STRUCTURED GENERATIVE MODELING FOR PIXEL-LEVEL MULTIPLEX PROTEIN PREDICTION FROM H&E WITH PRODICE

Anonymous authors
Paper under double-blind review

ABSTRACT

Multiplex spatial proteomics enables spatially resolved characterization of protein expression but remains costly and difficult to scale, whereas hematoxylin and eosin (H&E) imaging is routinely collected and widely available. Existing histology-to-spatial-omics methods predict gene expression at either the multicell spot level or the single-cell level, with single-cell prediction typically requiring cell segmentation to delineate cells before molecular profiles can be assigned. Pixel-level approaches avoid this requirement, but most produce deterministic predictions for individual protein channels rather than modeling the joint conditional distribution of the multiplex protein panel, limiting their ability to capture cross-protein dependencies. We introduce **ProDICE** (**Pro**tein **D**iffusion from **I**mage-**C**onditioned **E**mbeddings), a conditional generative framework for pixel-level prediction of multiplex protein expression from H&E. **ProDICE** performs diffusion by modeling the joint conditional protein distribution given histological features, enabling the simultaneous prediction of spatially organized protein maps while preserving cross-protein structure. We evaluate **ProDICE** on held-out subjects from colorectal cancer and heterogeneous multi-tumor cohorts against four competing methods. **ProDICE** achieves the best overall performance across Pearson and Spearman correlation, SSIM, Wasserstein distribution distance, and variance retention, while better preserving protein-protein correlation structure. We further demonstrate clinical relevance of **ProDICE** predicted protein profiles in downstream prediction tasks on TCGA-CRC. These results demonstrate the potential of joint generative modeling for scalable virtual spatial proteomics. *Code is available at* `https://anonymous.4open.science/r/prodice-iclr2027-review-25CD`.

1 INTRODUCTION

Spatial protein expression provides a direct view of cellular states, tissue compartments, and interactions within the tissue microenvironment. Highly multiplexed imaging platforms such as CODEX/PhenoCycler can measure dozens of proteins simultaneously while preserving their spatial organization(Goltsev et al., 2018; Black et al., 2021), enabling detailed characterization of heterogeneous tissue architecture and cellular phenotypes. However, these assays remain costly, require specialized instrumentation and tissue processing, and are difficult to deploy at the scale of routine histopathology. In contrast, hematoxylin and eosin (H&E) staining is inexpensive, routinely collected, and widely available across large clinical and research cohorts. Because tissue morphology captured by H&E reflects underlying cellular composition and organization, predicting spatial protein expression from H&E offers a potential route to extend molecular characterization to specimens for which multiplex protein measurements are unavailable. Such virtual spatial proteomics could enable retrospective analysis of existing histology cohorts and provide a scalable complement to experimentally measured spatial proteomics.

Predicting multiplex protein expression from H&E, however, is not simply a collection of independent image-to-image regression tasks. A multiplex protein panel is a structured spatial output: individual proteins exhibit distinct expression patterns, while their joint variation reflects shared cell types, tissue compartments, and spatial organization. A model may therefore produce plausible pre-

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.

dictions for individual proteins while still distorting spatial variation, expression distributions, or relationships across protein channels. Accurate virtual spatial proteomics should consequently recover not only individual protein maps, but also the spatial and cross-protein structure of multiplex panel.

Recent methods have established the feasibility of predicting dense multiplex protein expression from H&E, including ROSIE (Wu et al., 2025), and HEX (Li et al., 2026); we discuss these and related approaches in detail in Section 2. Most of these methods can predict multiple proteins simultaneously, but treat prediction as a deterministic mapping. This motivates our central question: can the full multiplex protein panel be modeled as a single structured generative target, such that protein-specific spatial patterns and cross-protein dependencies are learned jointly from H&E?

We address this problem with **ProDICE** (**Pro**tein **D**iffusion from **I**mage-**C**onditioned **E**mbeddings), a conditional generative framework for pixel-level prediction of multiplex protein expression from H&E. To our knowledge, **ProDICE** is the first diffusion method in the area of pixel-level histology-to-spatial-CODEX prediction. **ProDICE** models the complete protein tensor jointly and performs diffusion directly in normalized protein-pixel space, without a separately trained target encoder or decoder.All protein channels therefore participate in the same noisy state and shared denoising process, allowing the model to capture protein-specific spatial patterns together with dependencies across the panel. To condition this generative process on tissue morphology, **ProDICE** combines spatially aligned local H&E features with patch-level morphology representations and morphology-conditioned protein queries, providing complementary information about local appearance and broader tissue context.

We evaluate **ProDICE** on held-out subjects from a colorectal cancer cohort and a heterogeneous multi-tumor cohort against four competing methods. Across both cohorts, **ProDICE** achieves the strongest overall performance across complementary measures of pixel-level agreement, spatial structure, distributional similarity, and variance preservation, while more accurately recovering protein–protein correlation patterns within the multiplex panel. Ablation studies further examine the contributions of the histological conditioning components to the final model. We additionally apply **ProDICE** to 388 TCGA colorectal cancer whole-slide images comprising 710,812 tissue tiles and show that virtual protein representations retain information relevant to patient-level prediction of tumor stage and molecular subtype, supporting their utility beyond local image reconstruction.

Our contributions are:

- **Joint generative modeling of multiplex spatial proteins.** We formulate pixel-level H&E-to-protein prediction as conditional diffusion of the complete multiplex protein panel, with all protein channels generated in protein-pixel space rather than independently by marker or through a separately trained target representation.
- **Evaluation of multiplex structure beyond individual-protein accuracy.** We assess not only reconstruction of individual protein maps but also spatial variation, distributional agreement, variance retention, and preservation of protein–protein correlation structure across two independent cohorts, showing that methods with plausible individual-channel predictions can nevertheless produce distorted multiplex panels.
- **Clinical-scale downstream evaluation.** We applied case-averaged **ProDICE** virtual-protein features to 388 TCGA colorectal cancer whole-slide images comprising 710,812 tissue tiles and evaluate aggregated virtual-protein representations using patient-level nested cross-validation for tumor stage and molecular subtype prediction, demonstrating that predicted protein profiles retain informative structure beyond pixel-level reconstruction.

2 Related Work

Virtual spatial proteomics builds on a broader line of work that predicts molecular information from routine histology. In spatial transcriptomics, methods such as ST-Net, Hist2ST, BLEEP, and more recent generative approaches infer spatial gene expression from H&E (He et al., 2020; Zeng et al., 2022; Xie et al., 2023; Zhu et al., 2025; Hu et al., 2026). Collectively, these methods establish that tissue morphology can support prediction of spatially resolved molecular signals. However,

they typically operate at spot resolution and predict gene expression rather than dense multiplex protein images. We therefore focus on work most directly related to virtual spatial proteomics and conditional generative modeling.

Virtual spatial proteomics and multiplex image translation. A growing body of work has explored direct prediction of protein-associated images from histology, including DeepLIIF, HEMIT, VirtualMultiplexer, and HistoPlexer (Ghahremani et al., 2022; Bian et al., 2025; Pati et al., 2024; Andani et al., 2025). However, most methods remain a cell/spot/nuclei-level predictor, and few fully leverage the complete information which imaging-based proteomics provides. This reliance on model-defined units can limit downstream analyses of extracellular signals and broader tissue-compartment organization. Among methods trained to predict CODEX imagery at the pixel level, ROSIE performs patch-wise multitask regression with ConvNeXt and HEX uses pathology foundation-model representations for simultaneous biomarker prediction (Wu et al., 2025; Li et al., 2026). However, these two methods both produce a deterministic point estimate of the multiplex output.

Conditional diffusion. Diffusion models learn generative distributions through iterative denoising (Ho et al., 2020), with velocipy parameterization providing an alternative denoising objective (Salimans & Ho, 2022) and conditional image-to-image formulations enable generation conditioned on an input image (Saharia et al., 2022). For diffusion models, a key design choice is especially relevant to multiplex proteins predictions: the space in which denoising occurs. Latent diffusion reduces computational cost by first encoding the target into a lower-dimensional representation and performing denoising in that latent space. (Kingma & Welling, 2014; Rombach et al., 2022). However, this introduces a separate target representation and decoding stage, such that the fidelity of the final prediction depends not only on the conditional diffusion model but also on how accurately the target can be represented and reconstructed. For virtual spatial proteomics, where the target is itself a dense multi-channel protein image, **ProDICE** instead performs diffusion directly in the normalized protein-pixel space. This avoids a separately trained encoder and decoder and allows the model to directly optimize reconstruction of the complete protein panel, at the cost of denoising at full spatial resolution. **ProDICE** further conditions this process on complementary histological information through spatially aligned local features, patch-level morphology, and protein-specific queries.

3 METHODS

We formulate virtual spatial proteomics as conditional generation of a complete multiplex protein image from spatially aligned H&E morphology. **ProDICE** performs diffusion directly in normalized protein-pixel space without a separately trained protein encoder, decoder, or target-side representation-learning stage. This avoids two limitations of a two-stage target-representation pipeline: prediction accuracy is not capped by a separately trained protein autoencoder, and prediction and decoding errors are not compounded across models trained under different objectives. The trade-off is computational, since denoising occurs at full spatial resolution over all P protein channels. We mitigate this cost with a comparatively narrow denoiser and a 50-step sampler. Figure 1 summarizes morphology conditioning, diffusion training on paired protein tiles, and tiled full-region reconstruction.

3.1 PROBLEM FORMULATION

For tissue region i, let $x_i \in [0,1]^{3\times H_i\times W_i}$ and $y_i \in [0,1]^{P\times H_i\times W_i}$ denote the RGB H&E image and its spatially aligned multiplex protein image with P channels. Each protein channel and each H&E RGB channel is normalized independently within each region by clipping to its 1st and 99.5th percentiles and rescaling to $[0,1]$ (Appendix E). As a result, protein values represent relative within-region spatial signal rather than absolute cross-patient abundance. **ProDICE** operates on fixed 384×384 tiles. Let $\mathcal{C}_{ik}$ denote the operator that extracts tile k from region i. Applying it to the paired images yields the spatially aligned H&E tile $x_{ik} = \mathcal{C}_{ik}(x_i)$ and protein target $y_{ik} = \mathcal{C}_{ik}(y_i)$. Both are mapped from $[0,1]$ to $[-1,1]$ according to $\tilde{x}_{ik} = 2x_{ik} - 1$ and $\tilde{y}_{ik} = 2y_{ik} - 1$. Rather than regressing a single point estimate, we model the conditional distribution $p_\theta(\tilde{y}_{ik} \mid \tilde{x}_{ik})$ of the complete normalized protein tile given its spatially aligned H&E tile.

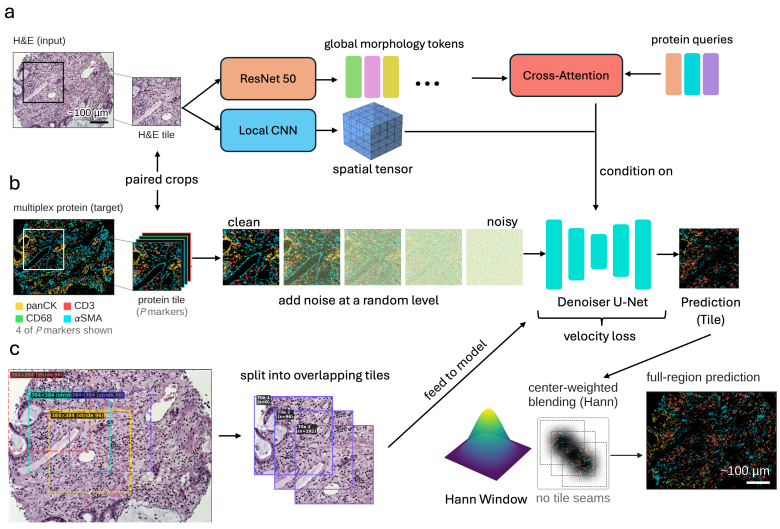

Figure 1: **ProDICE architecture, training, and full-region reconstruction.** **(a)** Each H&E tile supplies complementary global and local morphology information. A frozen ResNet-50 produces global morphology tokens that condition learned protein-aware queries through cross-attention, while a local convolutional encoder produces spatially aligned features. Both signals condition the denoising U-Net. **(b)** During training, a spatially paired multiplex protein tile is corrupted at a randomly sampled noise level. The H&E-conditioned U-Net jointly processes the noisy marker panel and predicts its diffusion velocity. Optimization uses the velocity-prediction loss. The resulting clean protein-space estimate provides the multiplex prediction. **(c)** For full-region inference, the H&E image is divided into overlapping 384×384 tiles at stride 96. Tile-level predictions are combined by center-weighted Hann blending, reducing boundary seams and producing a full-resolution multiplex protein panel.

3.2 H&E MORPHOLOGY CONDITIONING

Two complementary conditioning signals are derived from the same 384×384 H&E tile: a spatially aligned local representation that preserves correspondence with the protein-pixel diffusion state, and a patch-level representation that summarizes broader morphology within the tile. The model receives no additional or larger field of view.

Locally aligned morphology. A convolutional encoder produces a 64-channel feature map $L = c_{\text{local}}(\tilde{x}) \in \mathbb{R}^{64\times48\times48}$ at eightfold spatial reduction. This matches the $H/8$ denoiser resolution and preserves pixel-level correspondence between tissue morphology and the diffusion state. Inside the denoiser, L is projected by a 1×1 convolution and added at $H/8$.

Patch-level (global) morphology. A frozen ResNet-50 (He et al., 2016) pretrained on ImageNet (Deng et al., 2009) summarizes the same tile into morphology tokens $M = c_{\text{morph}}(\tilde{x}) \in \mathbb{R}^{4\times256}$. Its terminal feature map is adaptively pooled to a 2×2 grid, then passed through a trainable projection and normalization (Appendix F.1). Because these four tokens summarize the entire tile, we refer to them as global morphology, in contrast to the spatially aligned local features.

3.3 CONDITIONAL DIFFUSION IN PROTEIN-PIXEL SPACE

Suppressing region and tile indices, let $s_0 = \tilde{y}$ denote a clean normalized protein tile. The forward process uses $T = 1000$ timesteps with a cosine schedule (Nichol & Dhariwal, 2021) rescaled to zero terminal signal-to-noise ratio (Lin et al., 2024), ensuring that the terminal state carries no residual signal and matches sampling from pure noise. For $t \sim \text{Uniform}\{0, \ldots, T-1\}$ and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$,

the noisy state and velocity target (Salimans & Ho, 2022; Lin et al., 2024) are

$$s_t = \sqrt{\bar{\alpha}_t}\, s_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \tag{1}$$

$$v_t^\star = \sqrt{\bar{\alpha}_t}\, \epsilon - \sqrt{1-\bar{\alpha}_t}\, s_0, \tag{2}$$

The conditional model predicts $\hat{v}_t = v_\theta(s_t, t, \tilde{x})$ through the conditioning signals L and M. Because s_t remains in normalized protein units, no separate target-side representation is introduced.

3.4 Protein-aware conditional denoiser

The denoiser is a four-level U-Net that takes the complete noisy P-channel protein tile and returns a velocity estimate of the same shape. The denoiser therefore processes all protein channels simultaneously throughout the diffusion process rather than applying separate denoisers to individual markers. Residual blocks receive a sinusoidal timestep embedding, L is injected at $H/8$, and cross-attention at the $H/8$ bottleneck attends over the morphology and protein-conditioning tokens defined below (Appendix F.1).

Protein-query conditioning. Although all protein channels are denoised jointly, their relationships with tissue morphology are protein-specific. We therefore learn one query per protein, $Q \in \mathbb{R}^{P\times 256}$. For each tile, the queries attend to the morphology tokens, $Q' = \mathrm{LayerNorm}\big(Q + \mathrm{MHA}(Q, M, M)\big)$, so each marker's conditioning vector adapts to the tile's morphology. The context $[\,M; Q'\,] \in \mathbb{R}^{(4+P)\times 256}$ is used by bottleneck cross-attention, while projections of the mean morphology and query tokens are added to the timestep embedding that conditions the residual blocks.

Marker-specific output modulation. Let $\hat{v}_t^{\mathrm{raw}}$ denote the unconstrained P-channel U-Net output. Each conditioned query is projected to candidate scale and bias values for every output channel, giving $S, B \in \mathbb{R}^{P\times P}$. The final architecture retains only the marker-matched diagonal terms $s_p = S_{pp}$ and $b_p = B_{pp}$:

$$\hat{v}_{t,p} = \hat{v}_{t,p}^{\mathrm{raw}}\big(1 + \tanh s_p\big) + \tanh b_p. \tag{3}$$

This bounded affine transformation provides marker-specific gain and offset conditioning, while channels remain coupled through the joint diffusion state, shared U-Net hierarchy, common conditioning, and bottleneck attention.

3.5 Training objective

ProDICE is trained with the velocity-prediction objective

$$\mathcal{L}_{\mathrm{diffusion}} = \mathbb{E}_{t,\epsilon}\Big[\big\|v_\theta(s_t, t, \tilde{x}) - v_t^\star\big\|_2^2\Big], \tag{4}$$

with the expectation also taken over sampled training tiles. All components except the frozen ResNet-50 backbone are optimized jointly (Appendix F.3).

3.6 Inference and full-region reconstruction

At inference, each H&E tile conditions a reverse trajectory initialized at $s_T \sim \mathcal{N}(0, \mathbf{I})$, $s_T \in \mathbb{R}^{P\times 384\times 384}$, and denoised with $K = 50$ DDIM steps (Song et al., 2021). Stochasticity follows a cosine schedule over reverse steps $j \in \{0, \ldots, K-1\}$, $\eta_j = \big(1 - \cos(\pi j/(K-1))\big)/2$: the first transition is deterministic, then intermediate transitions inject Gaussian noise, and the terminal transition returns the predicted clean sample without additional noise (Appendix F.4). The terminal state $\hat{s}_0$ already lies in normalized protein space so it requires no decoding, and each tile is mapped back to the intensity space as $\hat{y} = \mathrm{clip}\big((\hat{s}_0 + 1)/2, 0, 1\big)$ before spatial blending. Full regions are reconstructed from overlapping 384×384 tiles at stride 96 (75% overlap), each sampled from an independently initialized trajectory, and combined by Hann-weighted overlap-add (Eq. (7), Appendix F.4).

4 EXPERIMENTAL SETUP

4.1 DATASETS AND SPLITS

We evaluate two cohorts in which H&E and multiplex CODEX images were acquired from the same physical tissue section in a shared coordinate frame (Appendix E, Table 6). To avoid data leakage, all regions from the same subject share a single training, validation, or test role. **CRC** is derived from the public `CRC_FFPE-CODEX_CellNeighs` collection associated with the colorectal-cancer CODEX study (Schürch et al., 2020a;b). It contains 35 subjects, four regions per subject, and 56 markers. Our subject-disjoint data split comprises 24 training subjects (96 regions), 4 validation subjects (16 regions), and 7 held-out test subjects (28 regions). **Multi-Tumor** contains 66 regions from 66 unique subjects and 58 markers. It is evaluated by strict five-fold subject-grouped outer cross-validation, so every subject is predicted exactly once and the 66 out-of-fold (OOF) predictions form the pooled evaluation set.

4.2 TRAINING AND MODEL SELECTION

Architecture selection uses CRC validation data only, and neither held-out CRC test regions nor Multi-Tumor outer-test regions are used to choose the final architecture or hyperparameters. The locked configuration (Appendix Table 7) is the model of Section 3, with explicit off-diagonal cross-marker output mixing disabled. CRC final models are trained with seeds 42, 314, and 2718. For Multi-Tumor, the same architecture is retrained independently within each outer fold using seed 42.

4.3 BASELINES

We compare **ProDICE** with two control baselines and two deep models. **Mean Prior Floor** broadcasts each marker's training-only pixel-weighted mean over the evaluation image. Since each predicted channel is spatially constant, Pearson and Spearman correlations are undefined rather than zero. **H&E Feature Ridge** fits multi-output ridge regression ($\alpha = 10$) from ten fixed per-pixel H&E color and local-intensity descriptors to the full panel, without validation-set hyperparameter search (Appendix G.1). **ROSIE** (Wu et al., 2025) predicts marker vectors from local H&E patches and reconstructs full regions with the upstream-faithful stride-8 assignment of central 8×8 target blocks, whereas **HEX** (Li et al., 2026) predicts marker abundance over 14×14 target footprints using larger H&E context windows. Both are cohort-adapted and fine-tuned on the same training splits (Appendix G.2). Each method retains its native inference footprint and full-region reconstruction procedure, and only the reconstructed full-resolution arrays are passed to the shared evaluator.

