NEIGHBORMIF: REFERENCE-GUIDED CROSS-SECTION MOLECULAR TRANSFER FOR MULTIPLEX IMMUNOFLUORESCENCE RECOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

In spatial multi-omics studies, different molecular modalities are often measured on separate tissue sections. Existing approaches mainly infer molecular expression from H&E morphology or learn cross-omics mappings, but their performance remains limited for protein markers that are weakly associated with morphology or other omics features. To address this, we propose NeighborMIF, a reference-guided framework for cross-section molecular information transfer. NeighborMIF performs cell-level matching to transfer experimentally measured protein signals from reference sections, reconstructs the transferred molecular representations in the spatial feature space of the target section, and integrates them with H&E features for missing mIF recovery. On the consecutive-section dataset, NeighborMIF achieves an average cell-level Pearson correlation of 82.83% ($r \times 100$), outperforming the best external method by 6.11 percentage points. On ORION CRC under the cross-sample non-contiguous-reference setting, reference molecular information consistently improves the same-architecture H&E-only baseline across all 15 proteins. These results demonstrate that experimentally measured reference molecular signals provide effective complementary evidence beyond H&E morphology, offering a new computational pathway for cross-section molecular information transfer and spatial multi-omics integration.

1 INTRODUCTION

Biological systems are highly complex, and no single modality is sufficient to fully characterize cellular states and their surrounding microenvironments. Conventional hematoxylin and eosin (H&E) staining provides rich morphological information, while recent advances in spatial omics technologies have enabled molecular profiling with preserved spatial context. Spatial transcriptomics reveals the spatial heterogeneity of gene expression, whereas spatial proteomic imaging techniques, such as immunohistochemistry and multiplex immunofluorescence, characterize protein expression, cellular phenotypes, immune states, and the tumor microenvironment. Ideally, morphological, transcriptomic, and proteomic information would be jointly acquired from the same tissue region, or even at the same cellular resolution, enabling a more comprehensive understanding of cellular states and their interactions with the tissue microenvironment, and providing richer molecular evidence for studies of disease mechanisms, patient stratification, and clinical diagnosis.

Different spatial omics technologies exhibit substantial differences in sample preparation procedures, experimental workflows, detection throughput, spatial resolution, and cost. Consequently, it remains challenging to acquire comprehensive multi-omics information from a single tissue section in large-scale tissue studies. Instead, a more common experimental strategy is to perform distinct molecular measurements on serial sections derived from the same tissue block, followed by computational integration across sections to recover missing spatial molecular modalities in the target section. To enable multi-omics integration within the same tissue, existing approaches typically recover missing molecular modalities through computational prediction.

These methods can be broadly categorized into two groups. The first group directly predicts protein or gene expression from H&E images , as illustrated in Fig. 1(a). However, their performance is inherently limited because molecular expression is not fully determined by observable tissue mor-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

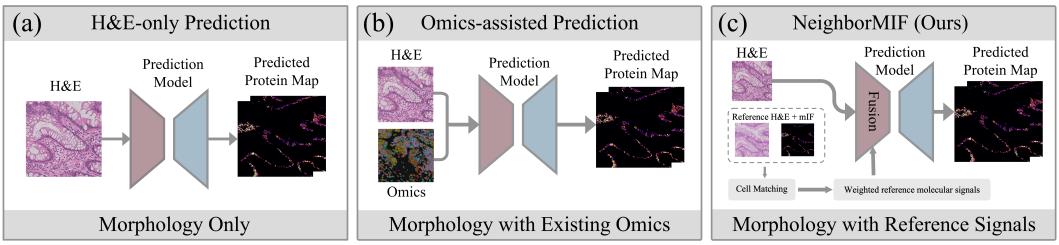

Figure 1: Comparison of protein prediction paradigms. (a) H&E-only prediction uses tissue morphology alone. (b) Omics-assisted prediction combines H&E with existing omics data. (c) The proposed NeighborMIF fuses H&E features with weighted reference molecular signals.

phology. Some biomarkers exhibit weak associations with morphological features, making their spatial expression difficult to accurately infer from H&E images alone. The second group predicts missing molecular modalities from already measured omics data by learning cross-modal mappings, as illustrated in Fig. 1(b). However, such cross-omics prediction relies on learned relationships between heterogeneous molecular modalities and may introduce additional biases when estimating intrinsic biological associations. For example, spatial transcriptomics combined with histological images has been used to predict spatial protein expression (Fig. 1b). Nevertheless, transcriptomic measurements do not always faithfully reflect protein abundance due to post-transcriptional regulation and other biological factors (Buccitelli & Selbach, 2020; Upadhya & Ryan, 2022), which may introduce additional uncertainty when transcriptomic signals are used to predict protein expression.

Serial sections provide an alternative source of molecular information. When a molecular modality missing in the target section has already been experimentally measured in an adjacent section, the reference molecular measurements can potentially be transferred across sections. Although adjacent sections do not exhibit strict one-to-one cellular correspondence, their shared tissue architecture and local cellular environments provide valuable structural constraints for cross-section molecular information transfer.

Based on this observation, we propose NeighborMIF, a reference-guided framework for cross-slice molecular information transfer , as illustrated in Fig. 1(c) and Fig. 2. NeighborMIF leverages shared tissue structures across consecutive sections to transfer experimentally measured molecular evidence from reference sections, enabling the recovery of missing protein signals in target sections. It extends morphology-driven molecular inference toward reference-guided recovery using measured molecular information, providing a new computational paradigm for spatial multi-omics integration across serial sections.

To summarize, our main contributions are as follows:

- We introduce a new reference-guided problem setting that explicitly incorporates experimentally measured molecular information from a neighboring section as inference-time evidence, rather than relying solely on the morphology of the target section.
- We propose **NeighborMIF**, a reference-guided molecular transfer framework that integrates experimentally measured cross-section molecular evidence into a multiscale H&E prediction network through fine-grained mask-aware cellular representations, tissue-constrained donor retrieval, and deep residual molecular feature fusion.
- We validate NeighborMIF in both serial-section and non-neighboring-reference settings. Across both settings, experimentally measured reference molecular evidence consistently improves the corresponding H&E-only baseline, demonstrating its complementary value beyond morphology-only molecular prediction.

