EPS QUANTIFIES THE PREDICTABILITY OF GENE EXPRESSION FROM HISTOLOGY IMAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Histology images and transcriptomic profiles capture complementary aspects of tissue biology: morphology and molecular state, respectively. Recent advances in multimodal learning have enabled models to predict the expression of thousands of genes directly from volumetric H&E images. However, the fundamental predictability of individual genes from histology—and hence the intrinsic limits of such models—remains poorly understood. Here, we introduce the Expression Predictability Score (EPS), a model-agnostic metric for quantifying the extent to which the expression of each gene can be inferred from another modality. Under a constrained Gaussian Markov random field model, we establish a theoretical connection between EPS and mutual information. In simulations where mutual information is analytically tractable, EPS closely tracks the true mutual information between modalities. Across diverse spatial omics datasets and tissue types, EPS consistently distinguishes highly predictable from poorly predictable genes, and its relationship with empirical prediction performance remains stable across substantially different predictive models. Genes with high and low EPS further exhibit distinct enrichment patterns across cellular components, revealing biological determinants of cross-modal predictability. Beyond histology-to-transcriptome inference, EPS shows consistent score–performance relationships when predicting RNA expression from surface proteins or chromatin accessibility. Together, these results establish EPS as a general framework for characterizing cross-modal information content, identifying predictable molecular features, and delineating the intrinsic scope of multimodal inference.

1 INTRODUCTION

Histology images and spatial transcriptomics describe tissue states at different levels. Routine hematoxylin and eosin (H&E) staining reveals tissue architecture, cellular morphology, and their spatial organization, whereas spatial transcriptomics measures gene expression at corresponding tissue locations. Inferring expression from histology images using paired data could extend the use of widely available tissue images to molecular analysis and broaden access to spatial expression information. Convolutional neural networks, cross-modal contrastive learning, and deep learning methods integrating local and spatial information have advanced this direction (He et al., 2020; Xie et al., 2023; Jia et al., 2024).

Tissue morphology and gene expression are related, but the information they capture does not fully overlap. Some expression changes accompany recognizable changes in cellular morphology or tissue architecture, whereas other molecular changes may lack discernible image features. Consequently, genes differ in how readily they can be predicted from the same images, and the predictability of a given gene may vary with tissue context. Systematic benchmarks have reported substantial variation in performance across genes and samples (Wang et al., 2025). These differences raise a question about the data themselves: which expression changes are reflected in histology images, and how can the correspondence between these modalities be quantified?

We introduce the Expression Predictability Score (EPS), which constructs a nearest-neighbor graph from image embeddings and treats gene expression as a signal on this graph. EPS increases as expression differences between image-similar locations decrease relative to the overall variation of the gene. EPS thus translates cross-modal consistency into a gene-level assessment of predictability without relying on the outputs of a particular expression prediction model. The Slice Predictability

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.

Score (SPS), defined as the mean gene-level EPS, summarizes the overall predictability of a common gene set across tissue sections.

Simulations with analytically tractable mutual information show that EPS reflects changes in cross-modal information dependence. In expression prediction experiments across tissues, low-EPS genes exhibit shared limitations in predictive performance, whereas the potential of high-EPS genes is realized to different degrees by different models. Analyses at the slice level, gene functional enrichment, and experiments using paired single-cell omics further extend this relationship to differences across samples, functional categories, and input modalities, establishing gene predictability as a quantifiable property for understanding cross-modal expression inference.

2 Related work

2.1 Predicting gene expression from histology images

Existing studies extract information at scales ranging from local image patches and spatial neighborhoods to whole tissue sections, and predict expression through supervised regression or cross-modal matching. ST-Net connects tissue morphology with spatial expression through deep learning (He et al., 2020); BLEEP learns cross-modal representations through contrastive learning between images and expression (Xie et al., 2023); and THItoGene combines dynamic convolution, capsule networks, and spatial relationship modeling to extract expression-related image features (Jia et al., 2024). These methods have improved the recovery of expression information from tissue images. Nevertheless, the differences in prediction difficulty across genes observed with multiple models require an explanation grounded in the correspondence between images and expression themselves.

2.2 Information and attainable predictive performance

Mutual information describes statistical dependence between two random variables and provides a basis for quantifying cross-modal information relationships. The amount of information a model can use also depends on its representational capacity and computational constraints, as formalized in the theory of usable information (Xu et al., 2020). Cross-modal dependence and the accuracy achieved by a particular model should therefore be assessed separately. EPS characterizes gene predictability through local consistency in paired data, while predictions from multiple models provide empirical tests of whether this data property is reflected in practice.