4.4 EVALUATION PROTOCOL

The benchmark evaluates the complete exported held-out raster at native resolution. For our main evaluation metrics (Pearson (r) and Spearman (ρ) correlation, RMSE, empirical 1-Wasserstein distance W_1 between measured and predicted marker-intensity distributions, and the variance ratio $\mathrm{VR} = \mathrm{Var}(\hat{y})/\mathrm{Var}(y)$ with ideal value 1), pixels are concatenated across all held-out regions separately for each marker, without per-region centering, and marker-level metrics are then macro-averaged across the panel. We refer to this estimand as *global-raw*. Spatial SSIM (Wang et al., 2004) is computed on native 2D marker maps and macro-averaged across markers. Inter-marker structure is assessed by comparing predicted and measured protein–protein spatial correlation matrices (Section 5.2). Supplementary metrics and sensitivity analyses are reported in Appendix B.

Subject-level uncertainty. The independent biological unit is the subject. For CRC, uncertainty is summarized with 10,000 subject-cluster bootstrap resamples (Field & Welsh, 2007) of the seven held-out subjects, retaining all four regions whenever a subject is sampled and using identical draws across methods for paired comparisons. Training-seed dispersion is reported separately (Appendix B). For Multi-Tumor, the primary point estimate is the pooled 66-subject OOF union, not the mean of five fold-specific metrics. Uncertainty is quantified with 10,000 subject bootstrap resamples of the fixed OOF predictions.

Table 1: **Held-out reconstruction benchmark.** Native-resolution, unmasked global-raw evaluation (Section 4.4); Multi-Tumor uses the pooled OOF union over five outer folds. Parentheses are 95% subject-level bootstrap CIs (10,000 resamples). Var. ratio: $\mathrm{Var}(\hat{y})/\mathrm{Var}(y)$ (ideal 1). N/A: undefined for the spatially constant Mean Prior Floor.

Method	Pearson ↑	Spearman ↑	SSIM ↑	RMSE ↓	Wasserstein W_1 ↓	Var. ratio → 1
Colorectal Cancer (CRC) ($N = 7$ held-out subjects, 28 regions, 56 markers)						
ProDICE (Ours)	**0.223** (0.198, 0.241)	**0.248** (0.213, 0.275)	**0.196** (0.173, 0.217)	0.155 (0.148, 0.163)	**0.041** (0.038, 0.046)	**0.307** (0.289, 0.330)
H&E Feature Ridge	0.188 (0.170, 0.207)	0.227 (0.196, 0.256)	0.115 (0.110, 0.122)	**0.148** (0.140, 0.156)	0.076 (0.073, 0.079)	0.050 (0.043, 0.057)
Mean Prior Floor	N/A	N/A	0.043 (0.040, 0.046)	0.152 (0.144, 0.160)	0.096 (0.091, 0.101)	0.000 (0.000, 0.000)
ROSIE	0.104 (0.089, 0.115)	0.105 (0.080, 0.127)	0.041 (0.040, 0.042)	0.163 (0.158, 0.168)	0.120 (0.116, 0.123)	0.001 (0.001, 0.001)
HEX	0.021 (-0.003, 0.044)	0.057 (0.019, 0.091)	0.038 (0.020, 0.055)	0.171 (0.157, 0.186)	0.072 (0.069, 0.085)	0.194 (0.096, 0.279)
Multi-Tumor (pooled OOF union; $N = 66$ subjects, 66 regions, 58 markers)						
ProDICE (Ours)	**0.225** (0.207, 0.243)	**0.242** (0.221, 0.261)	**0.191** (0.180, 0.203)	0.164 (0.160, 0.167)	**0.050** (0.049, 0.052)	**0.266** (0.249, 0.287)
H&E Feature Ridge	0.168 (0.151, 0.185)	0.145 (0.118, 0.170)	0.106 (0.098, 0.113)	**0.159** (0.155, 0.162)	0.087 (0.084, 0.089)	0.051 (0.045, 0.056)
Mean Prior Floor	N/A	N/A	0.039 (0.038, 0.040)	0.163 (0.159, 0.167)	0.108 (0.106, 0.110)	0.000 (0.000, 0.000)
ROSIE	0.017 (-0.017, 0.051)	0.015 (-0.029, 0.060)	0.041 (0.039, 0.044)	0.168 (0.165, 0.170)	0.107 (0.104, 0.112)	0.024 (0.020, 0.028)
HEX	-0.030 (-0.054, -0.007)	0.015 (-0.019, 0.051)	0.009 (0.001, 0.016)	0.187 (0.181, 0.194)	0.089 (0.085, 0.095)	0.161 (0.128, 0.198)

4.5 Exploratory TCGA-CRC downstream transfer

To assess transfer beyond co-registered CODEX regions, we applied a fixed **ProDICE** checkpoint to 710,812 eligible tiles from 388 primary TCGA-COAD/READ slides and averaged predicted marker intensities into 56-dimensional case representations. The TCGA feature extraction uses the mixing-enabled **ProDICE** configuration that preceded the final no-mixing architecture. We evaluate MSI-H versus MSS ($N = 325$), binary pathologic stage I/II versus III/IV ($N = 388$), regional lymph-node involvement ($N = 388$), and T3/T4 versus T1/T2 depth of invasion ($N = 387$). For each endpoint, the same frozen patient-level outer folds are reused across four classification models: elastic-net logistic regression, CatBoost, random forest, and XGBoost. Hyperparameter selection and sigmoid probability calibration are performed strictly within outer-training data. Full methods, measured-RPPA references, and paired uncertainty analyses are reported in Appendix D.

5 Results

5.1 Held-out reconstruction

Across both cohorts, **ProDICE** achieved the highest Pearson and Spearman correlations and SSIM, the lowest W_1, and the variance ratio closest to one among the five evaluated methods (Table 1). Its Pearson correlations were 0.223 on CRC and 0.225 on Multi-Tumor, compared with at most 0.188 and 0.168, respectively, for any competing method. H&E Feature Ridge instead achieved the lowest RMSE (CRC 0.148 vs. 0.155 for **ProDICE**; Multi-Tumor 0.159 vs. 0.164). Thus, **ProDICE**'s advantage is not simply lower pointwise squared error. Instead, its strength is concentrated in marker-specific correspondence, spatial structure, distributional agreement, and retention of spatial variance.

Paired subject-level bootstrap comparisons (Appendix Table 2) show that these differences persist under biological-unit resampling. For Pearson, Spearman, SSIM, W_1, and $|\mathrm{VR} - 1|$, paired 95% intervals favored **ProDICE** against Ridge, HEX, and ROSIE in both cohorts. In general and descriptively, **ProDICE** also showed improved Pearson, Spearman, SSIM, and W_1 in every Multi-Tumor outer fold (Appendix Table 11).

5.2 Pixel-level and inter-marker structure

The representative reconstructions in Figure 2A–B illustrate the same variance trade-off seen in Table 1: **ProDICE** retains more cell-scale and stromal contrast, whereas the deterministic comparators produce smoother maps with lower variance ratios. Pixel-level CRC diagnostics (Figure 2C–D) show the pronounced near-zero component of the measured CD3/CD8 distributions and attenuation of higher-intensity CD3 signal in the predictions; CD3 reconstruction achieved Pearson $r = 0.320$ and Spearman $\rho = 0.240$ across 15.8 million pixels. The panel-level analysis reveals a discrepancy not captured by individual-marker metrics: the deterministic comparators substantially overestimate correlations among predicted protein channels, whereas **ProDICE** more closely reproduces the magnitude and organization of measured inter-marker relationships (Figure 2E–F).

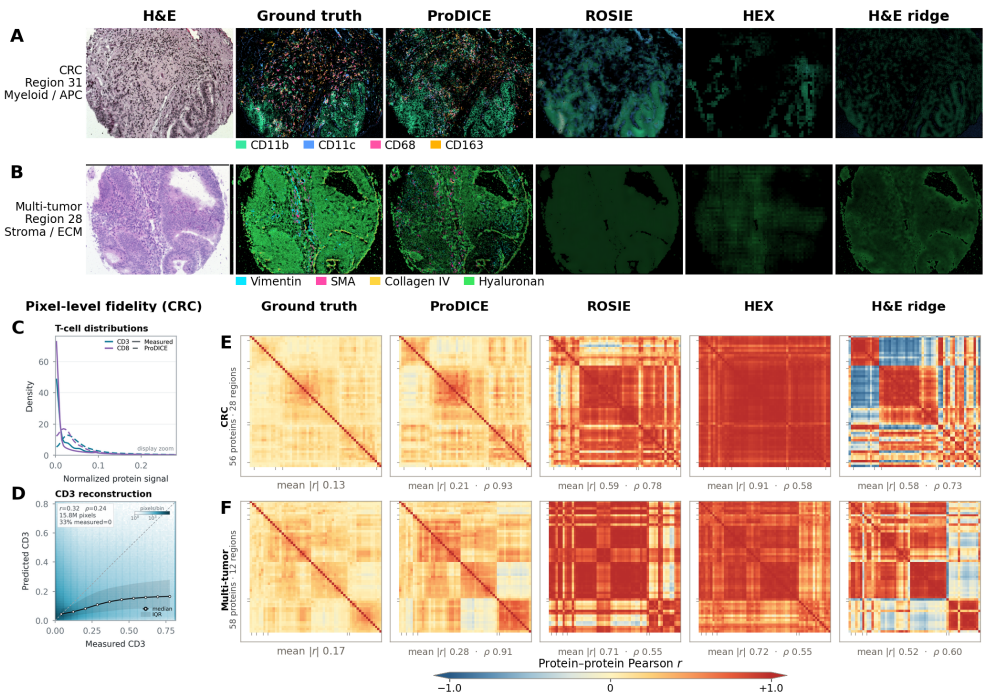

Figure 2: **Representative spatial reconstructions, pixel-level fidelity, and inter-marker correlation structure.** **(A)** Held-out CRC region 31, showing H&E and measured and predicted myeloid/antigen-presenting-cell markers. **(B)** Held-out Multi-Tumor region 28, showing stromal and extracellular-matrix markers rendered with a ground-truth-derived display transform applied identically to all methods; display transforms do not affect quantitative evaluation. **(C)** Measured (solid) and **ProDICE**-predicted (dashed) distributions of normalized CD3 and CD8 signal across the held-out CRC regions. **(D)** Pixel-level CD3 reconstruction across 15.8 million held-out CRC pixels. The black curve and shaded band show the median and interquartile range of predictions within measured-signal bins, and the diagonal denotes equality. **(E–F)** Protein–protein spatial Pearson correlation matrices for CRC (all 28 test regions) and a representative Multi-Tumor outer-test fold, respectively. Mean absolute off-diagonal correlation $|r|$ is reported below each matrix. ρ_{struct} denotes the Spearman correlation between the vectorized upper-triangular off-diagonal entries of the predicted and measured correlation matrices (higher is better).

5.3 Ablations

The final architecture was selected on CRC validation data only (4 subjects, 16 regions, three training seeds), with each ablation compared against an otherwise matched reference in which explicit output mixing was enabled (Appendix C, Table 4). For each validation region, we computed the 56×56 marker correlation matrix across spatial pixels and compared its 1,540 upper-triangle entries between prediction and measurement. The final no-mixing architecture achieved an upper-triangle correlation of 0.794 ± 0.004 across seeds. Removing explicit output-level mixing improved this diagnostic by $+0.007 \pm 0.002$ in all three seeds while leaving global Pearson essentially unchanged. This result is consistent with substantial cross-marker structure being captured through the shared multichannel diffusion state and denoiser without an explicit off-diagonal output-mixing head.

The remaining ablations show complementary conditioning effects, and more details can be found in Appendix C.

5.4 Exploratory whole-slide downstream evaluation

Representative TCGA slides and their spatially arranged tile-level ProDICE predictions are shown in Appendix Fig. 6. Case-level discrimination was endpoint- and learner-dependent: virtual proteins alone did not consistently outperform H&E morphology, and fusion produced only isolated

nominal improvements. Overall, the external analysis demonstrates scalable whole-slide feature extraction and model-dependent MSI signal, but further extensive experimentation is needed to establish consistent additive clinical discrimination over H&E morphology alone. Full results and measured-RPPA references are reported in Appendix D.

6 DISCUSSION

Among the five evaluated methods, **ProDICE** improved marker-wise Pearson and Spearman correlation, native spatial SSIM, Wasserstein distance, and variance retention on both cohorts, whereas H&E Feature Ridge achieved lower RMSE. This distinction matters because a smooth predictor can reduce squared error while attenuating the spatial dynamic range of measured protein maps. **ProDICE** itself remains under-dispersed, with variance ratios of 0.307 on CRC and 0.266 on Multi-Tumor, so joint generative modeling does not eliminate variance compression but mitigates it relative to the evaluated competing methods. More broadly, these results show why multiplex prediction should not be evaluated solely as a collection of independent regression problems: favorable per-marker error can coexist with attenuated spatial variance or distorted relationships across the predicted panel.

The ablations further clarify the model design. Global ResNet-50 morphology contributes primarily to fine spatial fidelity whilemorphology-conditioned protein queries provide a modest but broadly-distributed per-marker benefit. Explicit off-diagonal output marker mixing appears unnecessary, as removing it leaves reconstruction performance nearly unchanged and slightly improves the validation inter-marker correlation diagnostic. Although the present experiments do not establish that direct pixel-space diffusion is universally preferable to compressed target-representation approaches, it provides a single-stage alternative in which the protein image itself serves as the diffusion state, avoiding a separately trained protein representation and decoder at the cost of denoising at full spatial resolution.

The TCGA deployment shows that dense virtual-protein predictions can be generated over hundreds of whole-slide cases and aggregated into case-level features. The downstream signal is endpoint- and learner-dependent. In general, across tasks, tree-based models extracted more signal from the **ProDICE** summaries, but performance even for the measured-RPPA empirical references was also only modest, showing room for further improvement. Together, these results establish technical feasibility of cohort-scale deployment.

Limitations. All outputs are morphology-conditioned estimates rather than measured molecular values. Protein and H&E channels are normalized independently within each region, so predicted intensities represent relative within-region spatial signal and cannot be interpreted as absolute cross-patient protein abundance. The CRC test set contains only seven independent subjects despite millions of evaluated pixels, and subject-level intervals are therefore sensitive to a small number of biological units. Multi-Tumor broadens tissue heterogeneity but contains only one region per subject and uses a fixed 58-marker panel.

External-baseline adaptations may not reproduce the original ROSIE and HEX training environments exactly. Diffusion sampling is stochastic, and the current study does not claim calibrated predictive uncertainty. The exploratory TCGA analyses are feasibility analyses rather than clinical validation. No result in this study should be interpreted as a substitute for experimental spatial proteomics.

7 CONCLUSION

We present **ProDICE**, a single-stage conditional diffusion model that predicts dense multiplex protein panels directly in pixel space from routine H&E morphology. Across subject-disjoint CRC and strict Multi-Tumor OOF evaluation, **ProDICE** improved correlation, spatial structural similarity, distributional agreement, and variance retention relative to the four evaluated competing methods. ProDICE also more closely preserved the inter-marker correlation structure of measured multiplex protein panels, while exploratory whole-slide analyses demonstrated the feasibility of cohort-scale deployment. These results support conditional pixel-space diffusion as a promising framework

for virtual spatial proteomics while leaving room for improved calibration, variance recovery, and prospective biological validation.

REFERENCES

Sonali Andani, Boqi Chen, Joanna Ficek-Pascual, Simon Heinke, Ruben Casanova, Bernard Friedrich Hild, Bettina Sobottka, Bernd Bodenmiller, Tumor Profiler Consortium, Viktor H. Koelzer, and Gunnar Rätsch. Histopathology-based protein multiplex generation using deep learning. *Nature Machine Intelligence*, 7:1292–1307, 2025. doi: 10.1038/s42256-025-01074-y. URL https://www.nature.com/articles/s42256-025-01074-y.

Chang Bian, Beth Phillips, Tim Cootes, and Martin Fergie. HEMIT: H&E to multiplex-immunohistochemistry image translation with dual-branch pix2pix generator. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 Workshops*, volume 15401 of *Lecture Notes in Computer Science*, pp. 184–197. Springer Nature, 2025. doi: 10.1007/978-3-031-84525-3_16. URL https://link.springer.com/chapter/10.1007/978-3-031-84525-3_16.

Sarah Black, Darci Phillips, John W. Hickey, Julia Kennedy-Darling, Vishal G. Venkataraaman, Nikolay Samusik, Yury Goltsev, Christian M. Schürch, and Garry P. Nolan. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nature Protocols*, 16:3802–3835, 2021. doi: 10.1038/s41596-021-00556-8. URL https://www.nature.com/articles/s41596-021-00556-8.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848. URL https://doi.org/10.1109/CVPR.2009.5206848.

C. A. Field and A. H. Welsh. Bootstrapping clustered data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3):369–390, 2007. doi: 10.1111/j.1467-9868.2007.00593.x. URL https://academic.oup.com/jrsssb/article-abstract/69/3/369/7109361.

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*, 4:401–412, 2022. doi: 10.1038/s42256-022-00471-x. URL https://www.nature.com/articles/s42256-022-00471-x.

Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, Salil Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black, and Garry P. Nolan. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*, 174(4):968–981.e15, 2018. doi: 10.1016/j.cell.2018.07.010. URL https://doi.org/10.1016/j.cell.2018.07.010.

Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4:827–834, 2020. doi: 10.1038/s41551-020-0578-x. URL https://www.nature.com/articles/s41551-020-0578-x.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Jonathan Ho, Ajay N. Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Susu Hu, Qinghe Zeng, Nithya Bhasker, Jakob Nikolas Kather, and Stefanie Speidel. HistoPrism: Unlocking functional pathway analysis from pan-cancer histology via gene expression prediction. In *14th International Conference on Learning Representations (ICLR)*, 2026. URL https://arxiv.org/abs/2601.21560.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations (ICLR)*, 2014. URL https://arxiv.org/abs/1312.6114.

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*, 32:231–244, 2026. doi: 10.1038/s41591-025-04060-4. URL https://www.nature.com/articles/s41591-025-04060-4.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5404–5411, 2024. doi: 10.1109/WACV57701.2024.00532. URL https://openaccess.thecvf.com/content/WACV2024/html/Lin_Common_Diffusion_Noise_Schedules_and_Sample_Steps_Are_Flawed_WACV_2024_paper.html.

Jianfang Liu et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*, 173(2):400–416.e11, 2018. doi: 10.1016/j.cell.2018.02.052. URL https://pubmed.ncbi.nlm.nih.gov/29625055/. PMID: 29625055.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8162–8171. PMLR, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

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*, 6:1077–1093, 2024. doi: 10.1038/s42256-024-00889-5. URL https://www.nature.com/articles/s42256-024-00889-5.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022. doi: 10.1109/CVPR52688.2022.01042. URL https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html.

Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–10, 2022. doi: 10.1145/3528233.3530757. URL https://dl.acm.org/doi/10.1145/3528233.3530757.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *10th International Conference on Learning Representations (ICLR)*, 2022. URL https://arxiv.org/abs/2202.00512.

Christian M. Schürch, Salil S. Bhate, Graham L. Barlow, Darci J. Phillips, Luca Noti, Inti Zlobec, P. Chu, Sarah Black, Janos Demeter, David R. McIlwain, Nikolay Samusik, Yury Goltsev, and Garry P. Nolan. High-dimensional imaging of colorectal carcinoma and other tumors with 50+ markers, 2020a. URL https://www.cancerimagingarchive.net/collection/crc_ffpe-codex_cellneighs/. Data set.

Christian M. Schürch, Salil S. Bhate, Graham L. Barlow, Darci J. Phillips, Luca Noti, Inti Zlobec, Pauline Chu, Sarah Black, Janos Demeter, David R. McIlwain, Shigemi Kinoshita, Nikolay Samusik, Yury Goltsev, and Garry P. Nolan. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*, 182(5):1341–1359.e19, 2020b. doi: 10.1016/j.cell.2020.07.005. URL https://pubmed.ncbi.nlm.nih.gov/32763154/. PMID: 32763154.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *9th International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2010.02502.