2 Related Work

Morphology-driven molecular expression prediction Deep learning has been widely used to predict immunostaining, protein expression, and spatial gene expression from H&E images. Early virtual staining methods mainly focused on a limited number of immune markers. For example,

VirtualMultiplexer generates multiplex immunostaining from unpaired H&E and IHC images (Pati et al., 2024), whereas Multi-V-Stain, DeepLIIF, and HEMIT learn mappings from H&E images to protein expression using paired or co-stained data (Andani et al., 2023; Ghahremani et al., 2022; Bian et al., 2025). With the emergence of larger-scale spatial proteomics datasets, ROSIE and HistoPlexer extended this task to multichannel protein prediction (Wu et al., 2025; Andani et al., 2025). MIPHEI-ViT and HEX further incorporated representations from pathology foundation models to improve multiplex protein-expression prediction (Balezo et al., 2026; Li et al., 2026). Beyond protein expression, GHIST extended a similar morphology-to-molecular paradigm to single-cell spatial gene-expression prediction (Fu et al., 2025). In these approaches, real molecular measurements primarily serve as supervision during training, whereas inference mainly depends on morphological features from the target section itself. Consequently, predictive performance for molecular markers without prominent visual phenotypes remains fundamentally constrained by the strength of the morphology–molecular-state association.

Cross-omics inference of missing modalities To move beyond morphology-only prediction, another line of work predicts missing modalities from already measured omics modalities by learning statistical associations across molecular layers. Early studies primarily focused on predicting protein expression from RNA. cTP-net learns a cross-modal mapping from paired single-cell RNA–protein data and uses it to impute surface-protein expression for cells with only single-cell RNA-seq measurements (Zhou et al., 2020). sciPENN and totalVI further jointly model RNA and protein measurements for protein prediction, missing-modality imputation, and multi-omics integration (Lakkis et al., 2022; Gayoso et al., 2021). Subsequently, BABEL and scButterfly extended cross-modal prediction to more general combinations of single-cell modalities, including RNA, chromatin accessibility, and protein (Wu et al., 2021; Cao et al., 2024). This idea has also been extended to spatial omics. SpatialEx/SpatialEx+ addresses spatial diagonal integration, in which complementary molecular marker sets are measured on different serial tissue sections. It uses histological information as a common anchor and applies cross-omics consistency constraints to integrate marker sets and omics layers in space (Liu et al., 2026). However, computationally predicted omics information is inherently constrained by both the input modality and the learned mapping. When the downstream objective is itself to study genuine associations across omics layers, such computational mappings may further affect subsequent cross-omics association analyses.

These studies demonstrate that cross-section information can provide additional evidence for molecular prediction. However, existing approaches mainly transfer shared expression patterns across samples. In contrast, NeighborMIF directly utilizes experimentally measured protein expression from adjacent sections as molecular evidence during inference and investigates how such information can be transferred to target sections without requiring strict cell-to-cell correspondence, thereby improving the recovery of missing protein signals.

3 METHOD

In this section, the NeighborMIF framework is introduced in Section 3.1, followed by cell-level morphological feature extraction in Section 3.2. Reference molecular signal extraction is detailed in Section 3.3. Reference molecular feature fusion and mIF reconstruction are presented in Section 3.4.

3.1 FRAMEWORK OVERVIEW

As summarized in Fig. 2, given a target H&E whole-slide image (WSI) and a reference H&E WSI acquired from an adjacent section of the same tissue block, WSI-level registration is first performed to estimate the transformation parameters for tissue localization. A target H&E patch I^t_{HE} is then cropped from the target WSI, and its location is mapped to the reference WSI using the estimated transformation parameters to extract the corresponding reference H&E patch I^r_{HE}. The patch I^t_{HE} is paired with a spatially aligned measured multiplex immunofluorescence (mIF) patch Y^t_{mIF}, which is used only for training supervision. The reference H&E patch I^r_{HE} and its spatially aligned measured mIF patch Y^r_{mIF} jointly provide morphological and molecular reference information to guide target mIF prediction during both training and inference. Within each pair of H&E patches, cell-level morphological features establish associations between target and reference cells. These associations determine the weights used to aggregate molecular signals extracted from the measured reference

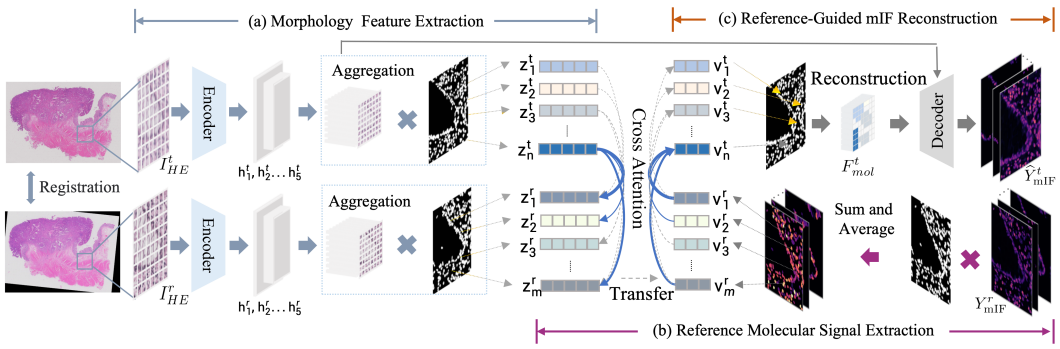

Figure 2: Overview of NeighborMIF. Following WSI-level registration, the framework comprises three stages: (a) Morphology feature extraction: H&E features from the target and reference patches are aggregated using cell masks to obtain cell-level morphological features. (b) Reference molecular signal extraction: measured reference mIF signals are averaged within cell masks, then weighted and transferred to target cells through cross-attention-based matching. (c) Reference-guided mIF reconstruction: transferred signals are reconstructed into a spatial molecular feature map, fused with target H&E features, and decoded to generate target mIF predictions.

mIF patch Y^r_{mIF} for each target cell, accommodating differences in cellular composition between adjacent tissue sections. The transferred signals are then reconstructed into a spatial molecular feature map $\mathbf{F}^t_{\mathrm{mol}}$ using the target cell masks. After channel projection, $\mathbf{F}^t_{\mathrm{mol}}$ is incorporated into the target H&E bottleneck through residual fusion, while multiscale target H&E features provide spatial detail for the decoder to generate the target mIF prediction $\widehat{Y}^t_{\mathrm{mIF}}$. This formulation combines reference protein measurements with the morphology and spatial organization of the target patch.

3.2 Cell-Level Morphological Feature Extraction

Target-Reference cell-level matching requires sensitivity to fine-grained morphological variations. Dense features are extracted from the frozen H-optimus-0 backbone (Saillard et al., 2024) and combined with whole-cell instance masks generated by VitaminP (Shokrollahi et al., 2026) to construct cell-level morphological features. The visual tokens are mapped back to the H&E patch coordinate system according to their spatial locations, forming a dense spatial feature grid. Weighted aggregation based on the overlap between token regions and cell masks improves the coverage of individual cells. This construction preserves the local semantic information encoded by the backbone and the spatial structure of cell instances, as illustrated in Fig. 2(a). Let C^s_i denote the mask of cell i in I^s_{HE}, with $s \in \{t, r\}$. The target and reference inputs contain N_t and M_r cells, respectively.