2.3 Graph smoothness and feature evaluation

The graph Laplacian quadratic form describes signal variation between adjacent nodes, and methods such as Laplacian Score use local structure to evaluate features (He et al., 2005). In cross-modal data, a graph constructed from one modality captures similarities between observations, while variation of a signal from the other modality on this graph indicates their consistency. EPS applies this relationship to gene expression predictability. Normalization by expression variance and total edge weight makes the score reflect local expression differences relative to overall variation.

3 Expression Predictability Score

3.1 Problem formulation

Consider a tissue section with N paired observation locations. At location i, let I_i denote the histology image patch, $z_i = \phi(I_i)$ its embedding extracted by a pretrained encoder, and x_{ig} the expression of gene g. For each gene, the image embeddings and corresponding expression vector jointly determine a scalar score that describes its relative predictability in the given tissue and image representation.

The score is computed over all paired locations, whereas expression prediction performance is evaluated at held-out test locations. EPS characterizes cross-modal dependence in the available paired data without using prediction model outputs; image embeddings are obtained from a fixed pretrained encoder.

3.2 EXPRESSION CONSISTENCY WITHIN IMAGE NEIGHBORHOODS

We construct a k-nearest-neighbor graph in the image embedding space. Let a_{ij} denote the nonnegative weight of the edge from location i to its neighbor j, and $s_0 = \sum_{i,j} a_{ij}$ the total edge weight. We use a binary nearest-neighbor graph. Define $A_s = (A + A^T)/2$, let D_s be its degree matrix, and set $L = D_s - A_s$. Because squared expression differences are symmetric in i and j, summation over directed edges is equivalent to the quadratic form of the symmetric graph Laplacian:

$$E_g = \boldsymbol{x}_g^T L \boldsymbol{x}_g = \frac{1}{2} \sum_{i,j} a_{ij} (x_{ig} - x_{jg})^2. \tag{1}$$

Let $T_g = \sum_i (x_{ig} - \bar{x}_g)^2$ be the total sum of squared deviations of gene expression. We define the normalized graph energy and EPS as

$$C_g = \frac{(N-1)\, E_g}{s_0\, T_g}, \qquad \mathrm{EPS}_g = -\log C_g. \tag{2}$$

This normalization makes the score invariant to expression shifts and nonzero linear rescaling, while adjusting neighborhood differences by total edge weight. Smaller expression differences between image-similar locations relative to overall variation yield a lower C_g and a higher EPS. The score is undefined for constant genes, for which $T_g = 0$. In the idealized case where $T_g > 0$ but $E_g = 0$, EPS tends to positive infinity.

Let $s_g^2 = T_g/(N-1)$, $\boldsymbol{u}_g = (\boldsymbol{x}_g - \bar{x}_g \mathbf{1})/s_g$, and $L = L/s_0$ (redefining L as the total-edge-weight-normalized Laplacian for the remainder of the paper). Then $\sum_i u_{ig}^2 = N - 1$, and Eq. (2) becomes

$$C_g = \boldsymbol{u}_g^T L \boldsymbol{u}_g, \qquad \mathrm{EPS}_g = -\log\left(\boldsymbol{u}_g^T L \boldsymbol{u}_g\right). \tag{3}$$

EPS therefore describes the consistency of target expression within image-defined neighborhoods relative to its overall variation. These neighborhoods are determined by image features and do not require spots to be adjacent in tissue coordinates.

3.3 CONNECTION TO MUTUAL INFORMATION

For the expression X_g of a gene at a single observation and the input representation Z, mutual information can be decomposed into the difference between marginal and conditional entropy:

$$I(X_g; Z) = H(X_g) - H(X_g \mid Z). \tag{4}$$

When marginal uncertainty is held constant, lower residual uncertainty in expression given the input implies higher mutual information. To connect graph energy to this conditional uncertainty, we introduce a Gaussian Markov random field (GMRF) model for the expression vector conditioned on the graph G (Rue & Held, 2005). We first consider a connected graph.

Because the graph Laplacian satisfies $L\mathbf{1} = 0$, we define the model on the zero-sum subspace $S = \{\boldsymbol{v} \in \mathbb{R}^N : \mathbf{1}^T \boldsymbol{v} = 0\}$, with dimension $r = N - 1$. On this subspace, the precision operator is $Q_g = \tau_g L$, where $\tau_g > 0$. Let $\boldsymbol{V}_g$ denote the random vector under this model and $\boldsymbol{u}_g$ the observed standardized expression. The conditional density is

$$p(\boldsymbol{v} \mid G, \tau_g) = (2\pi)^{-r/2} \tau_g^{r/2} |L|_*^{1/2} \exp\left(-\tfrac{1}{2} \tau_g\, \boldsymbol{v}^T L \boldsymbol{v}\right). \tag{5}$$

The density is defined with respect to the r-dimensional volume measure on S, and $|L|_*$ denotes the product of the nonzero eigenvalues. Standardized expression is treated as the observation used to fit this scale model; no additional fixed-sample-variance constraint is imposed on the model's random vector. This formulation avoids using the zero eigenvalue of the Laplacian in an ordinary matrix inverse or determinant.