The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407):330–337, 2012. doi: 10.1038/nature11252. URL https://pubmed.ncbi.nlm.nih.gov/22810696/. PMID: 22810696.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861. URL https://doi.org/10.1109/TIP.2003.819861. PMID: 15376593.

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*, 16:7633, 2025. doi: 10.1038/s41467-025-62346-0. URL https://www.nature.com/articles/s41467-025-62346-0.

Jinxi Xiang, Xiyue Wang, Xiaoming Zhang, Yinghua Xi, Feyisope Eweje, Yijiang Chen, Yuchen Li, Colin Bergstrom, Matthew Gopaulchan, Ted Kim, Kun-Hsing Yu, Sierra Willens, Francesca Maria Olguin, Jeffrey J. Nirschl, Joel Neal, Maximilian Diehn, Sen Yang, and Ruijiang Li. A vision–language foundation model for precision oncology. *Nature*, 638(8051):769–778, 2025. doi: 10.1038/s41586-024-08378-w. URL https://www.nature.com/articles/s41586-024-08378-w.

Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, and Gary Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 70626–70637, 2023. doi: 10.52202/075280-3095. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/df656d6ed77b565e8dcdfbf568aead0a-Abstract.html.

Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):bbac297, 2022. doi: 10.1093/bib/bbac297. URL https://academic.oup.com/bib/article/23/5/bbac297/6645485.

Sichen Zhu, Yuchen Zhu, Molei Tao, and Peng Qiu. Diffusion generative modeling for spatially resolved gene expression inference from histology images. In *13th International Conference on Learning Representations (ICLR)*, 2025. URL https://arxiv.org/abs/2501.15598.

A Appendix

AI Assistance

Generative AI tools were used to assist with literature search and identification of relevant prior work, software debugging and code editing, and drafting and editing portions of the manuscript for clarity and organization. Generative AI tools were not used to formulate the central research hypothesis, develop the core methodological framework, or determine the primary experimental design. All AI-assisted literature references, code modifications, analyses, and written content were

independently reviewed and, where applicable, verified or tested by the authors. The authors take full responsibility for the methodology, experimental results, interpretations, claims, and final content of the paper.

B Additional Reconstruction Results

Unless stated otherwise, metrics follow the native-resolution, unmasked global-raw protocol of Section 4.4: native pixels are pooled across all regions within a split separately for each marker, metrics are computed marker-wise, and markers are then macro-averaged with equal weight.

B.1 Rule-based qualitative robustness across held-out cases

To complement aggregate metrics without selecting only visually favorable examples, Figure 3 chooses held-out region–marker pairs by a prespecified rule. Within each cohort, eligible pairs require at least 5% measured foreground signal, and examples nearest the 90th, 50th, and 10th percentiles of region-level unmasked Pearson correlation are displayed. The histogram above each cohort shows the full eligible-pair distribution and the selected percentile locations.

Paired differences. Table 2 reports **ProDICE**-minus-comparator differences computed within identical subject-bootstrap resamples. In both cohorts, every interval for Pearson, Spearman, SSIM, W_1, and $|\mathrm{VR}-1|$ excludes zero in favor of **ProDICE** against Ridge, HEX, and ROSIE. RMSE remains the notable exception: H&E Feature Ridge has lower RMSE than **ProDICE** on both cohorts.

Table 2: **Paired ProDICE-minus-comparator differences.** Point differences with 95% subject-bootstrap CIs from identical resampled biological units. Positive differences favor **ProDICE** for Pearson, Spearman, and SSIM; negative differences favor **ProDICE** for RMSE, W_1, and $|\mathrm{VR}-1|$.

Comparison	ΔPearson ↑	ΔSpearman ↑	ΔSSIM ↑	ΔRMSE ↓	ΔW_1 ↓	Δ$\lvert\mathrm{VR}-1\rvert$ ↓
CRC held-out test ($N = 7$ subjects)						
vs. H&E Feature Ridge	+0.035 [+0.009,+0.056]	+0.021 [+0.005,+0.041]	+0.081 [+0.058,+0.099]	+0.007 [+0.006,+0.010]	-0.035 [-0.037,-0.032]	-0.257 [-0.280,-0.237]
vs. HEX	+0.202 [+0.175,+0.228]	+0.190 [+0.138,+0.237]	+0.158 [+0.142,+0.171]	-0.016 [-0.024,-0.006]	-0.031 [-0.041,-0.029]	-0.114 [-0.220,-0.031]
vs. ROSIE	+0.119 [+0.095,+0.143]	+0.143 [+0.104,+0.177]	+0.155 [+0.133,+0.175]	-0.007 [-0.010,-0.004]	-0.079 [-0.080,-0.075]	-0.306 [-0.329,-0.288]
Multi-Tumor pooled OOF ($N = 66$ subjects)						
vs. H&E Feature Ridge	+0.057 [+0.041,+0.073]	+0.097 [+0.078,+0.117]	+0.086 [+0.078,+0.094]	+0.005 [+0.004,+0.007]	-0.036 [-0.038,-0.034]	-0.216 [-0.236,-0.198]
vs. HEX	+0.256 [+0.223,+0.289]	+0.227 [+0.184,+0.269]	+0.183 [+0.166,+0.199]	-0.023 [-0.030,-0.017]	-0.039 [-0.045,-0.034]	-0.106 [-0.147,-0.062]
vs. ROSIE	+0.209 [+0.175,+0.241]	+0.227 [+0.186,+0.264]	+0.150 [+0.139,+0.161]	-0.004 [-0.006,-0.001]	-0.057 [-0.061,-0.053]	-0.242 [-0.263,-0.225]

Supplementary CRC metrics and training-seed dispersion. Table 3 complements the global-raw benchmark with metrics that separate local spatial pattern from region-level intensity shifts. Region-centered Pearson subtracts the spatial mean within each region before computing marker-wise spatial correlation; region-mean Pearson correlates measured and predicted region-level mean normalized intensities across the 28 held-out regions; and mean error is the absolute discrepancy between predicted and measured marker-wide means. For stochastic CRC methods, the reported point estimates summarize training seeds 42, 314, and 2718. Training-seed dispersion is reported directly in Table 3; the primary reconstruction table instead uses subject-level bootstrap uncertainty.

Table 3: **Supplementary CRC reconstruction diagnostics.** Values are macro-averaged across 56 markers on the 28-region held-out test set. Stochastic methods report mean ± sample SD across training seeds 42, 314, and 2718; deterministic controls are single fits. "Region-mean" refers to normalized within-region signal rather than absolute abundance.

Method	Centered Pearson ↑	Region-mean Pearson ↑	MAE ↓	R^2 ↑	Mean error ↓	$\lvert\mathrm{VR}-1\rvert$ ↓
ProDICE	**0.215** ± 0.009	**0.363** ± 0.017	0.091 ± 0.001	-0.056 ± 0.008	0.007 ± 0.000	**0.693** ± 0.028
H&E Feature Ridge	0.190	0.214	0.091	**0.042**	**0.006**	0.950
Mean Prior Floor	—	—	0.096	-0.003	0.007	1.000
ROSIE	0.103 ± 0.035	0.152 ± 0.039	0.121 ± 0.009	-0.144 ± 0.062	0.053 ± 0.012	0.999 ± 0.001
HEX	0.005 ± 0.029	0.138 ± 0.014	0.111 ± 0.012	-0.266 ± 0.207	0.031 ± 0.013	0.806 ± 0.129

Distribution fidelity and marker-level heterogeneity. Figure 4 compares marginal distributions for representative markers and summarizes per-marker W_1 differences; **ProDICE** had lower W_1 than every comparator for all 56 CRC and 58 Multi-Tumor markers. Figure 5 relates per-marker

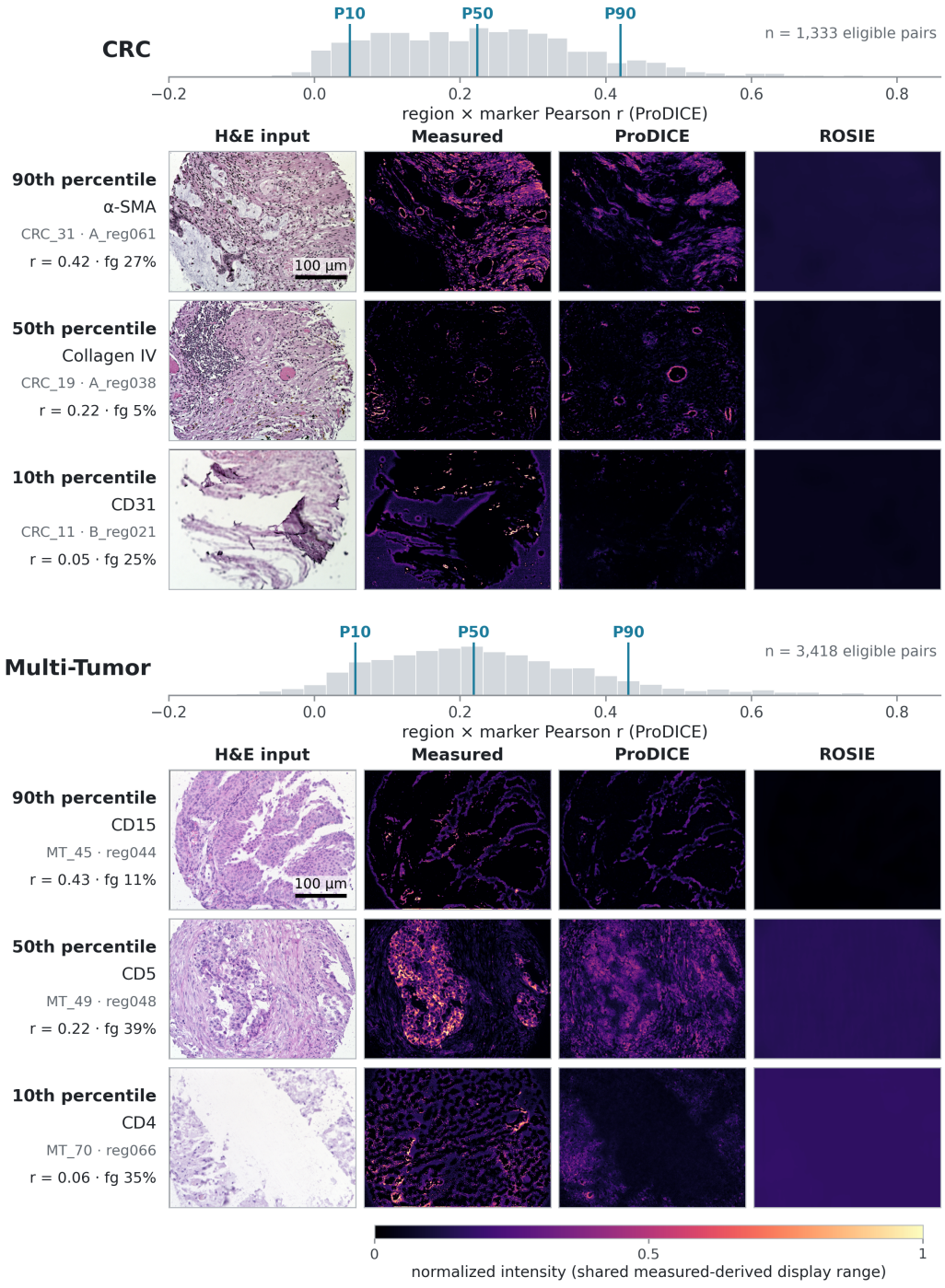

Figure 3: **Rule-based high-, median-, and challenging-fidelity reconstruction cases.** CRC and Multi-Tumor examples are selected mechanically from eligible held-out region–marker pairs (1,333 CRC and 3,418 Multi-Tumor pairs) nearest the 90th, 50th, and 10th percentiles of unmasked region-level Pearson correlation. Each row shows the H&E input, measured protein map, **ProDICE** prediction, and ROSIE prediction on the same region and marker. The measured and predicted protein panels within a row use the same measured-derived display range. These examples are qualitative diagnostics; quantitative comparisons use the complete held-out benchmark. HEX is excluded here.

differences to measured foreground fraction. Pearson differences favored **ProDICE** against ROSIE and HEX for 55 of 56 CRC markers (synaptophysin excepted) and all 58 Multi-Tumor markers, and against Ridge for 43 of 56 and 41 of 58. Gains over HEX increased with foreground fraction (descriptive Spearman $\rho = 0.54$ CRC, 0.72 Multi-Tumor), whereas gains over Ridge decreased ($\rho = -0.25, -0.44$).

B.2 Complete per-marker reconstruction fidelity

Cohort-level macro averages can obscure substantial heterogeneity across molecular targets. We therefore report every canonical protein individually rather than only selected functional groups. For each marker and method, we report Pearson correlation, Spearman correlation, Lin's concordance correlation coefficient (CCC), SSIM, RMSE, W_1, and variance ratio. Lin's CCC is

$$\rho_c = \frac{2\,\mathrm{cov}(y, \hat{y})}{\sigma_y^2 + \sigma_{\hat{y}}^2 + (\mu_y - \mu_{\hat{y}})^2}, \tag{5}$$

and complements Pearson correlation by penalizing both scale disagreement and location shift. The exhaustive tables below (Tables 8, 9, 10, 11) are the primary marker-level record.

C Conditioning Ablations

The conditioning study uses the CRC validation split (4 subjects, 16 regions, 56 markers) and three matched training seeds (42, 314, and 2718). Each row of Table 4 reports ablation minus an otherwise matched reference in which explicit cross-marker output mixing (Appendix F.2) is enabled. Variant C, which removes this mixing, is the final **ProDICE** architecture; the deltas for variants A and B are therefore measured relative to the mixing-enabled reference rather than to the final model.

Table 4: **CRC validation conditioning-ablation deltas** (ablation minus mixing-enabled reference; mean ± SD over seeds 42, 314, and 2718). Raw Pearson is the global-raw Pearson of Section 4.4, centered Pearson its region-centered counterpart, and r_{spatial} the inter-marker structure diagnostic of Section 5.3. Positive ΔW_1 is worse; positive values are better in all other columns.

Ablation	Δ Raw Pearson	Δ Centered Pearson	Δ SSIM	ΔW_1	$\Delta r_{\mathrm{spatial}}$
A. No global morphology	-0.001±0.004	-0.008±0.002	-0.010±0.003	+0.001±0.000	+0.010±0.007
B. Static protein queries	-0.003±0.005	-0.002±0.004	+0.002±0.004	-0.001±0.000	+0.006±0.002
C. No cross-marker output mixing	-0.001±0.004	-0.001±0.004	0.000±0.004	+0.000±0.001	**+0.007**±0.002

Removing global morphology worsened centered Pearson, SSIM, and W_1 in all three runs despite near-neutral global-raw Pearson, indicating a contribution to fine spatial fidelity. Static protein queries reduced mean global-raw Pearson by 0.003, and 39 of 56 markers had lower mean Pearson across runs; native RMSE and MAE each increased by approximately 0.001. Removing explicit cross-marker output mixing left the main metrics essentially unchanged while improving the inter-marker diagnostic. For the final no-mixing architecture, the validation upper-triangle correlation between predicted and measured inter-marker spatial-correlation matrices is 0.794 ± 0.004 across runs, and the no-mixing-minus-mixing delta is $+0.007 \pm 0.002$.

D Exploratory Downstream Evaluation

Qualitative whole-slide ProDICE deployment in TCGA-CRC. Although the downstream TCGA analyses summarize each case by the mean predicted abundance of each of the 56 proteins, the underlying ProDICE predictions retain substantial spatial variation across tissue tiles. To illustrate the representation prior to whole-case pooling, Fig. 6 shows H&E tissue together with tile-level virtual-protein abundance for four deterministically selected cases spanning MSI and stage strata.

Model provenance and scope. For the downstream analysis only, the archived TCGA extraction uses the mixing-enabled **ProDICE** configuration that preceded the final no-mixing architecture lock. This distinction is stated here once because the extraction was completed before the final architecture was frozen. The downstream results are therefore treated as exploratory evidence of deployment

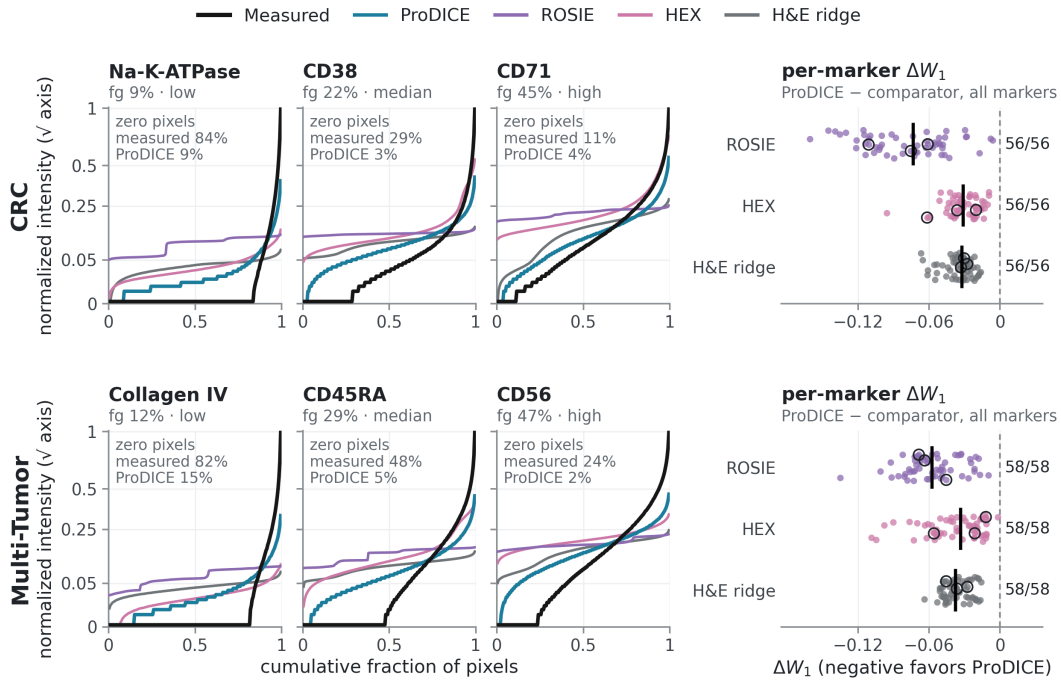

Figure 4: **Distribution fidelity across representative markers and the full panel.** Q–Q curves compare measured and predicted intensity quantiles for markers selected at low, median, and high measured foreground fractions in each cohort. The dashed diagonal denotes perfect marginal agreement. Right-hand panels show per-marker $\Delta W_1 = W_1(\textbf{ProDICE}) - W_1(\text{comparator})$; negative values favor **ProDICE**, and black ticks denote medians across markers.

