MoSPR: Histology-to-Gene Expression Prediction with Morpho-Spatial Macrostates and Low-Rank Molecular Programs

Anonymous authors
Paper under double-blind review

Abstract

Predicting molecular profiles from histopathology remains challenging because whole-slide images contain spatially organized, heterogeneous tissue patterns, while gene expression comprises thousands of correlated targets. We introduce MoSPR (Morpho-Spatial Program Regression), a linear framework that couples an adjacency-informed histology representation with a low-rank molecular basis. MoSPR clusters frozen patch embeddings into morphology microstates, aggregates their spatial adjacencies across the training cohort, and groups microstates with similar adjacency patterns into shared macrostates. Each slide is then represented by global morphology and macrostate-specific deviations, which are linearly mapped to coefficients of a training-derived low-rank gene-expression basis. Across three cancer cohorts from The Cancer Genome Atlas, MoSPR achieves the highest mean gene-expression prediction scores among all evaluated methods. Without pathway-level supervision, pathway scores derived from its predicted expression profiles rank first in eight of nine comparisons across three pathway collections. Ablation studies on the breast cancer cohort show complementary gains from adjacency-derived macrostate representation and low-rank molecular prediction. Moreover, with half of the training data on this cohort, MoSPR exceeds the full-data gene-prediction score of the strongest competing baseline. Finally, its linear formulation enables exact decomposition of each predicted expression profile into global and macrostate-specific molecular contributions, providing an interpretable link between spatially coherent macrostate regions and their associated molecular programs. Our code is available at `https://anonymous.4open.science/r/MoSPR-C957/README.md`.

1 Introduction

Molecular profiling is central to cancer diagnosis, biological stratification, and treatment selection, but it remains substantially more expensive and less routinely available than hematoxylin-and-eosin (H&E) histology. This has motivated methods that infer transcriptomic and other molecular measurements directly from whole-slide images (WSIs) (Schmauch et al., 2020; Alsaafin et al., 2023; Şenbabaoğlu et al., 2024; Pizurica et al., 2024; Nishimura et al., 2026). A successful model could provide a low-cost virtual molecular screen, prioritize samples for confirmatory assays, and expose image–molecule associations (Coudray et al., 2018; Saillard et al., 2023; Schmauch et al., 2020).

Despite rapid progress, histology-to-molecular prediction remains challenging because WSIs contain heterogeneous tissue patterns, while bulk gene expression comprises thousands of correlated targets. Global pooling and standard multiple-instance learning (MIL) summarize overall patch-level morphology but do not explicitly encode the spatial adjacency of recurring tissue states. Transformer- and graph-based methods (Shao et al., 2021; Chen et al., 2021) can model contextual or spatial interactions, but operate over patient-specific patch sets or graphs. This motivates a more compact representation that captures recurrent spatial organization at the cohort level.

We propose **MoSPR** (**Mo**rpho-**S**patial **P**rogram **R**egression), a linear framework that compresses whole-slide histology into a compact *morpho-spatial* representation, defined here by the spatial adjacency of recurring morphological states. To construct this representation, MoSPR first clusters

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.

frozen patch embeddings into morphology microstates. For each patient, spatial co-occurrences between neighboring patches are counted to construct a microstate adjacency matrix (Fig. 1(a)), which is normalized and aggregated across the training cohort to form a global microstate adjacency matrix. After row-normalizing this global matrix, eigendecomposition embeds microstates according to their adjacency patterns, and clustering these embeddings defines a shared set of macrostates (Fig. 1(b)). Each slide is then represented by a global morphology vector M and abundance-weighted macrostate deviations S, yielding $X = [M \mid S]$ (Fig. 1(c)).

To reduce the dimensionality of molecular prediction, MoSPR learns a low-rank expression basis $U_q \in \mathbb{R}^{q \times G}$ from the training gene-expression targets ($Y_z \in \mathbb{R}^{N \times G}$) and a regression matrix W_q mapping X to the corresponding molecular coefficients:

$$\widehat{Y}_z = XW_qU_q, \qquad X = [M \mid S], \tag{1}$$

where $q \ll G$. Thus, MoSPR combines morphology aggregation within adjacency-informed states with regression to low-dimensional molecular coefficients rather than directly to thousands of genes.

Across three cohorts from The Cancer Genome Atlas (TCGA; Weinstein et al., 2013)—breast invasive carcinoma (BRCA), kidney renal clear cell carcinoma (KIRC), and lung adenocarcinoma (LUAD)—MoSPR achieves the highest mean gene-expression prediction scores among all evaluated methods. Pathway scores derived from its predicted expression profiles rank first in eight of nine comparisons across Hallmark (Liberzon et al., 2015), Gene Ontology Biological Process (GO-BP) (The Gene Ontology Consortium, 2021), and KEGG collections (Kanehisa & Goto, 2000), without pathway-level supervision. Controlled ablations on BRCA show complementary improvements from the macrostate representation and low-rank output model. On the same cohort, MoSPR trained with half of the data exceeds the full-data gene-prediction score of the strongest competing baseline. Finally, the linear formulation separates the predicted molecular profile into global morphology and macrostate-specific contributions, supporting qualitative interpretation of macrostate-associated molecular programs.

Our contributions are:

- a cohort-level morphology dictionary learned from patch adjacency, yielding shared macrostates and a compact WSI representation without spatial-omics supervision or large-scale patient-specific graph construction.
- a linear framework coupling this representation with a training-derived low-rank expression basis, enabling exact decomposition of each predicted molecular profile into global and macrostate-specific molecular contributions.
- a systematic evaluation across three cancer cohorts, including gene and pathway reconstruction, component ablations, training-set scaling, and qualitative macrostate interpretation.

2 Related Work

Whole-slide gene-expression prediction from H&E images has progressed from patch-level aggregation (Schmauch et al., 2020) to attention-, transformer-, and clustering-based approaches (Graziani et al., 2022; Alsaafin et al., 2023; Şenbabaoğlu et al., 2024; Pizurica et al., 2024). General WSI aggregation methods further include attention-based MIL (Ilse et al., 2018), low-rank attention (Xiang & Zhang, 2023), structured state-space models (Fillioux et al., 2023), and Mamba-based sequence models (Yang et al., 2024; Zhang et al., 2025). CPNN (Nishimura et al., 2026) additionally incorporates molecular structure through single-cell-derived prototypes. MoSPR uses training-cohort patch adjacency to learn shared morphology states and couples state-wise aggregation with low-rank expression prediction.

