein supervision from the target samples. This transfer probably reflects conserved relationships among tissue organization, transcriptional programs and protein expression, while also demonstrating that the model can tolerate differences in tissue scale and image resolution. Cross-species reconstruction was stronger for relative spatial patterns than for absolute abundance agreement. This distinction is expected because concordance correlation is sensitive to shifts in both the mean and variance, which may be affected by species-dependent protein abundance, antibody affinity, panel composition and assay dynamic range. Orthologue mapping also does not guarantee equivalent regulation or function across species. Species-specific calibration, domain adaptation and explicit modelling of orthologue uncertainty may therefore improve the quantitative accuracy of future cross-species applications.

Although developed primarily for protein reconstruction, SPRINT is not intrinsically restricted to proteins as prediction targets. By training the framework directly on matched spatial metabolomic data, we reconstructed metabolite distributions in an independent mouse brain section, demonstrating that the architecture can be adapted to another molecular modality. This experiment represents an extension of the learning framework rather than transfer of a protein-trained model to metabolomics. The comparatively weaker abundance-scale concordance observed for some metabolites may reflect section- and run-specific variation in ionization efficiency, matrix effects, molecular stability and dynamic range, all of which can alter measured abundance without eliminating relative spatial organization. Modality-specific normalization, calibration objectives and measurement-aware likelihood functions could improve absolute agreement. More broadly, the same formulation could potentially be adapted to other spatially resolved molecular measurements, provided that paired training data and suitable target-specific pre-processing are available.

In summary, SPRINT links tissue morphology and spatial transcriptional states to reconstruct protein-specific molecular landscapes across variations in image resolution, tissue section, spatial location and species. Its predictions preserve both broad tissue architecture and local multivariate protein environments, and they provide complementary information for downstream spatial analysis. The successful adaptation of the framework to metabolomic targets further suggests that the underlying strategy is applicable beyond spatial proteomics. With broader paired training resources, improved calibration and explicit uncertainty modelling, SPRINT could provide a general computational interface for completing and integrating partially measured spatial multi-omics datasets.

Methods

The SPRINT model

SPRINT is a supervised multimodal regression framework that reconstructs spatially resolved molecular abundance from paired tissue images and spatial transcriptomic profiles. For each spatial location i , the model receives an image patch I_i centred on that location and a gene-expression vector $x_i \in R^G$ input genes. During training, the corresponding target vector $y_i \in R^P$, containing P measured proteins or other

500 molecular features, is used as supervision. At inference, only I_i and x_i are required. SPRINT comprises a
 501 resolution-dependent image encoder, a transcriptomic encoder, a multimodal self-attention module and a
 502 regression head.

503 **Resolution-dependent image encoding**

504 SPRINT uses separate image-processing branches for high- and low-resolution inputs. The appropriate
 505 branch is selected according to the resolution of the input image and can also be specified manually. Both
 506 branches transform an image patch into a sequence of d -dimensional tokens, where $d = 512$, allowing the
 507 downstream fusion module to use a common architecture across different imaging settings.

508 For high-resolution images, the image patch is encoded using a ResNet50 backbone initialized with
 509 ImageNet-pretrained weights. The image backbone is subsequently fine-tuned together with the remaining
 510 model components. The global average-pooling and classification layers of ResNet50 are removed, producing
 511 the feature map

$$512 \quad F_i^H = f_{R50}(I_i) \in R^{C_H \times H_H \times W_H} \quad (1)$$

513 Where $C_H = 2048$, and H_H and W_H denote the height and width of the feature map, respectively.

514 The spatial dimensions are then flattened while retaining their ordering, and the channel dimension is
 515 partitioned into groups of width d . The high-resolution image tokens are defined as

$$516 \quad T_i^H = \text{reshape}[\text{permute}(F_i^H)] \in R^{N_H \times d} \quad (2)$$

517 where

$$518 \quad N_H = H_H W_H \frac{C_H}{d} \quad (3)$$

519 Each spatial position therefore contributes one or more tokens that encode the local appearance of the
 520 corresponding image region. For an input patch producing a $2,048 \times 8 \times 8$ feature map, these operations
 521 generate 256 image tokens, each with a dimension of 512.

522 For low-resolution images, we use a lightweight residual network trained directly on the study data. This
 523 network contains residual blocks composed of 3×3 convolutional layers, batch normalization and rectified
 524 linear unit (ReLU) activation, with the channel dimension progressively increasing from 64 to 512. Its output
 525 feature map is expressed as

$$526 \quad F_i^L = f_{\text{smallResNet}}(I_i) \in R^{C_L \times H_L \times W_L} \quad (4)$$

527 Global average pooling is applied over the spatial dimensions:

$$g_i^L = \frac{1}{H_L W_L} \sum_{u=1}^{H_L} \sum_{v=1}^{W_L} F_i^L[:, u, v] \quad (5)$$

The pooled representation is subsequently partitioned into d – dimensional tokens:

$$T_i^L = \text{reshape}(g_i^L) \in R^{N_L \times d} \quad (6)$$

where

$$N_L = \frac{C_L}{d} \quad (7)$$

In the low-resolution implementation, $C_L = d = 512$, resulting in one global image token for each spatial location. This token summarizes the overall morphological context of the image patch and reduces sensitivity to unstable fine-scale variation in low-resolution images.