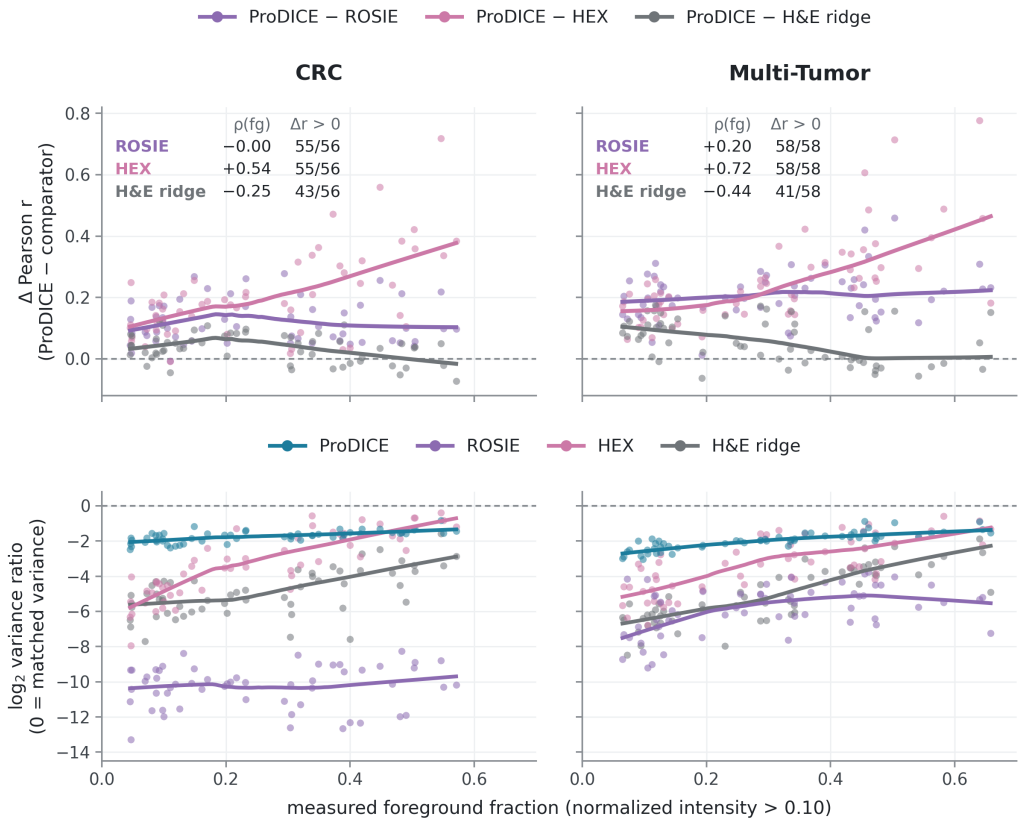

Figure 5: **Performance heterogeneity across measured foreground fraction.** Top panels plot per-marker Pearson improvement of **ProDICE** over each comparator against the measured fraction of pixels with normalized intensity above 0.10; lines are running medians. Bottom panels show each method's per-marker $\log_2$ variance ratio against the same foreground statistic, where zero indicates matched measured variance.

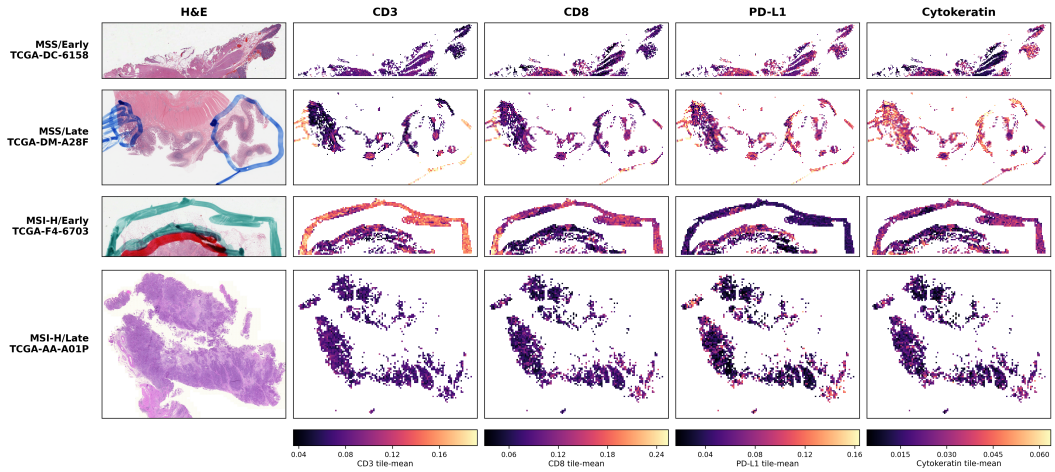

Figure 6: **Representative TCGA-CRC H&E slides and spatial ProDICE virtual-protein expression mosaics.** Four cases were selected randomly to span the four combinations of microsatellite-instability status (MSS or MSI-H) and binary pathologic stage (early, stage I/II; late, stage III/IV), without using ProDICE predictions during case selection. For each case, the H&E tissue region is shown alongside ProDICE predictions for CD3, CD8, PD-L1, and cytokeratin. Each colored location represents the mean predicted abundance of that marker within one eligible 384×384 tissue tile; the panels therefore depict spatially arranged tile-mean expression mosaics, rather than stitched native-resolution protein-pixel predictions. Color limits are fixed separately for each marker using the pooled 1st–99th percentiles across all four displayed cases, enabling within-marker comparison across subjects. The heterogeneous spatial patterns illustrate information present before the downstream TCGA representation is reduced to one whole-case mean abundance per marker. Note the TCGA data contains markings that influence the model's predictions.

feasibility and case-level transfer rather than confirmatory evidence for the final reconstruction architecture.

TCGA-CRC whole-slide deployment and case-level features. The exploratory cohort comprises 388 primary diagnostic slides from TCGA-COAD and TCGA-READ (The Cancer Genome Atlas Network, 2012; Liu et al., 2018). The prespecified tile set contains 710,812 level-0 512×512 H&E tiles after requiring tissue fraction ≥ 0.50 and at least 50 eligible tiles per case; tiles were resized bilinearly to 384×384 before inference. For each tile, the 56 predicted channels were mapped to $[0, 1]$ and spatially averaged, and case-level virtual-protein features were the arithmetic mean of these tile-level marker means. The H&E comparator is a 256-dimensional ResNet-50 morphology representation averaged over the same eligible tiles; early fusion concatenates the H&E and **ProDICE** representations.

Clinical sensitivity protocol. We evaluate four endpoints: MSI-H versus MSS ($N = 325$, 58 MSI-H), binary pathologic stage I/II versus III/IV ($N = 388$, 165 late-stage cases), N-positive versus N0 ($N = 388$, 160 N-positive cases), and T3/T4 versus T1/T2 depth of invasion ($N = 387$, 316 T-advanced cases). The same frozen patient-level outer folds are reused across elastic-net logistic regression, CatBoost, random forest, and XGBoost. Hyperparameter tuning is performed by inner stratified cross-validation on outer-training cases only, and sigmoid probability calibration is also fit strictly within outer-training data. Paired uncertainty uses 5,000 case-level bootstrap resamples. Fixed 50/50 probability-level late fusion is evaluated for MSI and stage. A prespecified 14-marker direct TCGA RPPA subset is modeled under the same fold structure as a measured-protein empirical reference.

MSI-H versus MSS. **ProDICE** case-level features were strongly learner-dependent: ROC-AUC ranged from 0.521 with elastic net to 0.626 with CatBoost, while H&E ranged from 0.663 to 0.693. Early H&E+**ProDICE** fusion ranged from 0.656 to 0.691, and fixed 50/50 late fusion likewise did

Table 5: **Pooled OOF ROC-AUC across learner families for exploratory TCGA-CRC endpoints.** Only H&E, **ProDICE**, their early fusion, and the measured-RPPA empirical reference are shown.

Endpoint	Learner	H&E	**ProDICE**	H&E + **ProDICE**	Measured RPPA
MSI	Elastic net	0.673	0.521	0.667	0.854
	CatBoost	0.674	0.626	0.665	0.852
	Random forest	0.663	0.612	0.656	0.869
	XGBoost	0.693	0.590	0.691	0.865
Stage	Elastic net	0.582	0.485	0.583	0.636
	CatBoost	0.569	0.534	0.570	0.631
	Random forest	0.552	0.506	0.564	0.638
	XGBoost	0.565	0.526	0.570	0.634
N-positive	Elastic net	0.526	0.539	0.565	0.625
	CatBoost	0.498	0.499	0.524	0.616
	Random forest	0.477	0.485	0.468	0.643
	XGBoost	0.505	0.474	0.533	0.637
T-advanced	Elastic net	0.541	0.576	0.566	0.570
	CatBoost	0.554	0.501	0.525	0.526
	Random forest	0.515	0.490	0.507	0.526
	XGBoost	0.540	0.469	0.559	0.557

not consistently exceed H&E alone. The measured direct-marker RPPA reference achieved 0.852–0.869, indicating substantial protein-level MSI signal in the measured panel and a remaining gap for the virtual case-level representation. The stronger tree-based result is evidence of classifier-family dependence, but it does not establish a particular biological mechanism.

Stage and TNM endpoints. For binary pathologic stage, **ProDICE** achieved ROC-AUC 0.485–0.534 versus 0.552–0.582 for H&E, while the measured-RPPA reference reached 0.631–0.638. For N-positive versus N0, **ProDICE** achieved 0.474–0.539 and H&E+**ProDICE** 0.468–0.565. Under elastic net, H&E+**ProDICE** showed a nominal paired AUC difference of $+0.039$ (95% CI $[0.002, 0.076]$) over H&E alone; this pattern was not reproduced consistently across the other learner families and is treated as exploratory within the broader family of comparisons. For T3/T4 versus T1/T2, **ProDICE** achieved 0.469–0.576, similar in range to the measured-RPPA reference (0.526–0.570). The weak measured-protein reference contextualizes this endpoint but does not distinguish limited endpoint signal from additional loss introduced by virtual prediction or case-level aggregation.

Measured-RPPA molecular grounding. Virtual-to-RPPA case-level concordance was evaluated on the prespecified direct-marker set: BCL-2, CD20, CD31, CD38, CD44, CD45, CD4, EGFR, GATA3, Granzyme B, IDO-1, p53, beta-catenin, and PD-L1. Proxy mappings were excluded. Across these 14 markers, **ProDICE** achieved macro mean Pearson $r = 0.062$ and macro mean Spearman $\rho = 0.066$. These correlations are weak in absolute magnitude and should be interpreted as partial case-level molecular grounding under substantial assay and domain differences, not as validation of quantitative equivalence between virtual CODEX and RPPA.

CD3/CD8 reconstruction diagnostic. On the held-out CRC test regions, a three-seed **ProDICE** ensemble was used for a descriptive CD3/CD8 reconstruction analysis. Across 28 regions from seven subjects, the composite immune score achieved Pearson $r = 0.550$, Spearman $\rho = 0.280$, and CCC $= 0.460$; the subject-clustered interval for Pearson correlation was wide and included zero. This is therefore treated as a spatial reconstruction diagnostic rather than evidence of patient-level clinical validity.

Interpretation boundaries. The clinical stress test supports three conclusions. First, case-level MSI information in **ProDICE** summaries depends materially on classifier family, but neither early nor fixed late fusion consistently improves on H&E morphology. Second, stage and TNM discrimination remains modest across learners; the measured-RPPA references are themselves modest for

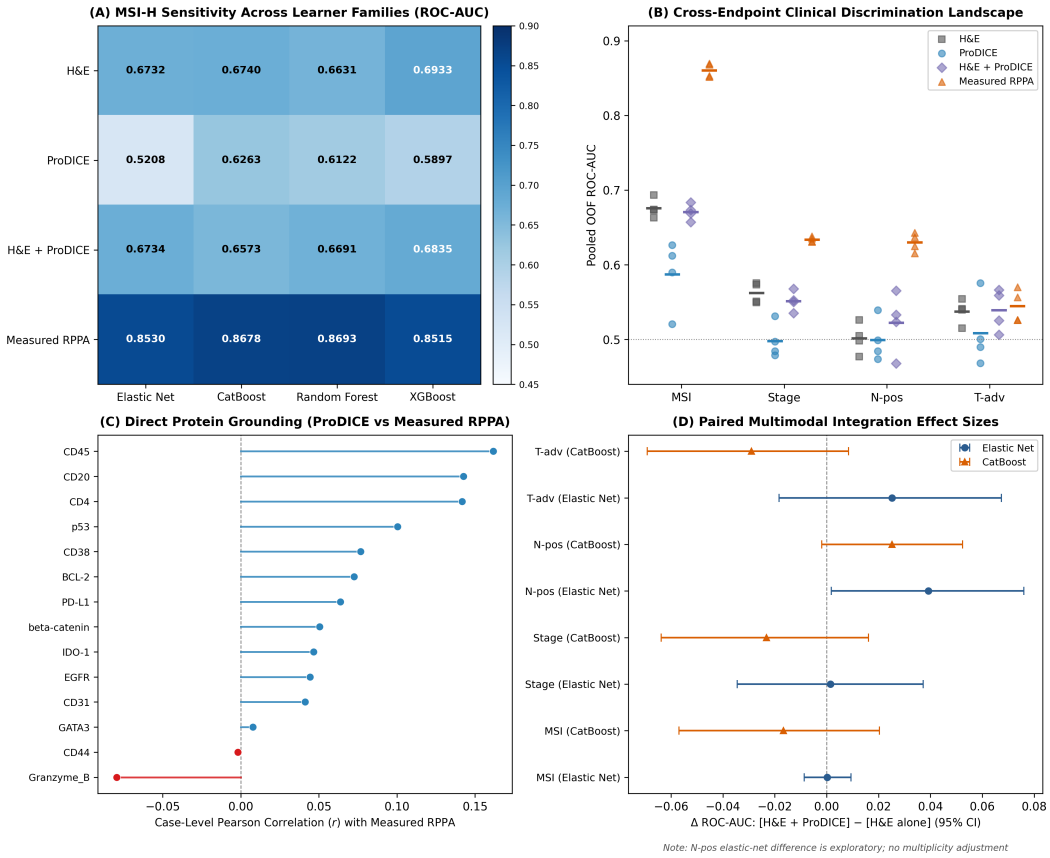

Figure 7: **Exploratory ProDICE whole-slide downstream sensitivity in TCGA-CRC.** (A) MSI-H versus MSS pooled out-of-fold ROC-AUC across four learner families for H&E morphology, **ProDICE** virtual-protein summaries, early H&E+**ProDICE** fusion, and the measured-RPPA empirical reference. (B) Cross-endpoint ROC-AUC landscape for MSI, binary pathologic stage, nodal involvement (N-positive), and depth of invasion (T-advanced); points denote learner-family results and horizontal bars denote their means. (C) Case-level Pearson concordance between **ProDICE** virtual-protein summaries and measured bulk RPPA for the 14 prespecified direct marker matches. (D) Paired patient-bootstrap differences in ROC-AUC for H&E+**ProDICE** versus H&E alone under elastic net and CatBoost; error bars show 95% confidence intervals. The positive N-positive elastic-net estimate is nominal and exploratory because no prespecified multiplicity adjustment was applied across the family of downstream comparisons.

stage and nodal status and weak for T-advanced under this restricted direct-marker protocol. Third, direct case-level agreement between virtual proteins and measured RPPA is weak in absolute magnitude. Whole-case mean aggregation may discard clinically useful spatial heterogeneity, but the present analysis does not isolate aggregation effects from domain shift or virtual-prediction error. No downstream result is interpreted as standalone diagnostic utility.

E Dataset and Preprocessing Details

Table 6 summarizes the two reconstruction cohorts. In both, H&E and multiplex CODEX images come from the same physical tissue section and share the native microscope-stage coordinate system, so no retrospective registration is applied.

Table 6: **Reconstruction cohorts and evaluation designs.** CRC uses a fixed train/validation/test split of 24/4/7 subjects; Multi-Tumor uses five-fold subject-grouped outer cross-validation. All regions from a subject share a single split role.

Cohort	Subjects	Regions	Markers	Design	Test unit
CRC	35	140	56	Fixed split	7 subjects / 28 regions
Multi-Tumor	66	66	58	5-fold outer CV	66 OOF subjects / regions

CRC. The CRC cohort is derived from the public CRC FFPE-CODEX CellNeighs collection (Schürch et al., 2020a;b). H&E brightfield and multiplex CODEX fluorescence were acquired from the same physical tissue section on a Keyence CODEX-MPI system at $20\times$ magnification (NA 0.75; calibrated pixel size approximately $377.4\,\mathrm{nm}$), and the brightfield image shares the native microscope-stage coordinate system of the fluorescence acquisition. No retrospective elastic or affine registration was applied in the fixed export. Paired arrays were cropped only to their common rectangular extent and exported at 720×960 pixels.

Multi-Tumor. The Multi-Tumor cohort is a multi-tumor tissue-microarray acquisition dated 2020-04-06 and contains 66 cores from 66 unique subjects with 58 canonical markers curated from 23 imaging cycles. The frozen manifest spans hematologic and solid-tumor, as well as non-neoplastic, tissue from multiple anatomical sites. Acquisition metadata identify a Keyence BZ-X710 CODEX-MPI run with a PlanApo $20\times$ air objective (NA 0.75) and calibrated resolution $377.442\,\mathrm{nm/pixel}$. H&E brightfield and multiplex fluorescence were acquired sequentially from the same physical section and share the native stage coordinate system. No retrospective elastix registration was used; the images retain their native scale and a 720×960-pixel canvas for each core. Donor-level clinical cross-walk records beyond the fixed 66-case manifest and detailed optical-QC logs are not recoverable from the available artifacts.

Region-wise intensity normalization. For both cohorts, the 1st and 99.5th percentiles are computed independently for each protein channel within each region; values are clipped to these bounds and rescaled to $[0, 1]$. H&E is normalized analogously, independently for each RGB channel within each region. Predicted and measured protein values therefore represent normalized within-region spatial intensity rather than absolute physical abundance across subjects. No Macenko, Reinhard, or other stain normalization is used.

Training crops. Each epoch draws 32 random paired 384×384 H&E–protein crops from every training region, with crop origins uniform over positions for which the full crop lies inside the region. To limit a known yellow annotation artifact, a proposed crop is accepted when fewer than 5% of its pixels have red and green intensities above 0.7 and blue intensity below 0.4; otherwise sampling is retried up to 10 times, after which the last sampled location is used. No spatial, scale, stain, or other augmentation is used in the final training protocol.

F Model, Training, and Inference Details

Table 7 summarizes the final architecture and training protocol used for all **ProDICE** results in Section 5.1 and Appendix B; the CRC and Multi-Tumor models differ only in the number of predicted markers. The subsections below provide additional methodological detail.

F.1 Denoiser and conditioning architecture

The denoiser is a four-level U-Net with channel widths $[96, 192, 384, 384]$ at resolutions H, $H/2$, $H/4$, and $H/8$; each level contains two residual blocks with GroupNorm and SiLU, and both train-

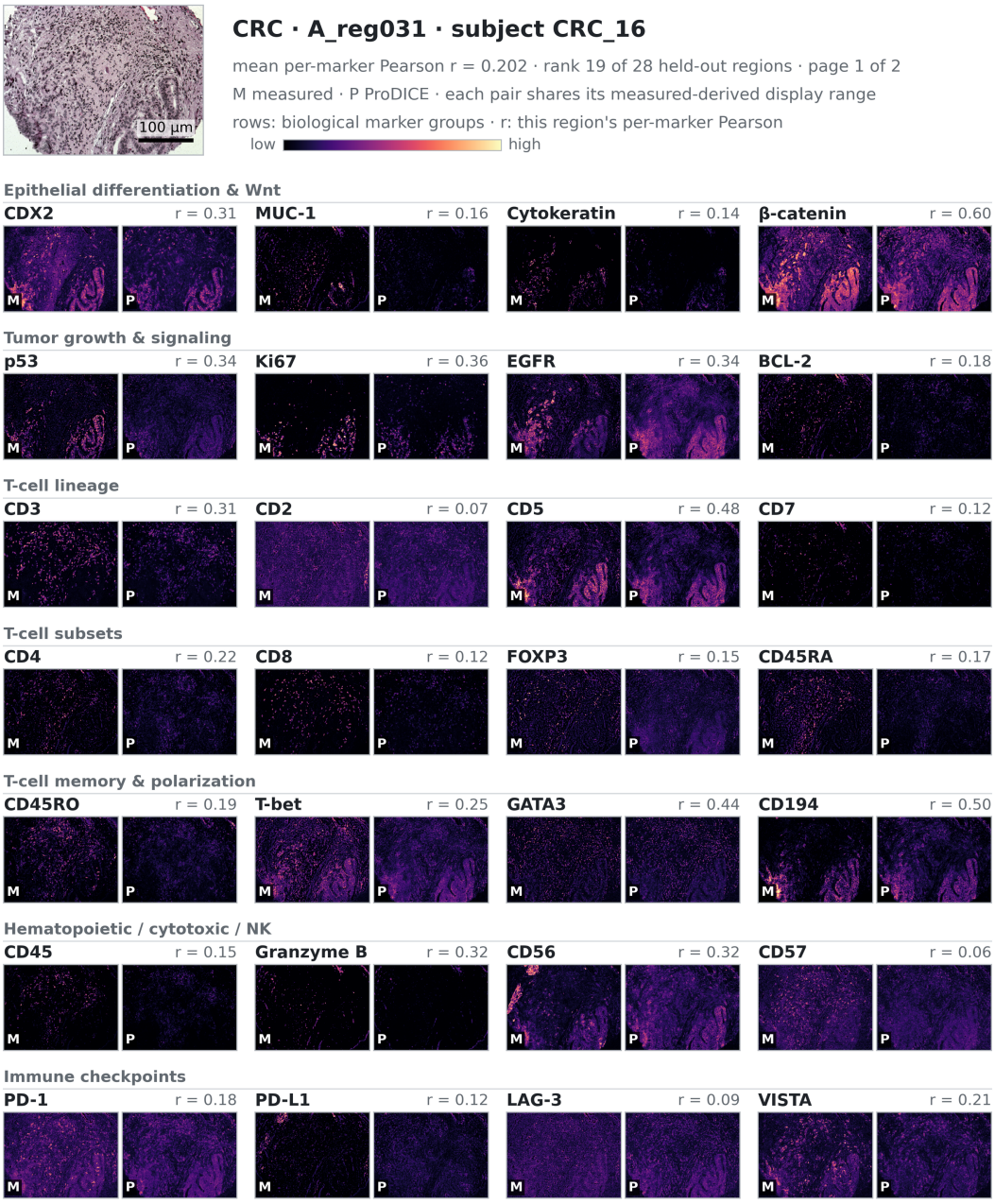

Figure 8: **Complete held-out CRC protein-panel atlas, page 1 of 2.** The same held-out region is used for all markers. For each marker, measured (M) and **ProDICE**-predicted (P) maps share the same measured-derived display range. Marker order follows the prespecified biological grouping used for visualization.

