Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Training-Free Correction of Gene Expression Predictions Using Cancer-Specific and Spatial Priors

Tae Joon Jun

taejoon@amc.seoul.kr

Asan Medical Center, University of Ulsan College of Medicine https://orcid.org/0000-0002-6808-5149

Young-Hak Kim

Asan Medical Center, University of Ulsan College of Medicine https://orcid.org/0000-0002-3610-486X

Yunha Kim

Asan Medical Center https://orcid.org/0000-0001-6713-1900

Gaeun Kee

Asan Medical Center

Yoobin Park

Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine

Bokyung Ahn

Asan Medical Center

Chang Ohk Sung

Asan Medical Center https://orcid.org/0000-0002-8567-456X

Article

Keywords:

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10272989/v1

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

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

Training-Free Correction of Gene Expression Predictions Using Cancer-Specific and Spatial Priors

Author information removed for anonymized review.

Abstract

Predicting spot-level gene expression from haematoxylin-eosin (H&E) histology promises to bring spatially resolved molecular profiling to routine pathology slides. However, existing predictors optimize an image-conditional likelihood and leave two forms of structure in the training expression data only partially exploited, namely cancer-specific gene co-expression and slide-level spatial smoothness. We introduce multi-prior posterior guidance, a training-free and model-agnostic correction applied directly at inference time. This method refines spot-level predictions by simultaneously projecting them toward two structural constraints, a cancer-specific biological profile for highly variable genes, and a spatial smoothness graph across the tissue coordinates. We evaluate it across seven baselines on two Visium cohorts, HEST-bench (10 cancers) for prior estimation and STimage-1K4M (8 cancers) for cross-cohort transfer. The correction yields a positive Pearson-correlation gain in all 14 (baseline, cohort) pairs, with 11 reaching Bonferroni-corrected significance, and improves the current public state of the art. A prior trained once on HEST-bench transfers to STimage without retraining, and the gain comes from the off-diagonal structure of the precision matrix, which recovers known cell-type co-expression programs at 10-40% overlap with curated marker lists. Substantial residual error in current spatial transcriptomics predictors therefore reflects biological structure the training data expose but architectures fail to encode.

Introduction

Spatial transcriptomics (ST) measures gene expression at thousands of spatially indexed tissue spots per slide, enabling characterization of tumor microenvironments with spatial molecular resolution[3,4]. The financial and logistical cost of ST profiling still limits its routine use in clinical-scale studies, motivating image-to-ST prediction, in which a haematoxylin-eosin (H&E) whole-slide image is used to infer gene expression on the corresponding spot grid. Accurate image-to-ST prediction could extend ST-like molecular readouts to existing H&E archives without additional wet-lab profiling. Recent benchmarks have helped standardize this task, with HEST-bench[1] providing 72 paired ST and H&E slides across 10 cancers under a per-cancer cross-validation protocol; reported mean per-gene Pearson correlation coefficient (PCC) values on HEST-bench typically fall within 0.30-0.42, with STPath[5] and STFlow[6] representing the current state of the art.

Existing image-to-ST predictors span a wide architectural spectrum, including linear probes on pretrained features[7], end-to-end convolutional neural networks[8], custom transformers[9,10], retrieval-based methods[11], graph neural networks[12,13], pretrained foundation models with multimodal tokenization[5,14], and more recently flow-matching generative models[6]. These approaches have substantially improved spot-level expression prediction through better image representation, spatial aggregation, and multimodal modeling. Their training objectives, however, are typically centered on image-conditioned prediction $\log p_\theta(y \mid x)$.This focus leaves two structural properties of the training expression data only partially exploited. One such property is cancer-specific gene co-expression.Highly variable gene (HVG) expression exhibits structured covariance patterns that reflect tissue composition and microenvironmental programs such as endothelial, immune, stromal, or epithelial gene modules. A predictor optimized primarily for per-gene accuracy may still produce joint expression profiles that violate these cancer-specific co-expression patterns.Another unexploited property is slide-level spatial smoothness. Although several architectures incorporate spatial information through graph layers, attention mechanisms, or image-derived neighborhoods, explicit coordinate-based smoothness can still provide complementary regularization, especially for predictors that produce spot-wise outputs.

Both forms of structure can be estimated from the training data alone. Gene co-expression is estimated from the training expression matrix, whereas spatial smoothness is estimated from spot coordinates. Existing methods commonly inject such information implicitly through archi-

tectural design or training losses, while recent efforts have increasingly explored model-agnostic, plug-and-play techniques that enhance performance across diverse networks without major architectural redesign[15,16]. In ST, structure-aware methods typically leverage such priors through dedicated models rather than as a correction on top of an existing predictor: Thor[28] refines measured spot-level expression into single-cell resolution via an anti-shrinking Markov-diffusion process over the tissue graph, and PASTA[29] is an imputation model that estimates pathway-level expression by borrowing information across spatially proximal and same-cell-type spots. However, these are standalone methods tied to specific tasks. We instead introduce multi-prior posterior guidance, a general inference-time correction that nudges the output of any spot-level gene predictor toward two priors learned from training data. The first is a cancer-conditional Gaussian prior that captures gene co-expression structure, and the second is a k-nearest-neighbor Gaussian Markov random field spatial prior over slide coordinates. The correction is inference-only (no retraining), model-agnostic (works on any predictor that outputs the HVG vector), and applied as a single closed-form Tikhonov-regularized step at prediction time (Fig. 1).

Rather than introducing yet another predictor or competing for the best absolute PCC, we ask how much of the residual error in current predictors reflects biological structure that the training data already contain but leaves unexploited. We evaluate multi-prior posterior guidance on seven architecturally distinct baselines spanning a Ridge linear probe, an ST-Net CNN, a graph convolutional network, a SEPAL graph model, a HisToGene transformer, the STPath foundation model, and the STFlow flow-matching generator across two Visium cohorts, HEST-bench (10 cancers) for prior estimation and STimage-1K4M[2] (8 cancers) for cross-cohort transfer. We find that the correction yields a positive mean PCC gain in all 14 (baseline, cohort) pairs, with 11 reaching Bonferroni-corrected significance, and that the bio prior trained once on HEST-bench transfers cleanly to STimage without retraining. We further show that the gain comes specifically from the off-diagonal structure of the precision matrix Σ_c^{-1}, which recovers known cell-type co-expression programs at 10-40% overlap with curated marker lists. Taken together, these results indicate that a meaningful portion of the gap between current predictors and ground-truth ST measurements can be recovered cheaply at inference time, and that future image-to-ST models should model biological co-expression and spatial tissue organization explicitly.

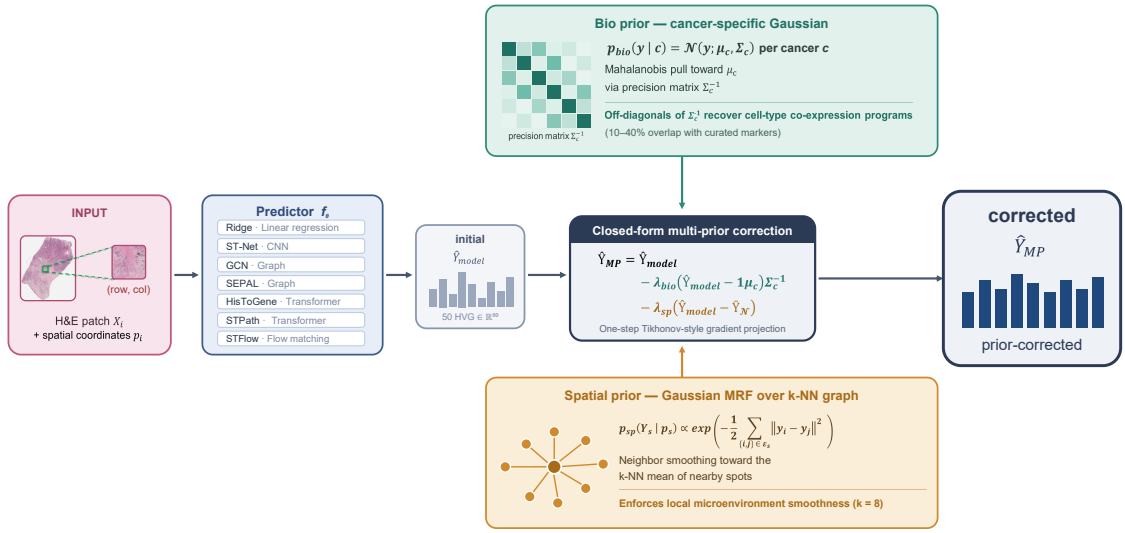


Figure 1: Overview of multi-prior posterior guidance. A spot-level image-to-ST predictor first maps an H&E patch and its spatial coordinate to an initial 50 HVG expression vector. Multi-prior posterior guidance then applies a training-free, one-step correction using two priors estimated from the training data. The cancer-specific Gaussian bio prior encodes gene co-expression through the precision matrix Σ_c^{-1}, whereas the Gaussian Markov random field spatial prior smooths predictions over the coordinate-defined k-NN graph. The corrected output can be obtained from heterogeneous predictors, including linear, CNN, graph-based, transformer and flow-matching baselines.