The image-token sequence generated by either branch is denoted by

$$T_i^I \in R^{N_I \times d} \quad (8)$$

Transcriptomic encoding and tokenization

The input transcriptomic profile is compressed into a compact latent representation using a MLP. Each intermediate layer comprises a linear transformation, batch normalization, ReLU activation and dropout. For layer l , the transformation is defined as

$$\phi_l(z) = \text{Dropout}_l\{\text{ReLU}[W_l z + b_l]\} \quad (9)$$

where W_l and b_l denote the trainable weight matrix and bias vector of layer l , respectively.

The complete transcriptomic encoder is expressed as

$$r_i = f_{RNA}(x_i) = \phi_L(\phi_{L-1}(\dots \phi_1(x_i))) \in R^{N_R \times d} \quad (10)$$

The encoded transcriptomic vector is reshaped into N_R tokens:

$$T_i^R = \text{reshape}(r_i) \in R^{N_R \times d} \quad (11)$$

For the high-resolution implementation, the RNA encoder maps the complete gene-expression vector through hidden layers of 4,096 and 2,048 dimensions and produces a 2,048-dimensional latent representation. This representation is divided into four 512-dimensions tokens. For the low-resolution implementation, the RNA encoder uses hidden layers of 2,048 and 1,024 dimensions and produces one 512-dimensional token. These

operations place the transcriptomic and image representations in the same token-embedding space.

Cross-modal Token Interaction

The image and transcriptomic tokens are concatenated and supplemented with a learnable positional embedding:

$$Z_i = [T_i^I; T_i^R] + P \in R^{(N_I+N_R) \times d} \quad (12)$$

Where p is a trainable positional embedding. Because the image and RNA tokens occupy distinct positions in the concatenated sequence, this embedding allows the model to distinguish their modality origin while retaining the spatial ordering of the image tokens.

Multimodal interactions are learned using an eight-head self-attention layer. For attention head m , the query, key and value matrices are calculated as

$$Q_i^{(m)} = Z_i W_m^Q, K_i^{(m)} = Z_i W_m^K, V_i^{(m)} = Z_i W_m^V \quad (13)$$

Where W_m^Q , W_m^K and W_m^V are trainable projection matrices.

The output of attention head m is

$$A_i^{(m)} = \text{softmax} \left(\frac{Q_i^{(m)} K_i^{(m)\top}}{\sqrt{d_h}} \right) \quad (14)$$

where

$$d_h = \frac{d}{8} = 64 \quad (15)$$

is the dimension of each attention head.

The outputs of the eight attention heads are concatenated and linearly projected:

$$A_i = \text{Concat}(A_i^{(1)}, A_i^{(2)}, \dots, A_i^{(8)}) W^O \quad (16)$$

Where W^O is a trainable output-projection matrix.

The self-attention operation allows each transcriptomic token to interact with all image and transcriptomic tokens. Because the final prediction is conditioned on the molecular state encoded by the RNA branch, only

the attention outputs corresponding to the transcriptomic-token positions are retained. A residual connection from the original RNA tokens is subsequently added:

$$H_i^R = A_i[-N_R:, :] + T_i^R \quad (17)$$

Here, $A_i[-N_R:, :]$ denotes the final N_R rows of the attention output, corresponding to the positions occupied by the transcriptomic tokens. This residual connection preserves the original transcriptomic representation while incorporating information from both input modalities.

580 Molecular abundance Reconstruction

The fused RNA-token representation is flattened and passed through a two-layer MLP regression head. The reconstructed molecular profile is calculated as

$$\hat{y}_1 = W_2 ReLU[W_1 vec(H_i^R) + b_1] + b_2 \quad (18)$$

Where $vec(H_i^R)$ denotes the flattened fused representation, W_1 and W_2 are trainable weight matrices, b_1 and b_2 are trainable bias vectors. The output

$$\hat{y}_1 \in \mathbb{R}^P \quad (19)$$

contains the predicted abundances of the P target molecular features.

For protein reconstruction, P is set to the number of proteins included in the corresponding experiment. For spatial metabolite reconstruction, P is set to the number of selected metabolite features. Apart from the resolution-dependent image encoder and the corresponding numbers of image and RNA tokens, the multimodal self-attention and molecular regression procedures are shared across experiments.