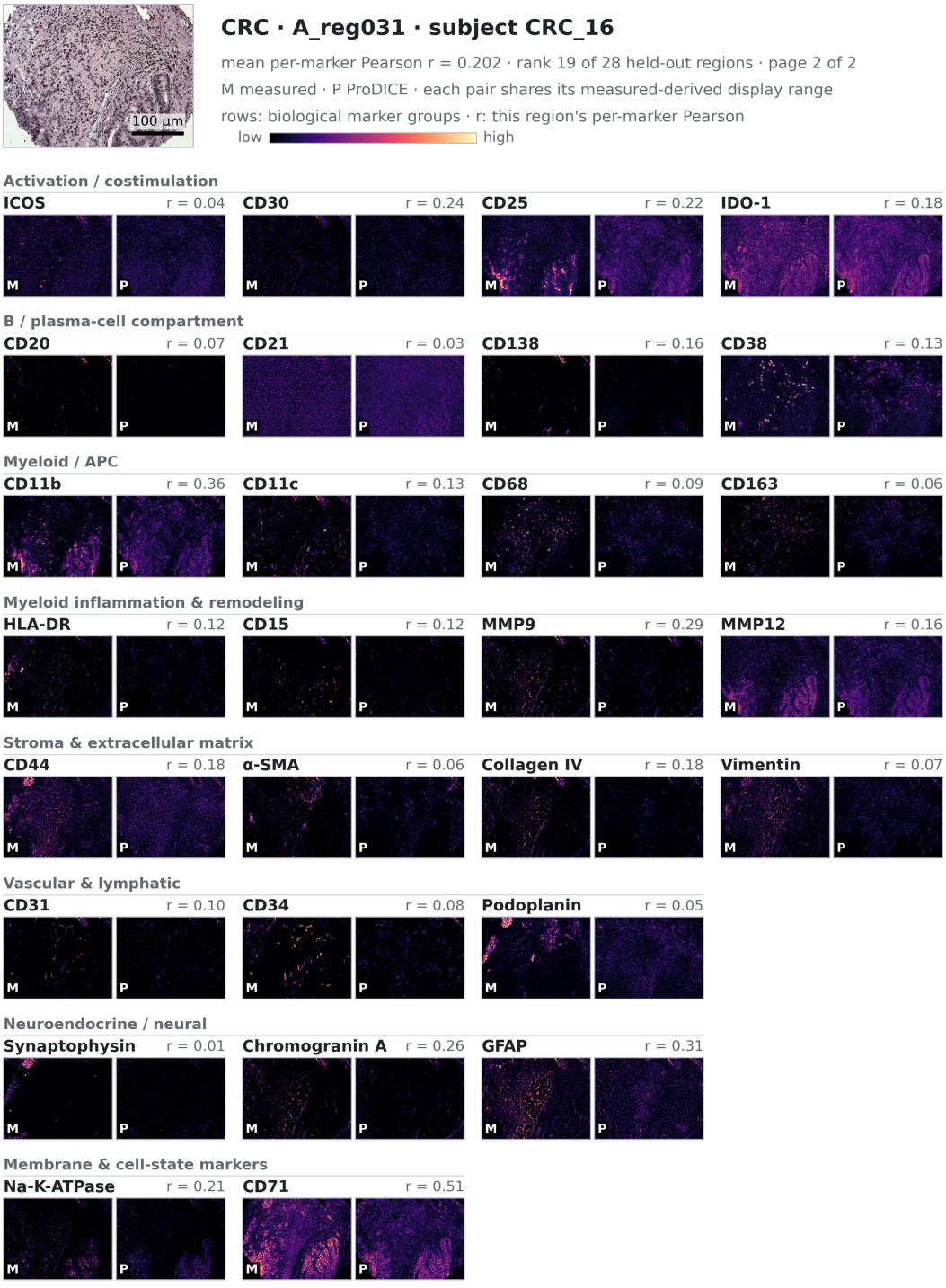

Figure 9: **Complete held-out CRC protein-panel atlas, page 2 of 2.** Continuation of Figure 8; measured and predicted maps use matched per-marker display ranges.

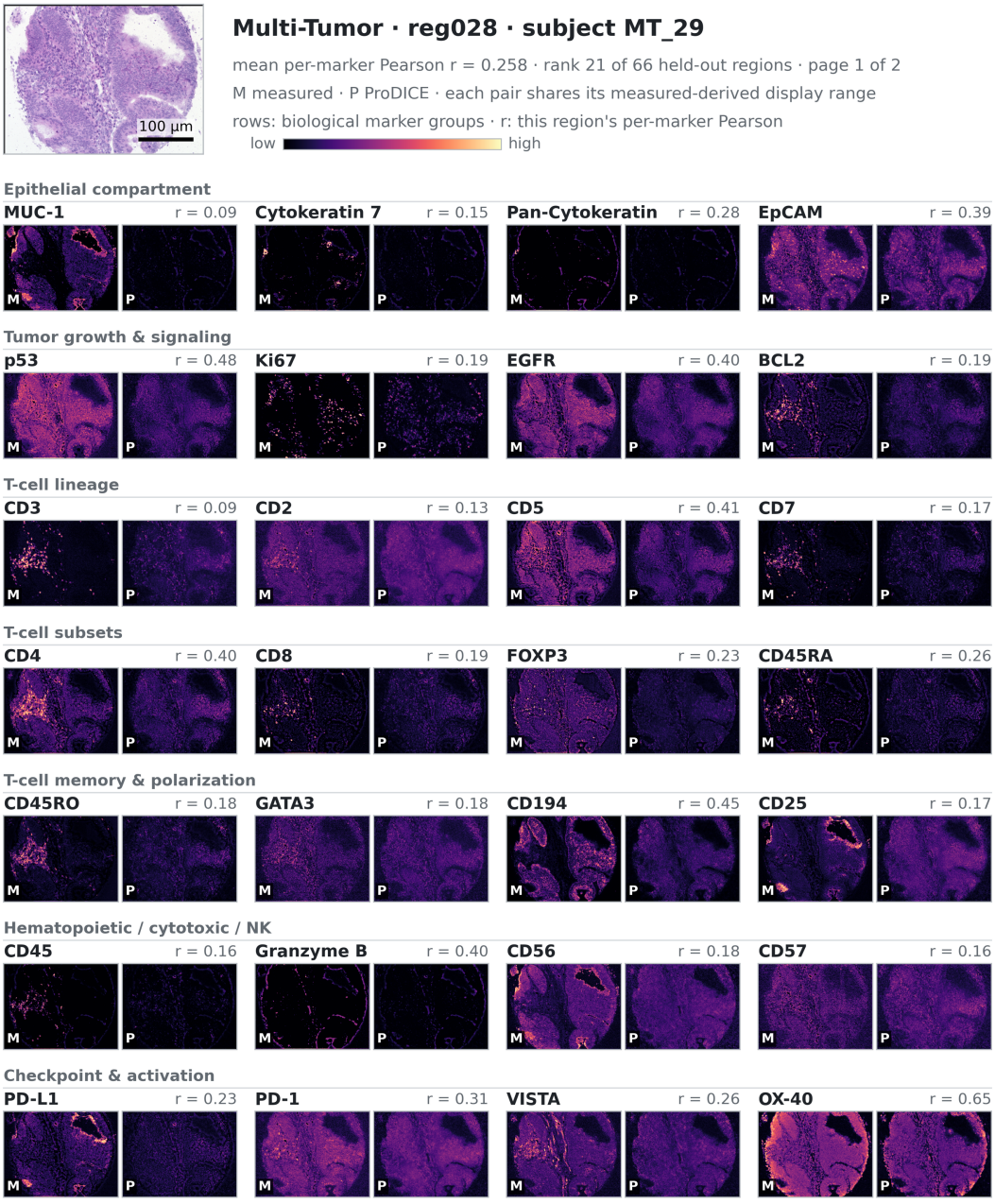

Figure 10: **Complete Multi-Tumor protein-panel atlas, page 1 of 2.** All markers are shown for the same held-out out-of-fold region. Each measured/predicted pair shares its measured-derived display range.

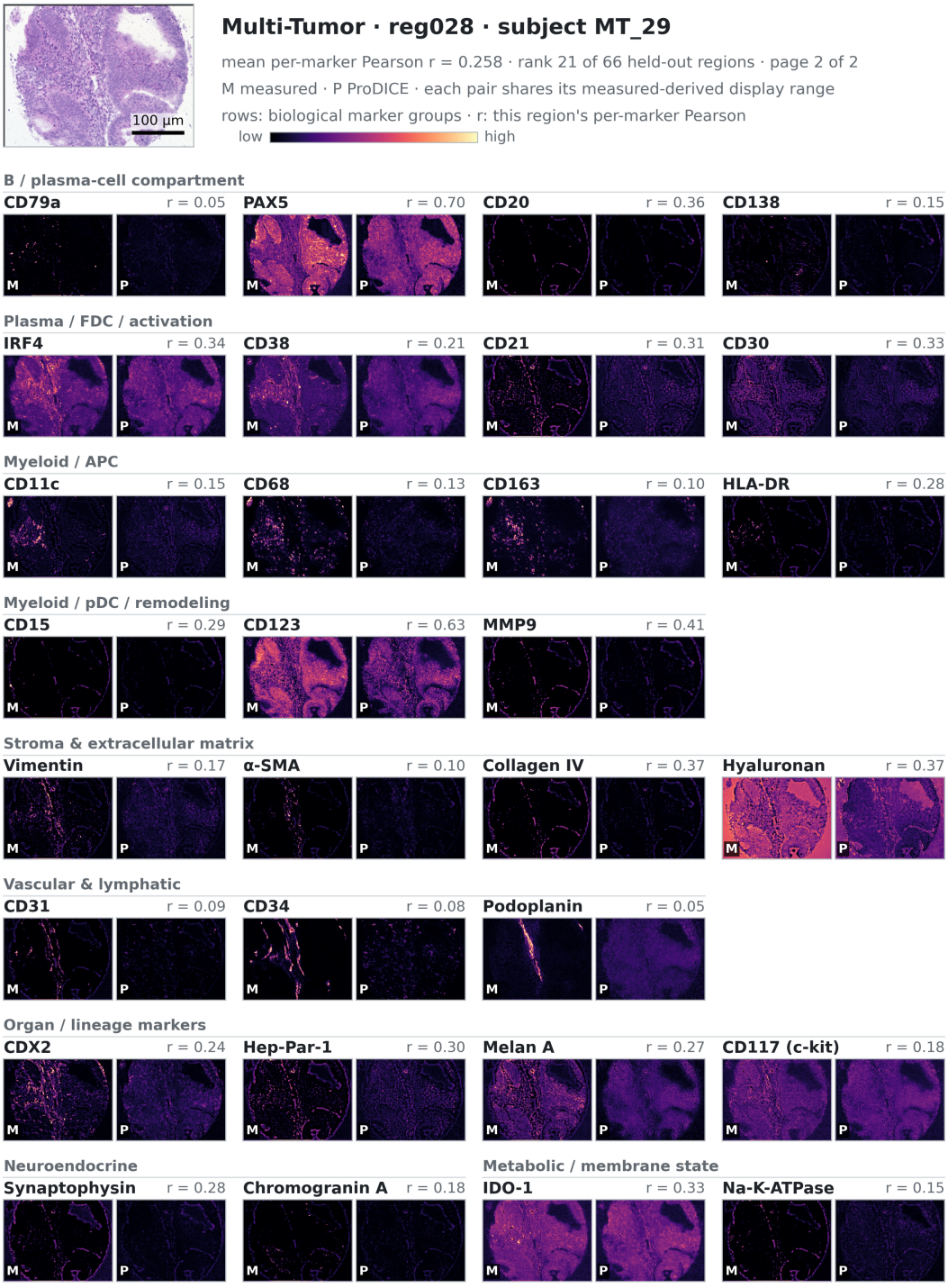

Figure 11: **Complete Multi-Tumor protein-panel atlas, page 2 of 2.** Continuation of Figure 10; the page boundary follows a biological-group boundary rather than splitting the marker group at the midpoint.

Table 7: **Final ProDICE architecture and training protocol.**

Component	Setting
Input and target	
H&E input	One 384×384 RGB tile per prediction; no additional or larger field of view
Protein target	One 384×384 tile per marker (56 CRC markers or 58 Multi-Tumor markers), region-normalized and mapped to $[-1, 1]$
Conditioning	
Local morphology	Convolutional encoder, $64 \times 48 \times 48$ features; 1×1 adapter to width 384, added at $H/8$
Global morphology	Frozen ImageNet ResNet-50 (input resized to 224×224); 2×2 pooling to 4 tokens; trainable projection to 256 dimensions and LayerNorm
Protein queries	One learned 256-dimensional query per marker; four-head attention to the morphology tokens
Timestep conditioning	256-dimensional sinusoidal embedding and two-layer MLP, plus projections of the mean morphology token and mean conditioned protein query
Denoiser	
U-Net	Four levels with widths $[96, 192, 384, 384]$; two residual blocks per level (GroupNorm, SiLU)
Spatial cross-attention	$H/8$ bottleneck only; four heads; morphology tokens and conditioned protein queries provide the context
Output modulation	Diagonal bounded affine modulation, Eq. (3); explicit cross-marker mixing disabled; conditioning strength 1.0
Parameters	CRC: 61,788,392 total, 38,280,360 trainable, 37,242,952 in the U-Net. Multi-Tumor: 61,793,618 total, 38,285,586 trainable, 37,248,178 in the U-Net
Diffusion and optimization	
Noise schedule	1,000 diffusion steps; cosine schedule rescaled to zero terminal SNR; velocity prediction
Objective	Velocity-prediction MSE only, Eq. (4)
Optimizer	AdamW; learning rate 10^{-4}; weight decay 10^{-5}
Batching	Batch size 1; gradient accumulation over 4 iterations; gradient clipping at norm 1.0
Stopping and selection	At most 80 epochs; early stopping after 20 epochs without validation improvement; validation-selected EMA model (decay 0.999)
Training data	32 random crops per training region per epoch; artifact-rejection sampling; no augmentation
Inference	
Sampler	50 DDIM steps; cosine η schedule from 0 to 1; no terminal noise
Tiling	384×384 tiles at stride 96; independently initialized trajectory per tile
Stitching	Per-tile mapping to $[0, 1]$ and clipping, then 2D Hann-weighted overlap-add in floating point
Random initialization	Fixed across benchmark inference

ing and prediction operate on the complete P-channel protein tensor. A 256-dimensional sinusoidal timestep embedding is transformed by a two-layer MLP, and trainable projections of the mean morphology token and the mean morphology-conditioned protein query are added to it to form the residual-block conditioning vector. The 64-channel local H&E features at $H/8$ pass through a 1×1 adapter to width 384 before addition at the deepest encoder level.

The global morphology branch applies an ImageNet-initialized ResNet-50 to the same H&E tile after internal resizing to 224×224 and ImageNet standardization. Its terminal feature map is adaptively pooled to 2×2, and the four resulting 2048-dimensional vectors are projected to 256 dimensions and normalized. The convolutional backbone is frozen, whereas the token projection and LayerNorm are trainable. The P learned 256-dimensional protein queries attend to the four morphology tokens through the four-head multi-head attention of Section 3.4, and the conditioned queries are concatenated with the morphology tokens to form the cross-attention context. Spatial cross-attention is active only in the $H/8$ bottleneck, with four heads over the 384-channel representation; the downsampling and upsampling paths contain no active cross-attention.

Each conditioned query is projected to candidate scale and bias values for every output channel, and the final model uses only the marker-matched diagonal terms of Eq. (3), with conditioning strength 1.0. The $P \times P$ proposal projections and the output-mixing logit matrix of the ablated variant (Appendix F.2) remain part of the trained model, but off-diagonal aggregation is inactive during prediction.

F.2 Ablated cross-marker output mixing

The mixing-enabled variant, which serves as the reference model in the conditioning ablations (Appendix C), replaces the diagonal selection by a learned convex combination of the proposals made by all protein queries. Let S_{qp} and B_{qp} denote the scale and bias proposed by query q for output channel p, and let $G \in \mathbb{R}^{P\times P}$ be a learned logit matrix. Then

$$A_{qp} = \frac{\exp(G_{qp})}{\sum_{q'=1}^{P} \exp(G_{q'p})}, \qquad s_p = \sum_{q=1}^{P} A_{qp} S_{qp}, \qquad b_p = \sum_{q=1}^{P} A_{qp} B_{qp}, \tag{6}$$

and the velocity is modulated as in Eq. (3). Purely marker-specific modulation corresponds to A concentrating on its diagonal; the final architecture instead fixes $s_p = S_{pp}$ and $b_p = B_{pp}$.

F.3 Optimization and model selection

ProDICE is trained with the velocity-prediction loss of Eq. (4) only. Optimization uses AdamW with learning rate 10^{-4} and weight decay 10^{-5}, physical batch size 1, gradient accumulation over 4 iterations, gradient clipping at norm 1.0, and at most 80 epochs, with early stopping after 20 epochs without improvement in validation velocity loss. An exponential moving average (EMA) of the weights with decay 0.999 is maintained, and the validation-selected EMA model is used for inference. For each Multi-Tumor outer fold, early stopping and model selection use only the internal-validation subjects assigned within that fold; outer-test subjects are never used for fitting or model selection. The local H&E encoder, U-Net, protein queries, and morphology projection and normalization layers are trainable; the ResNet-50 backbone remains frozen.

F.4 Sampling, stitching, and export precision

Inference uses 50 DDIM transitions under the cosine η schedule of Section 3.6. Transition 0 has $\eta = 0$ and is deterministic, and transitions 1–48 inject Gaussian noise. The final transition returns the predicted clean sample $\hat{s}_0$ directly, so no terminal noise is added. Benchmark inference uses fixed pseudorandom initialization while processing one tile at a time in a fixed order; this determines both the initial Gaussian states and the intermediate stochastic DDIM draws. Exact repeatability therefore assumes the same processing order, because changing the batch size or tile order changes which random draws are assigned to which tiles. The exploratory TCGA deployment instead uses deterministic initialization tied to tile location (Appendix D).

Each 720×960 region is covered by overlapping 384×384 tiles at stride 96, with terminal tiles anchored to the image boundaries when needed. Each tile is mapped back to $[0, 1]$ and clipped before two-dimensional Hann-weighted overlap-add,

$$\hat{y}_i(u) = \frac{\sum_{k\in\mathcal{K}_i(u)} w_{ik}(u)\, \hat{y}_{ik}(u)}{\sum_{k\in\mathcal{K}_i(u)} w_{ik}(u)}, \tag{7}$$

where $\mathcal{K}_i(u)$ is the set of tiles from region i covering coordinate u, $\hat{y}_{ik}(u)$ is the prediction of tile k, and $w_{ik}(u)$ is its shifted Hann weight. The stitched array remains floating point during blending.