Results

The ten cancer types across the two cohorts are clear cell renal cell carcinoma (CCRCC), colon adenocarcinoma (COAD), hepatocellular carcinoma (HCC), invasive ductal carcinoma (IDC), lung carcinoma (LUNG), lymph-node metastatic IDC (LYMPH_IDC), pancreatic adenocarcinoma (PAAD), prostate adenocarcinoma (PRAD), rectal adenocarcinoma (READ), and skin cutaneous melanoma (SKCM); STimage covers all but LUNG and PRAD. HEST-bench spans 72 slides in 29 folds and STimage 186 slides in 24 folds. Implementation details, baseline architectures, hyperparameter selection, and statistical testing are described in Methods.

Multi-prior posterior guidance is uniformly beneficial across seven baselines and two cohorts

Table 1 summarizes the headline result. For each baseline and cohort pair, we report the vanilla mean PCC, the multi-prior-corrected mean PCC, the signed improvement Δ, a 95% bootstrap confidence interval on Δ, and the Bonferroni-corrected paired Wilcoxon p-value. Statistical significance is computed by a paired Wilcoxon signed-rank test on the n paired observations and Bonferroni-corrected across the seven baselines per cohort. Furthermore, the confidence

intervals are obtained by 1000-iteration bootstrap resampling over the same paired units as detailed in the Methods section.

Multi-prior posterior guidance is uniformly beneficial, yielding a positive Δ for every baseline and cohort pairs, with 11 of 14 reaching Bonferroni-corrected significance at $\alpha = 0.05$. STPath is reported under its original per-cancer slide-concatenated protocol ($n =$ cancers) on both cohorts to align with its published evaluation, while all other baselines use per-(cancer, fold) PCC. The three non-significant pairs, all on the STimage cross-cohort, fall into two distinct categories. First, STPath with eight samples ($n = 8$) and STFlow with sixteen samples ($n = 16$, 6 cancers $\times$ 3 folds minus 2 COAD folds lacking test slides) on STimage likely reflects limited statistical power owing to the small number of paired observations, displaying positive point estimates of $+0.0026$ and $+0.0038$ but 95% confidence intervals that marginally cross zero, extending from $-0.00013, +0.0059$ and from $-0.0008, +0.0086$ respectively. Second, SEPAL on STimage shows a small effect with an adequate sample size of 24 and a Δ of $+0.0024$. In this case, both prior components are simultaneously small because the spatial contribution is near zero, where Δ_{sp} equals $+0.0002$ as shown in Table 3. This near-zero spatial effect happens because the graph-convolutional architecture of SEPAL already encodes the spatial-smoothness component, while the bio contribution, where Δ_{bio} equals $+0.0023$, attenuates under cross-cohort distributional shift.

Effect sizes on HEST-bench span from a Δ of $+0.0044$ for SEPAL, whose graph-based architecture already encodes the spatial-smoothness component, and $+0.0062$ for STFlow, the strongest vanilla baseline whose iterative flow-matching already captures much of the target structure, up to a Δ of $+0.0326$ for GCN. On the STimage cross-cohort, the four conventional baselines that reach Bonferroni significance, namely Ridge, ST-Net, GCN, and HisToGene, all show a tight Δ band ranging from $+0.0105$ to $+0.0156$. This indicates that the cross-cohort prior transfer is well-behaved across this family.

Per-cancer breakdown confirms broad applicability

Tables 2a and 2b report the per-cancer Δ at the canonical configuration. The multi-prior improves per-cancer accuracy almost universally. On HEST-bench, 68 of the 70 cancer$\times$baseline cells are positive, with gains spanning all ten cancer types. Only two cells are negative, namely SEPAL on READ ($\Delta = -0.014$) and STFlow on READ ($\Delta = -0.015$). Both negatives arise on the same cancer and on the two baselines that already capture spatial structure most strongly:

Table 1: Consistent improvement across seven baselines, at the single canonical configuration ($\lambda_{\text{bio}} = 0.15$ HEST / 0.10 STimage, $\lambda_{\text{spatial}} = 0.2$, $k = 8$; no per-baseline test-set tuning). STPath follows its original-paper per-cancer slide-concatenated PCC protocol (n = cancers). All other baselines, including STFlow on both cohorts, use per-(cancer, fold) PCC (STFlow STimage re-evaluated per fold, $n = 16$ over 6 cancers $\times$ 3 folds, minus 2 COAD folds lacking test slides). Paired Wilcoxon signed-rank, Bonferroni-corrected across 7 baselines per dataset; 95% CI by 1000-iter bootstrap. $^\star = p_{\text{Bonf}} < 0.05$.

Dataset	Baseline	n	Vanilla	+ Multi-prior	Δ	95% CI	p_{Bonf}
HEST-bench	Ridge	29	0.3179	0.3490	+0.0312*	[+0.0227, +0.0406]	2.61e-07
	ST-Net	29	0.2792	0.3056	+0.0264*	[+0.0215, +0.0316]	1.30e-07
	GCN	29	0.3473	0.3799	+0.0326*	[+0.0255, +0.0398]	5.22e-08
	HisToGene	29	0.3620	0.3839	+0.0220*	[+0.0175, +0.0267]	5.22e-08
	SEPAL	29	0.3611	0.3655	+0.0044*	[+0.0011, +0.0081]	3.11e-02
	STPath	10	0.3754	0.3969	+0.0215*	[+0.0118, +0.0317]	2.73e-02
	STFlow	29	0.3945	0.4007	+0.0062*	[+0.0024, +0.0096]	1.22e-03
STimage	Ridge	24	0.1497	0.1603	+0.0105*	[+0.0052, +0.0159]	5.92e-03
	ST-Net	24	0.1532	0.1687	+0.0156*	[+0.0106, +0.0205]	4.17e-06
	GCN	24	0.1693	0.1827	+0.0134*	[+0.0089, +0.0186]	3.59e-05
	HisToGene	24	0.1897	0.2037	+0.0141*	[+0.0085, +0.0198]	8.89e-04
	SEPAL	24	0.2037	0.2061	+0.0024	[-0.0018, +0.0067]	1.0
	STPath	8	0.0356	0.0382	+0.0026	[-0.0013, +0.0059]	1.0
	STFlow	16	0.1758	0.1796	+0.0038	[-0.0008, +0.0086]	1.0

SEPAL is a dedicated spatial-expression model and STFlow is the strongest baseline overall. On READ, where these spatially aware backbones reach markedly higher vanilla PCC than the non-spatial Ridge (0.141) and ST-Net (0.096), the canonical spatial prior is therefore largely redundant and slightly over-corrects. SEPAL and STFlow remain positive on the other nine cancers, and STPath on HEST is uniformly positive across all ten, with gains ranging from +0.003 to +0.024. Figure 2 visualizes the per-cancer effect-size variation across the seven baselines on HEST-bench.

On STimage cross-cohort, absolute PCC is lower throughout, as expected. STFlow lacks READ and SKCM, so 54 of 56 cancer$\times$baseline cells are evaluated; 46 are positive and 8 are negative (Table 2b). The eight negative cells concentrate on four cancers, HCC, IDC, LYMPH_IDC, and PAAD, which are the same cancers that improve solidly in-domain on HEST, indicating cross-cohort transfer weakness rather than intrinsic cancer-type effects. Five of the eight cells are floor effects with $|\Delta| < 0.01$, occurring on cells with near-zero vanilla PCC; the remaining three are SEPAL negatives at moderate vanilla PCC of 0.18-0.23, reflecting the spatial prior over-correcting a backbone that already captures the spatial structure strongly.

The magnitude of the gain tracks how much relevant structure a baseline leaves unexplained for the prior to supply. The largest per-cancer improvements all accrue to baselines that do not already encode the relevant structure strongly, namely Ridge, GCN, ST-Net, and HisToGene,

Table 2: Per-cancer Δ from the multi-prior. Each cell reports vanilla mean PCC → multi-prior mean PCC (Δ).

(a) HEST-bench, canonical $(\lambda_{\text{bio}}, \lambda_{\text{sp}}, k) = (0.15, 0.2, 8)$.

Cancer	Ridge	ST-Net	GCN	HisToGene	SEPAL	STPath	STFlow
CCRCC	0.240→0.253 (+0.013)	0.203→0.229 (+0.026)	0.271→0.290 (+0.019)	0.259→0.275 (+0.016)	0.316→0.320 (+0.004)	0.111→0.118 (+0.007)	0.315→0.328 (+0.013)
COAD	0.207→0.269 (+0.062)	0.224→0.258 (+0.034)	0.245→0.294 (+0.049)	0.282→0.320 (+0.037)	0.310→0.319 (+0.009)	0.395→0.411 (+0.017)	0.320→0.331 (+0.011)
HCC	0.046→0.065 (+0.019)	0.046→0.058 (+0.012)	0.070→0.100 (+0.030)	0.066→0.075 (+0.009)	0.091→0.093 (+0.002)	0.098→0.100 (+0.003)	0.110→0.115 (+0.005)
IDC	0.566→0.585 (+0.019)	0.464→0.485 (+0.021)	0.572→0.591 (+0.019)	0.567→0.582 (+0.016)	0.500→0.501 (+0.001)	0.661→0.666 (+0.005)	0.581→0.585 (+0.005)
LUNG	0.468→0.540 (+0.072)	0.490→0.523 (+0.033)	0.525→0.586 (+0.061)	0.547→0.583 (+0.036)	0.497→0.501 (+0.004)	0.640→0.651 (+0.011)	0.606→0.615 (+0.009)
LYMPH_IDC	0.229→0.253 (+0.024)	0.192→0.229 (+0.037)	0.230→0.265 (+0.035)	0.291→0.317 (+0.026)	0.298→0.314 (+0.016)	0.215→0.234 (+0.019)	0.271→0.278 (+0.006)
PAAD	0.410→0.455 (+0.045)	0.373→0.406 (+0.033)	0.442→0.476 (+0.035)	0.449→0.467 (+0.018)	0.399→0.404 (+0.005)	0.477→0.484 (+0.007)	0.501→0.505 (+0.004)
PRAD	0.258→0.280 (+0.022)	0.285→0.306 (+0.021)	0.325→0.386 (+0.061)	0.383→0.419 (+0.036)	0.354→0.356 (+0.003)	0.250→0.261 (+0.011)	0.375→0.377 (+0.002)
READ	0.141→0.152 (+0.011)	0.096→0.107 (+0.011)	0.172→0.176 (+0.004)	0.190→0.204 (+0.014)	0.235→0.221 (−0.014)	0.288→0.312 (+0.024)	0.221→0.206 (−0.015)
SKCM	0.564→0.638 (+0.074)	0.425→0.455 (+0.030)	0.618→0.668 (+0.050)	0.614→0.643 (+0.029)	0.606→0.614 (+0.008)	0.607→0.628 (+0.021)	0.687→0.699 (+0.012)

(b) STimage (cross-cohort transfer), canonical $(\lambda_{\text{bio}}, \lambda_{\text{sp}}, k) = (0.10, 0.2, 8)$. STFlow lacks READ and SKCM.