592 Loss function

SPRINT was optimized using a composite objective that combines mean squared error (MSE) with a Pearson correlation-based loss. Let y_{ip} and $\hat{y}_{ip}$ denote the measured and predicted abundance, respectively, of target p in sample i , where B is the batch size and P is the number of predicted targets. The MSE loss was defined as

$$L_{MSE} = \frac{1}{BP} \sum_{i=1}^B \sum_{p=1}^P (\hat{y}_{ip} - y_{ip})^2 \quad (20)$$

To encourage the model to recover the relative spatial variation of each target, the Pearson correlation coefficient was calculated independently for each target across the samples within a mini-batch:

$$r_p = \frac{\sum_{i=1}^B (\hat{y}_{ip} - \bar{\hat{y}}_p)(y_{ip} - \bar{y}_p)}{\sqrt{\sum_{i=1}^B (\hat{y}_{ip} - \bar{\hat{y}}_p)^2 \sum_{i=1}^B (y_{ip} - \bar{y}_p)^2 + \epsilon}} \quad (21)$$

where $\overline{\hat{y}_p}$ and $\overline{y_p}$ are the batch-wise mean predicted and measured abundances of target p , respectively, and ε is a small constant introduced for numerical stability. The PCC loss was then defined as

$$L_{PCC} = 1 - \frac{1}{P} \sum_{p=1}^P r_p \quad (22)$$

The overall training objective was

$$L_{\text{total}} = \lambda_{MSE} L_{MSE} + \lambda_{PCC} L_{PCC} \quad (23)$$

For the spatial-protein reconstruction experiments, the two components were assigned equal weights.

Datasets

All datasets used in this study were obtained from publicly available repositories and contained matched tissue images and spatial molecular measurements. No new human participants or animals were recruited for this study. The datasets were used exclusively for secondary computational analyses. The assignment of individual samples to training and evaluation sets was fixed before model evaluation and is described below.

Human tonsil dataset

Two publicly available human tonsil datasets generated using the 10x Genomics Visium CytAssist Spatial Gene and Protein Expression platform were used to evaluate cross-section protein reconstruction. Both datasets were derived from formalin-fixed, paraffin-embedded (FFPE) tissue and contained matched H&E images, spatial transcriptomic profiles and antibody-derived protein measurements processed using Space Ranger v.2.1.0.

The first dataset was generated using the standard antibody panel from a female donor in her 30s and contained 4,194 spots under tissue. This dataset was used for model training. The second dataset was generated from an independent male donor aged 25 years using the standard antibody panel supplemented with four custom add-on antibodies and contained 4,908 spots under tissue. This dataset was used exclusively for evaluation. Protein reconstruction was evaluated on the 29 proteins shared between the two antibody panels; proteins present in only one dataset were excluded. The measured protein profiles of the evaluation section were not provided to the model during inference and were used only as reference values.

The datasets are available from 10x Genomics as “Visium CytAssist Gene and Protein Expression Library of Human Tonsil, H&E, 6.5 mm (FFPE)” and “Visium CytAssist Gene and Protein Expression Library of Human Tonsil with Add-on Antibodies, H&E, 6.5 mm (FFPE)”.

Mouse spleen dataset

The mouse spleen data were obtained from the SPOTS study by Ben-Chetrit et al.[36] and downloaded from

the Gene Expression Omnibus under accession
 [GSE198353](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198353). The dataset was
 generated from fresh-frozen mouse spleen tissue using the 10x Genomics Visium platform supplemented with
 DNA-barcoded antibody staining and comprised matched tissue images, spatial transcriptomic profiles and
 measurements of 21 immune-related proteins.

Two biological replicates were used. After matching the transcriptomic, protein and spatial imaging data,
 replicate 1 contained 2,653 spots and replicate 2 contained 2,768 spots, with 32,285 gene features available
 in each sample. Replicate 1 was used for model training, whereas replicate 2 was retained for evaluation. The
 measured protein profiles from replicate 2 were excluded during inference and used only to quantify
 reconstruction performance. The relatively low-resolution image patches extracted from these data were also
 used to evaluate the low-resolution image-processing branch of SPRINT.

Human brain dataset

The human brain data were obtained from the public 10x Genomics “Visium CytAssist Gene and Protein
 Expression Library of Human Glioblastoma, IF, 11 mm (FFPE)” dataset. The dataset was generated from a
 5-μm FFPE tissue section from a female donor in her 60s and contained spatial transcriptomic profiles,
 antibody-derived protein measurements and matched immunofluorescence images. The complete dataset
 comprised 5,756 spots under tissue, 18,085 targeted genes and 31 measured proteins. The DAPI fluorescence
 channel was used as the image input in the present analysis.

To construct a spatially non-overlapping evaluation setting, the tissue section was divided diagonally
 according to its normalized spatial coordinates. Spots satisfying
 $(x_{\mathrm{norm}} > y_{\mathrm{norm}})$ were assigned to the training region, and the remaining spots
 were assigned to the evaluation region. This produced 3,283 training spots and 2,473 evaluation spots with
 no shared spatial locations. Owing to the diagonal partition, the two regions also differed in their cellular
 composition and relative cell-type abundance. Protein measurements from the evaluation region were masked
 during inference and used only as reference values.

The same human brain dataset was used to train the human model evaluated in the cross-species mouse brain
 analysis. For this experiment, the transcriptomic input was restricted to genes that could be matched between
 the human and mouse datasets.