The benchmark export then writes **ProDICE** predictions to PNG and repacks them into standardized `spatial_maps/*.npz` arrays; measured target maps follow the same PNG-based export. Ground-truth and **ProDICE** arrays consumed by the evaluator therefore lie on the 256-level grid $\{k/255\}_{k=0}^{255}$ (stored as floating-point arrays), whereas the Ridge, HEX, and ROSIE prediction arrays retain continuous floating-point values. The evaluator consumes these fixed arrays directly, and the precision difference is retained rather than harmonized post hoc.

G Baseline Details

G.1 Internal baselines

Mean Prior Floor. For each marker, this deterministic baseline computes the pixel-weighted mean over all training-region pixels and broadcasts it over each evaluation image. It has no access to H&E, and its Pearson and Spearman correlations are undefined because the predicted spatial variance is zero.

H&E Feature Ridge. Ten fixed per-pixel descriptors are used: RGB values (3), optical-density RGB (3), grayscale intensity (1), a 15×15 local grayscale mean (1), grayscale gradient magnitude (1), and RGB saturation range (1). A multi-output ridge model with $\alpha = 10$ maps these descriptors to the full protein panel. At most 5,000 spatially distributed training pixels are sampled from each region, with at most 500,000 pixels overall, and no validation-set hyperparameter search is performed.

G.2 External baselines

HEX. HEX (Li et al., 2026) uses the MUSK visual encoder (Xiang et al., 2025) to predict protein abundance over central 14×14 target footprints from 224×224 H&E context windows. We use cohort-specific output heads for the 56 CRC or 58 Multi-Tumor markers and fine-tune the upper four transformer blocks, the final normalization, and the regression head, keeping the remaining visual backbone frozen. Training uses the fixed subject-disjoint splits with validation-based checkpoint selection and no access to held-out test subjects. Full-region reconstruction assigns predictions to their native 14×14 footprints, adds boundary-completion footprints where needed, and averages overlapping contributions; the resulting continuous-valued 720×960 arrays are evaluated with the same unmasked global-raw implementation as **ProDICE**. Earlier common-support and matched-resolution analyses were development sensitivities only and do not define Table 1.

ROSIE. ROSIE (Wu et al., 2025) uses a ConvNeXt-Small backbone on 128×128 H&E context crops resized to 224×224 and predicts marker means for the central 8×8 target region. We replace the published 50-marker head with a cohort-specific 56- or 58-marker projection and fine-tune the adapted model on each fixed training split. The primary reconstruction follows the upstream-faithful policy: stride-8 predictions are assigned directly to non-overlapping central 8×8 blocks to form continuous-valued 720×960 arrays. A 16×16 Bartlett/tent overlap-add reconstruction was evaluated as a sensitivity analysis and changed the macro metrics negligibly; the upstream direct-assignment policy is used for the reported benchmark.

Table 8: **Per-marker reconstruction fidelity on the CRC held-out test.** Detailed evaluation across all 56 canonical markers for all five methods. Columns report Pearson correlation (r), Spearman rank correlation (ρ), Lin's concordance correlation coefficient (CCC), local spatial SSIM, root-mean-square error (RMSE), exact 1-Wasserstein transport distance (W_1), and spatial variance ratio (VR). Stochastic methods report the three-seed replicate mean.

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR ($\approx$ 1)
BCL-2 (apoptosis)	**ProDICE**	0.284	0.276	0.245	0.263	0.133	0.033	0.332
	H&E Feature Ridge	0.176	0.194	0.053	0.092	0.130	0.070	0.024
	Mean Prior Floor	N/A	N/A	0.000	0.049	0.132	0.082	0.000
	ROSIE	0.204	0.067	0.125	0.169	0.131	0.046	0.119
	HEX	0.118	0.016	0.060	0.110	0.134	0.046	0.077
CD11b (macrophages)	**ProDICE**	0.295	0.321	0.251	0.243	0.159	0.042	0.314
	H&E Feature Ridge	0.281	0.282	0.139	0.148	0.154	0.079	0.071
	Mean Prior Floor	N/A	N/A	0.000	0.037	0.161	0.107	0.000
	ROSIE	0.355	0.375	0.266	0.245	0.151	0.059	0.207
	HEX	0.323	0.331	0.218	0.189	0.157	0.052	0.171
CD11c (dendritic cells)	**ProDICE**	0.219	0.277	0.177	0.146	0.163	0.050	0.258
	H&E Feature Ridge	0.196	0.271	0.076	0.086	0.157	0.084	0.041
	Mean Prior Floor	N/A	N/A	0.000	0.038	0.160	0.109	0.000
	ROSIE	-0.022	-0.031	-0.014	0.073	0.171	0.076	0.128
	HEX	-0.001	-0.023	-0.000	0.045	0.173	0.067	0.067
CD138 (plasma cells)	**ProDICE**	0.165	0.214	0.137	0.151	0.130	0.022	0.280
	H&E Feature Ridge	0.190	0.271	0.066	0.103	0.122	0.048	0.032
	Mean Prior Floor	N/A	N/A	0.000	0.055	0.124	0.055	0.000
	ROSIE	0.036	-0.018	0.014	0.071	0.126	0.038	0.041
	HEX	-0.064	-0.107	-0.011	0.046	0.126	0.036	0.008
CD15 (granulocytes)	**ProDICE**	0.123	0.163	0.089	0.124	0.111	0.019	0.181
	H&E Feature Ridge	0.109	0.218	0.023	0.086	0.106	0.038	0.011
	Mean Prior Floor	N/A	N/A	0.000	0.070	0.107	0.040	0.000
	ROSIE	0.091	0.058	0.019	0.056	0.108	0.021	0.012
	HEX	-0.032	-0.090	-0.006	0.071	0.108	0.028	0.009
CD163 (macrophages)	**ProDICE**	0.091	0.125	0.071	0.101	0.128	0.027	0.237
	H&E Feature Ridge	0.070	-0.034	0.010	0.068	0.119	0.049	0.005
	Mean Prior Floor	N/A	N/A	0.000	0.061	0.119	0.052	0.000
	ROSIE	0.084	0.020	0.036	0.084	0.121	0.051	0.053
	HEX	0.076	-0.017	0.023	0.082	0.120	0.036	0.026

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD194/CCR4 (chemokine receptor)	**ProDICE**	0.356	0.352	0.308	0.280	0.157	0.040	0.337
	H&E Feature Ridge	0.304	0.299	0.160	0.158	0.156	0.076	0.082
	Mean Prior Floor	N/A	N/A	0.000	0.035	0.164	0.106	0.000
	ROSIE	0.393	0.396	0.341	0.292	0.154	0.039	0.351
	HEX	0.339	0.350	0.233	0.196	0.157	0.052	0.173
CD20 (B cells)	**ProDICE**	0.130	0.227	0.096	0.137	0.111	0.018	0.202
	H&E Feature Ridge	0.087	0.219	0.025	0.088	0.107	0.037	0.021
	Mean Prior Floor	N/A	N/A	0.000	0.073	0.107	0.040	0.000
	ROSIE	0.086	0.074	0.029	0.089	0.108	0.024	0.032
	HEX	0.076	-0.013	0.012	0.075	0.107	0.024	0.006
CD21 (dendritic cells)	**ProDICE**	0.038	0.046	0.033	0.049	0.225	0.077	0.351
	H&E Feature Ridge	0.052	0.056	0.007	0.039	0.196	0.144	0.005
	Mean Prior Floor	N/A	N/A	0.000	0.028	0.196	0.153	0.000
	ROSIE	0.050	0.082	0.036	0.073	0.212	0.124	0.194
	HEX	0.061	0.082	0.040	0.082	0.210	0.081	0.155
CD25 (IL-2 receptor alpha)	**ProDICE**	0.223	0.228	0.196	0.206	0.188	0.056	0.352
	H&E Feature Ridge	0.245	0.262	0.122	0.139	0.174	0.098	0.071
	Mean Prior Floor	N/A	N/A	0.000	0.030	0.180	0.134	0.000
	ROSIE	0.291	0.320	0.232	0.241	0.176	0.075	0.250
	HEX	0.247	0.277	0.184	0.179	0.185	0.063	0.235
CD2 (T cells)	**ProDICE**	0.121	0.117	0.105	0.121	0.216	0.070	0.337
	H&E Feature Ridge	0.151	0.112	0.044	0.069	0.195	0.131	0.023
	Mean Prior Floor	N/A	N/A	0.000	0.025	0.197	0.153	0.000
	ROSIE	0.154	0.168	0.114	0.131	0.203	0.099	0.195
	HEX	0.145	0.152	0.102	0.103	0.211	0.079	0.200
CD30 (costimulator)	**ProDICE**	0.256	0.300	0.215	0.220	0.158	0.044	0.299
	H&E Feature Ridge	0.179	0.222	0.058	0.093	0.154	0.088	0.027
	Mean Prior Floor	N/A	N/A	0.000	0.042	0.156	0.104	0.000
	ROSIE	0.106	-0.002	0.074	0.124	0.162	0.069	0.165
	HEX	0.092	-0.021	0.047	0.067	0.163	0.060	0.086

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD31 (vasculature)	**ProDICE**	0.125	0.177	0.101	0.132	0.109	0.018	0.259
	H&E Feature Ridge	0.049	0.265	0.013	0.078	0.103	0.036	0.019
	Mean Prior Floor	N/A	N/A	0.000	0.075	0.103	0.042	0.000
	ROSIE	0.045	-0.034	0.022	0.081	0.107	0.041	0.075
	HEX	0.001	-0.104	0.000	0.085	0.104	0.026	0.011
CD34 (vasculature)	**ProDICE**	0.159	0.232	0.130	0.143	0.120	0.024	0.272
	H&E Feature Ridge	0.049	0.250	0.015	0.063	0.115	0.049	0.024
	Mean Prior Floor	N/A	N/A	0.000	0.061	0.114	0.059	0.000
	ROSIE	0.076	0.001	0.043	0.095	0.118	0.047	0.104
	HEX	0.027	-0.054	0.009	0.072	0.116	0.035	0.031
CD38 (multifunctional)	**ProDICE**	0.220	0.208	0.179	0.177	0.142	0.049	0.277
	H&E Feature Ridge	0.137	0.204	0.048	0.106	0.138	0.079	0.034
	Mean Prior Floor	N/A	N/A	0.000	0.052	0.139	0.091	0.000
	ROSIE	0.197	0.273	0.137	0.173	0.140	0.065	0.175
	HEX	0.183	0.268	0.125	0.162	0.140	0.048	0.159
CD3 (T cells)	**ProDICE**	0.318	0.231	0.291	0.280	0.139	0.033	0.445
	H&E Feature Ridge	0.223	0.192	0.097	0.137	0.133	0.058	0.052
	Mean Prior Floor	N/A	N/A	0.000	0.050	0.137	0.077	0.000
	ROSIE	0.236	0.234	0.154	0.184	0.135	0.060	0.142
	HEX	0.172	0.211	0.086	0.134	0.137	0.043	0.076
CD44 (stroma)	**ProDICE**	0.291	0.294	0.251	0.252	0.170	0.052	0.329
	H&E Feature Ridge	0.208	0.179	0.090	0.109	0.167	0.092	0.053
	Mean Prior Floor	N/A	N/A	0.000	0.031	0.170	0.119	0.000
	ROSIE	0.243	0.207	0.187	0.194	0.171	0.082	0.226
	HEX	0.206	0.184	0.135	0.129	0.172	0.067	0.150
CD45RA (naive T cells)	**ProDICE**	0.229	0.247	0.195	0.200	0.143	0.040	0.318
	H&E Feature Ridge	0.153	0.155	0.044	0.080	0.137	0.078	0.022
	Mean Prior Floor	N/A	N/A	0.000	0.045	0.139	0.090	0.000
	ROSIE	0.182	0.096	0.121	0.145	0.141	0.070	0.155
	HEX	0.151	0.075	0.086	0.114	0.139	0.054	0.101

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD45RO (memory cells)	**ProDICE**	0.219	0.199	0.187	0.197	0.146	0.034	0.318
	H&E Feature Ridge	0.134	0.077	0.045	0.088	0.140	0.069	0.030
	Mean Prior Floor	N/A	N/A	0.000	0.045	0.141	0.085	0.000
	ROSIE	0.151	0.068	0.088	0.132	0.142	0.047	0.105
	HEX	0.147	0.102	0.081	0.105	0.143	0.051	0.095
CD45 (hematopoietic cells)	**ProDICE**	0.301	0.281	0.257	0.268	0.123	0.025	0.316
	H&E Feature Ridge	0.176	0.155	0.054	0.102	0.123	0.050	0.025
	Mean Prior Floor	N/A	N/A	0.000	0.055	0.125	0.058	0.000
	ROSIE	0.255	0.145	0.136	0.160	0.121	0.040	0.084
	HEX	0.190	0.122	0.075	0.112	0.123	0.042	0.043
CD4 (T helper cells)	**ProDICE**	0.270	0.269	0.239	0.245	0.147	0.036	0.371
	H&E Feature Ridge	0.182	0.121	0.067	0.098	0.142	0.075	0.037
	Mean Prior Floor	N/A	N/A	0.000	0.042	0.144	0.093	0.000
	ROSIE	0.265	0.176	0.175	0.188	0.140	0.050	0.146
	HEX	0.188	0.128	0.115	0.125	0.144	0.054	0.121
CD56 (NK cells)	**ProDICE**	0.291	0.361	0.262	0.267	0.161	0.045	0.405
	H&E Feature Ridge	0.286	0.375	0.192	0.210	0.153	0.078	0.152
	Mean Prior Floor	N/A	N/A	0.000	0.038	0.158	0.114	0.000
	ROSIE	0.333	0.420	0.294	0.292	0.158	0.065	0.389
	HEX	0.270	0.347	0.227	0.216	0.161	0.052	0.316
CD57 (NK cells)	**ProDICE**	0.086	0.086	0.071	0.087	0.211	0.072	0.278
	H&E Feature Ridge	0.112	0.085	0.029	0.054	0.192	0.127	0.018
	Mean Prior Floor	N/A	N/A	0.000	0.025	0.193	0.147	0.000
	ROSIE	0.081	0.101	0.056	0.087	0.203	0.108	0.160
	HEX	0.074	0.069	0.044	0.066	0.207	0.083	0.125
CD5 (T cells)	**ProDICE**	0.369	0.318	0.333	0.322	0.176	0.044	0.402
	H&E Feature Ridge	0.312	0.238	0.161	0.168	0.173	0.093	0.077
	Mean Prior Floor	N/A	N/A	0.000	0.029	0.182	0.129	0.000
	ROSIE	0.417	0.341	0.370	0.337	0.169	0.052	0.379
	HEX	0.344	0.296	0.253	0.227	0.177	0.059	0.217

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD68 (macrophages)	**ProDICE**	0.130	0.307	0.103	0.147	0.120	0.022	0.243
	H&E Feature Ridge	0.115	0.338	0.039	0.112	0.114	0.044	0.031
	Mean Prior Floor	N/A	N/A	0.000	0.069	0.114	0.053	0.000
	ROSIE	0.155	0.367	0.084	0.145	0.114	0.042	0.088
	HEX	0.122	0.327	0.052	0.118	0.114	0.035	0.051
CD71 (transferrin receptor)	**ProDICE**	0.437	0.573	0.395	0.365	0.156	0.040	0.404
	H&E Feature Ridge	0.403	0.525	0.290	0.277	0.155	0.073	0.180
	Mean Prior Floor	N/A	N/A	0.000	0.036	0.169	0.120	0.000
	ROSIE	0.436	0.544	0.378	0.324	0.154	0.056	0.337
	HEX	0.369	0.471	0.293	0.249	0.163	0.060	0.272
CD7 (T cells)	**ProDICE**	0.170	0.301	0.145	0.158	0.125	0.024	0.311
	H&E Feature Ridge	0.152	0.352	0.065	0.093	0.117	0.047	0.051
	Mean Prior Floor	N/A	N/A	0.000	0.059	0.119	0.059	0.000
	ROSIE	0.036	-0.048	0.020	0.080	0.123	0.044	0.095
	HEX	-0.033	-0.151	-0.008	0.071	0.121	0.037	0.016
CD8 (cytotoxic T cells)	**ProDICE**	0.149	0.160	0.122	0.142	0.132	0.029	0.276
	H&E Feature Ridge	0.089	0.033	0.022	0.079	0.124	0.059	0.015
	Mean Prior Floor	N/A	N/A	0.000	0.055	0.124	0.068	0.000
	ROSIE	0.141	0.099	0.090	0.127	0.128	0.060	0.140
	HEX	0.151	0.129	0.072	0.105	0.125	0.044	0.067
CDX2 (intestinal epithelia)	**ProDICE**	0.348	0.377	0.303	0.294	0.169	0.050	0.347
	H&E Feature Ridge	0.298	0.355	0.180	0.202	0.167	0.080	0.115
	Mean Prior Floor	N/A	N/A	0.000	0.032	0.175	0.124	0.000
	ROSIE	0.354	0.383	0.310	0.301	0.169	0.065	0.360
	HEX	0.287	0.337	0.212	0.211	0.177	0.066	0.234
Chromogranin A (neuroendocrine)	**ProDICE**	0.227	0.282	0.196	0.182	0.137	0.022	0.341
	H&E Feature Ridge	0.223	0.319	0.089	0.115	0.129	0.053	0.044
	Mean Prior Floor	N/A	N/A	0.000	0.047	0.132	0.062	0.000
	ROSIE	0.042	-0.059	0.019	0.057	0.134	0.047	0.051
	HEX	-0.054	-0.132	-0.009	0.037	0.135	0.039	0.007

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR ($\approx$ 1)
Collagen IV (basement membrane)	**ProDICE**	0.134	0.190	0.100	0.123	0.139	0.030	0.205
	H&E Feature Ridge	0.083	0.220	0.017	0.065	0.133	0.059	0.011
	Mean Prior Floor	N/A	N/A	0.000	0.049	0.134	0.064	0.000
	ROSIE	0.078	0.049	0.039	0.074	0.136	0.060	0.075
	HEX	0.044	0.022	0.011	0.046	0.136	0.046	0.017
Cytokeratin (epithelia)	**ProDICE**	0.178	0.193	0.138	0.170	0.113	0.019	0.229
	H&E Feature Ridge	0.098	0.199	0.018	0.080	0.110	0.039	0.009
	Mean Prior Floor	N/A	N/A	0.000	0.067	0.110	0.041	0.000
	ROSIE	0.101	0.073	0.034	0.062	0.111	0.022	0.032
	HEX	0.046	-0.042	0.009	0.074	0.111	0.026	0.011
EGFR (signaling)	**ProDICE**	0.351	0.329	0.318	0.299	0.189	0.052	0.404
	H&E Feature Ridge	0.313	0.277	0.160	0.156	0.184	0.105	0.076
	Mean Prior Floor	N/A	N/A	0.000	0.026	0.194	0.145	0.000
	ROSIE	0.394	0.371	0.354	0.311	0.184	0.060	0.401
	HEX	0.352	0.333	0.275	0.233	0.192	0.063	0.283
FOXP3 (regulatory T cells)	**ProDICE**	0.263	0.260	0.226	0.235	0.164	0.051	0.323
	H&E Feature Ridge	0.152	0.094	0.063	0.104	0.161	0.092	0.047
	Mean Prior Floor	N/A	N/A	0.000	0.038	0.162	0.117	0.000
	ROSIE	0.221	0.192	0.174	0.214	0.164	0.066	0.236
	HEX	0.192	0.169	0.137	0.154	0.168	0.056	0.196
GATA3 (Th2 helper T cells)	**ProDICE**	0.228	0.190	0.196	0.222	0.188	0.057	0.336
	H&E Feature Ridge	0.235	0.199	0.122	0.161	0.176	0.094	0.079
	Mean Prior Floor	N/A	N/A	0.000	0.029	0.181	0.131	0.000
	ROSIE	0.207	0.220	0.146	0.172	0.181	0.083	0.170
	HEX	0.180	0.175	0.114	0.123	0.188	0.073	0.150
GFAP (nerves)	**ProDICE**	0.253	0.296	0.226	0.201	0.174	0.045	0.380
	H&E Feature Ridge	0.155	0.283	0.047	0.068	0.166	0.101	0.025
	Mean Prior Floor	N/A	N/A	0.000	0.032	0.168	0.116	0.000
	ROSIE	0.087	-0.021	0.059	0.063	0.179	0.097	0.165
	HEX	0.060	-0.055	0.027	0.036	0.174	0.071	0.061