Multiscale H&E encoding. NeighborMIF uses a dual-stream H&E encoder composed of a shallow CNN and a frozen H-optimus-0 backbone, producing the five-scale features $\{\mathbf{h}^s_\ell\}_{\ell=1}^5$ for each H&E patch. The CNN branch directly processes the H&E patch to extract high-resolution local morphological features and produces $\mathbf{h}^s_1$ and $\mathbf{h}^s_2$. In parallel, H-optimus-0 extracts deep visual-token features from multiple Transformer layers. Trainable feature adapters align their channel dimensions, after which weighted fusion produces a unified deep semantic feature map $\mathbf{F}^s_{\mathrm{sem}}$. This map is resampled to multiple spatial scales to construct middle- and high-level features. At the $\mathbf{h}^s_3$ scale, it is fused with downsampled CNN features, whereas $\mathbf{h}^s_4$ and $\mathbf{h}^s_5$ are mapped from $\mathbf{F}^s_{\mathrm{sem}}$ at their corresponding resolutions. For the target patch, $\mathbf{h}^t_5$ serves as the decoder bottleneck, while $\mathbf{h}^t_1$–$\mathbf{h}^t_4$ provide multiscale skip features. H-optimus-0 remains frozen during training, while the remaining encoding and fusion modules are optimized.

Mask-based feature aggregation. Visual tokens from the frozen H-optimus-0 backbone are mapped to their spatial positions in each H&E patch to obtain cell-level morphological features. Let $\mathbf{u}^s_k$ be the feature of visual token k, and let G^s_k be its spatial support in the H&E patch coordi-

nate system. For cell i, define $\mathcal{T}_i^s$ as the set of tokens overlapping C_i^s. The feature is

$$\mathbf{z}_i^s = \frac{\sum_{k\in\mathcal{T}_i^s} \mathrm{Area}(C_i^s \cap G_k^s)\, \mathbf{u}_k^s}{\sum_{k\in\mathcal{T}_i^s} \mathrm{Area}(C_i^s \cap G_k^s) + \epsilon}, \qquad s \in \{t, r\}, \tag{1}$$

where $\mathrm{Area}(C_i^s \cap G_k^s)$ is the overlap area between cell mask C_i^s and token support G_k^s, and ϵ is a small constant for numerical stability. The same aggregation is applied to the target and reference H&E patches, yielding cell-level morphological features $\mathbf{z}_i^t$ and $\mathbf{z}_j^r$ in a shared feature space.

3.3 Reference Molecular Signal Extraction

Serial tissue sections exhibit local deformation, cell displacement, and cell appearance or disappearance, so absolute spatial coordinates alone cannot establish reliable one-to-one cellular correspondence. NeighborMIF therefore combines WSI-level registration with learnable cell-level matching to transfer reference molecular signals, as illustrated in Fig. 2(b).

Reference molecular signal extraction. Let P denote the number of protein channels and $Y_{\mathrm{mIF}}^r(p) \in \mathbb{R}^P$ the measured mIF signal at pixel p in the reference patch. Because the reference H&E and mIF patches are spatially aligned, the reference cell mask C_j^r identifies the same cell in both modalities. The reference molecular signal is obtained by averaging over the pixels in the mask:

$$\mathbf{v}_j^r = \frac{1}{|C_j^r|} \sum_{p\in C_j^r} Y_{\mathrm{mIF}}^r(p), \tag{2}$$

where $|C_j^r|$ is the number of pixels in the nonempty mask and $\mathbf{v}_j^r \in \mathbb{R}^P$ contains the measured protein intensities of reference cell j.

Candidate selection using WSI-level registration. Based on the morphology of the target and reference H&E WSIs, reliable shared tissue anchors are identified through global translation estimation and local multiscale cross-correlation. The displacements and matching confidences of nearby anchors are used for weighted interpolation to predict a reference search center for each target cell. Two candidate pools are then constructed: reference cells in the neighborhood of the predicted position, and reference cells retrieved from a larger spatial region according to their cell-level morphological features. The latter compensates for missed candidates caused by local correspondence failures or larger tissue deformations. These pools form the candidate set $\mathcal{A}_i$ for target cell i.

Cell-level matching and weighted transfer. The spatial constraint from WSI-level registration is used only to restrict the candidate set, while final donor selection is performed by learnable cell-level matching. For target cell i and candidate reference cell $j \in \mathcal{A}_i$, a shared feature encoder $\mathcal{E}$ and learnable query and key projections W_Q and W_K produce normalized matching features:

$$\mathbf{q}_i = \mathrm{Norm}\big(W_Q \mathcal{E}(\mathbf{z}_i^t)\big), \qquad \mathbf{k}_j = \mathrm{Norm}\big(W_K \mathcal{E}(\mathbf{z}_j^r)\big), \tag{3}$$

where Norm denotes L2 normalization. Their matching score is

$$s_{ij} = \min(\exp(\gamma), 100)\, \mathbf{q}_i^\top \mathbf{k}_j, \tag{4}$$

with learnable logit-scale parameter γ. The top-K candidates, with $K = 5$, form the donor set $\mathcal{D}_i$. Softmax normalization within this set yields the transfer weights and the target cell's transferred reference molecular signal:

$$\alpha_{ij} = \frac{\exp(s_{ij})}{\sum_{j'\in\mathcal{D}_i} \exp(s_{ij'})}, \qquad \mathbf{v}_i^t = \sum_{j\in\mathcal{D}_i} \alpha_{ij} \mathbf{v}_j^r, \qquad j \in \mathcal{D}_i. \tag{5}$$

Here, α_{ij} denotes the normalized matching weight and $\mathbf{v}_j^r$ is the experimentally measured mIF vector of reference donor cell j. Their weighted combination $\mathbf{v}_i^t$ represents the reference molecular signal transferred to target cell i. The cross-attention operation in Fig. 2(b) uses cell-level morphological features for matching and reference molecular signals for transfer.

Spatial molecular feature reconstruction. Because the transferred reference molecular signals are represented at discrete target cells, the target cell masks are used to reproject them onto a spatial grid aligned with $\mathbf{h}_5^t$, as shown in Fig. 2(c). Let B_k^t denote the region of the target H&E patch corresponding to grid location k. The transferred signals are aggregated according to the overlap of their cell masks with this region:

$$\mathbf{m}_k^t = \frac{\sum_{i=1}^{N_t} \mathrm{Area}(C_i^t \cap B_k^t)\, \mathbf{v}_i^t}{\sum_{i=1}^{N_t} \mathrm{Area}(C_i^t \cap B_k^t) + \epsilon}. \tag{6}$$

The summation includes target cells overlapping B_k^t, and $\mathrm{Area}(C_i^t \cap B_k^t)$ measures the overlap between a cell mask and the corresponding spatial region. If no cell covers B_k^t, its molecular feature is set to zero. Arranging $\mathbf{m}_k^t$ according to the grid locations yields the spatial molecular feature map $\mathbf{F}_{\mathrm{mol}}^t \in \mathbb{R}^{H_5 \times W_5 \times P}$, where H_5 and W_5 are the spatial dimensions of the target H&E bottleneck.