From Eq. (5), the log-likelihood terms involving τ_g and the maximum likelihood estimate are

$$l(\tau_g) = \frac{r}{2} \log \tau_g - \frac{\tau_g}{2} C_g + \text{const}, \qquad \hat{\tau}_g = \frac{r}{C_g}. \tag{6}$$

The conditional entropy of this Gaussian model on the subspace is

$$H_S(V_g \mid G; \hat{\tau}_g) = \frac{r}{2}\log(2\pi e) - \frac{r}{2}\log \hat{\tau}_g - \frac{1}{2}\log |L|_*. \tag{7}$$

Substituting $\hat{\tau}_g$ and using $\mathrm{EPS}_g = -\log C_g$ gives

$$H_S(V_g \mid G; \hat{\tau}_g) = K_G - \frac{r}{2}\mathrm{EPS}_g, \qquad K_G = \frac{r}{2}\log\left(\frac{2\pi e}{r}\right) - \frac{1}{2}\log |L|_*. \tag{8}$$

For a fixed graph, K_G is independent of the gene. EPS is therefore linearly related to the negative conditional entropy of the fitted graph model. If a common marginal entropy H_0 is additionally used as a reference, the corresponding information score is

$$J_g = H_0 - H_S(V_g \mid G; \hat{\tau}_g) = \frac{r}{2}\mathrm{EPS}_g + \text{const.} \tag{9}$$

Equation (9) shows that EPS can serve as a proxy for information dependence given a fixed graph and a common marginal entropy. Equal marginal entropy is an additional assumption; rescaling genes to the same variance alone does not guarantee it. The graph model describes the conditional distribution of an entire expression vector, rather than the single-observation variables in Eq. (4). The derivation thus provides an information-theoretic interpretation under the model, while agreement with true single-observation mutual information is assessed through analytical simulations.

If the graph has c connected components, a zero-sum constraint is imposed within each component, giving an effective dimension $r = N - c$ (R-INLA). Projecting $\boldsymbol{u}_g$ onto this subspace leaves its Laplacian energy unchanged, so the same scale derivation applies. Because graph structure and effective dimension differ across tissue sections, K_G is not a common constant across sections. The interpretation of SPS across sections is evaluated empirically below.

3.4 Slice-level score

To characterize overall slice predictability, we aggregate gene-level scores over a specified gene set $\mathcal{G}$ and define

$$\mathrm{SPS} = \frac{1}{|\mathcal{G}|}\sum_{g \in \mathcal{G}} \mathrm{EPS}_g. \tag{10}$$

SPS summarizes the consistency between expression and images over a common gene set. Cross-slice analyses use the same gene set, expression preprocessing, and graph construction parameters. We compare SPS with the mean test-set predictive performance across genes to assess how gene-level predictability accumulates at the sample level.

4 Experiments and results

4.1 Data and evaluation settings

We used seven representative 10x Genomics spatial transcriptomics sections (brain, bowel, lung, breast, ovary, prostate, uterus) for gene-level analyses, and 510 human sections from HEST (Jaume et al., 2024) for large-scale slice-level analyses. Histology image embeddings were obtained with UNI (Chen et al., 2024) to construct $k = 5$ nearest-neighbor graphs, and expression was normalized and restricted to a common set of 2,292 genes used for prediction, evaluation, and SPS. We evaluated five published prediction methods (BLEEP, ST-Net, DeepPT, HisToGene, THItoGene) and four local baselines (Ensemble, MLP, Ridge, Linear) on an 80/20 within-slice split, and compared EPS with a k-nearest-neighbor mutual information estimator computed on reduced UNI embeddings for BLEEP, ST-Net, and DeepPT. Single-cell multi-omics experiments used 10x Multiome and CITE-seq data from the NeurIPS 2021 challenge (Lance et al., 2022) (13 and 12 batches, respectively), with RNA prediction evaluated by an MLP from graphs built on protein or ATAC embeddings (the latter via SnapATAC2 (Zhang et al., 2024)). Full preprocessing, normalization, and evaluation details are given in Supplementary Appendix D.