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR ($\approx$ 1)
Granzyme B (cytotoxicity)	**ProDICE**	0.221	0.205	0.183	0.194	0.116	0.018	0.284
	H&E Feature Ridge	0.177	0.257	0.059	0.108	0.111	0.037	0.030
	Mean Prior Floor	N/A	N/A	0.000	0.066	0.113	0.042	0.000
	ROSIE	0.125	0.082	0.030	0.029	0.114	0.022	0.016
	HEX	-0.052	-0.108	-0.008	0.059	0.114	0.029	0.006
HLA-DR (MHC-II)	**ProDICE**	0.199	0.272	0.160	0.174	0.130	0.026	0.253
	H&E Feature Ridge	0.175	0.315	0.055	0.097	0.125	0.054	0.026
	Mean Prior Floor	N/A	N/A	0.000	0.054	0.127	0.061	0.000
	ROSIE	0.094	0.013	0.041	0.087	0.128	0.042	0.054
	HEX	-0.006	-0.074	-0.001	0.049	0.129	0.043	0.018
ICOS (costimulator)	**ProDICE**	0.042	0.055	0.036	0.055	0.213	0.060	0.312
	H&E Feature Ridge	0.064	0.080	0.010	0.035	0.189	0.125	0.006
	Mean Prior Floor	N/A	N/A	0.000	0.026	0.189	0.135	0.000
	ROSIE	0.063	0.101	0.028	0.059	0.193	0.102	0.056
	HEX	0.054	0.069	0.027	0.054	0.198	0.087	0.079
IDO-1 (metabolism)	**ProDICE**	0.295	0.316	0.262	0.270	0.202	0.064	0.370
	H&E Feature Ridge	0.367	0.386	0.239	0.233	0.187	0.103	0.137
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.201	0.158	0.000
	ROSIE	0.374	0.407	0.328	0.303	0.192	0.083	0.354
	HEX	0.311	0.347	0.257	0.232	0.205	0.073	0.339
Ki67 (proliferation)	**ProDICE**	0.160	0.105	0.122	0.143	0.135	0.041	0.214
	H&E Feature Ridge	0.106	0.099	0.025	0.081	0.130	0.057	0.015
	Mean Prior Floor	N/A	N/A	0.000	0.056	0.131	0.064	0.000
	ROSIE	0.087	0.085	0.040	0.099	0.133	0.061	0.062
	HEX	0.061	0.051	0.018	0.090	0.133	0.042	0.024
LAG-3 (checkpoint)	**ProDICE**	0.147	0.134	0.127	0.144	0.213	0.067	0.338
	H&E Feature Ridge	0.200	0.163	0.080	0.098	0.193	0.120	0.043
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.197	0.151	0.000
	ROSIE	0.190	0.211	0.145	0.152	0.201	0.097	0.215
	HEX	0.168	0.179	0.115	0.113	0.209	0.077	0.190

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
MMP12 (matrix metalloproteinase)	**ProDICE**	0.091	0.091	0.077	0.085	0.217	0.067	0.305
	H&E Feature Ridge	0.127	0.106	0.030	0.046	0.197	0.128	0.014
	Mean Prior Floor	N/A	N/A	0.000	0.023	0.198	0.144	0.000
	ROSIE	0.116	0.125	0.061	0.073	0.200	0.109	0.081
	HEX	0.102	0.090	0.053	0.063	0.206	0.091	0.089
MMP9 (matrix metalloproteinase)	**ProDICE**	0.248	0.303	0.216	0.198	0.141	0.027	0.343
	H&E Feature Ridge	0.220	0.331	0.095	0.112	0.134	0.058	0.052
	Mean Prior Floor	N/A	N/A	0.000	0.045	0.137	0.072	0.000
	ROSIE	0.019	-0.056	0.010	0.049	0.142	0.065	0.072
	HEX	-0.052	-0.132	-0.010	0.034	0.141	0.045	0.010
MUC-1 (epithelia)	**ProDICE**	0.172	0.231	0.129	0.148	0.145	0.033	0.211
	H&E Feature Ridge	0.138	0.246	0.037	0.082	0.139	0.062	0.019
	Mean Prior Floor	N/A	N/A	0.000	0.046	0.141	0.071	0.000
	ROSIE	0.085	0.023	0.035	0.070	0.141	0.056	0.047
	HEX	0.008	-0.020	0.002	0.033	0.144	0.050	0.012
Na-K-ATPase (membranes)	**ProDICE**	0.177	0.237	0.141	0.153	0.129	0.027	0.256
	H&E Feature Ridge	0.112	0.224	0.025	0.073	0.124	0.055	0.013
	Mean Prior Floor	N/A	N/A	0.000	0.054	0.125	0.059	0.000
	ROSIE	0.069	0.021	0.040	0.070	0.129	0.054	0.108
	HEX	0.078	-0.027	0.023	0.067	0.125	0.038	0.024
PD-1 (checkpoint)	**ProDICE**	0.303	0.284	0.263	0.270	0.184	0.057	0.338
	H&E Feature Ridge	0.324	0.308	0.182	0.195	0.176	0.095	0.095
	Mean Prior Floor	N/A	N/A	0.000	0.028	0.186	0.139	0.000
	ROSIE	0.347	0.369	0.305	0.292	0.181	0.072	0.369
	HEX	0.312	0.334	0.243	0.218	0.189	0.066	0.288
PD-L1 (checkpoint)	**ProDICE**	0.195	0.224	0.160	0.165	0.165	0.050	0.275
	H&E Feature Ridge	0.124	0.149	0.035	0.068	0.158	0.092	0.021
	Mean Prior Floor	N/A	N/A	0.000	0.038	0.159	0.107	0.000
	ROSIE	0.114	0.074	0.075	0.120	0.164	0.063	0.147
	HEX	0.069	0.045	0.040	0.077	0.167	0.062	0.109

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
Podoplanin (lymphatics)	**ProDICE**	0.088	0.111	0.067	0.094	0.160	0.058	0.218
	H&E Feature Ridge	0.078	0.145	0.018	0.061	0.149	0.088	0.015
	Mean Prior Floor	N/A	N/A	0.000	0.041	0.150	0.097	0.000
	ROSIE	0.097	0.162	0.056	0.101	0.153	0.081	0.104
	HEX	0.093	0.164	0.049	0.088	0.153	0.059	0.087
Synaptophysin (neuroendocrine)	**ProDICE**	0.156	0.310	0.114	0.102	0.133	0.030	0.191
	H&E Feature Ridge	0.201	0.354	0.085	0.095	0.127	0.053	0.049
	Mean Prior Floor	N/A	N/A	0.000	0.055	0.129	0.067	0.000
	ROSIE	-0.154	-0.069	-0.066	0.030	0.137	0.061	0.051
	HEX	-0.108	-0.122	-0.022	0.051	0.134	0.043	0.012
T-bet (Th1 cells)	**ProDICE**	0.353	0.338	0.315	0.307	0.186	0.055	0.387
	H&E Feature Ridge	0.317	0.289	0.179	0.184	0.181	0.100	0.096
	Mean Prior Floor	N/A	N/A	0.000	0.026	0.191	0.144	0.000
	ROSIE	0.392	0.374	0.324	0.303	0.180	0.065	0.294
	HEX	0.336	0.321	0.262	0.233	0.190	0.066	0.275
VISTA (costimulator)	**ProDICE**	0.269	0.293	0.229	0.219	0.156	0.049	0.313
	H&E Feature Ridge	0.213	0.263	0.100	0.139	0.151	0.081	0.062
	Mean Prior Floor	N/A	N/A	0.000	0.042	0.155	0.106	0.000
	ROSIE	0.257	0.326	0.206	0.225	0.157	0.071	0.271
	HEX	0.244	0.306	0.181	0.186	0.156	0.051	0.212
Vimentin (cytoplasm)	**ProDICE**	0.146	0.139	0.117	0.142	0.142	0.034	0.250
	H&E Feature Ridge	0.094	0.017	0.024	0.070	0.134	0.066	0.017
	Mean Prior Floor	N/A	N/A	0.000	0.048	0.135	0.076	0.000
	ROSIE	0.166	0.122	0.103	0.129	0.136	0.065	0.125
	HEX	0.116	0.110	0.054	0.088	0.136	0.051	0.062
Alpha-SMA (smooth muscle)	**ProDICE**	0.207	0.185	0.160	0.150	0.146	0.037	0.222
	H&E Feature Ridge	0.172	0.115	0.038	0.067	0.143	0.065	0.013
	Mean Prior Floor	N/A	N/A	0.000	0.043	0.145	0.072	0.000
	ROSIE	0.216	0.248	0.126	0.114	0.142	0.070	0.104
	HEX	0.124	0.131	0.036	0.062	0.146	0.052	0.023

Table 8 (CRC, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
Beta-catenin (Wnt signaling)	**ProDICE**	0.612	0.664	0.587	0.552	0.161	0.036	0.564
	H&E Feature Ridge	0.561	0.631	0.487	0.448	0.166	0.071	0.336
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.201	0.155	0.000
	ROSIE	0.597	0.652	0.554	0.485	0.164	0.048	0.486
	HEX	0.499	0.556	0.435	0.370	0.181	0.067	0.395
p53 (tumor suppressor)	**ProDICE**	0.337	0.363	0.279	0.253	0.152	0.044	0.290
	H&E Feature Ridge	0.369	0.370	0.203	0.196	0.147	0.071	0.091
	Mean Prior Floor	N/A	N/A	0.000	0.038	0.158	0.107	0.000
	ROSIE	0.378	0.397	0.322	0.281	0.149	0.047	0.317
	HEX	0.360	0.383	0.246	0.199	0.152	0.054	0.179

Table 9: **Per-marker reconstruction fidelity on the Multi-Tumor pooled OOF test.** Detailed evaluation across all 58 canonical markers for all five methods. Columns report Pearson correlation (r), Spearman rank correlation (ρ), Lin's concordance correlation coefficient (CCC), local spatial SSIM, root-mean-square error (RMSE), exact 1-Wasserstein transport distance (W_1), and spatial variance ratio (VR).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
BCL-2 (apoptosis)	**ProDICE**	0.207	0.211	0.167	0.162	0.167	0.053	0.258
	H&E Feature Ridge	0.195	0.091	0.079	0.093	0.160	0.090	0.045
	Mean Prior Floor	N/A	N/A	0.000	0.034	0.163	0.111	0.000
	ROSIE	0.187	0.099	0.122	0.114	0.163	0.088	0.140
	HEX	0.170	0.136	0.090	0.091	0.163	0.074	0.085
CD117/c-Kit (stem cells)	**ProDICE**	0.254	0.269	0.215	0.199	0.203	0.071	0.303
	H&E Feature Ridge	0.284	0.289	0.155	0.140	0.192	0.112	0.088
	Mean Prior Floor	N/A	N/A	0.000	0.026	0.200	0.154	0.000
	ROSIE	0.265	0.295	0.196	0.171	0.196	0.104	0.195
	HEX	0.275	0.315	0.187	0.152	0.197	0.092	0.161
CD11c (dendritic cells)	**ProDICE**	0.164	0.190	0.127	0.131	0.150	0.045	0.225
	H&E Feature Ridge	0.024	-0.039	0.003	0.045	0.145	0.087	0.004
	Mean Prior Floor	N/A	N/A	0.000	0.043	0.145	0.091	0.000
	ROSIE	0.008	-0.158	0.004	0.047	0.149	0.075	0.072
	HEX	0.015	-0.072	0.006	0.041	0.149	0.064	0.043
CD123 (plasmacytoid dendritic cells)	**ProDICE**	0.587	0.531	0.561	0.507	0.174	0.046	0.545
	H&E Feature Ridge	0.433	0.446	0.318	0.243	0.190	0.097	0.193
	Mean Prior Floor	N/A	N/A	0.000	0.022	0.211	0.163	0.000
	ROSIE	0.416	0.466	0.312	0.240	0.192	0.103	0.205
	HEX	0.438	0.485	0.326	0.226	0.191	0.087	0.206
CD138 (plasma cells)	**ProDICE**	0.186	0.219	0.126	0.149	0.134	0.039	0.151
	H&E Feature Ridge	0.075	0.115	0.014	0.064	0.133	0.062	0.009
	Mean Prior Floor	N/A	N/A	0.000	0.050	0.134	0.067	0.000
	ROSIE	-0.065	-0.155	-0.017	0.040	0.136	0.056	0.017
	HEX	-0.092	-0.138	-0.020	0.033	0.137	0.052	0.012
CD15 (myeloid cells)	**ProDICE**	0.205	0.230	0.133	0.155	0.119	0.027	0.136
	H&E Feature Ridge	0.098	0.126	0.021	0.078	0.119	0.045	0.011
	Mean Prior Floor	N/A	N/A	0.000	0.060	0.120	0.049	0.000
	ROSIE	-0.015	-0.073	-0.003	0.053	0.121	0.040	0.014
	HEX	-0.046	-0.079	-0.006	0.046	0.122	0.033	0.005

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD163 (macrophages)	**ProDICE**	0.048	0.090	0.036	0.087	0.147	0.046	0.200
	H&E Feature Ridge	0.111	0.151	0.031	0.079	0.136	0.071	0.020
	Mean Prior Floor	N/A	N/A	0.000	0.054	0.137	0.080	0.000
	ROSIE	0.115	0.139	0.050	0.095	0.137	0.062	0.054
	HEX	0.102	0.153	0.046	0.094	0.138	0.059	0.058
CD194/CCR4 (chemokine receptor)	**ProDICE**	0.309	0.347	0.270	0.249	0.168	0.051	0.348
	H&E Feature Ridge	0.322	0.319	0.202	0.183	0.160	0.083	0.124
	Mean Prior Floor	N/A	N/A	0.000	0.032	0.169	0.118	0.000
	ROSIE	0.317	0.344	0.194	0.179	0.161	0.082	0.118
	HEX	0.324	0.374	0.208	0.176	0.163	0.065	0.141
CD20 (B cells)	**ProDICE**	0.267	0.257	0.191	0.205	0.123	0.029	0.177
	H&E Feature Ridge	0.145	0.150	0.041	0.092	0.125	0.049	0.021
	Mean Prior Floor	N/A	N/A	0.000	0.054	0.126	0.055	0.000
	ROSIE	-0.029	-0.085	-0.006	0.047	0.128	0.040	0.010
	HEX	-0.055	-0.086	-0.008	0.034	0.129	0.035	0.006
CD21 (follicular dendritic cells)	**ProDICE**	0.229	0.226	0.184	0.189	0.168	0.058	0.254
	H&E Feature Ridge	0.057	-0.042	0.012	0.045	0.166	0.106	0.011
	Mean Prior Floor	N/A	N/A	0.000	0.033	0.166	0.114	0.000
	ROSIE	0.031	-0.091	0.019	0.049	0.175	0.093	0.126
	HEX	0.043	-0.023	0.018	0.030	0.172	0.079	0.051
CD25 (IL-2 receptor alpha)	**ProDICE**	0.260	0.273	0.223	0.221	0.191	0.059	0.322
	H&E Feature Ridge	0.263	0.214	0.143	0.150	0.181	0.105	0.086
	Mean Prior Floor	N/A	N/A	0.000	0.027	0.188	0.139	0.000
	ROSIE	0.267	0.230	0.175	0.172	0.183	0.096	0.141
	HEX	0.285	0.287	0.183	0.160	0.184	0.083	0.141
CD2 (T cells)	**ProDICE**	0.189	0.192	0.158	0.160	0.222	0.083	0.291
	H&E Feature Ridge	0.246	0.245	0.119	0.120	0.206	0.129	0.067
	Mean Prior Floor	N/A	N/A	0.000	0.021	0.213	0.168	0.000
	ROSIE	0.217	0.233	0.151	0.143	0.213	0.115	0.168
	HEX	0.225	0.247	0.154	0.137	0.215	0.102	0.170

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD30 (activation)	**ProDICE**	0.219	0.216	0.178	0.191	0.161	0.052	0.264
	H&E Feature Ridge	0.064	-0.079	0.015	0.058	0.158	0.098	0.014
	Mean Prior Floor	N/A	N/A	0.000	0.036	0.158	0.109	0.000
	ROSIE	0.037	-0.118	0.025	0.071	0.168	0.083	0.151
	HEX	0.066	-0.013	0.036	0.058	0.163	0.075	0.090
CD31 (vasculature)	**ProDICE**	0.093	0.157	0.065	0.108	0.112	0.028	0.164
	H&E Feature Ridge	0.019	0.093	0.002	0.076	0.108	0.044	0.003
	Mean Prior Floor	N/A	N/A	0.000	0.074	0.108	0.046	0.000
	ROSIE	-0.045	-0.190	-0.022	0.057	0.113	0.048	0.068
	HEX	-0.031	-0.122	-0.007	0.069	0.109	0.037	0.013
CD34 (vasculature)	**ProDICE**	0.082	0.087	0.066	0.097	0.128	0.028	0.251
	H&E Feature Ridge	0.017	-0.195	0.003	0.069	0.118	0.052	0.007
	Mean Prior Floor	N/A	N/A	0.000	0.063	0.118	0.055	0.000
	ROSIE	0.049	-0.159	0.016	0.077	0.119	0.048	0.028
	HEX	0.068	-0.064	0.020	0.076	0.119	0.041	0.022
CD38 (multifunctional)	**ProDICE**	0.279	0.347	0.239	0.208	0.185	0.060	0.318
	H&E Feature Ridge	0.317	0.366	0.189	0.179	0.175	0.097	0.110
	Mean Prior Floor	N/A	N/A	0.000	0.031	0.185	0.135	0.000
	ROSIE	0.306	0.363	0.227	0.206	0.178	0.088	0.198
	HEX	0.333	0.400	0.236	0.200	0.178	0.071	0.186
CD3 (T cells)	**ProDICE**	0.155	0.165	0.123	0.157	0.189	0.067	0.248
	H&E Feature Ridge	0.123	0.147	0.037	0.074	0.180	0.109	0.023
	Mean Prior Floor	N/A	N/A	0.000	0.033	0.181	0.127	0.000
	ROSIE	0.124	0.165	0.076	0.110	0.184	0.097	0.118
	HEX	0.121	0.174	0.062	0.099	0.185	0.084	0.081
CD45RA (naive cells)	**ProDICE**	0.223	0.225	0.182	0.187	0.165	0.052	0.270
	H&E Feature Ridge	0.108	-0.004	0.031	0.066	0.161	0.098	0.022
	Mean Prior Floor	N/A	N/A	0.000	0.036	0.162	0.109	0.000
	ROSIE	0.101	0.005	0.065	0.081	0.167	0.090	0.133
	HEX	0.097	0.038	0.047	0.065	0.167	0.071	0.071

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD45RO (memory)	**ProDICE**	0.172	0.184	0.141	0.158	0.171	0.052	0.275
	H&E Feature Ridge	0.058	-0.048	0.014	0.054	0.164	0.099	0.015
	Mean Prior Floor	N/A	N/A	0.000	0.036	0.164	0.110	0.000
	ROSIE	0.061	-0.052	0.033	0.076	0.169	0.079	0.089
	HEX	0.089	0.026	0.049	0.070	0.169	0.071	0.097
CD45 (hematopoietic cells)	**ProDICE**	0.186	0.229	0.145	0.148	0.137	0.036	0.229
	H&E Feature Ridge	0.086	-0.005	0.021	0.066	0.133	0.063	0.016
	Mean Prior Floor	N/A	N/A	0.000	0.049	0.134	0.071	0.000
	ROSIE	0.021	-0.096	0.006	0.054	0.135	0.058	0.022
	HEX	0.043	0.002	0.014	0.049	0.135	0.054	0.027
CD4 (T helper cells)	**ProDICE**	0.246	0.242	0.215	0.217	0.182	0.053	0.347
	H&E Feature Ridge	0.219	0.128	0.101	0.118	0.173	0.101	0.060
	Mean Prior Floor	N/A	N/A	0.000	0.030	0.178	0.127	0.000
	ROSIE	0.190	0.102	0.120	0.130	0.177	0.091	0.126
	HEX	0.212	0.185	0.131	0.122	0.176	0.082	0.123
CD56 (NK cells)	**ProDICE**	0.183	0.221	0.146	0.164	0.195	0.072	0.251
	H&E Feature Ridge	0.233	0.263	0.114	0.131	0.183	0.108	0.068
	Mean Prior Floor	N/A	N/A	0.000	0.028	0.189	0.140	0.000
	ROSIE	0.228	0.262	0.148	0.157	0.187	0.099	0.138
	HEX	0.227	0.280	0.146	0.154	0.188	0.090	0.141
CD57 (NK cells)	**ProDICE**	0.210	0.206	0.173	0.178	0.210	0.079	0.277
	H&E Feature Ridge	0.233	0.223	0.111	0.109	0.199	0.122	0.064
	Mean Prior Floor	N/A	N/A	0.000	0.023	0.205	0.157	0.000
	ROSIE	0.215	0.229	0.125	0.123	0.205	0.109	0.110
	HEX	0.232	0.259	0.147	0.131	0.204	0.100	0.138
CD5 (T cells)	**ProDICE**	0.241	0.227	0.212	0.223	0.197	0.056	0.356
	H&E Feature Ridge	0.237	0.143	0.124	0.127	0.185	0.107	0.080
	Mean Prior Floor	N/A	N/A	0.000	0.026	0.191	0.140	0.000
	ROSIE	0.216	0.124	0.139	0.137	0.189	0.099	0.137
	HEX	0.240	0.208	0.165	0.140	0.189	0.086	0.164