3.4 Reference Molecular Feature Fusion and mIF Reconstruction

The spatial molecular feature map is fused with the target H&E features at a deep semantic level, as shown in Fig. 2(c). The target H&E patch provides the spatial structure, while the transferred reference molecular signals serve as supplementary evidence. Spatial reconstruction aligns $\mathbf{F}_{\mathrm{mol}}^t$ with $\mathbf{h}_5^t$ in spatial resolution. A learnable molecular projector $\mathcal{P}_{\mathrm{mol}}$ maps its P protein channels to the channel dimension of $\mathbf{h}_5^t$:

$$\widetilde{\mathbf{F}}_{\mathrm{mol}}^t = \mathcal{P}_{\mathrm{mol}}(\mathbf{F}_{\mathrm{mol}}^t). \tag{7}$$

The target H&E bottleneck and projected molecular features are concatenated along the channel dimension. A learnable residual fusion module $\mathcal{G}_{\mathrm{res}}$ learns the correction contributed by the reference molecular signals:

$$\Delta\mathbf{F}^t = \mathcal{G}_{\mathrm{res}}\Big([\mathbf{h}_5^t; \widetilde{\mathbf{F}}_{\mathrm{mol}}^t]\Big), \qquad \mathbf{F}_{\mathrm{fused}}^t = \mathbf{h}_5^t + \Delta\mathbf{F}^t, \tag{8}$$

where $[\cdot;\cdot]$ denotes channel concatenation, $\Delta\mathbf{F}^t$ is the learned correction, and $\mathbf{F}_{\mathrm{fused}}^t$ is the bottleneck feature after incorporating reference molecular signals. The parameters of the final layer of $\mathcal{G}_{\mathrm{res}}$ are initialized to zero. The model therefore preserves the original target H&E features at the beginning of training and gradually learns the incremental correction contributed by the reference molecular signals.

The UNet3+ decoder (Huang et al., 2020), denoted by $\mathcal{D}_{\mathrm{mIF}}$, combines the fused bottleneck with the target multiscale skip features:

$$\widehat{Y}_{\mathrm{mIF}}^t = \mathcal{D}_{\mathrm{mIF}}\big(\mathbf{F}_{\mathrm{fused}}^t, \mathbf{h}_1^t, \mathbf{h}_2^t, \mathbf{h}_3^t, \mathbf{h}_4^t\big). \tag{9}$$

The output contains P predicted protein channels in the target H&E patch coordinate system. The skip pathway shown in Fig. 2 supplies target morphological information at multiple scales during reconstruction.

Training objective. NeighborMIF is trained end-to-end with cell-level supervision. Predicted and measured mIF intensities from $\widehat{Y}_{\mathrm{mIF}}^t$ and Y_{mIF}^t are averaged within each target cell mask C_i^t to obtain cell-level protein vectors. The training objective combines their Smooth L1 loss, with $\beta = 0.05$, and a Pearson correlation loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{SmoothL1}}^{\mathrm{cell}} + 0.5\, \mathcal{L}_{\mathrm{Pearson}}^{\mathrm{cell}}. \tag{10}$$

The Pearson term is omitted when fewer than two valid cells are available or when either the predicted or measured values have insufficient variance. No additional pixel-level or branch-specific auxiliary loss is used. The cell-matching module uses no correspondence labels or contrastive loss; its shared feature encoder and query and key projections are optimized through the downstream mIF reconstruction objective. Top-5 donor selection is discrete, while gradients propagate through the Softmax weights of the selected donors and the subsequent reconstruction and fusion modules.

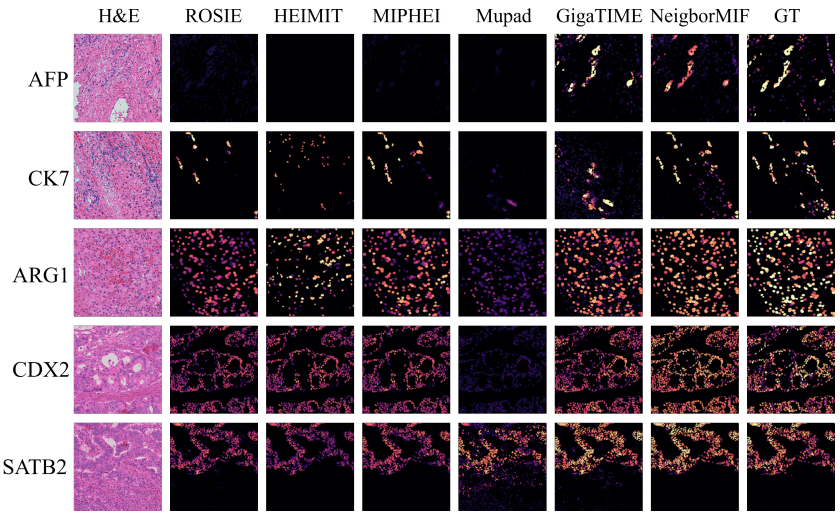

Figure 3: Representative qualitative comparison on the serial-section dataset for AFP, CK7, ARG1, CDX2, and SATB2. Each row corresponds to one protein marker. From left to right, the columns show the input H&E image, predictions from ROSIE, HEMIT, MIPHEI-ViT, Mupad, GigaTIME, and NeighborMIF, followed by the ground-truth mIF signal.

4 Experiments

4.1 Experimental Setup and Implementation Details

Datasets. We evaluate NeighborMIF on an in-house consecutive-section dataset and the public ORION CRC dataset (Lin et al., 2023). The former contains seven physically adjacent section pairs from colorectal adenocarcinoma and hepatocellular carcinoma cases, while ORION CRC is used for cross-sample non-contiguous-reference evaluation. All splits are patient- or case-disjoint to prevent information leakage. Detailed dataset statistics, preprocessing, and reference-bank construction protocols are provided in Appendix A.1.

Comparison Methods and Evaluation Metrics. NeighborMIF is compared with five publicly available molecular prediction or virtual staining methods: MIPHEI-ViT (Balezo et al., 2026), ROSIE (Wu et al., 2025), HEMIT (Bian et al., 2025), MUPAD (Xiang et al., 2026), and GigaTIME (Valanarasu et al., 2026). Cell-level Pearson correlation is used as the primary evaluation metric, computed between predicted and measured mIF expression values after aggregation within individual cell instances. A detailed comparison with the same-architecture H&E-only baseline is provided in the Appendix.