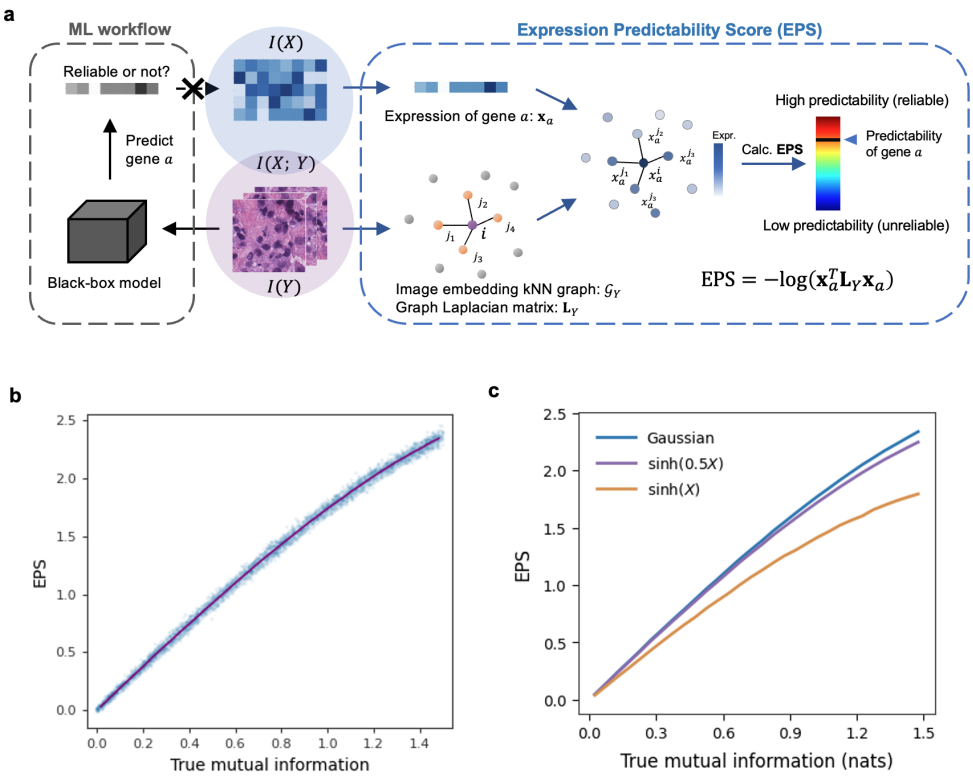


Figure 1: **EPS and its relationship with mutual information.** **a**, Overview of EPS. **b–c**, EPS closely tracks analytical mutual information in controlled Gaussian simulations, including under invertible non-linear transformations of expression. Full legend in Supplementary Appendix C.

4.2 EPS REFLECTS MUTUAL INFORMATION IN CONTROLLED SIMULATIONS

Each baseline simulation generated 2,000 spots and 500 simulated genes. Four-dimensional standard Gaussian latent variables were mapped invertibly to five-dimensional unit-sphere representations. Each gene's expression was generated as a linear combination of a unit-direction projection of the latent variables and independent Gaussian noise. By controlling the signal-to-noise ratio, we set analytical mutual information at evenly spaced values between 0 and 1.5 nats and computed EPS with $k = 5$. Across 30 random seeds, EPS was strongly positively correlated with mutual information, with mean Pearson and Spearman correlations of 0.9961 and 0.9986, respectively (Fig. 1b). Thus, when the strength of cross-modal dependence is known, expression consistency on the graph accurately reflects relative differences in information across genes.

Applying the invertible transformations $\sinh(0.5X)$ and $\sinh(X)$ to the same expression values produced non-Gaussian distributions while preserving analytical mutual information. Under these transformations, the mean Spearman correlations between EPS and mutual information were 0.9984 and 0.9880, and the mean Pearson correlations were 0.9953 and 0.9853, respectively (Fig. 1c). The stronger transformation altered the numerical relationship between the score and mutual information, but gene rankings by information content remained largely preserved, indicating robustness to the non-Gaussian expression distributions tested.

When the mutual information range was extended to 0–3 nats, EPS gradually flattened at high mutual information. Increasing sample size reduced the bias associated with finite nearest-neighbor distances (Extended Data Fig. 1a). Adding independent, irrelevant dimensions weakened the correspondence between EPS and mutual information (Extended Data Fig. 1b). Together, these results show that EPS reflects changes in cross-modal dependence, while its resolution also depends on sample size and the neighborhoods induced by the representation.

4.3 EPS REFLECTS THE PREDICTIVE POTENTIAL OF INDIVIDUAL GENES

Gene-level analyses across the seven tissues revealed consistent trends between EPS and predictive performance among the published methods. Higher-EPS genes generally achieved higher prediction Pearson correlations for BLEEP, ST-Net, and DeepPT, while the supplementary local baselines showed the same overall pattern (Fig. 2a). Prediction Spearman correlations and RMSE provided corresponding evidence (Extended Data Figs. 2–3). Although these models learned expression relationships from the same data in different ways, their performance varied consistently with EPS, indicating a shared component of gene-level prediction difficulty explained by the image–expression relationship.