Cancer	Ridge	ST-Net	GCN	HisToGene	SEPAL	STPath	STFlow
CCRCC	0.224→0.240 (+0.016)	0.196→0.219 (+0.023)	0.237→0.259 (+0.022)	0.259→0.284 (+0.026)	0.277→0.287 (+0.010)	0.026→0.028 (+0.002)	0.265→0.273 (+0.008)
COAD	0.087→0.095 (+0.007)	0.067→0.077 (+0.010)	0.076→0.083 (+0.007)	0.102→0.114 (+0.012)	0.130→0.131 (+0.001)	0.039→0.043 (+0.004)	0.103→0.110 (+0.007)
HCC	0.093→0.091 (−0.003)	0.170→0.190 (+0.020)	0.163→0.183 (+0.019)	0.196→0.209 (+0.013)	0.231→0.225 (−0.006)	0.052→0.059 (+0.007)	0.202→0.207 (+0.004)
IDC	0.163→0.166 (+0.003)	0.146→0.149 (+0.004)	0.175→0.177 (+0.002)	0.189→0.191 (+0.001)	0.185→0.183 (−0.002)	0.033→0.037 (+0.004)	0.174→0.176 (+0.002)
LYMPH_IDC	0.174→0.185 (+0.011)	0.176→0.191 (+0.015)	0.180→0.186 (+0.006)	0.187→0.187 (−0.000)	0.193→0.182 (−0.011)	0.030→0.024 (−0.006)	0.183→0.190 (+0.007)
PAAD	0.030→0.030 (−0.000)	0.051→0.061 (+0.009)	0.050→0.059 (+0.009)	0.077→0.090 (+0.013)	0.099→0.105 (+0.006)	−0.045→ − 0.050 (−0.004)	0.079→0.085 (+0.007)
READ	0.099→0.129 (+0.030)	0.061→0.084 (+0.023)	0.111→0.138 (+0.027)	0.131→0.162 (+0.031)	0.133→0.152 (+0.019)	0.050→0.055 (+0.005)	—
SKCM	0.326→0.347 (+0.021)	0.358→0.379 (+0.021)	0.363→0.377 (+0.015)	0.377→0.394 (+0.017)	0.381→0.384 (+0.002)	0.100→0.110 (+0.010)	—

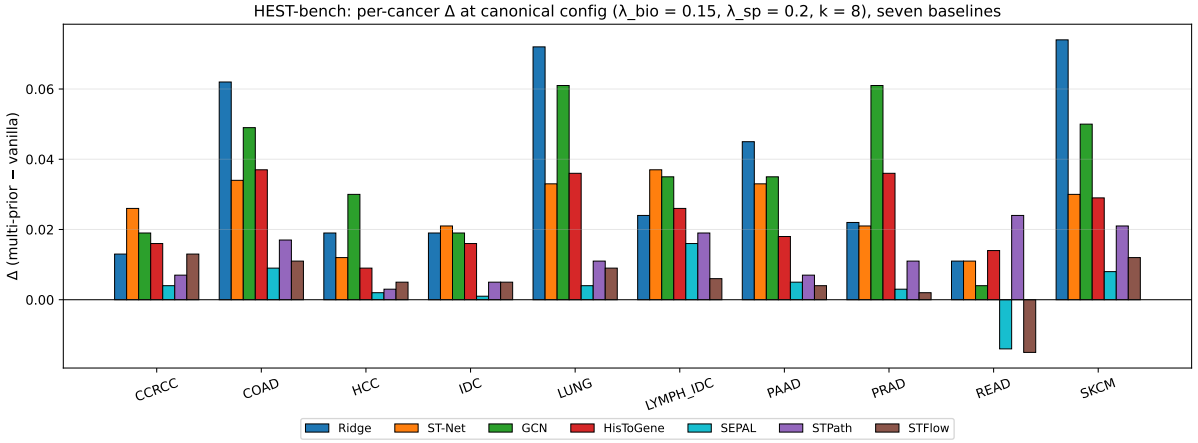


Figure 2: Per-cancer effect of the multi-prior on HEST-bench. Grouped bars show Δ (multi-prior − vanilla mean PCC) at the canonical configuration $(\lambda_{\text{bio}}, \lambda_{\text{sp}}, k) = (0.15, 0.2, 8)$ for each of the ten HEST-bench cancers, with the seven baselines colour-coded as in the legend. Positive bars indicate improvement: 68 of the 70 cancer×baseline cells are positive, the only negatives being SEPAL and STFlow on READ, where the spatial prior over-corrects two already spatially-aware baselines.

illustrated by Ridge SKCM (+0.074) and LUNG (+0.072) and by GCN PRAD (+0.061) and LUNG (+0.061). The baselines that already capture structure strongly gain least, with SEPAL improving by +0.001 to +0.016, STPath by +0.003 to +0.024, and STFlow by +0.002 to +0.015. PAAD and COAD, both heterogeneous adenocarcinomas for which the cancer-conditional biological prior has the most to contribute, also show sizable absolute gains, with Ridge improving by +0.045 and +0.062, respectively.

Component decomposition reveals an architecture-dependent compensation pattern

Table 3 decomposes the multi-prior gain for each (baseline, cohort) pair into bio-only, spatial-only, and joint contributions at the canonical configuration. The bio component is the primary contributor to the joint gain in 12 of the 14 pairs. Two pairs depart from this pattern on HEST-bench. Specifically, STFlow demonstrates comparable bio and spatial components, where Δ_{bio} is +0.0026 and Δ_{sp} is +0.0042. Additionally, STPath exhibits a near-zero joint gain and a slightly negative bio component with a Δ_{bio} of -0.0019. The bio component is positive for every baseline except STPath on HEST-bench. Furthermore, the ratio $\Delta_{\text{bio}}/\Delta_{\text{both}}$ spans from 0.42 to 1.14 with a median of 0.75. The two priors interact weakly and are well-approximated as additive. Excluding the degenerate pair for STPath on HEST-bench, the synergy term defined as $\Delta_{\text{both}} - \Delta_{\text{bio}} - \Delta_{\text{sp}}$ is non-positive on every pair. It is also at most 10% of the magnitude of $|\Delta_{\text{both}}|$ with a median of 4%, meaning the priors can be treated as effectively additive in practice.

The spatial column of Table 3 reveals a per-architecture pattern. Baselines with a spatial-aware inductive bias, namely graph-native baselines like GCN and SEPAL alongside the meta-information-aware STPath, and whole slide-aware STFlow, gain little from the spatial prior on HEST-bench, with Δ_{sp} falling between -0.0005 and +0.0062. Notably, SEPAL is the only baseline with a marginally negative spatial contribution, where Δ_{sp} equals -0.0005. This result reflects the over-smoothing of its 3-layer GCN architecture. Conversely, baselines without an explicit spatial inductive bias, such as Ridge, ST-Net, and HisToGene, gain meaningfully with Δ_{sp} ranging from +0.0089 to +0.0114. STPath conditions on slide-level meta-information including the organ, sequencing technology, tissue type, and the spatial arrangement of spots. This information also acts as a coarse cancer-type covariate. Meanwhile, STFlow uses a Zero-Inflated Negative Binomial likelihood whose parameters are grid-searched per cancer type. Both architectures therefore encode some biological and spatial structure internally, yielding comparatively small prior gains.

Figure 3 aggregates these observations across baselines. The bio gain decreases moderately with vanilla strength, exhibiting a Pearson correlation of -0.58. This trend is consistent with the intuition that weaker baselines miss more biological structure, leaving a larger gap for the prior to fill in. Conversely, the spatial gain decreases more strongly with a correlation of -0.73

Table 3: Component-wise contribution of the bio prior and the spatial prior at the canonical configuration (HEST-bench: $\lambda_{\text{bio}} = 0.15$, $\lambda_{\text{sp}} = 0.2$; STimage: $\lambda_{\text{bio}} = 0.10$, $\lambda_{\text{sp}} = 0.2$). Synergy is $\Delta_{\text{both}} - \Delta_{\text{bio}} - \Delta_{\text{sp}}$, so $\Delta_{\text{both}} \approx \Delta_{\text{bio}} + \Delta_{\text{sp}}$.