Unlike spatial transcriptomics, which directly measures spatially resolved molecular variation (Ståhl et al., 2016), MoSPR uses bulk molecular supervision and does not reconstruct spatial expression maps. WSI methods such as TransMIL (Shao et al., 2021) and Patch-GCN (Chen et al., 2021) model contextual or spatial dependencies over patient-specific patch sequences or graphs, whereas MoSPR constructs a cohort-level adjacency structure over shared morphology states and derives a compact macrostate representation.

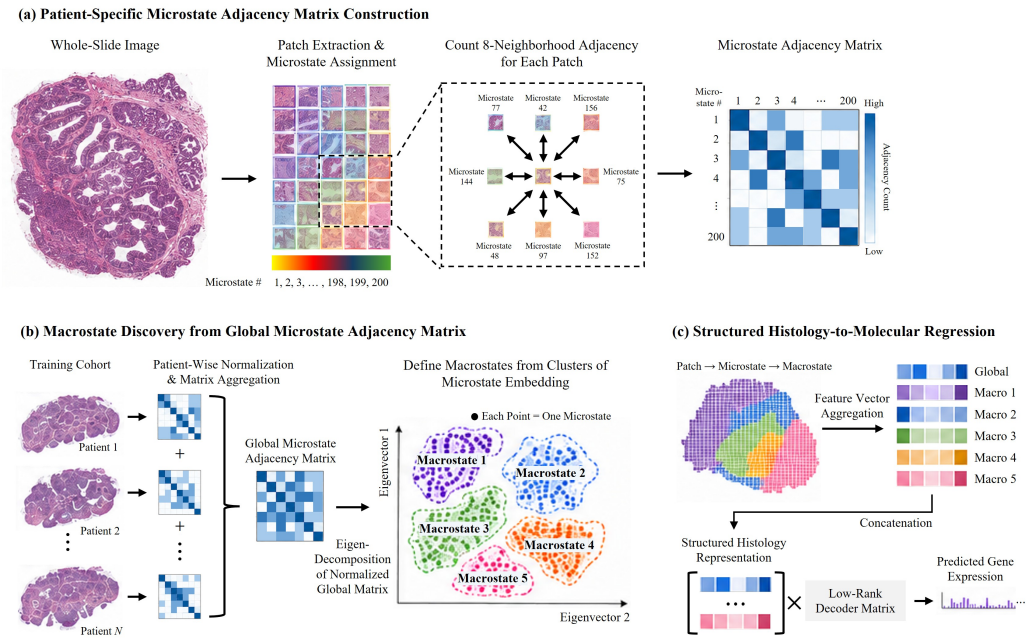


Figure 1: Overview of MoSPR.

3 METHODS

3.1 PATCH FEATURE EXTRACTION

For patient n, let $\mathcal{X}^{(n)} = \{x_i^{(n)}\}_{i=1}^{N_n}$ denote the tissue patches and let $Y^{(n)} \in \mathbb{R}^G$ denote the paired gene-expression vector. A frozen pathology encoder (f_{enc}) maps each patch to

$$z_i^{(n)} = f_{\text{enc}}(x_i^{(n)}) \in \mathbb{R}^{D_0}, \tag{2}$$

where $D_0 = 512$ for CONCH (Lu et al., 2024).

3.2 MICROSTATE-TO-MACROSTATE CONSTRUCTION

Microstate assignment and adjacency matrix construction. To discretize the continuous patch-embedding space into recurrent morphological patterns, we apply k-means clustering with $J = 200$ clusters to a patient-balanced sample of training patch embeddings. We define each resulting cluster as a *microstate*, represented by its centroid

$$\{\mu_j\}_{j=1}^{J}, \qquad \mu_j \in \mathbb{R}^{D_0}.$$

Each tissue patch is then assigned to the nearest microstate,

$$a_i^{(n)} = \arg\min_{j \in \{1,\dots,J\}} \left\| z_i^{(n)} - \mu_j \right\|_2^2, \tag{3}$$

where $a_i^{(n)} \in \{1, \dots, J\}$ denotes the microstate assigned to patch i of patient n.

We then characterize how these microstates are spatially arranged within each WSI using 8-neighbor patch adjacency (Fig. 1(a)). For each pair of neighboring tissue patches (p, q) assigned to microstates $a_p^{(n)} = u$ and $a_q^{(n)} = v$, the corresponding entries of the patient-level *microstate adjacency matrix* $C^{(n)} \in \mathbb{R}_+^{J \times J}$ are incremented:

$$C_{uv}^{(n)} \leftarrow C_{uv}^{(n)} + 1, \qquad C_{vu}^{(n)} \leftarrow C_{vu}^{(n)} + 1. \tag{4}$$

Thus, $C_{uv}^{(n)}$ quantifies how frequently patches belonging to microstates u and v occur as spatial neighbors within patient n, and $C^{(n)}$ is symmetric by construction.

To prevent patients with larger tissue areas from dominating the cohort-level spatial structure, each patient-level microstate adjacency matrix is normalized by its total adjacency mass:

$$\widehat{C}^{(n)} = \frac{C^{(n)}}{\sum_{u,v} C_{uv}^{(n)}}. \tag{5}$$

The global microstate adjacency matrix is then defined as the average of the normalized patient-level matrices over the training cohort $\mathcal{D}_{\mathrm{tr}}$ (Fig. 1(b)):

$$C_{\mathrm{global}} = \frac{1}{|\mathcal{D}_{\mathrm{tr}}|} \sum_{n \in \mathcal{D}_{\mathrm{tr}}} \widehat{C}^{(n)}. \tag{6}$$

This averaging gives equal weight to each training patient while preserving symmetry and unit total mass ($C_{\mathrm{global}} = C_{\mathrm{global}}^{\top}$ and $\sum_{u,v}(C_{\mathrm{global}})_{uv} = 1$).

Macrostate discovery. To define macrostates according to their spatial neighborhood patterns, we first row-normalize the global microstate adjacency matrix. Let

$$d_u = \sum_v (C_{\mathrm{global}})_{uv}, \qquad P = \Delta^{-1} C_{\mathrm{global}}, \qquad \Delta = \mathrm{diag}(d_1, \ldots, d_J). \tag{7}$$

Here, $P \in \mathbb{R}^{J \times J}$ and P_{uv} represents the relative frequency with which microstate v occurs adjacent to microstate u.