HisToGene and THItoGene showed lower overall predictive performance under the current settings and failed to accurately predict some high-EPS genes. However, their weaker overall correlations with EPS did not translate into high accuracy in the low-EPS region. Their scatter plots were concentrated at lower performance levels and remained consistent with the shared difficulty of accurately predicting low-EPS genes. These weaker methods thus provide complementary evidence: a model may fail to realize the predictive potential supported by the data, but its low accuracy does not change a gene's relative predictability determined by cross-modal dependence.

In the gene-level scatter plots, high predictive performance was uncommon in the low-EPS region, whereas the high-EPS region accommodated a wider range of performance. This asymmetry supports interpreting EPS as an empirical constraint on attainable performance: low scores are associated with shared performance limitations across methods, whereas at high scores, the extent to which a model exploits the relevant information determines the outcome. Unlike an overall correlation coefficient, this distribution retains evidence of predictive potential that models have not fully realized (Fig. 2a and Extended Data Fig. 4). As a direct dependence baseline, kNN mutual information was also positively associated with gene-level prediction Pearson performance for BLEEP, ST-Net, and DeepPT across all seven tissues. EPS showed comparable or stronger rank associations in most tissue–model combinations (Fig. 2b), indicating that its relationship with prediction performance is consistent with cross-modal dependence measured by a direct mutual-information estimator.

The difficulty of gene expression prediction is therefore not determined solely by the chosen algorithm; expression information reflected in the images imposes a shared constraint across models. EPS quantifies these differences for individual genes, helping distinguish prediction difficulty associated with weak cross-modal dependence from a model's incomplete use of available information. This provides an empirical basis for identifying genes likely to benefit from model improvements and expression changes that still require direct measurement.

4.4 SPS REFLECTS DIFFERENCES IN PREDICTIVE PERFORMANCE ACROSS TISSUE SECTIONS

Across the 510 human tissue sections in HEST, SPS was positively correlated with slice-averaged predictive performance for multiple models (Fig. 2c). The Pearson correlations between SPS and the slice-averaged prediction Pearson correlation were 0.98 for Ensemble, 0.93 for MLP, 0.91 for Ridge, 0.87 for Linear, 0.92 for ST-Net, 0.95 for DeepPT, and 0.94 for BLEEP. Averaging gene-level scores thus preserved differences in performance across samples, indicating that gene predictability accumulates consistently at the slice level.

Comparing performance gains over linear regression showed that Ensemble, MLP, ST-Net, and DeepPT generally achieved larger improvements on higher-SPS sections, whereas gains on low-SPS sections were smaller. Differences between Ridge and Linear were modest overall. The image–expression relationships in high-SPS data provide information that improved models can exploit, whereas low-SPS data present prediction difficulties shared across methods.

The gene- and slice-level results reinforce each other: when expression variation in more genes is reflected in image structure, multiple prediction methods tend to achieve higher accuracy on the same section. Performance differences across samples therefore capture information differences between the paired modalities as well as differences in model suitability. SPS summarizes this data-level variation in a comparable score, supporting the interpretation of cross-sample predictive performance and the identification of data better suited to image-based expression inference.

4.5 PREDICTABILITY IS ASSOCIATED WITH GENE FUNCTION AND TISSUE CONTEXT

To investigate the biological features associated with predictability, we aggregated gene EPS across HEST and performed Gene Ontology cellular component enrichment analysis on genes in the top