Dataset	Baseline	Vanilla	Bio-only	Sp-only	Both	Δ_{bio}	Δ_{sp}	Δ_{both}	Syn.
HEST-bench	Ridge	0.3179	0.3412	0.3268	0.3490	+0.0233	+0.0089	+0.0311	−0.0011
	ST-Net	0.2792	0.2957	0.2905	0.3056	+0.0166	+0.0114	+0.0264	−0.0016
	GCN	0.3473	0.3757	0.3535	0.3799	+0.0285	+0.0062	+0.0326	−0.0021
	SEPAL	0.3611	0.3661	0.3606	0.3655	+0.0050	−0.0005	+0.0044	−0.0001
	HisToGene	0.3620	0.3780	0.3701	0.3839	+0.0160	+0.0081	+0.0220	−0.0021
	STPath	0.3741	0.3722	0.3775	0.3743	−0.0019	+0.0034	+0.0002	−0.0013
	STFlow	0.3945	0.3971	0.3987	0.4007	+0.0026	+0.0042	+0.0062	−0.0006
STimage	Ridge	0.1497	0.1584	0.1517	0.1603	+0.0087	+0.0020	+0.0106	−0.0001
	ST-Net	0.1532	0.1648	0.1577	0.1687	+0.0117	+0.0045	+0.0156	−0.0006
	GCN	0.1693	0.1814	0.1707	0.1827	+0.0121	+0.0015	+0.0134	−0.0002
	SEPAL	0.2037	0.2060	0.2039	0.2061	+0.0023	+0.0002	+0.0024	−0.0001
	HisToGene	0.1897	0.2002	0.1944	0.2037	+0.0105	+0.0048	+0.0141	−0.0012
	STPath	0.0356	0.0381	0.0357	0.0382	+0.0025	+0.0001	+0.0026	0.0000
	STFlow	0.1677	0.1702	0.1698	0.1721	+0.0025	+0.0021	+0.0044	−0.0002

and is dominated by the architecture family rather than by vanilla strength alone. For instance, the four spatially-aware baselines, namely SEPAL, STPath, STFlow, and GCN, show small Δ_{sp} values ($\leq +0.006$ on HEST-bench, with SEPAL uniquely reaching -0.0005), while the non-graph baselines sit meaningfully above at $\Delta_{\text{sp}} \geq +0.008$.These two trends refine the popular claim that weaker baselines gain more into the more precise statement that each prior helps specifically when the baseline lacks the corresponding inductive bias. Furthermore, STFlow sits between the two camps. Since its flow-matching sampler operates in a coordinate-conditioned attention space that partially encodes spatial structure, its spatial gain is correspondingly intermediate. Furthermore, STFlow sits within the spatially-aware group but at its upper edge ($\Delta_{\text{sp}} = +0.0042$), since its flow-matching sampler operates in a coordinate-conditioned attention space that only partially encodes spatial structure.

The gain comes from off-diagonal precision-matrix structure, not mean shrinkage

To isolate the functional mechanics of the biological prior, we evaluate three reduced regularizers against the full multi-prior on HEST-bench at the canonical configuration. The first shrinks predictions toward the empirical mean μ_c alone, analogous to a James-Stein estimator. The second uses a diagonal-only Σ_c^{-1}, which retains the conditional precision of each gene but drops the off-diagonal partial correlations. The third is the full Σ_c^{-1} formulation. All three variants use an identical correction protocol, isolating the covariance topology of the prior as the sole experimental variable (Supplementary Table S1).

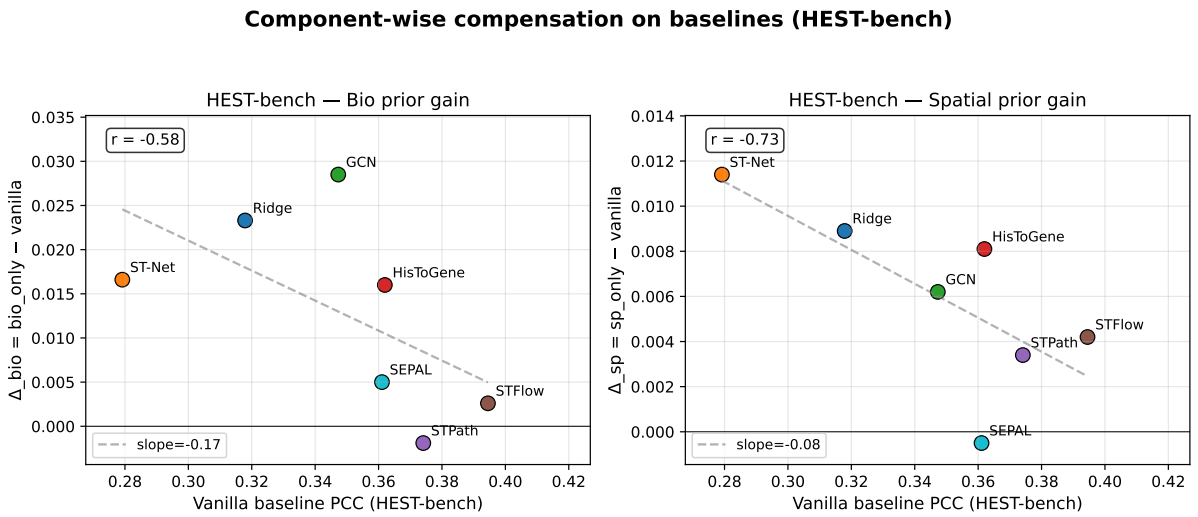


Figure 3: Component-wise compensation pattern. Per-baseline bio-prior gain (left) and spatial-prior gain (right) plotted against vanilla mean PCC on HEST-bench across all seven baselines. Bio gain falls moderately with vanilla strength (Pearson $r = -0.58$); spatial gain falls more strongly (Pearson $r = -0.73$) and is concentrated in the non-graph baselines, while the spatially-aware architectures (GCN, SEPAL, STPath) sit at or near $\Delta_{\text{sp}} = 0$.

Our analysis yields two complementary findings. First, James-Stein-style shrinkage toward μ_c alone yields no incremental gain ($\Delta = 0$) across all baselines, because the vanilla predictions are inherently distributed near the empirical mean, rendering naive mean reversion uninformative. Second, the diagonal-only Σ_c^{-1} configuration actively degrades performance, systematically underperforming the vanilla baselines by 0.05 to 0.07 in mean PCC. Adding only the diagonal of Σ_c^{-1} does worse than nothing because it rescales each gene independently, pulling co-regulated genes apart and pushing the prediction further off the cancer-mode manifold. Only when the off-diagonal terms are restored does the correction move predictions along the co-expression structure rather than across it. The off-diagonal entries of Σ_c^{-1} are therefore not an incremental refinement but the sole active ingredient driving the efficacy of the bio prior.

Top partial correlations recover known cell-type co-expression programs

To test whether the precision matrix Σ_c^{-1} captures genuine cancer-specific co-expression rather than mathematical noise, we rank gene pairs by the magnitude of their partial correlation (the off-diagonal entries of Σ_c^{-1}) and cross-reference the top 20 pairs per cancer against curated cell-type and functional-program marker lists from PanglaoDB[17], CellMarker 2.0[18], and the MSigDB Hallmark collection[19]. Figure 4 visualizes $|\Sigma_c^{-1}|$ for four cancers (IDC breast, PRAD prostate, SKCM melanoma, LUNG lung adenocarcinoma); a pair is counted as recovered only when these sources support a shared program assignment, with full per-pair

Table 4: Match rate between the top-20 partial-correlation pairs and validated marker programs, against a random-pairing baseline (1000 iterations). †PRAD's matches are housekeeping families (ribosomal, metallothionein) only and reflect no cell-type-specific program.

Cancer	Total match	Cross-pathway	Random baseline	× above random
IDC	40%	20%	0.7%	57×
PRAD	10%†	0%	0.2%	50×†
SKCM	30%	20%	1.2%	25×
LUNG	30%	20%	0.9%	33×

annotations released in `bio_interpretation_top_pairs.csv` (Supplementary Table S2). The recovered programs span the major compartments of each tumor, with IDC matches including endothelial (AQP1-VWF), mast cell (KIT-TPSAB1), T-cell (CD3E-TRAC), basal/myoepithelial keratin (KRT5-KRT14, KRT5-KRT6B), adipocyte (ADH1B-ADIPOQ), and ER/luminal (ESR1-GATA3) programs; SKCM matches including fibroblast/CAF-ECM (MGP-POSTN, HAPLN1-MGP), keratinocyte (DSC1-KRT2, KRT16-KRT17), and endothelial (MMRN1-VWF) programs; and LUNG matches including alveolar macrophage (MARCO-MCEMP1, CD163-MARCO), B-cell (CD79A-MZB1), and fibroblast/ECM (COL5A2-CTSK) programs. PRAD is the negative case: only ribosomal (RPL37A-RPS29) and metallothionein (MT1E-MT1X) housekeeping pairs survive verification, with no cell-type-specific program.