Implementation Details. NeighborMIF is trained using the AdamW optimizer with an initial learning rate of 1×10^{-4}, together with linear warm-up and automatic mixed-precision training. We use a batch size of 4 and train the model for 20 epochs. The H-optimus-0 backbone remains frozen throughout training, while all remaining learnable components, including the CNN encoder, feature adaptation modules, molecular projection module, residual fusion module, and decoder, are jointly optimized using the training objective described above.

4.2 Evaluation on Serial Sections

Evaluation Setting. We evaluate NeighborMIF on the in-house consecutive-section dataset across five protein markers: CDX2, SATB2, CK7, AFP, and ARG1. NeighborMIF is compared with five publicly available molecular prediction or virtual staining methods: HEMIT, ROSIE, MIPHEI-ViT, MUPAD, and GigaTIME. Detailed dataset statistics and preprocessing procedures are provided in Appendix A.1.

Results. As shown in Table 1, NeighborMIF achieves the best performance across all five protein markers, with a mean cell-level Pearson correlation of 82.83% ($r \times 100$). Compared with the best-performing external method, which achieves 76.72% on average, NeighborMIF improves the mean correlation by 6.11 percentage points. The improvement is consistent across all five markers, demonstrating the effectiveness of NeighborMIF for molecular recovery from consecutive tissue sections. A detailed comparison with the same-architecture H&E-only baseline is provided in Appendix A.2.

Table 1: Cell-level Pearson correlation comparison on the consecutive-section dataset. All Pearson correlation coefficients are reported as percentages ($r \times 100$), and bold denotes the best result for each protein. Δ denotes the absolute improvement of NeighborMIF over the best-performing external method, measured in percentage points.

Method	AFP	CK7	ARG1	CDX2	SATB2	Mean
HEMIT	37.67	26.63	57.32	45.78	35.28	40.54
ROSIE	50.13	63.79	63.79	47.65	79.59	60.99
MIPHEI-ViT	56.62	74.70	77.11	90.47	84.69	76.72
MUPAD	45.30	47.34	53.12	56.42	53.81	51.20
GigaTIME	64.47	67.36	75.60	80.28	76.57	72.86
NeighborMIF	**73.29**	**76.58**	**85.94**	**91.27**	**87.05**	**82.83**
Δ	+8.82	+1.88	+8.83	+0.80	+2.36	+6.11

4.3 Cross-Sample Evaluation with Non-Contiguous References

Dataset and Preprocessing. To evaluate NeighborMIF without physically adjacent reference sections, we conduct cross-sample non-contiguous-section experiments on the public ORION CRC dataset, which provides cell-level registered H&E and mIF measurements for 15 protein markers. Following the split used by MIPHEI-ViT, we use 37 sections for training, 2 for validation, and 2 for testing, with patient-level disjoint splitting to prevent information leakage.

Cross-sample Reference Construction. Because ORION CRC does not provide physically adjacent references with spatial correspondence to the target section, we disable the tissue-level localization used in the consecutive-section setting, including global translation estimation, local multiscale cross-correlation, and anchor-based displacement interpolation. Instead, for each target H&E region, morphologically similar reference regions are retrieved from non-contiguous training sections using H&E information only. Mask-aware cellular descriptors are then extracted from the target and retrieved reference regions, and the learned Query–Key matcher selects the Top-5 reference donor cells for each target cell. Their measured mIF signals are aggregated as reference molecular evidence. The subsequent spatial reconstruction, molecular projection, residual fusion, and decoding modules remain unchanged.

Results. As shown in Table 2, NeighborMIF achieves a mean cell-level Pearson correlation of 50.40% ($r \times 100$) in the cross-sample non-contiguous-reference setting, closely approaching the best external result of 51.74%. NeighborMIF outperforms GigaTIME (44.33%), ROSIE (36.88%), and MUPAD (31.16%) in mean performance, and achieves the best result among the compared methods for CD45 and PDL1. These results indicate that reference-guided molecular transfer remains effective even when physically adjacent reference sections are unavailable. The contribution of reference molecular evidence relative to the same-architecture H&E-only baseline is further analyzed in Appendix.

4.4 Ablation on Molecular Feature Fusion

Fusion Strategies and Results. We compare the proposed zero-initialized ResidualFusion with DirectFusion, which directly replaces the H&E bottleneck representation after concatenating H&E and molecular features. ResidualFusion instead preserves the H&E pathway and learns an additive molecular correction initialized from zero. As shown in Table 3, ResidualFusion improves the mean cell-level Pearson correlation from 80.46% to 82.83%, corresponding to a gain of 2.37 per-

Table 2: Cell-level Pearson correlation comparison on the ORION CRC dataset under the cross-sample non-contiguous-section setting. All values are Pearson correlation coefficients reported as percentages ($r \times 100$). Bold denotes the best result for each protein. "–" indicates an undefined Pearson correlation due to zero variance in the predicted values.

Protein	HEMIT	ROSIE	MIPHEI-ViT	MUPAD	GigaTIME	NeighborMIF
CD31	–	23.73	43.27	22.78	**51.05**	37.35
CD45	–	59.93	74.92	48.17	66.36	**77.89**
CD68	–	25.30	39.28	23.71	**50.48**	37.93
CD4	–	51.06	**62.93**	38.03	50.18	57.99
FOXP3	–	10.53	**17.57**	9.37	4.59	15.12
CD8a	–	24.15	**29.98**	10.97	18.76	26.62
CD45RO	–	46.26	**58.76**	44.55	33.59	57.04
CD20	–	32.17	**64.41**	25.94	37.47	60.83
PDL1	–	21.38	37.16	22.14	32.96	**50.78**
CD3e	–	57.02	**65.42**	53.72	39.62	62.30
CD163	–	29.10	**43.25**	21.56	39.69	41.45
E-Cadherin	15.47	37.37	55.58	28.79	**83.06**	60.67
Ki67	–	28.71	43.52	31.16	**61.86**	39.96
Pan-CK	61.53	75.82	**83.34**	50.20	79.31	80.25
SMA	–	30.60	**56.78**	36.23	16.02	49.78
Mean	–	36.88	**51.74**	31.16	44.33	50.40

centage points. The largest gains are observed for CK7, CDX2, and SATB2, supporting the residual incorporation of reference molecular evidence.

Table 3: Ablation study of molecular feature fusion strategies on the consecutive-section dataset. All values are cell-level Pearson correlation coefficients reported as percentages ($r \times 100$). Bold denotes the better result between the two fusion strategies. Δ denotes the absolute change from DirectFusion to ResidualFusion, measured in percentage points.