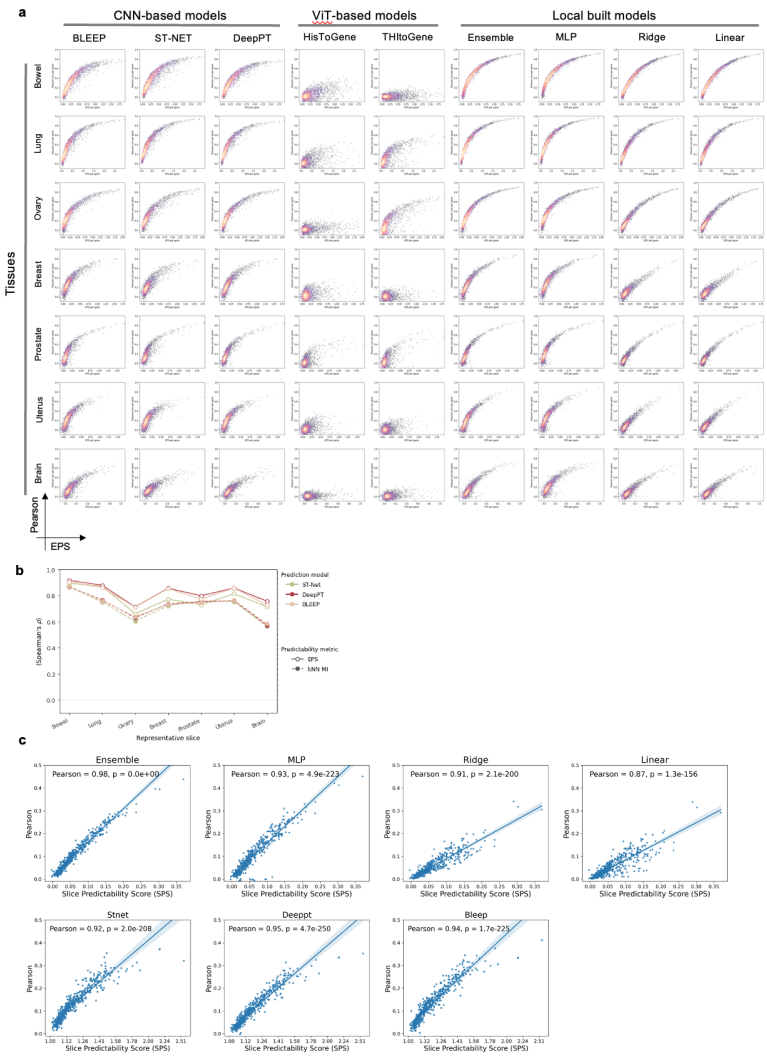


Figure 2: **EPS, kNN mutual information, and SPS reflect predictive performance.** **a**, Gene-level EPS versus prediction performance across seven tissues and nine methods. **b**, EPS shows rank associations with performance comparable to or stronger than a direct kNN mutual information estimator. **c**, Slice-level SPS versus slice-averaged prediction performance. Full legend in Supplementary Appendix C.

and bottom 5% by mean score. High-EPS genes were enriched in structures such as the collagen-containing extracellular matrix, as well as blood microparticles and secretory granule lumens. Low-EPS genes were more frequently associated with endocytic vesicles, membrane rafts, and membrane microdomains (Fig. 3a–b). Highly and poorly predictable genes thus exhibit distinct functional distributions, indicating systematic biological differences in the correspondence between images and expression.

The extracellular matrix and associated tissue structures can produce recognizable morphological features in H&E images, providing indirect cues to their expression patterns. Some membrane- and vesicle-related molecular processes may not be accompanied by equally evident morphological changes. These enrichment patterns are consistent with the range of information captured by routine staining and suggest that expression information from different functional categories is reflected in tissue images to different degrees. Staining properties, cellular composition, and tissue state may all contribute to this relationship, and their individual contributions remain to be distinguished.

Gene scores also exhibited tissue dependence. High-EPS genes were not identical across tissues, indicating that gene predictability must be interpreted within a specific tissue context (Extended Data Fig. 5d). The same gene may be associated with different cell populations, expression ranges,

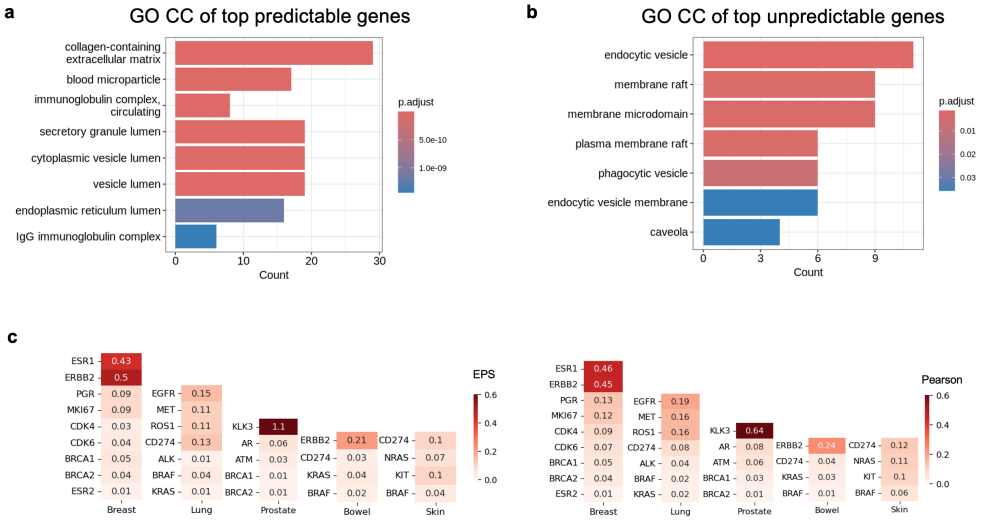


Figure 3: **Functional associations and biomedical examples of gene predictability. a–b**, Gene Ontology cellular component enrichment for high- and low-EPS genes. **c**, EPS and prediction performance for selected genes of biomedical interest. Full legend in Supplementary Appendix C.

and morphological changes in different tissues, leading to different relationships with image information. Scores aggregated across tissues reveal shared trends, whereas tissue-specific scores retain differences relevant to particular applications.