We then compute the eigendecomposition of P,

$$P\psi_\ell = \lambda_\ell \psi_\ell, \qquad \psi_\ell \in \mathbb{R}^J. \tag{8}$$

Following diffusion-map and spectral-clustering constructions (Coifman & Lafon, 2006; Ng et al., 2002), we exclude the trivial stationary component associated with $\lambda_0 = 1$ and retain the next $L = 20$ nontrivial components. Each microstate u is then represented by the eigenvalue-weighted spectral embedding

$$\phi_u = [\lambda_1 \psi_1(u), \ldots, \lambda_L \psi_L(u)] \in \mathbb{R}^L. \tag{9}$$

The eigenvalue weighting gives greater influence to dominant spectral components of the adjacency structure. Each embedding is subsequently normalized to unit length,

$$\bar{\phi}_u = \frac{\phi_u}{\|\phi_u\|_2 + \varepsilon}, \tag{10}$$

where $\varepsilon = 10^{-12}$ is a small numerical constant. Finally, k-means clustering is applied to the $J = 200$ normalized spectral embeddings, grouping the microstates into $K = 8$ macrostates. This clustering defines a fixed microstate-to-macrostate mapping

$$\rho : \{1, \ldots, J\} \to \{1, \ldots, K\}, \tag{11}$$

where $\rho(u) = k$ indicates that microstate u is assigned to macrostate k (Fig. 1(b)).

3.3 Morpho-spatial low-rank regression

Structured histology representation. We represent each patient's WSI by combining global morphology with macrostate-specific deviations that summarize within-slide morphological heterogeneity across the learned states (Fig. 1(c)). The global morphology vector is the mean of the original patch features,

$$M_n = \frac{1}{N_n} \sum_{i=1}^{N_n} z_i^{(n)} \in \mathbb{R}^{D_0}. \tag{12}$$

For macrostate k, let $\mathcal{I}_{nk} = \{i : \rho(a_i^{(n)}) = k\}$, $N_{nk} = |\mathcal{I}_{nk}|$, and $p_{nk} = N_{nk}/N_n$. When $N_{nk} > 0$, its mean patch feature and abundance-weighted deviation from the slide mean are

$$h_{nk} = \frac{1}{N_{nk}} \sum_{i \in \mathcal{I}_{nk}} z_i^{(n)}, \qquad s_{nk} = p_{nk}(h_{nk} - M_n) \in \mathbb{R}^{D_0}. \tag{13}$$

If macrostate k is absent from patient n ($N_{nk} = 0$), we set $s_{nk} = 0$. We concatenate the state blocks as $S_n = [s_{n1} \mid \cdots \mid s_{nK}] \in \mathbb{R}^{KD_0}$ and form

$$X_n = [M_n \mid S_n] \in \mathbb{R}^{(K+1)D_0}. \tag{14}$$

The resulting vector X_n serves as the structured histological representation of patient n. Stacking these representations across N patients yields the matrix $X \in \mathbb{R}^{N\times(K+1)D_0}$.

Low-rank molecular factorization. Gene targets are standardized per gene using training-fold statistics, yielding $Y_z \in \mathbb{R}^{N\times G}$ (Supplementary Section A.1). Rather than predicting all G genes independently, we exploit the strong correlation structure of gene expression by learning a compact molecular basis directly from the training cohort. PCA is fitted exclusively to the standardized training-fold expression matrix, and the top q principal directions define

$$U_q \in \mathbb{R}^{q\times G}. \tag{15}$$

Each molecular profile is represented by a low-dimensional coefficient vector, and the corresponding coefficient matrix is

$$A_q = Y_z U_q^\top \in \mathbb{R}^{N\times q}. \tag{16}$$

The resulting approximation

$$Y_z \approx A_q U_q \tag{17}$$

reduces the output regression problem from G genes to $q \ll G$ molecular coefficients.

Molecular regression and reconstruction. We fit a ridge regression matrix W_q that maps the structured histology representation $X = [M \mid S]$ to the low-dimensional molecular coefficient space:

$$W_q^* = \arg\min_{W_q} \|A_q - XW_q\|_F^2 + \lambda_{\text{ridge}}\|W_q\|_F^2, \tag{18}$$

where $W_q^* \in \mathbb{R}^{(K+1)D_0\times q}$ is the fitted regression matrix.

The predicted molecular coefficients are projected back through the learned basis to reconstruct the standardized gene-expression profile:

$$\widehat{Y}_z = XW_q^* U_q. \tag{19}$$

Intercept terms, which are included in the implementation, are omitted from the notation for clarity.

Thus, the original high-dimensional histology-to-gene regression problem is reduced to predicting only q molecular coefficients from the morpho-spatial representation.

3.4 Macrostate-level molecular interpretation

According to the block structure of $X = [M \mid S]$, the fitted regression matrix can be partitioned as

$$W_q^* = \begin{bmatrix} W_{q,\text{global}}^* \\ W_{q,1}^* \\ \vdots \\ W_{q,K}^* \end{bmatrix}, \qquad W_{q,\text{global}}^*,\, W_{q,k}^* \in \mathbb{R}^{D_0\times q}. \tag{20}$$

For patient n, the macrostate-k contribution to the predicted gene-expression profile is

$$\Gamma_{nk} = s_{nk} W_{q,k}^* U_q \in \mathbb{R}^G, \tag{21}$$

and the global morphology contribution is

$$\Gamma_{n,\text{global}} = M_n W_{q,\text{global}}^* U_q \in \mathbb{R}^G. \tag{22}$$

Ignoring the intercept term for notational simplicity, the predicted profile can therefore be decomposed as

$$\widehat{Y}_{z,n} = \Gamma_{n,\text{global}} + \sum_{k=1}^{K} \Gamma_{nk}. \tag{23}$$

This provides an exact algebraic decomposition of the prediction into contributions from global morphology and individual macrostates, without requiring a post-hoc attribution method.

4 EXPERIMENTAL SETUP

4.1 DATASETS AND EVALUATION PROTOCOL

We evaluate MoSPR on paired H&E WSIs and bulk RNA-sequencing profiles from TCGA-BRCA, TCGA-KIRC, and TCGA-LUAD (Weinstein et al., 2013). Performance is assessed at two levels: gene-expression prediction and pathway-score reconstruction. We use four-fold patient-level cross-validation, where held-out patients in each outer fold are reserved exclusively for final evaluation and model fitting and hyperparameter selection are performed using only the corresponding training and validation partitions. Within each cancer type, all competing methods use identical patient-level splits and the same target-gene universe. Cohort statistics and fold-specific training, validation, and test sample counts are provided in Supplementary Table S1.

4.2 EVALUATION METRICS

For gene-level evaluation, Pearson correlation coefficient (PCC) and Spearman rank correlation coefficient (SCC) are computed across held-out patients separately for each gene and then averaged across genes. For pathway-level evaluation, pathway scores are computed independently from the predicted and observed gene-expression profiles as the mean gene-wise z-score across the member genes of each pathway (Lee et al., 2008). SCC is then computed across held-out patients separately for each pathway. Performance within each pathway collection is summarized by the median SCC across pathways in the Hallmark (Liberzon et al., 2015), GO-BP (The Gene Ontology Consortium, 2021), and KEGG (Kanehisa & Goto, 2000) collections.