Human and mouse breast datasets

The human breast dataset was obtained from the public 10x Genomics “Visium CytAssist Gene and Protein
 Expression Library of Human Breast Cancer, IF, 6.5 mm (FFPE)” dataset. It was generated from an FFPE
 tissue section from a female donor in her 50s and contained matched immunofluorescence imaging, spatial
 transcriptomic profiles and antibody-derived protein measurements. After matching the spatial locations
 across modalities, 4,165 spots and 18,085 gene features were retained.

The mouse breast dataset was obtained from the SPOTS study under GEO accession

665 [GSE198353](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198353). The data were generated
666 from an MMTV-PyMT mouse mammary tumour and contained matched spatial gene-expression, antibody-
667 derived protein and tissue imaging measurements. After preprocessing and modality matching, the dataset
668 contained 1,978 spots, 32,286 gene features and 11 measured proteins.

669 For cross-species reconstruction, SPRINT was trained on the human breast dataset and evaluated directly on
670 the mouse breast dataset. Gene symbols were matched between species without regard to letter case, resulting
671 in 15,462 shared transcriptomic features. Evaluation was restricted to the 11 proteins measured in the mouse
672 dataset: CD27, CD19, PDCD1, PTPRC, CD4, ITGAM, CD8A, EPCAM, PECAM1, ITGAX and CD274.
673 Mouse protein measurements were not used during model training or inference and served only as reference
674 values for performance evaluation.

675 **Mouse brain spatial transcriptomic and metabolomic datasets**

676 The mouse brain data used for cross-species protein inference, spatial-domain identification and spatial
677 metabolite reconstruction were obtained from [Zenodo record
678 15706278](https://zenodo.org/records/15706278) and downloaded as 'PD_mouse_brain.zip'. The repository
679 contained matched tissue images, spatial transcriptomic profiles, spatial metabolomic measurements and
680 anatomical region annotations for multiple mouse brain sections.

681 Two sections, denoted V11L12-109 A1 and V11L12-109 B1, were used for spatial metabolite reconstruction.
682 After alignment and spot matching, sections A1 and B1 contained 2,658 and 3,098 spots, respectively. Each
683 section contained 32,285 transcriptomic features and 1,538 aligned metabolite features. SPRINT was trained
684 directly using the imaging, transcriptomic and metabolomic measurements from section A1 and evaluated on
685 section B1, for which the measured metabolite profiles were withheld during inference.

686 To exclude poorly detected, near-constant or excessively noisy metabolite features, evaluation was restricted
687 to a quality-controlled subset defined exclusively from the measured metabolomic profiles of section B1. For
688 each metabolite, we calculated its mean abundance, detection fraction, 90th-percentile abundance and
689 coefficient of variation (CV) across all B1 spots and converted each statistic to a percentile rank among the
690 1,538 measured features. Metabolites were retained if their mean-abundance percentile was between 0.84 and
691 0.995, their detection-fraction percentile was at least 0.82, their 90th-percentile abundance was at or above
692 the 0.92 percentile, and their CV percentile was between 0.07 and 0.16. These criteria yielded 26 metabolite
693 features. The selection procedure used only properties of the measured target signals and did not use
694 predictions from SPRINT or any baseline method.

695 Section B1 was also used as the target sample in the human-to-mouse brain transfer experiment. A protein
696 reconstruction model trained using the human brain dataset was applied to the matched image and
697 transcriptomic measurements from this mouse section without mouse protein supervision. The resulting
698 predictions comprised 31 protein features and were subsequently evaluated for their utility in spatial-domain
699 identification. Of the 3,098 spots in section B1, 2,819 had available anatomical annotations and were used to
700 calculate agreement between inferred spatial domains and the reference regions. The anatomical annotations
701 were used only for evaluation and parameter selection and were not supplied as input features to SPRINT or

702 to the clustering algorithm.

703 **Baseline Methods**

704 We compared SPRINT with nine published methods using their publicly available implementations. GHIST:
705 <https://github.com/SydneyBioX/GHIST>. GigaTIME: <https://github.com/prov-gigatime/GigaTIME>. HEX:
706 <https://github.com/lilab-stanford/HEX>. HIPI: <https://github.com/RonZeira/HIPI>. ROSIE:
707 <https://gitlab.com/enable-medicine-public/rosie>. DGAT: <https://github.com/osmanbeyoglulab/DGAT>.
708 sciPENN: <https://github.com/jlakkis/sciPENN>. scTranslator:
709 <https://github.com/TencentAILabHealthcare/scTranslator>. SpatialEx: [https://github.com/KEAML-](https://github.com/KEAML-JLU/SpatialEx)
710 [JLU/SpatialEx](https://github.com/KEAML-JLU/SpatialEx).

711 **Evaluation metrics**

712 Protein-reconstruction performance was evaluated using the Pearson correlation coefficient (PCC), Spearman
713 rank correlation coefficient and concordance correlation coefficient (CCC). Let y_{ip} and $\hat{y}_{ip}$ denote the
714 measured and predicted abundance, respectively, of protein p at spot i , and let N denote the number of test
715 spots.