To distinguish biologically meaningful matches from gene-family artifacts, we report two rates (Table 4). The total match rate is the fraction of top pairs that map to a validated program. The more conservative cross-pathway rate counts only pairs whose two genes belong to different gene families but the same program, such as the endothelial pair AQP1-VWF, a correspondence that gene-family similarity alone cannot produce. Same-family pairs are counted toward the total rate only when they are genuine cell-type markers (keratins for keratinocytes, CD3E/TRAC for T-cells); generic housekeeping and cell-state axes (ribosomal, metallothionein, the cell-cycle pair MKI67-TOP2A) and the contractile module (MYH11-MYLK) are reported as intra-family and not claimed as cell-type structure. For IDC, SKCM, and LUNG, the top pairs match validated programs at 30–40%, or 25–57× above the random-pairing baseline, with a cross-pathway subset of 20%. We do not claim the prior discovers these gene programs; it inherits them from the training expression data. The interpretability matters instead because it makes explicit what structure the multi-prior injects into the output of a predictor, since a practitioner can inspect Σ_c^{-1} to see which gene-pair constraints the corrected predictions will respect.

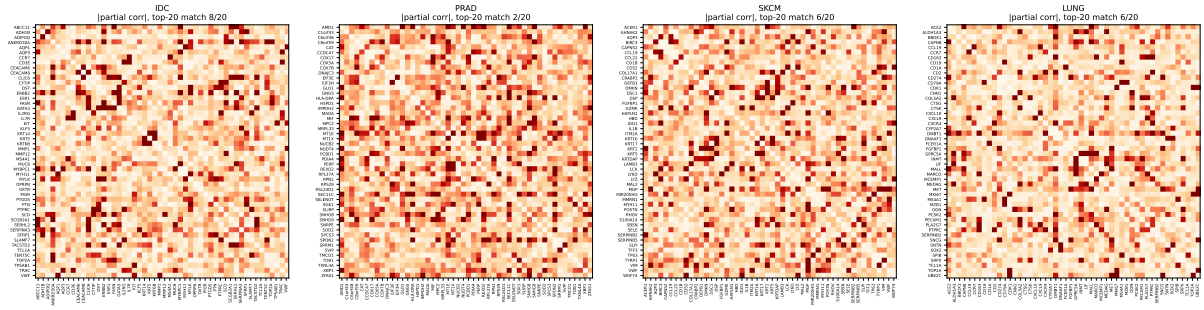


Figure 4: Absolute partial-correlation heatmaps of Σ_c^{-1} for four cancers (IDC breast, PRAD prostate, SKCM melanoma, LUNG lung). Rows and columns are labeled by gene symbol; the visible block-sparse off-diagonal structure encodes known cell-type and functional-program co-expression.

The bio prior trained on HEST-bench transfers to STimage without retraining

We next ask whether a prior estimated on a source cohort can effectively transfer to an independent target cohort under different sequencing conditions. Table 5 reports a prior-source ablation across three architecturally distinct baselines, a linear probe (Ridge), a transformer (HisToGene), and a graph model (SEPAL), evaluated on STimage cross-cohort under the canonical configuration. The STimage-trained prior is evaluated on the 6 cancers with clean STimage coverage, while the HEST-bench-trained prior covers all 8 STimage cancers. Because the prior is model-agnostic, computed from expression data rather than from any predictor, the conclusion is a property of the prior and generalizes across baselines.

Empirically, the HEST-bench-trained prior outperforms the domain-specific STimage prior by a substantial margin on every tested baseline. The STimage-specific prior suffers severe statistical degradation, directly attributable to truncated HVG coverage (25/50 target transcripts) in STimage assays. To accommodate missing modalities, automated data loaders implement zero-padding, which artificially pulls the empirical mean μ_c toward zero and introduces spurious correlations in the estimated Σ_c^{-1}. The HEST-bench reference prior circumvents this corruption because its μ_c and Σ_c^{-1} are estimated on full 50-HVG coverage, preserving an uncompromised representation of cancer-mode co-expression. The seamless transfer of the HEST-bench-trained prior indicates that these core spatial-transcriptomic structures are conserved across disparate Visium cohorts to first order, and consequently the statistical prior can be computed once on a full-coverage reference assembly and universally deployed on partial-coverage target registries without cohort-specific recalibration.

Table 5: Cross-dataset prior transfer on STimage at the fixed canonical configuration (λ_{bio}=0.10, λ_{sp}=0.2, k_{sp}=8). Prior-source ablation across three architecturally distinct baselines—linear (Ridge), transformer (HisToGene), and graph (SEPAL). STimage-trained prior is evaluated on the 6 cancers with a clean STimage prior (CCRCC, COAD, HCC, IDC, LYMPH_IDC, PAAD) and the HEST-bench-trained prior covers all 8 STimage cancers.

Baseline	Prior source	Vanilla	Config	+ Multi-prior	Δ
HisToGene	STimage-trained prior	0.1682	bio0.10_sp0.2	0.1407	-0.0275
	HEST-bench-trained prior	0.1897	bio0.10_sp0.2	0.2037	+0.0140
Ridge	STimage-trained prior	0.1288	bio0.10_sp0.2	0.1040	-0.0248
	HEST-bench-trained prior	0.1497	bio0.10_sp0.2	0.1603	+0.0106
SEPAL	STimage-trained prior	0.1859	bio0.10_sp0.2	0.1427	-0.0432
	HEST-bench-trained prior	0.2037	bio0.10_sp0.2	0.2061	+0.0024

Hyperparameter robustness and STFlow temporal schedule

The correction has three hyperparameters, the bio-prior weight λ_{bio}, the spatial-prior weight λ_{sp}, and the neighborhood size k of the spatial k-NN graph, all selected on validation folds rather than learned as described in Methods. Figure 5 shows the $\lambda_{\text{bio}} \times \lambda_{\text{sp}}$ sensitivity heatmaps for all seven baselines on both cohorts, sweeping the grid $\lambda_{\text{bio}} \in \{0, 0.05, 0.10, 0.15, 0.20\} \times \lambda_{\text{sp}} \in \{0, 0.1, 0.2, 0.3\}$ with k fixed at 8. Four baselines, namely Ridge, ST-Net, GCN, and HisToGene, behave identically: Δ increases monotonically as either prior is strengthened and peaks in a shared region containing the per-cohort default $(\lambda_{\text{bio}}, \lambda_{\text{sp}}, k) = (0.15, 0.2, 8)$ for HEST-bench and $(0.10, 0.2, 8)$ for STimage. The remaining three baselines, SEPAL, STPath, and STFlow, depart from this and all three departures follow the compensation principle above, namely that a baseline that already encodes a given structure strongly gains little from the matching prior and is actively hurt when that prior is pushed too far. SEPAL and STFlow show over-correction at high λ_{sp} on STimage, and STPath peaks at a markedly lower $\lambda_{\text{bio}} \approx 0.05$ because UNI features already encode cancer-conditional structure. The per-cohort default was fixed by a single grid sweep on the HEST-bench Ridge baseline and reused unchanged across all other baselines without per-baseline tuning. This single shared default lies within 0.01 mean PCC of the per-baseline optimum on 6 of the 7 HEST-bench cells and within 0.005 of the per-baseline best for Ridge, ST-Net, GCN, HisToGene, and STPath on STimage cross-cohort. The correction is also robust to the covariance ridge η of Σ_c^{-1} over $[10^{-4}, 10^{-1}]$ (per-η spread ≤ 0.0032 across all seven baselines, Supplementary Fig. S2 and Supplementary Table S3) and to the spatial neighborhood size k over $\{4, 6, 8, 12, 16, 20\}$ (variation < 0.001, Supplementary Fig. S1). The per-fold canonical $\Delta = +0.0002$ for STPath shown in Fig. 5 is the only baseline-default gap

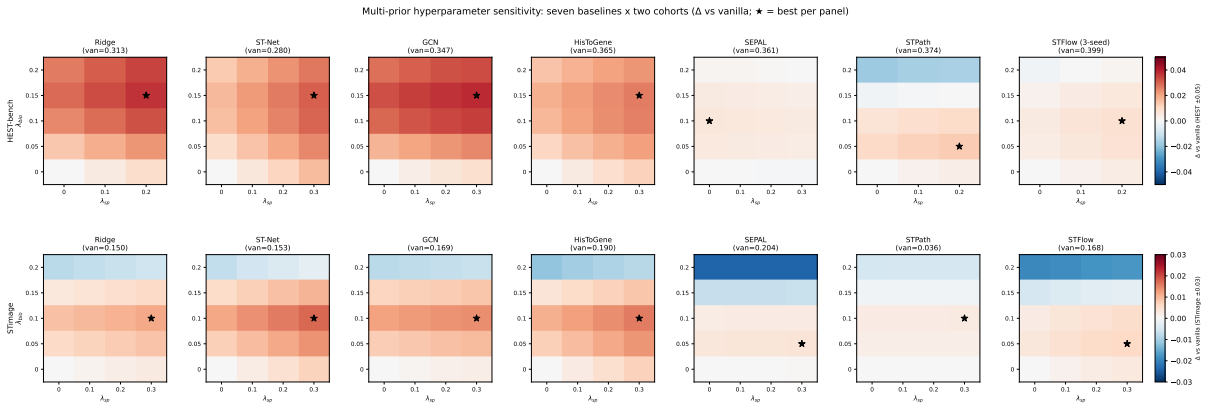


Figure 5: Multi-prior hyperparameter sensitivity across seven baselines and two cohorts. Each panel is a heatmap of the mean-PCC gain over vanilla, Δ, as a function of the biological-prior weight λ_{bio} (rows within a panel) and the spatial-prior weight λ_{sp} (columns within a panel). The two panel rows correspond to the cohorts (top, HEST-bench; bottom, STimage cross-cohort) and the seven panel columns to the baselines, with the vanilla PCC of each baseline reported in the panel title. Red indicates a positive Δ (improvement) and blue a negative Δ (degradation); the colour scale is separate for each cohort (HEST ± 0.05, STimage ± 0.03). The black star marks the best configuration in each panel. Most baselines improve over a broad region and the per-cohort default lies close to their optima.

exceeding 0.01 under per-fold aggregation and reflects an aggregation artifact addressed by the concatenated protocol reported in Table 1. Extended per-baseline sensitivity analysis is provided in Supplementary Note 3.