4.3 BENCHMARK METHODS

We compare MoSPR with both general-purpose WSI aggregation methods and models developed specifically for transcriptomic prediction from histology. WSI aggregation baselines include max and mean pooling (Wang et al., 2018), ABMIL (Ilse et al., 2018), ILRA (Xiang & Zhang, 2023), S4MIL (Fillioux et al., 2023), MambaMIL and SRMambaMIL (Yang et al., 2024), and 2DMamba (Zhang et al., 2025). Transcriptomic prediction baselines include HE2RNA (Schmauch et al., 2020), AbReg (Graziani et al., 2022), tRNAformer (Alsaafin et al., 2023), MOSBY (Şenbabaoğlu et al., 2024), SEQUOIA VIS (Pizurica et al., 2024), and CPNN (Nishimura et al., 2026).

4.4 IMPLEMENTATION DETAILS

Structural hyperparameters are fixed across all cohorts at $J = 200$ microstates, $L = 20$ spectral components, and $K = 8$ macrostates. The molecular rank q and ridge penalty λ_{ridge} are selected independently within each fold using the validation reconstruction error. Hyperparameter ranges and selection details are provided in Supplementary Section A.2.

4.5 ADDITIONAL ANALYSES

Ablation analysis. We isolate the contributions of the macrostate representation and low-rank molecular prediction using matched model variants on TCGA-BRCA. *Spatial Low-Rank* corresponds to the full MoSPR formulation, combining macrostate-based features with low-rank regression. *Spatial Direct* retains the same macrostate representation but replaces the low-rank output model with direct gene-wise regression, whereas *Global Low-Rank* retains low-rank molecular prediction but removes the macrostate-specific features. As a control, *Shuffled-State Low-Rank* and *Shuffled-State Direct* permute the learned microstate-to-macrostate mapping ρ across microstates, preserving the number and sizes of macrostates while disrupting their adjacency-derived composition. All variants use the same patient-level four-fold cross-validation splits as the main benchmarks (Supplementary Table S1). Performance is evaluated using mean gene-level SCC and median Hallmark pathway SCC.

Data-efficiency analysis. To assess performance under limited paired histology–transcriptomic supervision, we conduct a training-set scaling analysis across the three cancer cohorts using the same patient-level four-fold splits as the main benchmarks (Supplementary Table S1). Within

each outer fold, the training and validation partitions are randomly subsampled at fractions $f \in \{0.1, 0.25, 0.5, 0.75, 1.0\}$, while the held-out test partition remains unchanged. For each fold and fraction, the sampled training and validation patients are fixed and shared across all competing methods. All data-derived representations and prediction models are recomputed using only the corresponding subsample. Performance is evaluated using mean gene-level SCC and median Hallmark pathway SCC on the unchanged test folds.

Macrostate interpretation. To qualitatively illustrate how learned macrostates map to tissue regions and associate with molecular programs, we select a representative held-out patient WSI from TCGA-BRCA and analyze it using the model trained in a single cross-validation fold. We map the patch assignments of three selected macrostates onto this WSI. For each macrostate, its model-derived gene-expression contribution (Γ_{nk} in Section 3.4) is used for Hallmark pathway enrichment analysis. The three pathways with the largest absolute enrichment z-scores are reported for each macrostate.

5 RESULTS

5.1 MOSPR CONSISTENTLY IMPROVES CROSS-CANCER GENE-EXPRESSION PREDICTION

Table 1 summarizes gene-level prediction performance. MoSPR achieves the highest mean PCC and SCC in all six cohort–metric combinations across BRCA, KIRC, and LUAD. Although the strongest competing baseline varies across cohorts and metrics, MoSPR consistently exceeds the strongest competing baseline in each setting, with PCC/SCC gains of $0.043/0.048$ in BRCA, $0.022/0.029$ in KIRC, and $0.039/0.040$ in LUAD. The consistent gains across both linear and rank correlation indicate that the improvement is not specific to a single correlation metric or cancer type. The best- and worst-predicted individual genes are reported in Supplementary Table S2. In addition, MoSPR achieves these gains with the lowest parameter count (0.96M parameters) among the evaluated methods, while requiring only 2.02 inference GFLOPs after frozen CONCH feature extraction (Supplementary Table S3).

Table 1: Cross-cancer gene-expression prediction. Entries report mean gene-wise PCC and SCC over four cross-validation folds. Best and second-best results are shown in bold and underlined, respectively.

	BRCA		KIRC		LUAD	
Method	PCC	SCC	PCC	SCC	PCC	SCC
Max (Wang et al., 2018)	0.294	0.283	0.236	0.244	0.258	0.278
Mean (Wang et al., 2018)	0.336	0.338	0.285	0.293	0.292	0.315
ABMIL (Ilse et al., 2018)	<u>0.370</u>	<u>0.363</u>	0.293	0.299	0.308	0.324
HE2RNA (Schmauch et al., 2020)	0.300	0.293	0.256	0.273	0.276	0.297
AbReg (Graziani et al., 2022)	0.355	0.349	0.289	0.293	0.296	0.311
tRNAformer (Alsaafin et al., 2023)	0.341	0.332	<u>0.312</u>	<u>0.319</u>	0.316	<u>0.336</u>
ILRA (Xiang & Zhang, 2023)	0.238	0.271	0.191	0.248	0.280	0.304
S4MIL (Fillioux et al., 2023)	0.328	0.330	0.269	0.274	0.299	0.315
MambaMIL (Yang et al., 2024)	0.363	0.356	0.290	0.290	<u>0.319</u>	0.335
SRMambaMIL (Yang et al., 2024)	0.361	0.352	0.292	0.294	0.311	0.326
MOSBY (Şenbabaoğlu et al., 2024)	0.343	0.349	0.264	0.289	0.289	0.308
SEQUOIA VIS (Pizurica et al., 2024)	0.353	0.337	0.310	0.314	0.317	0.331
2DMamba (Zhang et al., 2025)	0.352	0.348	0.287	0.292	0.300	0.313
CPNN (Nishimura et al., 2026)	0.360	0.356	0.293	0.310	0.305	0.325
MoSPR (ours)	**0.413**	**0.411**	**0.334**	**0.348**	**0.358**	**0.376**

5.2 GENE PREDICTIONS PRESERVE PATHWAY-LEVEL FUNCTIONAL VARIATION

MoSPR achieves the highest median pathway-wise SCC in eight of nine comparisons across the three cohorts and the Hallmark, GO-BP, and KEGG collections, and ranks second in the remaining comparison (Table 2). Relative to the strongest competing method in each setting, MoSPR improves median pathway-wise SCC by up to 0.051, with gains across all three pathway collections in BRCA and KIRC. Notably, pathway scores are derived solely from predicted gene-expression profiles and are not used for training. These results indicate that improvements in gene-expression prediction extend to coordinated variation across biological programs.