Method	AFP	CK7	ARG1	CDX2	SATB2	Mean
DirectFusion	**73.35**	73.74	**86.00**	87.54	81.65	80.46
ResidualFusion	73.29	**76.58**	85.94	**91.27**	**87.05**	**82.83**
Δ	-0.06	+2.84	-0.06	+3.73	+5.40	+2.37

5 Conclusion and Future Work

We presented NeighborMIF, a novel reference-guided framework that transfers experimentally measured molecular information across tissue sections to recover missing multiplex immunofluorescence signals. In the consecutive-section setting, NeighborMIF achieves overall superior performance across five protein markers compared with MIPHEI-ViT, ROSIE, HEMIT, MUPAD, and GigaTIME. In the cross-sample non-contiguous-reference setting, NeighborMIF also achieves competitive prediction performance, demonstrating that the proposed transfer mechanism can extend beyond physically adjacent reference sections.

NeighborMIF extends morphology-only molecular prediction by incorporating measured reference molecular evidence during inference. Our results show that existing molecular measurements can provide complementary information beyond H&E morphology, offering a new computational strategy for cross-section molecular information transfer and spatial multi-omics integration. In future work, we aim to generalize this reference-guided transfer paradigm beyond mIF and protein measurements to broader spatial omics modalities, enabling cross-section transfer and recovery of heterogeneous molecular information across different assays.

REFERENCES

Sonali Andani, Boqi Chen, Joanna Ficek-Pascual, Simon Heinke, Ruben Casanova, Bettina Sobottka, Bernd Bodenmiller, Viktor Koelzer, and Gunnar Ratsch. Multi-v-stain: Multiplexed virtual staining of histopathology whole-slide images. In *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023. URL https://openreview.net/forum?id=3tu9EnRoWl.

Sonali Andani, Boqi Chen, Joanna Ficek-Pascual, Simon Heinke, Ruben Casanova, Bernard Friedrich Hild, Bettina Sobottka, Bernd Bodenmiller, Viktor H. Koelzer, and Gunnar Rätsch. Histopathology-based protein multiplex generation using deep learning. *Nature Machine Intelligence*, 7(8):1292–1307, August 2025. ISSN 2522-5839. doi: 10.1038/s42256-025-01074-y.

Guillaume Balezo, Roger Trullo, Albert Pla Planas, Etienne Decencière, and Thomas Walter. MIPHEI-ViT: Multiplex immunofluorescence prediction from H&E images using ViT foundation models. *Computers in Biology and Medicine*, 206:111564, April 2026. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2026.111564.

Chang Bian, Beth Phillips, Tim Cootes, and Martin Fergie. HEMIT: H&E to Multiplex-Immunohistochemistry Image Translation with Dual-Branch Pix2pix Generator. In Anna Schroder, Xiang Li, Tanveer Syeda-Mahmood, Neil P. Oxtoby, Alexandra Young, Alessa Hering, Tejas S. Mathai, Pritam Mukherjee, Sven Kuckertz, Tiantian He, Isaac Llorente-Saguer, Andreas Maier, Satyananda Kashyap, Hayit Greenspan, and Anant Madabhushi (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 Workshops*, pp. 184–197, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-84525-3. doi: 10.1007/978-3-031-84525-3_16.

Christopher Buccitelli and Matthias Selbach. mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*, 21(10):630–644, October 2020. ISSN 1471-0064. doi: 10.1038/s41576-020-0258-4.

Yichuan Cao, Xiamiao Zhao, Songming Tang, Qun Jiang, Sijie Li, Siyu Li, and Shengquan Chen. scButterfly: A versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. *Nature Communications*, 15(1):2973, April 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-47418-x.

Xiaohang Fu, Yue Cao, Beilei Bian, Chuhan Wang, Dinny Graham, Nirmala Pathmanathan, Ellis Patrick, Jinman Kim, and Jean Yee Hwa Yang. Spatial gene expression at single-cell resolution from histology using deep learning with GHIST. *Nature Methods*, 22(9):1900–1910, September 2025. ISSN 1548-7105. doi: 10.1038/s41592-025-02795-z.

Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L. Nazor, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods*, 18(3):272–282, March 2021. ISSN 1548-7105. doi: 10.1038/s41592-020-01050-x.

Parmida Ghahremani, Yanyun Li, Arie Kaufman, Rami Vanguri, Noah Greenwald, Michael Angelo, Travis J. Hollmann, and Saad Nadeem. Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification. *Nature Machine Intelligence*, 4(4):401–412, April 2022. ISSN 2522-5839. doi: 10.1038/s42256-022-00471-x.

Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, May 2020. doi: 10.1109/ICASSP40776.2020.9053405.

Justin Lakkis, Amelia Schroeder, Kenong Su, Michelle Y.Y. Lee, Alexander C. Bashore, Muredach P. Reilly, and Mingyao Li. A multi-use deep learning method for CITE-seq and single-cell RNA-seq data integration with cell surface protein prediction and imputation. *Nature machine intelligence*, 4(11):940–952, November 2022. ISSN 2522-5839. doi: 10.1038/s42256-022-00545-w.

Zhe Li, Yuchen Li, Jinxi Xiang, Xiyue Wang, Sen Yang, Xiaoming Zhang, Feyisope Eweje, Yijiang Chen, Xiangde Luo, Yuanyuan Li, Jonathan Mulholland, Colin Bergstrom, Ted Kim, Francesca Maria Olguin, Sierra Willens, Jeffrey J. Nirschl, Robert West, Joel Neal, Maximilian Diehn, and Ruijiang Li. AI-enabled virtual spatial proteomics from histopathology for interpretable biomarker discovery in lung cancer. *Nature Medicine*, 32(1):231–244, January 2026. ISSN 1546-170X. doi: 10.1038/s41591-025-04060-4.

Jia-Ren Lin, Yu-An Chen, Daniel Campton, Jeremy Cooper, Shannon Coy, Clarence Yapp, Juliann B. Tefft, Erin McCarty, Keith L. Ligon, Scott J. Rodig, Steven Reese, Tad George, Sandro Santagata, and Peter K. Sorger. High-plex immunofluorescence imaging and traditional histology of the same tissue section for discovering image-based biomarkers. *Nature Cancer*, 4(7): 1036–1052, July 2023. ISSN 2662-1347. doi: 10.1038/s43018-023-00576-1.

Yonghao Liu, Chuyao Wang, Zhikang Wang, Liang Chen, Zhi Li, Jiangning Song, Qi Zou, Rui Gao, Bin-Zhi Qian, Xiaoyue Feng, Renchu Guan, and Zhiyuan Yuan. High-parameter spatial multiomics through histology-anchored integration. *Nature Methods*, 23(2):373–386, February 2026. ISSN 1548-7105. doi: 10.1038/s41592-025-02926-6.