716 For each protein, the PCC was calculated across all test spots as

$$717 \quad PCC_p = \frac{\sum_{i=1}^N (y_{ip} - \bar{y}_p)(\hat{y}_{ip} - \bar{\hat{y}}_p)}{\sqrt{\sum_{i=1}^N (y_{ip} - \bar{y}_p)^2} \sqrt{\sum_{i=1}^N (\hat{y}_{ip} - \bar{\hat{y}}_p)^2}} \quad (24)$$

718 where $\bar{y}_p$ and $\bar{\hat{y}}_p$ are the mean measured and predicted abundances of protein p , respectively. PCC
719 evaluates the linear association between predicted and measured spatial abundance profiles.

720 The Spearman correlation coefficient was calculated by applying the PCC to the ranks of the measured and
721 predicted values:

$$722 \quad \rho_p = PCC[R(y_p), R(\hat{y}_p)] \quad (25)$$

723 where $R(\cdot)$ denotes rank transformation, with tied values assigned their average rank. Spearman correlation
724 therefore evaluates whether the relative ordering of protein abundance across spots is preserved without
725 assuming a linear relationship.

726 CCC was used to assess agreement between the predicted and measured abundance values:

$$CCC_p = \frac{2cov(y_p, \hat{y}_p)}{\sigma_{y_p}^2 + \sigma_{\hat{y}_p}^2 + (\bar{y}_p - \bar{\hat{y}}_p)} \quad (26)$$

where $\sigma_{y_p}^2$ and $\sigma_{\hat{y}_p}^2$ are the variances of the measured and predicted values, respectively. Unlike correlation coefficients, CCC penalizes differences in both the mean and variance and thus quantifies agreement in abundance scale in addition to association.

All three metrics were calculated independently for each protein or metabolite across the test spots. Overall performance was summarized using the median value across all measured targets. For CCC, the first and third quartiles were additionally reported to describe variability across targets. Higher PCC, Spearman correlation and CCC values indicate better reconstruction performance. Quantile matching was used only for visual comparisons and was not applied when calculating these metrics.

For local-neighbourhood analyses, spot-level prediction accuracy was calculated across all measured targets. For spot i , the spot-level PCC was defined as

$$PCC_i^{spot} = \frac{\sum_{p=1}^P (y_{ip} - \bar{y}_i)(\hat{y}_{ip} - \bar{\hat{y}}_i)}{\sqrt{\sum_{p=1}^P (y_{ip} - \bar{y}_i)^2} \sqrt{\sum_{p=1}^P (\hat{y}_{ip} - \bar{\hat{y}}_i)^2}} \quad (27)$$

where P is the number of evaluated targets and the means are calculated across targets within the corresponding spot. Spots were ranked according to this value to select representative neighbourhoods at different accuracy percentiles.

Spatial-domain identification was evaluated using the adjusted Rand index (ARI) and normalized mutual information (NMI). Given a contingency table in which n_{jk} is the number of spots assigned to true domain j and predicted domain k , with row sums a_j , column sums b_k and total spot number N , ARI was calculated as

$$ARI = \frac{\sum_{jk} C(n_{jk}, 2) - \frac{[\sum_j C(a_j, 2) \sum_k C(b_k, 2)]}{C(N, 2)}}{0.5[\sum_j C(a_j, 2) + \sum_k C(b_k, 2)] - \frac{[\sum_j C(a_j, 2) \sum_k C(b_k, 2)]}{C(N, 2)}} \quad (28)$$

ARI measures agreement between the predicted and annotated spatial partitions while correcting for agreement expected by chance.

NMI was calculated as

$$NMI(U, V) = \frac{2I(U; V)}{H(U) + H(V)} \quad (29)$$

where U and V denote the annotated and predicted spatial-domain assignments, $I(U; V)$ is their mutual

information, and $H(U)$ and $H(V)$ are their respective entropies. NMI measures the information shared between the two partitions independently of the numerical domain labels. Higher ARI and NMI values indicate greater agreement with the annotated spatial domains.

Ablation experiments

We performed modality-specific ablation experiments using the human tonsil datasets to quantify the respective contributions of the histological and transcriptomic branches of SPRINT. Two single-modality variants, termed H&E_only and RNA_only, were trained and evaluated using the same training–test split, target-protein panel and evaluation procedure as the complete model.

In the RNA_only model, the image-processing branch and multimodal self-attention module were removed. The transcriptomic expression vector was processed by the RNA encoder and divided into four 512-dimensional tokens, which were flattened and passed directly to the same MLP regression head used in SPRINT to predict protein abundance. In the H&E_only model, the transcriptomic branch was removed. The H&E image was processed by the ResNet50 image encoder and converted into 256 image tokens. Learnable positional embeddings and an eight-head self-attention layer were retained to model relationships among the image regions. The resulting image-token representations were averaged, replicated into four 512-dimensional tokens and passed to the same regression head. Thus, the output dimensionality and regression architecture were held constant, while the available input modality differed among the complete, H&E_only and RNA_only models.