Table 2: Gene-derived pathway reconstruction across three TCGA cohorts. Entries report the median pathway-wise SCC within each collection, averaged over four cross-validation folds. Pathway scores are derived solely from predicted gene-expression profiles and are not used for training. Best and second-best results are shown in bold and underlined, respectively.

	BRCA			KIRC			LUAD		
Method	Hallmark	GO-BP	KEGG	Hallmark	GO-BP	KEGG	Hallmark	GO-BP	KEGG
Max (Wang et al., 2018)	0.356	0.337	0.321	0.281	0.284	0.309	0.420	0.379	0.388
Mean (Wang et al., 2018)	0.469	0.448	0.415	0.367	0.379	0.415	0.498	0.446	0.452
ABMIL (Ilse et al., 2018)	<u>0.500</u>	<u>0.476</u>	<u>0.446</u>	0.378	0.388	0.418	0.519	0.465	0.474
HE2RNA (Schmauch et al., 2020)	0.333	0.329	0.300	0.299	0.321	0.340	0.388	0.349	0.365
AbReg (Graziani et al., 2022)	0.481	0.453	0.425	<u>0.397</u>	<u>0.389</u>	<u>0.420</u>	0.507	0.443	0.462
tRNAformer (Alsaafin et al., 2023)	0.412	0.381	0.345	0.367	0.368	0.404	0.467	0.415	0.424
ILRA (Xiang & Zhang, 2023)	0.385	0.358	0.330	0.301	0.317	0.340	0.463	0.418	0.430
S4MIL (Fillioux et al., 2023)	0.434	0.414	0.383	0.301	0.329	0.343	0.481	0.428	0.436
MambaMIL (Yang et al., 2024)	0.479	0.458	0.423	0.308	0.326	0.345	**0.529**	<u>0.481</u>	<u>0.487</u>
SRMambaMIL (Yang et al., 2024)	0.463	0.448	0.411	0.306	0.315	0.338	0.514	0.455	0.472
MOSBY (Şenbabaoğlu et al., 2024)	0.400	0.381	0.354	0.257	0.263	0.293	0.348	0.312	0.324
SEQUOIA VIS (Pizurica et al., 2024)	0.444	0.413	0.385	0.384	0.388	0.420	0.488	0.441	0.450
2DMamba (Zhang et al., 2025)	0.460	0.439	0.408	0.381	0.370	0.388	0.501	0.443	0.457
CPNN (Nishimura et al., 2026)	0.391	0.377	0.327	0.282	0.293	0.349	0.427	0.380	0.388
MoSPR (ours)	**0.551**	**0.520**	**0.486**	**0.402**	**0.415**	**0.447**	<u>0.528</u>	**0.482**	**0.491**

5.3 Ablation of spatial and low-rank structure

On TCGA-BRCA, Global Low-Rank improves mean gene SCC from 0.348 (Global Direct) to 0.380, while Spatial Direct reaches 0.395 (Table 3). Combining the macrostate representation with low-rank regression (Spatial Low-Rank; MoSPR) yields the highest mean gene SCC (0.411). Importantly, applying baseline-specific low-rank output controls to ABMIL and tRNAformer does not reproduce MoSPR's performance (Supplementary Table S4), suggesting that low-rank output compression alone is insufficient to explain the gain. Randomizing the microstate-to-macrostate assignments reduces performance to 0.360 for Shuffled-State Direct and 0.379 for Shuffled-State Low-Rank, both below their learned-macrostate counterparts (0.395 and 0.411, respectively). For Hallmark reconstruction, the shuffled-state variants both reach 0.499, whereas Spatial Direct (0.552) and MoSPR (0.551) perform similarly and both exceed Global Low-Rank (0.481).

Beyond predictive performance, the learned macrostate organization was also stable under patient resampling. Across 300 patient-level bootstrap resamples per cohort with fixed microstates, mean Adjusted Rand Index (ARI) values relative to the corresponding full-cohort partitions were 0.769, 0.653, and 0.786 for TCGA-BRCA, TCGA-KIRC, and TCGA-LUAD, respectively (Supplementary Table S5). Together, these results indicate complementary contributions of the macrostate representation and low-rank output model to gene prediction, with pathway-level gains primarily associated with the learned macrostate representation.

Table 3: Component ablation on TCGA-BRCA. Best and second-best results are shown in bold and underlined, respectively.

Variant	Histology Input	Molecular Output	Gene SCC	Hallmark SCC
Global Direct	M	Direct	0.348	0.458
Global Low-Rank	M	U_q	0.380	0.481
Shuffled-State Direct	$[M \mid S_{\text{shuffle}}]$	Direct	0.360	0.499
Shuffled-State Low-Rank	$[M \mid S_{\text{shuffle}}]$	U_q	0.379	0.499
Spatial Direct	$[M \mid S]$	Direct	<u>0.395</u>	**0.552**
Spatial Low-Rank (MoSPR)	$[M \mid S]$	U_q	**0.411**	<u>0.551</u>

5.4 Data efficiency under limited paired supervision

Across reduced training-set sizes, MoSPR consistently preserves its performance advantage, including against ABMIL, the strongest competing baseline on TCGA-BRCA (Fig. 2). For gene-expression prediction, MoSPR reaches the full-data ABMIL performance using only approximately 44% of the available training data. Similarly, for Hallmark pathway reconstruction, MoSPR trained with approximately 48% of the data matches the performance of ABMIL trained with 100%. These results indicate that MoSPR can recover the performance of the strongest full-data baseline with roughly

half of the paired histology–transcriptomic supervision. Corresponding training-set scaling results for TCGA-KIRC and TCGA-LUAD are provided in Supplementary Figs. S1 and S2.