For iterative samplers such as STFlow, the correction can be applied either post-hoc to the sampler output or at each of the 5 denoising steps as sampling-time guidance. We compared a parameter-free constant per-step schedule against post-hoc correction and several alternative per-step schedules (linear-increasing, linear-decreasing, late-only, cosine; Supplementary Fig. S3). All per-step schedules outperform post-hoc correction, with the constant schedule yielding $+0.0021$ over post-hoc ($n = 29$ cancer-fold cells, paired Wilcoxon $p = 0.033$). We adopt the parameter-free constant schedule as the headline configuration throughout, following the same fixed-default protocol applied to $(\lambda_{\text{bio}}, \lambda_{\text{sp}}, k)$, and treat post-hoc correction as the only applicable option for the six non-iterative baselines.

Discussion

We propose multi-prior posterior guidance, a closed-form training-free inference-time correction that significantly improves the performance of diverse spot-level gene predictors. Across seven baselines, two Visium cohorts, and ten cancers, the correction yields a positive Δ in all 14 (baseline, cohort) pairs, with 11 of 14 reaching Bonferroni significance. The magnitude of Δ varies from $+0.0024$ for SEPAL on STimage to $+0.0326$ for GCN on HEST-bench, and the pos-

itive sign is unanimous across all 14 pairs. We interpret this through the lens of inductive bias. Each predictor architecture encodes its own biases, such as local image correlations in CNNs, spot-graph smoothness in GCNs, position-conditioned aggregation in transformers, and a learnable prior over expression vectors in flow matching. However, none of these models encodes an explicit joint prior over the output. Our multi-prior correction supplies structural priors that these predictors lack. Regarding gene co-expression, vanilla architectures learn these relationships only implicitly through marginal per-gene losses, failing to encode them as an explicit joint prior over the 50-dimensional output. Furthermore, while models like GCN, SEPAL, and the spatial attention of STPath partially capture slide-level spatial smoothness other prominent architectures such as Ridge, ST-Net, and HisToGene transformers do not adequately account for this spatial context. The component ablation in Table 3 confirms that the spatial-only Δ is comparatively attenuated for the baselines that already encode spatial structure, namely GCN, SEPAL, and STPath. Conversely, this gain remains positive between $+0.008$ and $+0.011$ for architectures that lack this structure, such as Ridge, ST-Net, and HisToGene models. Furthermore, the bio-only Δ is meaningful for every baseline since none explicitly models the joint expression distribution conditioned on cancer. This conclusion is unsurprising in hindsight beca use the predictors optimize a marginal likelihood while the prior encodes joint structure. However, the size of the gain, reaching up to 10% of the vanilla PCC, suggests that current spot-level predictors leave more biological structure unused than is generally appreciated.

The HEST-bench-trained prior transfers cleanly to STimage, and the methodological implication of this cross-cohort result is that one can estimate the prior once on the cleanest full-coverage source and reuse it on partial-coverage target cohorts. This approach is effective because the cancer-mode co-expression structure captured by Σ_c^{-1} appears conserved across Visium cohorts to first order, aligning with the broader transcriptomics literature which demonstrates that dominant gene programs remain stable across patient cohorts within a subtype. A further benefit is that the injected structure is interpretable. The top partial-correlation pairs in Σ_c^{-1} recover known cell-type and functional-program markers, and the pairs that do not match a curated program are mostly genuine cell-state cooperatives rather than artifacts. Unlike attention maps, this form of interpretability describes the structure of the output rather than the salience of the input, letting a practitioner inspect in advance which gene-pair constraints the corrected predictions will respect. We are careful not to overclaim that the prior discovers these programs, since it inherits them from the training expression data, making our

true contribution the principled injection of that structure into arbitrary predictors at inference time.

Several considerations bound the scope of our results. The prior is indexed by cancer c, so deployment requires a pre-classification step. In clinical practice, the cancer type is typically known at inference, but mixed-cancer or unknown-primary slides would require a marginalized prior $p(y) = \sum_c \pi_c \mathcal{N}(\mu_c, \Sigma_c)$ or a soft cancer-classifier head, which we do not evaluate. Our method is demonstrated for $G = 50$ HVG following the HEST-bench protocol; scaling to full-transcriptome panels including Visium HD or single-cell ST would require low-rank approximation or sparse-precision estimation. Both cohorts are Visium-derived, so generalization to other platforms such as Stereo-seq or Slide-seq v2 remains untested, and a cancer absent from the training cohort needs its own (μ_c, Σ_c^{-1}) before the method applies.

Two further caveats concern measurement rather than method. First, the STimage cross-cohort pairs for STPath ($n = 8$) and STFlow ($n = 16$) are statistically underpowered, with positive point estimates but 95% bootstrap CIs that marginally cross zero and raw paired Wilcoxon p-values of 0.31 and 0.231, far from the Bonferroni-corrected threshold. Evaluating on a third dataset such as HER2ST may be needed to increase sample size and potentially resolve these marginal effects. Second, the partial HVG coverage of STimage inflates absolute PCCthrough the zero-padding convention that treats missing genes as expression zero, so STimage absolute numbers should be read as upper bounds while the gain Δ remains valid because both vanilla and corrected predictions are scored against the same ground truth. Finally, our STFlow reproduction reaches 0.399 (per-fold mean across 3 seeds in Supplementary Table S3) or 0.395 (under (cancer, fold)-averaged units of Table 1) versus the originally reported 0.416 owing to our use of UNI v1 rather than UNI v1.5 features.

Several extensions follow directly from these limitations. A cell-type-aware spatial prior that replaces the Euclidean k-NN graph with one respecting cell-type clusters from a pretrained patch classifier would address over-smoothing at tissue boundaries; moving from the current per-cancer pooled prior to slide- or patient-level priors could capture patient-specific microenvironments at the cost of higher-variance estimates; and a universal prior pretrained across cancers would remove the requirement that each target cancer appear in the training cohort. A more speculative direction involves scaling beyond the 50-HVG panel via low-rank or sparse-precision estimation of Σ_c^{-1}, which would make multi-prior guidance feasible on 1,000-gene panels from Visium HD or single-cell ST and broaden the range of downstream applications. We leave

empirical validation of these directions to future work.

Taken together, our findings demonstrate that existing spot-level ST predictors can be improved by enforcing biological and spatial structure already present in the training data, that the improvement does not come from simple mean shrinkage but depends specifically on the off-diagonal structure of the cancer-specific precision matrix, and that the cross-cohort transfer of this output structure captures conserved cancer-mode gene co-expression rather than dataset-specific correction artifacts. The consistent direction of improvement across architectures spanning linear regression to flow-matching generative models indicates that residual error in current image-to-ST predictors reflects a shared output-structure problem rather than a weakness of any single architecture. Multi-prior posterior guidance therefore provides both a practical plug-in calibration method that improves the current public state of the art and evidence that future image-to-ST models should model biological co-expression and spatial tissue organization explicitly, rather than expecting marginal per-gene prediction losses to recover them implicitly.

Methods

Problem formulation

Consider a histology slide composed of N tissue spots $\{(x_i, p_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^{H\times W\times 3}$ is an H&E image patch and $p_i \in \mathbb{R}^2$ is the corresponding spatial coordinate. Each spot has an associated log1p-normalized expression vector $y_i \in \mathbb{R}^G$ over a set of G HVG. Throughout this work we use the HEST-bench protocol with $G = 50$ HVG selected per cancer. A spot-level gene predictor is a function $f_\theta : (\mathbb{R}^{H\times W\times 3}, \mathbb{R}^2) \to \mathbb{R}^G$ that maps image and coordinate to predicted expression. The predictor is trained per cancer c on a held-out fold split, and we evaluate per-gene PCC averaged across genes within a fold and across folds within a cancer. Existing predictors f_θ optimize image-conditional log-likelihood $\mathbb{E}_{(x,y)}[\log p_\theta(y \mid x)]$ but rarely exploit the two structural biases that the training expression data exhibit and that motivate our correction, namely cancer-specific gene co-expression (within a cancer, the joint distribution of HVG expression occupies a low-dimensional manifold shaped by cancer-mode gene programs such as cell-type co-expression, metabolic state, and immune infiltration) and slide-level spatial smoothness (spatially neighboring spots share microenvironment context such as tumor-stroma boundaries, immune cell pockets, and vascular regions, so predictions for adjacent spots should

be similar in expression). Both biases are observed in the training data, and rather than retraining f_θ to encode them we apply an inference-time prior-based correction that nudges the output of the predictor toward two structural priors.