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
CD68 (macrophages)	**ProDICE**	0.095	0.105	0.076	0.107	0.130	0.031	0.251
	H&E Feature Ridge	0.104	-0.000	0.030	0.084	0.121	0.052	0.021
	Mean Prior Floor	N/A	N/A	0.000	0.060	0.121	0.061	0.000
	ROSIE	0.117	0.028	0.042	0.090	0.121	0.050	0.035
	HEX	0.106	0.038	0.035	0.083	0.122	0.042	0.030
CD79a (B and plasma cells)	**ProDICE**	0.185	0.117	0.139	0.148	0.130	0.037	0.206
	H&E Feature Ridge	0.132	0.056	0.043	0.086	0.126	0.058	0.028
	Mean Prior Floor	N/A	N/A	0.000	0.052	0.127	0.070	0.000
	ROSIE	0.106	0.013	0.050	0.081	0.128	0.064	0.063
	HEX	0.061	0.022	0.021	0.062	0.128	0.052	0.034
CD7 (T cells)	**ProDICE**	0.167	0.181	0.134	0.166	0.161	0.045	0.254
	H&E Feature Ridge	0.116	0.003	0.032	0.069	0.153	0.086	0.020
	Mean Prior Floor	N/A	N/A	0.000	0.041	0.154	0.099	0.000
	ROSIE	0.107	0.002	0.052	0.085	0.156	0.074	0.067
	HEX	0.121	0.061	0.053	0.080	0.157	0.064	0.058
CD8 (cytotoxic T cells)	**ProDICE**	0.154	0.171	0.120	0.146	0.160	0.048	0.234
	H&E Feature Ridge	0.114	0.022	0.037	0.080	0.153	0.086	0.028
	Mean Prior Floor	N/A	N/A	0.000	0.040	0.154	0.101	0.000
	ROSIE	0.113	0.054	0.058	0.091	0.155	0.079	0.076
	HEX	0.107	0.059	0.050	0.077	0.156	0.070	0.064
CDX2 (intestinal epithelia)	**ProDICE**	0.227	0.302	0.190	0.185	0.159	0.046	0.293
	H&E Feature Ridge	0.227	0.299	0.105	0.117	0.151	0.082	0.060
	Mean Prior Floor	N/A	N/A	0.000	0.040	0.156	0.103	0.000
	ROSIE	0.212	0.218	0.122	0.131	0.153	0.075	0.101
	HEX	0.227	0.294	0.120	0.119	0.154	0.067	0.087
Chromogranin A (neuroendocrine)	**ProDICE**	0.221	0.265	0.152	0.168	0.135	0.038	0.161
	H&E Feature Ridge	0.106	0.151	0.024	0.068	0.135	0.066	0.013
	Mean Prior Floor	N/A	N/A	0.000	0.048	0.136	0.073	0.000
	ROSIE	-0.034	-0.107	-0.018	0.028	0.145	0.079	0.084
	HEX	-0.039	-0.064	-0.010	0.033	0.139	0.053	0.017

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
Collagen IV (basement membrane)	**ProDICE**	0.243	0.291	0.176	0.192	0.129	0.034	0.184
	H&E Feature Ridge	0.085	0.156	0.017	0.067	0.130	0.062	0.010
	Mean Prior Floor	N/A	N/A	0.000	0.052	0.131	0.068	0.000
	ROSIE	-0.043	-0.126	-0.021	0.039	0.136	0.065	0.064
	HEX	-0.024	-0.039	-0.006	0.042	0.133	0.048	0.015
Cytokeratin 7 (epithelia)	**ProDICE**	0.185	0.222	0.115	0.134	0.128	0.035	0.122
	H&E Feature Ridge	0.102	0.123	0.022	0.074	0.128	0.056	0.012
	Mean Prior Floor	N/A	N/A	0.000	0.053	0.129	0.061	0.000
	ROSIE	-0.019	-0.054	-0.004	0.051	0.130	0.048	0.011
	HEX	-0.044	-0.077	-0.008	0.038	0.131	0.043	0.009
EGFR (therapy target)	**ProDICE**	0.222	0.203	0.196	0.203	0.210	0.062	0.361
	H&E Feature Ridge	0.234	0.143	0.130	0.132	0.196	0.115	0.092
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.202	0.152	0.000
	ROSIE	0.239	0.137	0.156	0.145	0.199	0.105	0.142
	HEX	0.234	0.204	0.162	0.138	0.203	0.082	0.176
EpCAM (epithelia)	**ProDICE**	0.405	0.362	0.370	0.351	0.186	0.055	0.425
	H&E Feature Ridge	0.309	0.307	0.178	0.156	0.186	0.104	0.100
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.196	0.148	0.000
	ROSIE	0.277	0.307	0.205	0.179	0.191	0.098	0.195
	HEX	0.314	0.336	0.207	0.159	0.191	0.081	0.157
FOXP3 (regulatory T cells)	**ProDICE**	0.200	0.164	0.160	0.163	0.149	0.042	0.249
	H&E Feature Ridge	0.159	0.104	0.056	0.088	0.144	0.078	0.033
	Mean Prior Floor	N/A	N/A	0.000	0.045	0.146	0.093	0.000
	ROSIE	0.138	0.042	0.075	0.101	0.146	0.068	0.087
	HEX	0.153	0.132	0.081	0.086	0.146	0.060	0.085
GATA3 (Th2 helper T cells)	**ProDICE**	0.205	0.177	0.170	0.180	0.202	0.076	0.285
	H&E Feature Ridge	0.210	0.206	0.086	0.097	0.191	0.120	0.046
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.195	0.149	0.000
	ROSIE	0.169	0.179	0.111	0.115	0.197	0.108	0.141
	HEX	0.172	0.196	0.105	0.105	0.198	0.100	0.122

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
Granzyme B (cytotoxicity)	**ProDICE**	0.264	0.282	0.188	0.194	0.127	0.033	0.177
	H&E Feature Ridge	0.106	0.165	0.024	0.074	0.129	0.058	0.014
	Mean Prior Floor	N/A	N/A	0.000	0.053	0.130	0.064	0.000
	ROSIE	-0.051	-0.139	-0.026	0.033	0.137	0.066	0.081
	HEX	-0.010	-0.047	-0.003	0.041	0.132	0.045	0.020
HLA-DR (MHC-I)	**ProDICE**	0.224	0.272	0.166	0.185	0.132	0.036	0.197
	H&E Feature Ridge	0.074	0.069	0.015	0.065	0.132	0.064	0.010
	Mean Prior Floor	N/A	N/A	0.000	0.050	0.132	0.069	0.000
	ROSIE	-0.033	-0.126	-0.011	0.043	0.135	0.061	0.030
	HEX	-0.050	-0.089	-0.012	0.037	0.135	0.051	0.015
Hep Par 1 (liver)	**ProDICE**	0.213	0.206	0.171	0.183	0.174	0.061	0.253
	H&E Feature Ridge	0.049	-0.064	0.010	0.044	0.170	0.110	0.011
	Mean Prior Floor	N/A	N/A	0.000	0.032	0.171	0.118	0.000
	ROSIE	0.035	-0.098	0.022	0.049	0.179	0.095	0.122
	HEX	0.032	-0.040	0.014	0.028	0.177	0.085	0.055
Hyaluronan (extracellular matrix)	**ProDICE**	0.251	0.300	0.219	0.244	0.210	0.066	0.346
	H&E Feature Ridge	0.099	0.154	0.035	0.081	0.205	0.130	0.034
	Mean Prior Floor	N/A	N/A	0.000	0.029	0.205	0.156	0.000
	ROSIE	-0.140	-0.139	-0.091	-0.045	0.238	0.107	0.162
	HEX	-0.139	-0.119	-0.077	-0.048	0.246	0.102	0.135
IDO-1 (metabolism)	**ProDICE**	0.352	0.378	0.317	0.285	0.215	0.068	0.398
	H&E Feature Ridge	0.386	0.399	0.266	0.234	0.203	0.112	0.159
	Mean Prior Floor	N/A	N/A	0.000	0.020	0.220	0.174	0.000
	ROSIE	0.356	0.375	0.290	0.253	0.209	0.107	0.269
	HEX	0.376	0.406	0.304	0.250	0.210	0.089	0.282
IRF4 (plasma cells)	**ProDICE**	0.342	0.379	0.301	0.275	0.198	0.065	0.357
	H&E Feature Ridge	0.356	0.380	0.232	0.214	0.190	0.105	0.137
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.203	0.158	0.000
	ROSIE	0.340	0.374	0.258	0.232	0.194	0.101	0.215
	HEX	0.364	0.399	0.260	0.212	0.197	0.082	0.208

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
Ki67 (proliferation)	**ProDICE**	0.096	0.082	0.078	0.108	0.146	0.042	0.272
	H&E Feature Ridge	0.089	0.118	0.021	0.069	0.134	0.060	0.014
	Mean Prior Floor	N/A	N/A	0.000	0.053	0.135	0.068	0.000
	ROSIE	0.028	0.068	0.007	0.065	0.136	0.053	0.016
	HEX	0.003	0.039	0.001	0.061	0.137	0.053	0.019
MMP9 (matrix metalloproteinase)	**ProDICE**	0.294	0.322	0.228	0.235	0.132	0.033	0.228
	H&E Feature Ridge	0.118	0.182	0.030	0.075	0.135	0.064	0.017
	Mean Prior Floor	N/A	N/A	0.000	0.049	0.136	0.071	0.000
	ROSIE	-0.057	-0.140	-0.026	0.031	0.142	0.067	0.057
	HEX	-0.065	-0.111	-0.014	0.032	0.140	0.049	0.012
MUC-1 (epithelia)	**ProDICE**	0.175	0.215	0.119	0.139	0.131	0.036	0.154
	H&E Feature Ridge	0.072	0.067	0.013	0.064	0.130	0.062	0.008
	Mean Prior Floor	N/A	N/A	0.000	0.051	0.130	0.067	0.000
	ROSIE	-0.015	-0.081	-0.004	0.049	0.132	0.054	0.021
	HEX	-0.060	-0.090	-0.013	0.036	0.133	0.050	0.012
Melan-A (melanocytes)	**ProDICE**	0.240	0.220	0.203	0.194	0.197	0.065	0.303
	H&E Feature Ridge	0.232	0.133	0.118	0.122	0.188	0.112	0.074
	Mean Prior Floor	N/A	N/A	0.000	0.026	0.194	0.142	0.000
	ROSIE	0.220	0.129	0.151	0.142	0.192	0.096	0.159
	HEX	0.247	0.207	0.160	0.132	0.194	0.079	0.151
Na-K-ATPase (membranes)	**ProDICE**	0.168	0.202	0.123	0.148	0.149	0.046	0.191
	H&E Feature Ridge	0.037	-0.042	0.006	0.049	0.146	0.080	0.006
	Mean Prior Floor	N/A	N/A	0.000	0.042	0.146	0.084	0.000
	ROSIE	0.024	-0.075	0.012	0.048	0.150	0.072	0.071
	HEX	0.024	-0.025	0.008	0.037	0.148	0.061	0.029
OX-40 (activation)	**ProDICE**	0.605	0.624	0.575	0.527	0.164	0.044	0.522
	H&E Feature Ridge	0.450	0.522	0.336	0.260	0.182	0.095	0.201
	Mean Prior Floor	N/A	N/A	0.000	0.023	0.204	0.159	0.000
	ROSIE	0.453	0.538	0.350	0.272	0.182	0.099	0.225
	HEX	0.463	0.568	0.355	0.253	0.182	0.086	0.227

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR (≈ 1)
PAX5 (B cells)	**ProDICE**	0.564	0.634	0.538	0.485	0.188	0.048	0.540
	H&E Feature Ridge	0.511	0.563	0.420	0.366	0.192	0.096	0.274
	Mean Prior Floor	N/A	N/A	0.000	0.020	0.224	0.180	0.000
	ROSIE	0.481	0.530	0.425	0.374	0.198	0.102	0.362
	HEX	0.508	0.570	0.409	0.347	0.203	0.083	0.316
PD-1 (checkpoint)	**ProDICE**	0.269	0.298	0.234	0.219	0.208	0.067	0.342
	H&E Feature Ridge	0.294	0.314	0.169	0.167	0.196	0.115	0.099
	Mean Prior Floor	N/A	N/A	0.000	0.024	0.205	0.158	0.000
	ROSIE	0.280	0.302	0.229	0.213	0.203	0.103	0.268
	HEX	0.302	0.332	0.210	0.191	0.203	0.088	0.189
PD-L1 (checkpoint)	**ProDICE**	0.182	0.173	0.147	0.150	0.173	0.058	0.255
	H&E Feature Ridge	0.094	-0.021	0.026	0.055	0.167	0.104	0.020
	Mean Prior Floor	N/A	N/A	0.000	0.034	0.167	0.117	0.000
	ROSIE	0.081	-0.068	0.052	0.075	0.173	0.092	0.133
	HEX	0.104	0.035	0.051	0.057	0.172	0.079	0.076
Pan-Cytokeratin (epithelia)	**ProDICE**	0.194	0.204	0.122	0.151	0.119	0.029	0.126
	H&E Feature Ridge	0.109	0.123	0.024	0.082	0.119	0.044	0.012
	Mean Prior Floor	N/A	N/A	0.000	0.060	0.119	0.049	0.000
	ROSIE	0.003	-0.042	0.000	0.056	0.120	0.037	0.007
	HEX	-0.075	-0.096	-0.011	0.045	0.121	0.035	0.005
Podoplanin (lymphatics)	**ProDICE**	0.106	0.115	0.079	0.083	0.199	0.078	0.205
	H&E Feature Ridge	0.142	0.157	0.044	0.061	0.187	0.118	0.025
	Mean Prior Floor	N/A	N/A	0.000	0.027	0.189	0.138	0.000
	ROSIE	0.131	0.155	0.058	0.071	0.192	0.100	0.059
	HEX	0.135	0.175	0.068	0.071	0.192	0.094	0.078
SMA (smooth muscle cells)	**ProDICE**	0.087	0.140	0.057	0.091	0.128	0.037	0.140
	H&E Feature Ridge	0.010	-0.088	0.001	0.060	0.123	0.056	0.004
	Mean Prior Floor	N/A	N/A	0.000	0.057	0.123	0.060	0.000
	ROSIE	0.028	-0.067	0.008	0.063	0.124	0.052	0.020
	HEX	-0.007	-0.074	-0.002	0.058	0.125	0.046	0.016

Table 9 (Multi-Tumor, continued).

Marker	Method	Pearson ↑	Spearman ↑	Lin's CCC ↑	SSIM ↑	RMSE ↓	W_1 ↓	VR ($\approx$ 1)
Synaptophysin (neuroendocrine)	**ProDICE**	0.160	0.233	0.107	0.128	0.146	0.047	0.148
	H&E Feature Ridge	0.057	0.122	0.012	0.057	0.144	0.076	0.011
	Mean Prior Floor	N/A	N/A	0.000	0.043	0.144	0.081	0.000
	ROSIE	-0.062	-0.167	-0.032	0.016	0.153	0.080	0.080
	HEX	-0.033	-0.054	-0.008	0.028	0.148	0.058	0.018
VISTA (activation)	**ProDICE**	0.242	0.287	0.203	0.200	0.174	0.056	0.293
	H&E Feature Ridge	0.270	0.322	0.146	0.155	0.165	0.091	0.086
	Mean Prior Floor	N/A	N/A	0.000	0.033	0.172	0.124	0.000
	ROSIE	0.246	0.316	0.176	0.184	0.169	0.084	0.177
	HEX	0.269	0.341	0.181	0.173	0.168	0.075	0.155
Vimentin (cytoplasm)	**ProDICE**	0.129	0.118	0.098	0.117	0.146	0.045	0.211
	H&E Feature Ridge	0.107	-0.032	0.027	0.065	0.139	0.076	0.017
	Mean Prior Floor	N/A	N/A	0.000	0.044	0.140	0.085	0.000
	ROSIE	0.091	-0.048	0.054	0.083	0.145	0.078	0.114
	HEX	0.084	0.022	0.033	0.068	0.142	0.062	0.043
p53 (tumor suppressor)	**ProDICE**	0.221	0.246	0.188	0.171	0.174	0.052	0.308
	H&E Feature Ridge	0.215	0.171	0.110	0.126	0.165	0.090	0.076
	Mean Prior Floor	N/A	N/A	0.000	0.034	0.170	0.117	0.000
	ROSIE	0.163	0.094	0.116	0.140	0.172	0.080	0.175
	HEX	0.254	0.260	0.169	0.140	0.166	0.072	0.150

Table 10: **CRC ProDICE leave-one-subject-out sensitivity.** Each row omits all four regions of one held-out test subject; "None" is the full 28-region test set of Table 1.

Omitted	Pearson	Spearman	SSIM	RMSE	W_1	VR
None	0.223	0.248	0.196	0.155	0.041	0.307
CRC-05	0.222	0.244	0.195	0.158	0.042	0.304
CRC-11	0.232	0.259	0.205	0.156	0.041	0.304
CRC-16	0.222	0.248	0.193	0.156	0.041	0.307
CRC-18	0.215	0.239	0.189	0.156	0.041	0.308
CRC-19	0.224	0.249	0.197	0.155	0.040	0.307
CRC-29	0.225	0.254	0.201	0.152	0.039	0.319
CRC-31	0.218	0.239	0.192	0.156	0.042	0.304

Table 11: **Multi-Tumor outer-fold heterogeneity.** Mean $\pm$ sample SD across the five outer-test folds; brackets give the fold minimum and maximum.

Method	Pearson	Spearman	SSIM	RMSE	W_1	VR
ProDICE	0.230±0.022 [0.201,0.251]	0.246±0.020 [0.220,0.265]	0.192±0.011 [0.178,0.206]	0.164±0.003 [0.158,0.166]	0.052±0.002 [0.050,0.055]	0.270±0.030 [0.230,0.302]
H&E Feature Ridge	0.182±0.014 [0.167,0.201]	0.164±0.027 [0.141,0.204]	0.106±0.005 [0.099,0.113]	0.158±0.005 [0.150,0.163]	0.088±0.002 [0.085,0.091]	0.050±0.006 [0.041,0.057]
Mean Prior Floor	N/A	N/A	0.039±0.002 [0.037,0.042]	0.163±0.006 [0.154,0.168]	0.110±0.004 [0.105,0.113]	0.000±0.000
ROSIE	0.097±0.029 [0.055,0.128]	0.119±0.037 [0.077,0.158]	0.042±0.006 [0.034,0.047]	0.167±0.008 [0.157,0.175]	0.119±0.013 [0.106,0.136]	0.002±0.001 [0.001,0.003]
HEX	-0.033±0.030 [-0.061,0.018]	0.009±0.042 [-0.016,0.084]	0.008±0.015 [-0.004,0.027]	0.187±0.012 [0.175,0.200]	0.092±0.006 [0.088,0.102]	0.155±0.047 [0.103,0.222]