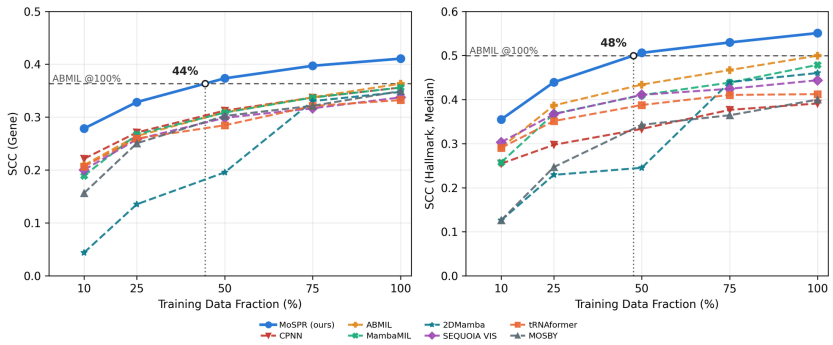


Figure 2: Training-set scaling on TCGA-BRCA.

5.5 Qualitative characterization of morpho-spatial macrostates

Three representative macrostates localize to distinct, spatially coherent regions in a representative WSI (Figure 3). Their corresponding gene-expression contributions show distinct Hallmark enrichment patterns, including oxidative phosphorylation, epithelial–mesenchymal transition, and interferon gamma response. This qualitative example suggests that the learned macrostates capture morphologically distinct tissue regions associated with different molecular programs.

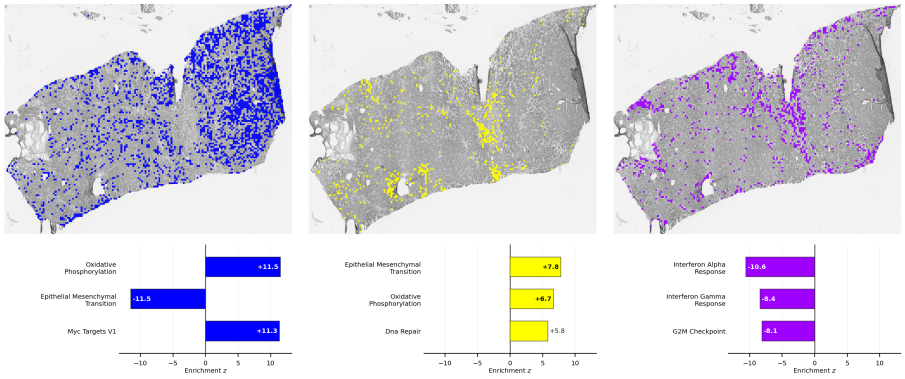


Figure 3: Representative macrostates and their associated Hallmark pathway enrichments.

6 Discussion and Conclusion

MoSPR couples a shared morphology dictionary informed by training-cohort adjacency with low-rank gene-expression regression, summarizing within-slide morphological heterogeneity across learned macrostates while reducing the dimension of supervised regression. The gene- and pathway-level results, together with the ablation and training-set scaling analyses, support this combination for bulk expression prediction. The fitted linear formulation further permits model-based interpretation of global and macrostate-specific molecular contributions.

Several limitations remain. Macrostates summarize recurrent local adjacency patterns rather than complete two-dimensional tissue geometry, and bulk RNA measurements cannot validate whether macrostate-associated molecular programs are spatially localized to the corresponding tissue regions. Accordingly, the biological characterization remains qualitative. Evaluation within TCGA does not establish generalization to independent cohorts or institutions, and external cohorts and spatially resolved molecular measurements will be important for validating both predictive robustness and the biological interpretation of the learned macrostates.

AI USE STATEMENT

Generative AI tools were used for translation, LaTeX formatting and conversion of manuscript elements such as tables and figures, grammatical proofreading, and code refactoring. Generative AI tools were not used to generate experimental measurements, fabricate numerical results, or determine the reported scientific findings. All AI-assisted text, code, analyses, and manuscript components were reviewed and verified by the authors, who take full responsibility for the final content of the submission.

ETHICS STATEMENT

This study uses de-identified, publicly available data from The Cancer Genome Atlas (TCGA), including the TCGA-BRCA, TCGA-KIRC, and TCGA-LUAD cohorts, and does not involve new participant recruitment or the collection of new clinical specimens. The study is intended for research and hypothesis generation rather than direct clinical decision-making. Potential sources of bias include cohort composition, tissue processing, scanner and staining differences, and under-representation of demographic groups. All TCGA data were used in accordance with the applicable data-access and data-use requirements.

REPRODUCIBILITY STATEMENT

An anonymous implementation of MoSPR is provided with the submission. The repository includes the code required to reproduce the proposed method and main experimental analyses, together with patient-level data splits, preprocessing procedures, model configurations, and scripts for generating the reported tables and figures. Fold-specific model selection and evaluation procedures are described in the manuscript and supplementary material. Data-derived quantities are constructed independently within each training fold to avoid information leakage. Where redistribution of derived data or intermediate artifacts is restricted by the terms of the original datasets, scripts for deterministic regeneration are provided where possible.

REFERENCES

Areej Alsaafin, Amir Safarpoor, Milad Sikaroudi, Jason D Hipp, and Hamid R Tizhoosh. Learning to predict rna sequence expressions from whole slide images with applications for search and classification. *Communications Biology*, 6(1):304, 2023.

Richard J. Chen, Ming Y. Lu, Jingwen Wang, Drew F. K. Williamson, Scott J. Rodig, Neal I. Lindeman, and Faisal Mahmood. Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pp. 339–349, 2021. doi: 10.1007/978-3-030-87237-3_33.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. doi: 10.1016/j.acha.2006.04.006.

Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigos. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature Medicine*, 24(10):1559–1567, 2018.

Leo Fillioux, Joseph Boyd, Maria Vakalopoulou, Paul-Henry Cournede, and Stergios Christodoulidis. Structured state space models for multiple instance learning in digital pathology. In *Medical Image Computing and Computer Assisted Intervention*, pp. 594–604. Springer, 2023.

Mara Graziani, Niccolo Marini, Nicolas Deutschmann, Nikita Janakarajan, Henning Müller, and María Rodríguez Martínez. Attention-based interpretable regression of gene expression in histology. In *MICCAI Workshop on Computational Pathology*, pp. 44–60. Springer, 2022.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International Conference on Machine Learning*, pp. 2127–2136. PMLR, 2018.

Minoru Kanehisa and Susumu Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

Eunjung Lee, Han-Yu Chuang, Jong-Won Kim, Trey Ideker, and Doheon Lee. Inferring pathway activity toward precise disease classification. *PLOS Computational Biology*, 4(11):e1000217, 2008.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdottir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell Systems*, 1(6): 417–425, 2015.

Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024.

Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, volume 14, 2002.