Pushpak Pati, Sofia Karkampouna, Francesco Bonollo, Eva Compérat, Martina Radić, Martin Spahn, Adriano Martinelli, Martin Wartenberg, Marianna Kruithof-de Julio, and Marianna Rapsomaniki. Accelerating histopathology workflows with generative AI-based virtually multiplexed tumour profiling. *Nature Machine Intelligence*, 6(9):1077–1093, September 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00889-5.

Charlie Saillard, Rodolphe Jenatton, Felipe Llinares-López, Zelda Mariet, David Cahané, Eric Durand, and Jean-Philippe Vert. H-optimus-0. `https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0`, 2024. Foundation model release.

Yasin Shokrollahi, Karina B. Pinao Gonzales, Elizve N. Barrientos Toro, Paul Acosta, Patient Mosaic Team, Pingjun Chen, Yinyin Yuan, and Xiaoxi Pan. VitaminP: Cross-modal learning enables whole-cell segmentation from routine histology, April 2026.

Swathi Ramachandra Upadhya and Colm J. Ryan. Experimental reproducibility limits the correlation between mRNA and protein abundances in tumor proteomic profiles. *Cell Reports Methods*, 2(9), September 2022. ISSN 2667-2375. doi: 10.1016/j.crmeth.2022.100288.

Jeya Maria Jose Valanarasu, Hanwen Xu, Naoto Usuyama, Chanwoo Kim, Cliff Wong, Peniel Argaw, Racheli Ben Shimol, Angela Crabtree, Kevin Matlock, Alexandra Q. Bartlett, Jaspreet Bagga, Yu Gu, Sheng Zhang, Tristan Naumann, Bernard A. Fox, Bill Wright, Ari Robicsek, Brian Piening, Carlo Bifulco, Sheng Wang, and Hoifung Poon. Multimodal AI generates virtual population for tumor microenvironment modeling. *Cell*, 189(2):386–400.e19, January 2026. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2025.11.016.

Eric Wu, Matthew Bieniosek, Zhenqin Wu, Nitya Thakkar, Gregory W. Charville, Ahmad Makky, Christian M. Schürch, Jeroen R. Huyghe, Ulrike Peters, Christopher I. Li, Li Li, Hannah Giba, Vivek Behera, Arjun Raman, Alexandro E. Trevino, Aaron T. Mayer, and James Zou. ROSIE: AI generation of multiplex immunofluorescence staining from histopathology images. *Nature Communications*, 16(1):7633, August 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-62346-0.

Kevin E. Wu, Kathryn E. Yost, Howard Y. Chang, and James Zou. BABEL enables cross-modality translation between multiomic profiles at single-cell resolution. *Proceedings of the National Academy of Sciences*, 118(15):e2023070118, April 2021. doi: 10.1073/pnas.2023070118.

Jinxi Xiang, Mingjie Li, Siyu Hou, Yijiang Chen, Xiangde Luo, Yuanfeng Ji, Xiang Zhou, Ehsan Adeli, Akshay Chaudhari, Curtis P. Langlotz, Kilian M. Pohl, and Ruijiang Li. A generative foundation model for multimodal histopathology, 2026. URL `https://arxiv.org/abs/2604.03635`.

Zilu Zhou, Chengzhong Ye, Jingshu Wang, and Nancy R. Zhang. Surface protein imputation from single cell transcriptomes by deep neural networks. *Nature Communications*, 11(1):651, January 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-14391-0.

A Appendix

A.1 Dataset Details and Evaluation Splits

We summarize the datasets, preprocessing pipelines, data splits, and reference-bank construction protocols used in this study.

Table 4: Datasets used in this study. The in-house dataset evaluates molecular transfer between physically adjacent consecutive tissue sections, whereas ORION CRC evaluates cross-sample transfer using non-contiguous references.

Dataset	Access	Scale	Primary Role
In-house consecutive-section	Private	7 cases / 7 section pairs	Consecutive-section transfer
ORION CRC	Public	40 patients / 41 WSIs	Cross-sample reference transfer

In-house Consecutive-Section Dataset. The in-house dataset contains seven pairs of physically adjacent tissue sections collected from three colorectal adenocarcinoma and four hepatocellular carcinoma cases. Each pair provides H&E images and experimentally measured mIF signals from adjacent sections. CDX2 and SATB2 are evaluated on the colorectal adenocarcinoma cases, whereas CK7, AFP, and ARG1 are evaluated on the hepatocellular carcinoma cases.

After tissue-level registration, the sections are cropped into image patches, and whole-cell instance masks are generated using VitaminP for cell-level representation construction and molecular transfer. We adopt bidirectional section transfer with leave-one-case-out evaluation and ensure case-level disjointness between training, validation, and testing. For the CDX2/SATB2 tasks, the training, validation, and test sets contain 2,482, 619, and 3,027 patches, respectively. For CK7/AFP/ARG1, the corresponding numbers are 2,417, 625, and 2,653. These patch counts are reported for a single transfer direction.

ORION CRC Dataset and Preprocessing. ORION CRC (Lin et al., 2023) contains paired H&E and multiplex immunofluorescence whole-slide images acquired from the same tissue section. We use the preprocessed ORION release associated with MIPHEI-ViT (Balezo et al., 2026), rather than directly cropping the raw fluorescence WSIs. In this release, WSIs are tiled into 512×512 image patches at approximately $20\times$ magnification (0.5 μm/pixel). The mIF preprocessing pipeline includes autofluorescence removal, mIF artifact filtering, percentile clipping, log transformation, and 8-bit normalization. Nuclear instance masks are generated using Cellpose.

The dataset contains 41 H&E–mIF WSI pairs from 40 patients and covers 15 protein markers. Following the MIPHEI-ViT split, 37 WSIs from 36 patients are used for training, while two patients are used for validation and two patients for testing. The resulting training, validation, and test sets contain 295,096, 12,402, and 10,952 image patches, respectively. Patient-level disjointness is maintained across all three splits.

ORION Reference-Bank Protocol. To avoid information leakage, experimentally measured mIF signals used as reference evidence are drawn only from the training split. For validation and test queries, candidate reference patches are restricted to training WSIs, and no mIF information from the validation or test split is used for reference retrieval or molecular transfer.

For a training query, candidate references are selected from other eligible training sections; the query's own physical section and duplicate copies are excluded, while another section from the same patient may be used when available. For each query H&E patch, morphology-based retrieval is performed over the eligible training reference bank, and the Top-3 reference patches are retained. Valid cells from these three patches are pooled, after which the learned Query–Key matcher selects the Top-5 donor cells for each query cell. The experimentally measured mIF signals of these donor cells are aggregated and transferred to the query cell as reference molecular evidence.