Bio prior

To incorporate cancer-specific gene co-expression structure, we define the bio prior as a Gaussian density over expression space for each cancer c, $p_{\text{bio}}(y \mid c) = \mathcal{N}(y; \mu_c, \Sigma_c)$. The mean and covariance are estimated from the training fold,

$$\mu_c = \mathbb{E}_{(x,y)\in D_c^{\text{train}}}[y], \qquad \Sigma_c = \text{Cov}_{(x,y)\in D_c^{\text{train}}}(y), \tag{1}$$

where D_c^{train} denotes all train-spot expression vectors for cancer c. The negative log-prior is a Mahalanobis quadratic,

$$-\log p_{\text{bio}}(y \mid c) = \tfrac{1}{2}(y-\mu_c)^\top \Sigma_c^{-1}(y-\mu_c) + \text{const}, \tag{2}$$

where the off-diagonal entries of the precision matrix Σ_c^{-1} are proportional to the negative partial correlations between genes conditional on all other genes, thereby recovering cell-type-specific co-expression programs. Gaussian graphical models with Σ^{-1} as the partial correlation matrix are a classical tool for inferring gene-gene regulatory structure[20,21], and our use of Σ_c^{-1} as the quadratic form of an inference-time prior is closer to shrinkage estimation in regression[22,23] than to genomic graphical-model inference. Given the limited dataset per cancer (50 HVG and a few hundred to a few thousand spots per cancer), the empirical covariance Σ_c can be ill-conditioned, and we stabilize the inversion with a small ridge regularization $\Sigma_c^{-1} = (\Sigma_c + \eta I)^{-1}$ with $\eta = 10^{-3}$. Robustness of Δ to η over $[10^{-4}, 10^{-1}]$ is reported in Supplementary Table S3.

Spatial prior

On the Visium platform, adjacent spots lie roughly 100 µm apart and are likely to fall within the same histological region. We therefore impose a soft spatial smoothness prior on the predicted expression field. For each slide s, we construct a separate k-nearest-neighbor graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ using the two-dimensional spot coordinates, symmetrizing the graph by adding an undirected edge $\{i, j\}$ whenever $j \in \mathcal{N}_k(i)$ or $i \in \mathcal{N}_k(j)$ so that the resulting graph Laplacian is symmetric and the idealized prior is a proper Gaussian Markov random field (GMRF). We use $k = 8$ by

default, motivated by the Visium lattice geometry, since the immediate hexagonal neighborhood contains six spots but missing spots after quality control and small coordinate distortions can move true local neighbors beyond the first six nearest ranks. Graphs are built independently per slide, with no cross-slide edges, since slide coordinate systems are unregistered and physically separated tissue sections share no spatial continuity.

Given this graph, we define the spatial prior as a GMRF over the predicted expression matrix $Y_s \in \mathbb{R}^{N_s \times G}$,

$$p_{\mathrm{sp}}(Y_s \mid p_s) \propto \exp\Big(-\tfrac{1}{2} \sum_{\{i,j\} \in \mathcal{E}_s} \|y_i - y_j\|^2 \Big) = \exp\Big(-\tfrac{1}{2} \operatorname{tr}(Y_s^\top L_s Y_s) \Big), \tag{3}$$

where $L_s = D_s - A_s$ is the graph Laplacian of the symmetrized slide-specific graph. This prior couples spots through the spatial graph but does not couple genes to one another, since gene-gene structure is handled separately by the bio prior. The corresponding gradient for spot i is $(L_s Y_s)_i = d_i(y_i - \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} y_j)$, proportional to the deviation of y_i from its local neighborhood average. In implementation we use the row-normalized form $y_i - \bar{y}_{\mathcal{N}(i)}$, folding the per-spot degree into a single λ_{sp}, which makes the spatial operator non-symmetric so it does not correspond to a valid Gaussian precision. We therefore treat it as a Tikhonov-style smoother rather than a GMRF MAP step. We deliberately use an unweighted coordinate-based graph rather than weighting edges by image similarity, since image-derived weights would reuse information already available to the predictor f_θ and make the spatial prior less independent of the image model; the resulting prior is therefore image-orthogonal.

Tikhonov-regularized one-step projection

Starting from the output of a predictor $\hat{Y}_{\mathrm{model}} = f_\theta(x, p)$, we apply a one-step Tikhonov-style projection toward the two priors by taking a single first-order gradient step from $\hat{Y}_{\mathrm{model}}$ on the sum of negative log-priors. The negative log-prior gradient is

$$\nabla_Y \big[-\log p_{\mathrm{bio}}(Y) - \log p_{\mathrm{sp}}(Y \mid p) \big] \Big|_{\hat{Y}_{\mathrm{model}}} = (\hat{Y}_{\mathrm{model}} - \mathbf{1}\mu_c)\, \Sigma_c^{-1} + L_s\, \hat{Y}_{\mathrm{model}}, \tag{4}$$

using a row-vector convention throughout. A single step with two empirically tuned weights λ_{bio} and λ_{sp}, with the spatial Laplacian implemented in its degree-normalized neighbor-mean

form and the remaining scale absorbed into λ_{sp}, yields the closed-form multi-prior correction

$$\hat{Y}_{\text{MP}} = \hat{Y}_{\text{model}} - \lambda_{\text{bio}} \left(\hat{Y}_{\text{model}} - \mathbf{1}\mu_c\right) \Sigma_c^{-1} - \lambda_{\text{sp}} \left(\hat{Y}_{\text{model}} - \bar{Y}_{\mathcal{N}}\right). \tag{5}$$

This is a Tikhonov-style regularization toward the prior, not a formal MAP estimate, since the MAP correspondence holds exactly only as $\lambda \to 0$ and we do not claim Bayesian optimality. The correction is inference-only, has no learnable parameters, requires only the precomputed (μ_c, Σ_c^{-1}) from training and a single k-NN-mean computation, and adds $O(NG^2 + NkG)$ arithmetic per slide. It applies either once to the output of a non-iterative predictor (post-hoc) or at each step of an iterative sampler (sampling-time).

Relation to inference-time guidance for generative models

Our method is conceptually related to inference-time guidance for generative models. Classifier-free guidance[24] steers diffusion samples by extrapolating conditional and unconditional model outputs; Diffusion Posterior Sampling[25] guides denoising by computing the gradient of the measurement log-likelihood with respect to the noisy estimate at each step; and Flow Matching Posterior Sampling[26] extends this to flow-matching models. Multi-prior posterior guidance differs from these methods in two ways. First, the guidance source is a pair of fixed prior-based regularizers over outputs rather than a measurement likelihood. Second, the correction is a closed-form one-step update rather than an iterative gradient-based posterior sampler. Two technical points make a formal MAP framing inappropriate even though the form of Eq. (5) is gradient-based. The spatial term uses a random-walk Laplacian $I - P$ that is in general non-symmetric and therefore does not correspond to a valid Gaussian precision matrix. The bio and spatial corrections are also applied sequentially in our implementation, which agrees with a joint posterior mode only to first order in $\lambda_{\text{bio}}\lambda_{\text{sp}}$; the small non-positive synergy in Table 3 is the empirical signature of this $O(\lambda^2)$ gap. The DPS/FMPS literature is concurrent inspiration for applying corrections at every sampling step, which we adopt for STFlow integration; the constant per-step schedule is reported throughout (Supplementary Fig. S3).

Dataset and cohort selection

We evaluate on two independent spatial-transcriptomic cohorts with complementary roles. HEST-bench[1], a spatial-transcriptomic registry spanning 10 cancer types across 29 folds, is

used for prior estimation because it provides complete 50/50 HVG coverage across all slides, so the empirical mean and precision matrix are computed without missing-gene imputation. STimage-bench, curated from STimage-1K4M[2] as an independent cross-cohort test domain, requires specimens to be human, flagged as cancer-involving, sequenced on the 10x Visium platform, and to cover at least 25 of the HEST-bench target HVG; the resulting cohort spans 8 cancers across 24 folds and is a partial-coverage setting in which slides cover only 25/50 of the target HVG with the remainder zero-padded by the standard data loaders. Detailed filtering criteria and per-cancer slide counts are in Supplementary Note 5.

Competing baselines

To comprehensively evaluate our proposed method, we compare it against seven baselines spanning a linear probe, convolutional neural networks, graph-based models, and recent transformer-based foundation models. For feature extraction, we use UNI[7], a ViT-Large self-supervised on 100M H&E patches that produces 1024-d embeddings and is the default encoder in HEST-bench and our work, for all baselines except STPath, which uses GigaPath[27] (1536-d) following its original specification. Each baseline is described below together with its architecture family, trainable parameter count, and the reference to the original method. Ridge and GCN are new implementations for this work; the remaining five baselines are re-implementations or reproductions of published methods adapted to the HEST-bench protocol.

- **Ridge** (linear regression, per-gene, closed form). A simple linear probe applying per-gene Ridge regression ($\alpha = 100$) to map standardized UNI features to the 50 HVG, with a closed-form solution that bypasses stochastic iterative optimization.

- **ST-Net**[8] (end-to-end CNN, $\sim$ 24M trainable parameters). A pretrained ResNet50 is fine-tuned on WSI patches with a gene-prediction head for 60 epochs using a mean-squared-error loss and the AdamW optimizer with learning rate 1×10^{-4}.

- **GCN** (k-NN over coordinates + 2-layer GCN over UNI features, $\sim$ 0.4M trainable parameters). Since existing spatial-transcriptomics graph models use data formats incompatible with our gene panels, we implement a custom GCN baseline. A k-nearest-neighbor graph is built from spatial coordinates and a lightweight two-layer GCN is applied over UNI features to predict the 50 HVG.

- **SEPAL**[13] (k-NN + 3-layer GCN ported to HEST-bench, $\sim$ 0.7M trainable parameters). As the original release targets HER2ST and CSCC formats incompatible with our setting, we retrain from scratch on the extracted features with a 50-HVG output head while preserving the published hyperparameters.