Genes of biomedical interest also differed in predictability. In the tissues analyzed, KLK3 in prostate and ERBB2 and ESR1 in breast showed high EPS and correspondingly strong predictive performance, whereas some other relevant genes in the same tissues had lower scores and poorer performance (Fig. 3c). Biological or clinical importance therefore does not determine how much information about a gene is visible in tissue images. Considering EPS alongside the intended use of a gene can help identify suitable targets for image-based expression inference and molecular information for which direct measurement should be retained.

4.6 Assessing RNA predictability from other paired modalities

The relationships of surface proteins and chromatin accessibility with RNA expression differ from that between tissue morphology and expression, providing an opportunity to assess whether EPS captures a broader notion of cross-modal predictability. In CITE-seq and 10x Multiome data, we constructed cell nearest-neighbor graphs from surface protein or ATAC embeddings, computed EPS for RNA genes on these graphs, and compared the scores with MLP test-set predictive performance.

Across all 12 CITE-seq batches, gene-level scatter plots showed a general increase in RNA prediction Pearson correlation with EPS. Genes in the low-EPS region were typically predicted poorly, whereas the high-EPS region contained more genes with high prediction correlations, with performance gains gradually leveling off at high scores (Extended Data Fig. 6). Thus, although surface protein measurements capture only a subset of molecular features, the cell similarities they define still reflect differences in RNA gene predictability.

The displayed 10x Multiome batches also showed positive gene-level score–performance relationships (Extended Data Fig. 7). Compared with protein inputs, some ATAC batches exhibited greater performance dispersion at low to intermediate EPS, but higher EPS still generally corresponded to higher prediction correlations. Despite their distinct measurement technologies and biological meanings, both input modalities supported the same relationship: target expression that is more consistent within input-defined neighborhoods is more readily predicted.

At the sample level, the Pearson correlation between mean gene EPS and mean prediction Pearson correlation was 0.95 in CITE-seq ($P = 2.99 \times 10^{-6}$) and 0.97 in 10x Multiome ($P = 4.73 \times 10^{-8}$) (Fig. 4). EPS therefore reflects not only gene-level differences within a sample but, after aggregation, also differences in overall RNA prediction difficulty across batches.

Gene- and sample-level evidence remained consistent as the input modality changed from tissue images to surface proteins and chromatin accessibility. The central relationship quantified by EPS

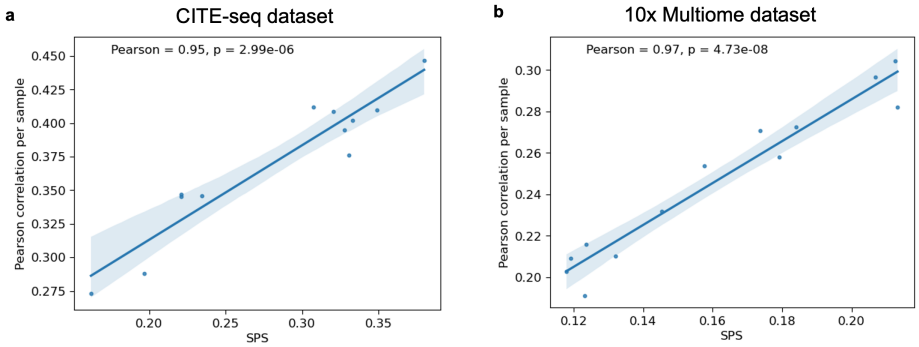


Figure 4: **Sample-level RNA predictability from surface proteins and chromatin accessibility. a–b**, Sample-level EPS (SPS) correlates strongly with mean prediction performance for CITE-seq surface proteins ($r = 0.95$) and 10x Multiome ATAC ($r = 0.97$). Full legend in Supplementary Appendix C.

is therefore the correspondence between target expression and the structure of the input modality. These results support its use in comparing the expression information provided by different paired modalities and extend the interpretation of predictability from histology images to other cross-modal inference tasks.