To examine whether the contributions of the two modalities depended on the extent to which protein expression was reflected in tissue morphology, we selected representative H&E-informative and weakly H&E-informative protein groups. The H&E-informative group comprised BCL2 and CD68, whose expression was associated with histologically distinguishable tissue or cellular features. The weakly H&E-informative group comprised CD14, CD27 and CR2, whose spatial expression was less directly apparent from H&E morphology. These representative groups were used only for stratified evaluation and did not contribute to model training or parameter selection. For each group and model, performance was summarized as the median per-protein PCC, with the interquartile range calculated across the proteins within that group.

Data availability

The human tonsil datasets are publicly available from 10x Genomics: [Visium CytAssist Gene and Protein Expression Library of Human Tonsil, H&E, 6.5 mm (FFPE)](<https://www.10xgenomics.com/datasets/gene-protein-expression-library-of-human-tonsil-cytassist-ffpe-2-standard>) and [Visium CytAssist Gene and Protein Expression Library of Human Tonsil with Add-on Antibodies, H&E, 6.5 mm (FFPE)](<https://www.10xgenomics.com/datasets/visium-cytassist-gene-and-protein-expression-library-of-human-tonsil-with-add-on-antibodies-h-e-6-5-mm-ffpe-2-standard>). The mouse spleen and mouse breast datasets from the SPOTS study are available through the Gene Expression Omnibus under accession [GSE198353](<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198353>). The human brain dataset,

787 Visium CytAssist Gene and Protein Expression Library of Human Glioblastoma, IF, 11 mm (FFPE), is
 788 publicly available from [10x Genomics]([https://www.10xgenomics.com/datasets/gene-and-protein-](https://www.10xgenomics.com/datasets/gene-and-protein-expression-library-of-human-glioblastoma-cytassist-ffpe-2-standard)
 789 [expression-library-of-human-glioblastoma-cytassist-ffpe-2-standard](https://www.10xgenomics.com/datasets/gene-and-protein-expression-library-of-human-glioblastoma-cytassist-ffpe-2-standard)). The human breast dataset, Visium
 790 CytAssist Gene and Protein Expression Library of Human Breast Cancer, IF, 6.5 mm (FFPE), is publicly
 791 available from [10x Genomics]([https://www.10xgenomics.com/datasets/gene-and-protein-expression-](https://www.10xgenomics.com/datasets/gene-and-protein-expression-library-of-human-breast-cancer-cytassist-ffpe-2-standard)
 792 [library-of-human-breast-cancer-cytassist-ffpe-2-standard](https://www.10xgenomics.com/datasets/gene-and-protein-expression-library-of-human-breast-cancer-cytassist-ffpe-2-standard)). The mouse brain spatial transcriptomic and
 793 metabolomic datasets are available from [Zenodo record 15706278](<https://zenodo.org/records/15706278>)
 794 as 'PD_mouse_brain.zip'.

795 Code availability

796 The codes were implemented in Python and are released at GitHub (<https://github.com/YuzhiSun/SPRINT>).
 797 Other relevant analysis and figure plotting code is available on Google Drive
 798 (<https://drive.google.com/file/d/1mtPcxuNwSnV1NZBW6PhWcFNiSvMRd-MS/view?usp=sharing>).

799 Competing interests

800 The authors declare no competing interests.

801 Reference

- 802 [1] E. Lundberg and G. H. H. Borner, "Spatial proteomics: a powerful discovery tool for cell biology,"
 803 2019. doi: 10.1038/s41580-018-0094-y.
- 804 [2] Y. Liu, A. Beyer, and R. Aebersold, "On the Dependency of Cellular Protein Levels on mRNA
 805 Abundance," 2016. doi: 10.1016/j.cell.2016.03.014.
- 806 [3] C. Buccitelli and M. Selbach, "mRNAs, proteins and the emerging principles of gene expression
 807 control," 2020. doi: 10.1038/s41576-020-0258-4.
- 808 [4] M. Angelo *et al.*, "Multiplexed ion beam imaging of human breast tumors," *Nat. Med.*, vol. 20, no. 4,
 809 2014, doi: 10.1038/nm.3488.
- 810 [5] Y. Goltsev *et al.*, "Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging,"
 811 *Cell*, vol. 174, no. 4, 2018, doi: 10.1016/j.cell.2018.07.010.
- 812 [6] L. Keren *et al.*, "A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer
 813 Revealed by Multiplexed Ion Beam Imaging," *Cell*, vol. 174, no. 6, 2018, doi:
 814 10.1016/j.cell.2018.08.039.

815 [7] J. R. Lin *et al.*, “Highly multiplexed immunofluorescence imaging of human tissues and tumors using
816 t-CyCIF and conventional optical microscopes,” *Elife*, vol. 7, 2018, doi: 10.7554/eLife.31657.

817 [8] M. Stoeckius *et al.*, “Simultaneous epitope and transcriptome measurement in single cells,” *Nat.*
818 *Methods*, vol. 14, no. 9, 2017, doi: 10.1038/nmeth.4380.