Kazuya Nishimura, Ryoma Bise, Shinnosuke Matsuo, Haruka Hirose, and Yasuhiro Kojima. Cell-type prototype-informed neural network for gene expression estimation from pathology images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Marija Pizurica, Yuanning Zheng, Francisco Carrillo-Perez, Humaira Noor, Wei Yao, Christian Wohlfart, Antoaneta Vladimirova, Kathleen Marchal, and Olivier Gevaert. Digital profiling of gene expression from histology images with linearized attention. *Nature Communications*, 15(1): 9886, 2024.

Charlie Saillard, Rémy Dubois, Oussama Tchita, Nicolas Loiseau, Thierry Garcia, Aurélie Adriansen, Séverine Carpentier, Joelle Reyre, Diana Enea, Katharina von Loga, Aurélie Kamoun, Stéphane Rossat, Corentin Wiscart, Meriem Sefta, Michaël Auffret, Lionel Guillou, Arnaud Fouillet, Jakob Nikolas Kather, and Magali Svrcek. Validation of MSIntuit as an AI-based pre-screening

tool for MSI detection from colorectal cancer histology slides. *Nature Communications*, 14(1): 6695, 2023.

Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, et al. A deep learning model to predict rna-seq expression of tumours from whole slide images. *Nature communications*, 11(1):3877, 2020.

Yasin Şenbabaoğlu, Vignesh Prabhakar, Aminollah Khormali, Jeff Eastham, Evan Liu, Elisa Warner, Barzin Nabet, Minu Srivastava, Marcus Ballinger, and Kai Liu. Mosby enables multi-omic inference and spatial biomarker discovery from whole slide images. *Scientific Reports*, 14(1): 18271, 2024.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing Zhang. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2136–2147, 2021.

Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O. Westholm, Mikael Huss, Annelie Mollbrink, Sten Linnarsson, Simone Codeluppi, Åke Borg, Fredrik Pontén, Paul I. Costea, Pelin Sahlén, Jan Mulder, Olaf Bergmann, Joakim Lundeberg, and Jonas Frisén. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016. doi: 10.1126/science.aaf2403.

The Gene Ontology Consortium. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Research*, 49(D1):D325–D334, 2021.

Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, and Wenyu Liu. Revisiting multiple instance neural networks. *Pattern Recognition*, 74:15–24, 2018.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Jinxi Xiang and Jun Zhang. Exploring low-rank property in multiple instance learning for whole slide image classification. In *International Conference on Learning Representations*, 2023.

Shu Yang, Yihui Wang, and Hao Chen. MambaMIL: Enhancing long sequence modeling with sequence reordering in computational pathology. In *Medical Image Computing and Computer Assisted Intervention*, pp. 296–306. Springer, 2024.

Jingwei Zhang, Anh Tien Nguyen, Xi Han, Vincent Quoc-Huy Trinh, Hong Qin, Dimitris Samaras, and Mahdi S Hosseini. 2DMamba: Efficient state space model for image representation with applications on giga-pixel whole slide image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3583–3592, 2025.

A Supplementary Methods

A.1 Gene-expression preprocessing

Let $T \in \mathbb{R}_+^{N\times G}$ denote the TPM expression matrix restricted to the G target genes used for evaluation. For each sample, the target-gene TPM profile was normalized within this gene subset to a total of 10^4 and log-transformed:

$$Y_{ng} = \log\left(1 + 10^4 \frac{T_{ng}}{\sum_{g'=1}^{G} T_{ng'}}\right). \tag{24}$$

Thus, $Y \in \mathbb{R}^{N\times G}$ represents log1p-transformed CP10K expression within the target-gene subset.

Within each cross-validation fold, these values were further standardized gene-wise using statistics estimated exclusively from the training partition. Let $\mu_{\mathrm{tr},g}$ and $\sigma_{\mathrm{tr},g}$ denote the training-fold mean and standard deviation of gene g. The regression targets were

$$(Y_z)_{ng} = \frac{Y_{ng} - \mu_{\mathrm{tr},g}}{\sigma_{\mathrm{tr},g}}. \tag{25}$$

The same training-fold mean and standard deviation were applied to the validation and test partitions.

A.2 Hyperparameter selection

The structural hyperparameters were fixed across cohorts at $J = 200$ microstates, $K = 8$ macrostates, and $L = 20$ spectral components. Additional searches over K and L did not improve validation performance and frequently selected less stable macrostate partitions. We therefore retained the fixed structural configuration across all cohorts.

The molecular rank q and ridge penalty λ_{ridge} were selected independently within each fold from

$$q \in \{4, 8, 16, 32, 48, 64, 96, 128\}, \qquad \lambda_{\mathrm{ridge}} \in \{1, 10, 10^2, 10^3, 10^4, 10^5\}. \tag{26}$$

For each candidate pair $(q, \lambda_{\mathrm{ridge}})$, model selection was based on the mean squared reconstruction error over the validation samples and genes:

$$\mathcal{L}_{\mathrm{val}} = \frac{1}{N_{\mathrm{val}} G} \left\| \widehat{Y}_{z,\mathrm{val}} - Y_{z,\mathrm{val}} \right\|_F^2. \tag{27}$$

After selecting $(q, \lambda_{\mathrm{ridge}})$, MoSPR and its fold-dependent transformations were refitted on the combined training and validation partitions before a single evaluation on the held-out test partition.

B SUPPLEMENTARY TABLES

Table S1: Dataset summary and fold-specific sample counts used for four-fold cross-validation.

	Cohort Summary				Train		Validation		Test	
Cohort	Patients	Slides	Genes	Fold	Patients	Slides	Patients	Slides	Patients	Slides
TCGA-BRCA	1,037	1,467	14,042	0	519	734	259	366	259	367
				1	519	733	259	367	259	367
				2	518	733	259	367	260	367
				3	518	734	260	367	259	366
TCGA-KIRC	342	681	14,295	0	170	340	86	170	86	171
				1	171	340	86	171	85	170
				2	172	341	85	170	85	170
				3	171	341	85	170	86	170
TCGA-LUAD	482	756	14,514	0	241	378	121	189	120	189
				1	242	378	120	189	120	189
				2	241	378	120	189	121	189
				3	240	378	121	189	121	189

Table S2: Per-gene MoSPR performance on TCGA-BRCA, TCGA-KIRC, and TCGA-LUAD. Shown are the 10 best- and worst-predicted genes ranked by mean SCC across four cross-validation folds.