5 Discussion and Conclusions

EPS quantifies differences in prediction difficulty across genes and samples through expression consistency within image neighborhoods, supported by controlled mutual information simulations, cross-model gene-level analyses, and large-scale slice-level results. Low scores reflect prediction limitations shared across methods, while high scores mark conditions for accurate inference, giving a data-based explanation of model performance that does not depend on any single set of predictions.

EPS is bounded by how well the input representation preserves relevant information: image features not captured by the encoder, and irrelevant embedding dimensions, can distort the neighborhood graph, and simulations show that under strong information dependence, residual signal differences between neighbors limit score growth at finite sample sizes. The GMRF derivation gives this an entropy-based interpretation under explicit distributional assumptions, but real-data scores are also shaped by expression distributions and measurement noise; the resulting performance constraint is therefore an empirical pattern for the given data and representation, not a rigorous upper bound for arbitrary prediction models.

EPS is computed here from paired data within entire tissue sections and evaluated on held-out test spots from the same slice, i.e., within-slice prediction; extending it to cross-patient or cross-tissue prediction will require accounting for input distribution shifts, though the single-cell multi-omics results, supported by MLP experiments, already indicate applicability across input modalities beyond histology. High- and low-EPS genes further show distinct functional associations with tissue morphology and molecular processes, and SPS aggregates these gene-level differences into overall sample-level prediction difficulty, with consistent surface protein and ATAC results supporting their cross-modal relevance.

Practically, EPS offers a quantitative basis for deciding how imaging and expression measurements can complement each other: for highly predictable genes, image-based inference can broaden the use of expression information, whereas for poorly predictable genes, direct measurements retain information current image representations cannot replace. Incorporating cross-modal dependence into evaluation can thus guide target selection and delineate the scope of image-based expression prediction.

AI use statement

In this work, we used generative AI tools (a large language model, Claude) for two purposes. First, during the research process, we used generative AI assistance to help generate the simulated datasets used in the controlled mutual-information experiments (Section 4.2 and Appendix B), for example

in drafting and debugging the data-generating code. Second, we used generative AI to polish and copyedit the manuscript text, including reformatting it into the ICLR submission template. We have not used generative AI tools for other tasks requiring disclosure, such as experimental design, running or analyzing the reported models, or generating the reported research results. We reviewed and verified all AI-assisted code and text, including checking that the simulated data satisfy the analytical properties (e.g., the target mutual information values) required for the experiments in which they are used. We take responsibility for the final content of this work, including all text, claims, and reported results.

Ethics statement

This study uses only publicly available, de-identified spatial transcriptomics, single-cell multi-omics, and histology imaging datasets (10x Genomics public datasets, HEST, and the NeurIPS 2021 multimodal single-cell data integration challenge datasets), together with an openly released pre-trained image encoder (UNI). No new human-subjects data were collected, and no personally identifiable information is used or generated. We are not aware of any direct ethical concerns, dual-use risks, or fairness issues raised by this work.

Reproducibility statement

We provide the full derivation of EPS and its connection to mutual information under a Gaussian Markov random field model in Section 3, including all modeling assumptions. Dataset sources, preprocessing steps, graph construction parameters, model training/evaluation splits, and the mutual information estimator are summarized in Section 4 and given in full in Supplementary Appendix D, and the exact data-generating process, transformations, and parameter settings used in the controlled simulations are given in Appendix B. All datasets used are publicly available, and the simulation procedure is fully specified in Appendix B so that the controlled experiments can be regenerated directly from the paper.

References

R. J. Chen et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30:850–862, 2024.

B. He et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4:827–834, 2020.

X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. In *Advances in Neural Information Processing Systems*, volume 18, 2005.

G. Jaume et al. HEST-1k: A dataset for spatial transcriptomics and histology image analysis. In *Advances in Neural Information Processing Systems*, volume 37, pp. 53798–53833, 2024.

Y. Jia, J. Liu, L. Chen, T. Zhao, and Y. Wang. THItoGene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1):bbad464, 2024.

C. Lance et al. Multimodal single cell data integration challenge: results and lessons learned. *bioRxiv*, 2022.

R-INLA. Besag model for spatial effects. Model documentation. `https://www.inla.r-inla-download.org/r-inla.org/doc/latent/besag.pdf`.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, 2005.

C. Wang et al. Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications*, 16:1544, 2025.

R. Xie, K. Pang, G. D. Bader, and B. Wang. Spatially resolved gene expression prediction from h&e histology images via bi-modal contrastive learning. *arXiv preprint arXiv:2306.01859*, 2023.

Y. Xu, S. Zhao, J. Song, R. Stewart, and S. Ermon. A theory of usable information under computational constraints. In *International Conference on Learning Representations*, 2020.

K. Zhang et al. A fast, scalable and versatile tool for analysis of single-cell omics data. *Nature Methods*, 21:217–227, 2024.