A.2 Contribution of Reference Molecular Evidence

Consecutive-Section Reference Contribution. To isolate the contribution of transferred molecular information from the reference section, we construct an H&E-only baseline using exactly the same prediction architecture as NeighborMIF but removing all reference molecular inputs. The H&E encoder, multiscale feature hierarchy, decoder, training objective, and evaluation protocol are otherwise kept unchanged. Therefore, the performance difference between the two variants directly measures the additional contribution of reference molecular evidence.

As shown in Table 5, incorporating reference molecular information improves the cell-level Pearson correlation for all five protein markers. The mean correlation increases from 78.46% with the H&E-only baseline to 82.83% with NeighborMIF, corresponding to an absolute improvement of 4.37 percentage points. The largest gains are observed for AFP, CK7, and ARG1, while positive improvements are also obtained for CDX2 and SATB2. These results demonstrate that experimentally measured molecular signals from reference sections provide complementary information beyond target H&E morphology.

Table 5: Effect of reference molecular evidence on the consecutive-section dataset. All values are cell-level Pearson correlation coefficients reported as percentages ($r \times 100$). The H&E-only baseline uses the same prediction architecture and training objective as NeighborMIF but excludes reference molecular information. Δ denotes the absolute improvement of NeighborMIF over the H&E-only baseline, measured in percentage points.

Method	AFP	CK7	ARG1	CDX2	SATB2	Mean
H&E-only	65.19	70.89	80.01	90.31	85.88	78.46
NeighborMIF	**73.29**	**76.58**	**85.94**	**91.27**	**87.05**	**82.83**
Δ	+8.10	+5.69	+5.93	+0.96	+1.17	+4.37

Qualitative Comparison. We additionally provide representative qualitative comparisons on the consecutive-section dataset to complement the quantitative results. As shown in Fig. 4, the visualization compares the predicted spatial protein-expression patterns produced by NeighborMIF and multiple existing methods with the corresponding ground-truth mIF signals.

Cross-Sample Reference Contribution on ORION CRC. To further evaluate whether reference molecular evidence remains beneficial when physically adjacent sections are unavailable, we compare NeighborMIF with its same-architecture H&E-only baseline on the ORION CRC dataset. Both variants use the same H&E encoder, decoder, training objective, and data split; the only difference is whether cross-sample reference molecular information is introduced.

As shown in Table 6, incorporating cross-sample reference molecular evidence improves the cell-level Pearson correlation for all 15 protein markers. The mean correlation increases from 47.80 to 50.40 ($r \times 100$), corresponding to an absolute improvement of 2.60 percentage points. The largest gains are observed for CD8a, CD45RO, Pan-CK, and PDL1, with improvements of 7.80, 5.45, 4.65, and 4.25 percentage points, respectively. These results indicate that morphologically retrieved reference molecular information can provide complementary evidence even when physically adjacent reference sections are unavailable.

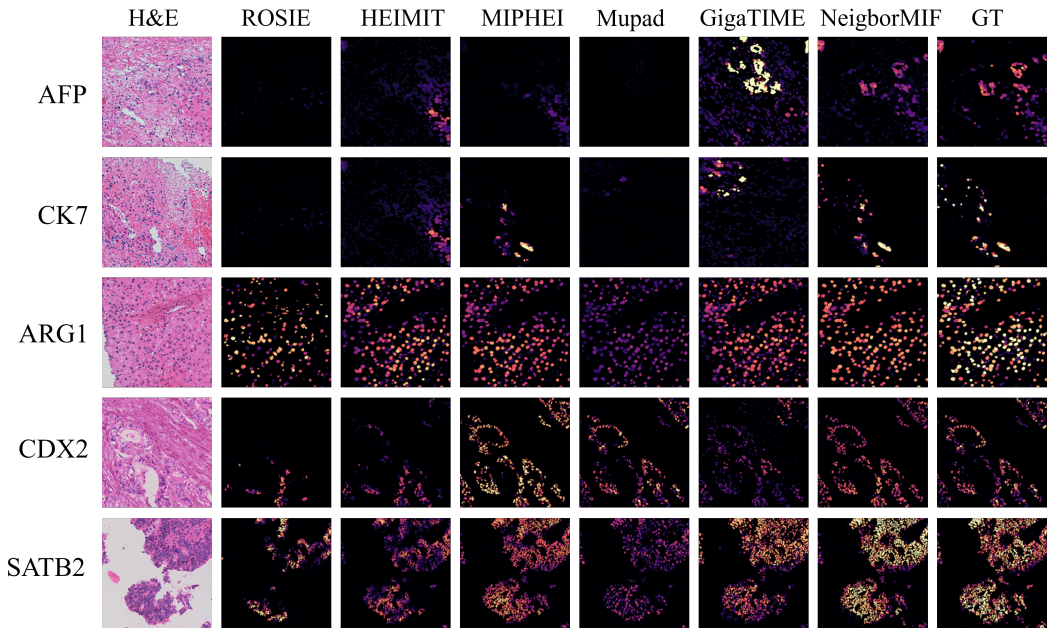

Figure 4: Representative qualitative comparison on the consecutive-section dataset for AFP, CK7, ARG1, CDX2, and SATB2. From left to right, the columns show the input H&E image, predictions from ROSIE, HEMIT, MIPHEI-ViT, MUPAD, GigaTIME, and NeighborMIF, followed by the corresponding ground-truth mIF signal.

Table 6: Effect of cross-sample reference molecular evidence on the ORION CRC dataset. All values are cell-level Pearson correlation coefficients reported as percentages ($r \times 100$). The H&E-only baseline uses the same prediction architecture and training objective as NeighborMIF but excludes reference molecular information. Δ denotes the absolute improvement of NeighborMIF over the H&E-only baseline, measured in percentage points.

Protein	**H&E-only Baseline**	**NeighborMIF**	**Δ**
CD31	36.77	**37.35**	+0.58
CD45	77.12	**77.89**	+0.77
CD68	36.65	**37.93**	+1.28
CD4	57.12	**57.99**	+0.87
FOXP3	13.17	**15.12**	+1.95
CD8a	18.82	**26.62**	+7.80
CD45RO	51.59	**57.04**	+5.45
CD20	59.58	**60.83**	+1.25
PDL1	46.53	**50.78**	+4.25
CD3e	61.64	**62.30**	+0.66
CD163	37.60	**41.45**	+3.85
E-Cadherin	56.82	**60.67**	+3.85
Ki67	39.40	**39.96**	+0.56
Pan-CK	75.60	**80.25**	+4.65
SMA	48.62	**49.78**	+1.16
Mean	47.80	**50.40**	**+2.60**