	BRCA		KIRC		LUAD	
Rank	Gene	SCC	Gene	SCC	Gene	SCC
Best-predicted						
1	PODN	0.781	ACAA2	0.728	TPX2	0.750
2	HTRA1	0.762	LINC01507	0.719	PRR11	0.742
3	ZCCHC24	0.759	EMX2OS	0.715	CENPA	0.736
4	COL14A1	0.759	LDB2	0.715	CCNA2	0.736
5	LRRC17	0.751	AGTR1	0.713	TROAP	0.735
6	MFAP4	0.751	TMEM204	0.709	NCAPH	0.734
7	SPARCL1	0.751	PITPNC1	0.708	KIF4A	0.734
8	GLT8D2	0.749	ERG	0.707	BIRC5	0.730
9	COL1A2	0.747	FRMD3	0.704	CDCA5	0.729
10	SFRP2	0.742	ESAM	0.703	KPNA2	0.727
Worst-predicted						
1	LINC01287	-0.124	AC093326.1	-0.120	LILRA2	-0.167
2	MAEL	-0.072	PLIN5	-0.101	NEFL	-0.083
3	PCDH9	-0.063	NEURL1	-0.098	HIST1H1B	-0.074
4	MRPL53	-0.061	ZFR2	-0.082	KCNE5	-0.051
5	PRSS21	-0.055	HIST1H3F	-0.081	HIST2H2AB	-0.042
6	DLK1	-0.054	HIST2H2AB	-0.081	LATS2-AS1	-0.036
7	HIST1H4A	-0.033	RNF212	-0.076	ADCYAP1	-0.035
8	PCSK1	-0.028	ACSBG1	-0.075	L1TD1	-0.035
9	HOXB8	-0.027	HIST1H2AJ	-0.074	CADM2	-0.035
10	LINC00221	-0.024	CAGE1	-0.073	HIST1H1E	-0.031

Table S3: Model size and inference complexity on TCGA-BRCA after frozen CONCH feature extraction. Parameter counts exclude the shared CONCH encoder and buffers such as standardization statistics. The MoSPR count includes the microstate centroids and the PCA basis, which are fixed during training and required at inference. GFLOPs are computed using the cohort median of $9{,}609$ patches for full-bag methods; HE2RNA, tRNAformer, and SEQUOIA VIS use their architecture-specific clustered inputs of 100, 49, and 100 super-patches, respectively.

Method	# Params	GFLOPs
Max	7.47M	5.14
Mean	7.47M	5.14
ABMIL	7.86M	12.84
HE2RNA	15.97M	22.40
AbReg	7.66M	149.86
tRNAformer	5.22M	**1.86**
ILRA	6.77M	33.42
S4MIL	1.91M	<u>1.96</u>
MambaMIL	2.00M	3.12
SRMambaMIL	2.02M	3.76
MOSBY	8.25M	161.12
SEQUOIA VIS	20.68M	2.76
2DMamba	3.99M	516.02
CPNN	<u>1.51M</u>	28.76
MoSPR (ours)	**0.96M**	2.02

Table S4: Low-rank output controls for ABMIL and tRNAformer. Values report mean gene SCC over four cross-validation folds. *Low-Rank* fits each baseline in its own validation-selected q-dimensional PCA target space, whereas *Post-Hoc* projects the full-rank predictions onto the same subspace ($U_q^\top U_q$) without retraining. The rank q is selected independently for each baseline and fold; MoSPR is shown for reference.

Cohort	Model	Full-Rank	Low-Rank	Post-Hoc
TCGA-BRCA	ABMIL	0.363	0.354	0.380
	tRNAformer	0.332	0.327	0.338
	MoSPR (ours)		**0.411**	
TCGA-KIRC	ABMIL	0.299	0.322	0.314
	tRNAformer	0.319	0.308	0.324
	MoSPR (ours)		**0.348**	
TCGA-LUAD	ABMIL	0.324	0.326	0.345
	tRNAformer	0.336	0.324	0.344
	MoSPR (ours)		**0.376**	

Table S5: Macrostate partition stability under patient-level bootstrap resampling. Microstates were fixed, and 300 patient-level bootstrap resamples were generated for each cancer cohort. For each resample, the global microstate adjacency matrix was reconstructed and re-clustered into $K = 8$ macrostates. Partition stability was quantified using the Adjusted Rand Index (ARI; Hubert & Arabie, 1985) between the bootstrap-derived macrostate partition and the reference partition obtained from the full cohort.

Cohort	Patients	Mean ARI	Median ARI
TCGA-BRCA	1,037	0.769	0.801
TCGA-KIRC	342	0.653	0.646
TCGA-LUAD	482	0.786	0.798

C SUPPLEMENTARY FIGURES

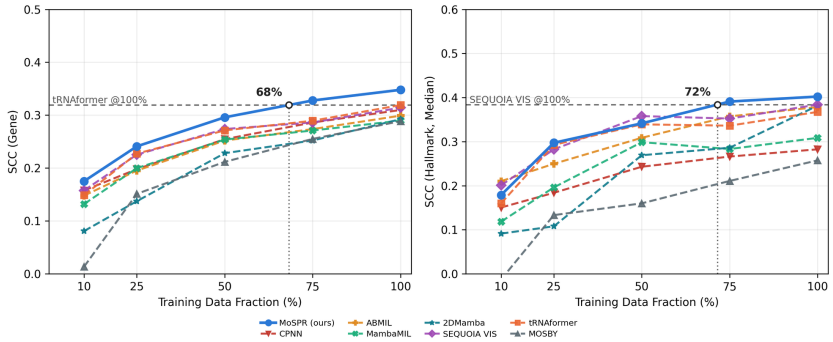


Figure S1: Training-set scaling on TCGA-KIRC. Gene-expression prediction performance (left) and Hallmark pathway reconstruction performance (right) are shown across increasing fractions of the training data.

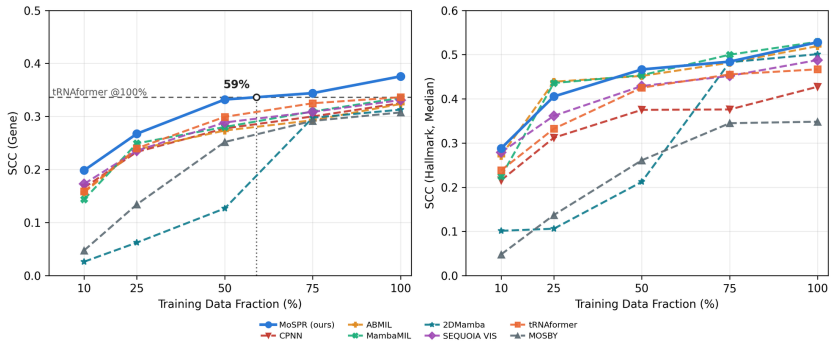


Figure S2: Training-set scaling on TCGA-LUAD. Gene-expression prediction performance (left) and Hallmark pathway reconstruction performance (right) are shown across increasing fractions of the training data.