819 [9] Y. Liu *et al.*, “High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial
820 CITE-seq,” *Nat. Biotechnol.*, vol. 41, no. 10, 2023, doi: 10.1038/s41587-023-01676-0.

821 [10] J. W. Hickey *et al.*, “Spatial mapping of protein composition and tissue organization: a primer for
822 multiplexed antibody-based imaging,” *Nat. Methods*, vol. 19, no. 3, 2022, doi: 10.1038/s41592-021-
823 01316-y.

824 [11] J. M. J. Valanarasu *et al.*, “Multimodal AI generates virtual population for tumor microenvironment
825 modeling,” *Cell*, vol. 189, no. 2, 2026, doi: 10.1016/j.cell.2025.11.016.

826 [12] Z. Li *et al.*, “AI-enabled virtual spatial proteomics from histopathology for interpretable biomarker
827 discovery in lung cancer,” *Nat. Med.*, vol. 32, no. 1, 2026, doi: 10.1038/s41591-025-04060-4.

828 [13] X. Fu *et al.*, “Spatial gene expression at single-cell resolution from histology using deep learning with
829 GHIST,” *Nat. Methods*, vol. 22, no. 9, 2025, doi: 10.1038/s41592-025-02795-z.

830 [14] R. Zeira, L. Anavy, Z. Yakhini, E. Rivlin, and D. Freedman, “HIPI: Spatially resolved multiplexed
831 protein expression inferred from H&E WSIs,” *PLoS Comput. Biol.*, vol. 20, no. 9, 2024, doi:
832 10.1371/journal.pcbi.1012501.

833 [15] E. Wu *et al.*, “ROSIE: AI generation of multiplex immunofluorescence staining from histopathology
834 images,” *Nature Communications*, vol. 16, no. 1, 2025, doi: 10.1038/s41467-025-62346-0.

835 [16] J. Lakkis *et al.*, “A multi-use deep learning method for CITE-seq and single-cell RNA-seq data
836 integration with cell surface protein prediction and imputation,” *Nat. Mach. Intell.*, vol. 4, no. 11, 2022,
837 doi: 10.1038/s42256-022-00545-w.

838 [17] L. Liu *et al.*, “A pre-trained large generative model for translating single-cell transcriptomes to
839 proteomes,” *Nat. Biomed. Eng.*, 2025, doi: 10.1038/s41551-025-01528-z.

840 [18] H. Wang *et al.*, “DGAT: a dual-graph attention network for inferring spatial protein landscapes from
841 transcriptomics,” *Nat. Commun.*, 2026, doi: 10.1038/s41467-026-73114-z.

842 [19] Y. Liu *et al.*, “High-parameter spatial multi-omics through histology-anchored integration,” *Nat.*
843 *Methods*, vol. 23, no. 2, 2026, doi: 10.1038/s41592-025-02926-6.

844 [20] Z. Wang *et al.*, “NicheTrans: Spatial-aware Cross-omics Translation,” 2024.

845 [21] P. L. Ståhl *et al.*, “Visualization and analysis of gene expression in tissue sections by spatial
846 transcriptomics,” 2016. doi: 10.1126/science.aaf2403.

847 [22] S. Vickovic *et al.*, “High-definition spatial transcriptomics for in situ tissue profiling,” *Nat. Methods*,
848 vol. 16, no. 10, 2019, doi: 10.1038/s41592-019-0548-y.

849 [23] S. G. Rodriques *et al.*, “Slide-seq: A scalable technology for measuring genome-wide expression at
850 high spatial resolution,” *Science (1979)*., vol. 363, no. 6434, 2019, doi: 10.1126/science.aaw1219.

851 [24] R. R. Stickels *et al.*, “Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-
852 seqV2,” *Nat. Biotechnol.*, vol. 39, no. 3, 2021, doi: 10.1038/s41587-020-0739-1.

853 [25] E. Zhao *et al.*, “Spatial transcriptomics at subspot resolution with BayesSpace,” *Nat. Biotechnol.*, vol.
854 39, no. 11, 2021, doi: 10.1038/s41587-021-00935-2.

855 [26] 10x Genomics, “Visium CytAssist Gene and Protein Expression Library of Human Tonsil with Add-
856 on Antibodies, H\&E, 6.5 mm (FFPE),” Aug. 2023.

857 [27] S. M. Albelda, W. A. Muller, C. A. Buck, and P. J. Newman, “Molecular and cellular properties of
858 PECAM-1 (endoCAM/CD31): A novel vascular cell-cell adhesion molecule,” *Journal of Cell Biology*,
859 vol. 114, no. 5, 1991, doi: 10.1083/jcb.114.5.1059.

860 [28] G. Spizzo *et al.*, “EpCAM expression in primary tumour tissues and metastases: An
861 immunohistochemical analysis,” *J. Clin. Pathol.*, vol. 64, no. 5, 2011, doi: 10.1136/jcp.2011.090274.

862 [29] P. E. Purkis, J. B. Steel, I. C. Mackenzie, W. B. J. Nathrath, I. M. Leigh, and E. B. Lane, “Antibody
863 markers of basal cells in complex epithelia,” *J. Cell Sci.*, vol. 97, no. 1, 1990, doi: 10.1242/jcs.97.1.39.

864 [30] J. Saylor *et al.*, “Spatial Mapping of Myeloid Cells and Macrophages by Multiplexed Tissue Staining,”
865 *Front. Immunol.*, vol. 9, 2018, doi: 10.3389/fimmu.2018.02925.

866 [31] M. F. Kramer, B. Mack, and G. Rasp, “Immunohistological expression of interleukin 16 in human
867 tonsils,” *Archives of Otolaryngology - Head and Neck Surgery*, vol. 127, no. 9, 2001, doi:
868 10.1001/archotol.127.9.1120.

869 [32] V. Avramović, V. Petrović, M. Jović, and P. Vlahović, “Quantification of cells expressing markers of
870 proliferation and apoptosis in chronic tonsillitis,” *Acta Otorhinolaryngologica Italica*, vol. 35, no. 4,
871 2015.

872 [33] C. Cobaleda, A. Schebesta, A. Delogu, and M. Busslinger, “Pax5: The guardian of B cell identity and
873 function,” 2007. doi: 10.1038/ni1454.

874 [34] A. H. Sharpe and K. E. Pauken, “The diverse functions of the PD1 inhibitory pathway,” 2018. doi:

875 10.1038/nri.2017.108.

876 [35] P. J. Murray and T. A. Wynn, “Protective and pathogenic functions of macrophage subsets,” 2011. doi:
877 10.1038/nri3073.

878 [36] N. Ben-Chetrit *et al.*, “Integration of whole transcriptome spatial profiling with protein markers,” *Nat.*
879 *Biotechnol.*, vol. 41, no. 6, 2023, doi: 10.1038/s41587-022-01536-3.

880 [37] N. V. Goncharov *et al.*, “Markers of Endothelial Cells in Normal and Pathological Conditions,” 2020.
881 doi: 10.1134/S1990747819030140.

882 [38] R. N. Germain, “t-cell development and the CD4-CD8 lineage decision,” 2002. doi: 10.1038/nri798.

883 [39] O. Gires, M. Pan, H. Schinke, M. Canis, and P. A. Baeuerle, “Expression and function of epithelial cell
884 adhesion molecule EpCAM: where are we after 40 years?,” 2020. doi: 10.1007/s10555-020-09898-3.

885 [40] H. K. Lim and H. C. O’Neill, “Identification of Stromal Cells in Spleen Which Support Myelopoiesis,”
886 *Front. Cell Dev. Biol.*, vol. 7, 2019, doi: 10.3389/fcell.2019.00001.

887 [41] D. A. Kaminski, C. Wei, Y. Qian, A. F. Rosenberg, and I. Sanz, “Advances in human B cell phenotypic
888 profiling,” 2012. doi: 10.3389/fimmu.2012.00302.

889 [42] S. Gordon, A. Plüddemann, and F. Martinez Estrada, “Macrophage heterogeneity in tissues:
890 Phenotypic diversity and functions,” *Immunol. Rev.*, vol. 262, no. 1, 2014, doi: 10.1111/imr.12223.

891 [43] Y. Y. Hey, J. K. H. Tan, and H. C. O’Neill, “Redefining myeloid cell subsets in murine spleen,” *Front.*
892 *Immunol.*, vol. 6, no. JAN, 2016, doi: 10.3389/fimmu.2015.00652.

893 [44] 10x Genomics, “Visium CytAssist Gene and Protein Expression Library of Human Glioblastoma, IF,
894 11 mm (FFPE),” May 2023.

895 [45] P. A. Roche and K. Furuta, “The ins and outs of MHC class II-mediated antigen processing and
896 presentation,” 2015. doi: 10.1038/nri3818.

897 [46] A. Etzerodt and S. K. Moestrup, “CD163 and inflammation: Biological, diagnostic, and therapeutic
898 aspects,” 2013. doi: 10.1089/ars.2012.4834.

899 [47] A. Woods, “Syndecans: Transmembrane modulators of adhesion and matrix assembly,” 2001. doi:
900 10.1172/JCI12802.

901 [48] G. A. Bishop and B. S. Hostager, “Signaling by CD40 and its mimics in B cell activation,” *Immunol.*
902 *Res.*, vol. 24, no. 2, 2001, doi: 10.1385/IR:24:2:097.

- 903 [49] C. G. Vinuesa, S. G. Tangye, B. Moser, and C. R. Mackay, “Follicular B helper T cells in antibody
904 responses and autoimmunity,” 2005. doi: 10.1038/nri1714.
- 905 [50] 10x Genomics, “Visium CytAssist Gene and Protein Expression Library of Human Breast Cancer, IF,
906 6.5 mm (FFPE),” May 2023.

907

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supp20260729.pdf](#)