- **HisToGene-style**[9] (spot-attention transformer over UNI, $\sim$ 2.5M trainable parameters; hereafter **HisToGene** in tables, figures, and body text for brevity). A re-implementation of the spot-attention transformer from HisToGene, adapted to the HEST-bench framework with rank-bucket positional encoding for spot coordinates, a four-layer transformer encoder, and a linear head for the 50-gene prediction.

- **STPath**[5] (pretrained transformer-based foundation model, zero-shot inference; encoder frozen). A recent generative model with a spatially aware architecture via $E(2)$-invariant self-attention. We use the publicly available checkpoint with the original evaluation protocol in a zero-shot inference setting; it generalizes via its 38,984-gene pretraining and is originally evaluated on a per-cancer basis.

- **STFlow**[6] (flow-matching denoiser, 5-step sampler, $\sim$ 1.1M trainable parameters). A frame-averaging transformer with a flow-matching denoiser that uses k-NN spatial attention to ensure $E(2)$-invariance. We reproduce the experimental setup of the original study using its public code release with UNI v1 features and report results across three random seeds.

Multi-prior parameters, protocols, and prior estimation

The bio prior (μ_c, Σ_c^{-1}) is computed once per cancer from HEST-bench training spots. The spatial prior uses a per-slide k-NN graph over spatial coordinates with $k = 8$. The per-cohort default coefficients are $(\lambda_{\text{bio}}, \lambda_{\text{sp}}) = (0.15, 0.2)$ on HEST-bench and $(0.10, 0.2)$ on STimage cross-cohort, fixed by a single grid sweep on the HEST-bench Ridge baseline and reused unchanged across all other baselines without per-baseline tuning. The smaller λ_{bio} on STimage is the only cross-cohort change and reflects mild HVG-distribution drift. Hyperparameter sensitivity, robustness sweeps in $\lambda_{\text{bio}} \times \lambda_{\text{sp}}$, the k neighborhood-size sweep, and the η covariance-ridge sweep are reported in Fig. 5 (main text) and Supplementary Note 3 with Supplementary Figs. S1–S2 and Table S3.

The correction is applied according to inference structure of each predictor. Non-iterative predictors (Ridge, ST-Net, GCN, SEPAL, HisToGene, STPath) produce y_{model} in a single for-

ward pass, so the correction is applied once to the final output (post-hoc). Iterative samplers (STFlow with 5-step flow matching) construct the prediction over a trajectory; we apply the correction at each step s, replacing y_s with y_s^{MP} (Eq. (5)) before advancing, so the prior guides the whole trajectory rather than only its endpoint (sampling-time). A constant per-step schedule is used throughout; alternative schedules are compared in Supplementary Fig. S3.

To prevent leakage, the prior is estimated without test spots. For HEST-bench we use a per-fold prior, where for each test fold of the K-fold cross-validation (μ_c, Σ_c^{-1}) is estimated only from training spots of that fold. For STimage cross-cohort we pool across all HEST-bench training spots of cancer c, which is leakage-free by construction since STimage is disjoint from HEST-bench, and the larger pool gives a more stable estimate.

Statistical analysis

We report three measures per (baseline, cohort) pair. The per-gene PCC is first averaged over genes and then reduced to a single score using intended evaluation unit of each baseline. For Ridge, ST-Net, GCN, HisToGene, and SEPAL on both cohorts the score is averaged over the (cancer, fold) pair. For STFlow on HEST-bench, per-seed PCC values are averaged within each (cancer, fold) pair and the (cancer, fold) pair is the resulting score unit, with $n = 29$ paired observations across 3 seeds $\times$ 10 cancers $\times$ per-fold splits. For STFlow on STimage the score is computed per-(cancer, fold), with $n = 16$ paired observations (6 cancers $\times$ 3 folds, minus 2 COAD folds lacking test slides), since READ and SKCM are excluded because all their slides have partial HVG coverage, which biases the zero-padded training targets of STFlow toward zero and makes its strict-load training loop impractical. For STPath the PCC is computed once per cancer under the original slide-concatenated protocol, with $n = 10$ on HEST and $n = 8$ on STimage.

Paired Wilcoxon signed-rank tests are computed over the per-unit PCC difference (vanilla vs. multi-prior), where the unit is (cancer, fold) for Ridge, ST-Net, GCN, HisToGene, SEPAL, and STFlow on HEST, and per-cancer for STPath and STFlow on STimage. The raw p-value is Bonferroni-corrected by multiplying by 7 and clipping to $[0, 1]$. Bootstrap 95% CIs on the mean Δ are obtained by 1000 iterations resampling the paired units defined above with replacement; reported CIs are the 2.5-97.5 percentile of the bootstrap mean-Δ distribution at seed 42.

Evaluation protocol and reproducibility

To ensure a fair and rigorous comparison, we standardize the inference and evaluation pipeline. For each baseline, predictions are precomputed per cancer type and test fold and saved in a uniform format containing the predicted HVG expression, ground-truth HVG expression, spot coordinates, and spot sizes. A single shared evaluation script applies the multi-prior correction and computes the final metrics, ensuring an identical assessment. All experiments are run on NVIDIA B200 GPUs with cached UNI v1 (1024-d) features for HEST-bench (and GigaPath, 1536-d, for STPath). While baseline training and inference times range from minutes to hours, the multi-prior correction requires no gradient computation and adds negligible overhead, about 1 ms per slide once $(\Sigma_c, \Sigma_c^{-1})$ are precomputed.

Data availability

The datasets used in this study are publicly available. HEST-bench and STimage-1K4M were obtained from their respective public repositories[1,2]. No new patient data were collected for this study.

Code availability

The source code generated for this study will be made available upon peer review at a public repository with a persistent DOI.

Acknowledgements

Institution name removed for anonymized review.

Author contributions

Author information removed for anonymized review.

Competing interests

The authors declare no competing interests.

Declaration of generative AI use

During the preparation of this work the authors used Claude Opus 4.7 from Anthropic to assist with developing source code for the experiments and data analysis, and Claude Opus 4.8 to assist with drafting the manuscript. After using these tools, the authors reviewed and edited all content and independently verified every citation and reference. The authors take full responsibility for the content of the published article.

References

[1] Jaume, G. et al. Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *Advances in Neural Information Processing Systems* **37**, 53798-53833 (2024).

[2] Chen, J. et al. Stimage-1k4m: A histopathology image-gene expression dataset for spatial transcriptomics. *Advances in neural information processing systems* **37**, 35796-35823 (2024).

[3] Moncada, R. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature biotechnology* **38**, 333-342 (2020).

[4] Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).

[5] Huang, T., Liu, T., Babadi, M., Ying, R. & Jin, W. STPath: A generative foundation model for integrating spatial transcriptomics and whole-slide images. *NPJ Digital Medicine* **8**, 659 (2025).

[6] Huang, T., Liu, T., Babadi, M., Jin, W. & Ying, R. Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. Preprint at https://arxiv.org/abs/2506.05361 (2025).

[7] Chen, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine* **30**, 850-862 (2024).

[8] He, B. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering* **4**, 827-834 (2020).

[9] Pang, M., Su, K. & Li, M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *BioRxiv*, 2021-11 (2021).

[10] Zeng, Y. et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics* **23**, bbac297 (2022).

[11] Xie, R. et al. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Advances in Neural Information Processing Systems* **36**, 70626-70637 (2023).

[12] Li, B. et al. Gene expression prediction from histology images via hypergraph neural networks. *Briefings in Bioinformatics* **25**, bbae500 (2024).

[13] Mejia, G., Cardenas, P., Ruiz, D., Castillo, A. & Arbelaez, P. SEPAL: Spatial gene expression prediction from local graphs. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2294-2303 (2023).

[14] Xu, Y. et al. A multimodal knowledge-enhanced whole-slide pathology foundation model. *Nature Communications* (2025).

[15] Hasan, M. K., Yang, G. & Yap, C. H. An efficient, scalable, and adaptable plug-and-play temporal attention module for motion-guided cardiac segmentation with sparse temporal labels. *Medical Image Analysis*, 103981 (2026).

[16] Huang, S. et al. Robust simultaneous multislice MRI reconstruction using slice-wise learned generative diffusion priors. *Medical Image Analysis*, 103851 (2025).

[17] Franzen, O., Gan, L.-M. & Bjorkegren, J. L. M. PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**, baz046 (2019).

[18] Hu, C. et al. CellMarker 2.0: An updated database of manually curated cell markers in human and mouse and web tools based on scRNA-seq data. *Nucleic Acids Research* **51**, D870-D876 (2023).

[19] Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Systems* **1**, 417-425 (2015).

[20] Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432-441 (2008).

[21] Meinshausen, N. & Buhlmann, P. High-dimensional graphs and variable selection with the lasso. (2006).

[22] James, W. & Stein, C. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 361-379 (1961).

[23] Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67 (1970).

[24] Ho, J. & Salimans, T. Classifier-free diffusion guidance. Preprint at https://arxiv.org/abs/2207.12598 (2022).

[25] Chung, H., Kim, J., Mccann, M. T., Klasky, M. L. & Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. Preprint at https://arxiv.org/abs/2209.14687 (2024).

[26] Pokle, A., Muckley, M. J., Chen, R. T. Q. & Karrer, B. Training-free linear image inverses via flows. Preprint at https://arxiv.org/abs/2310.04432 (2024).

[27] Xu, H. et al. A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**, 181-188 (2024).

[28] Zhang, P. et al. Thor: A platform for cell-level investigation of spatial transcriptomics and histology. *Nature Communications* **16**, 7178 (2025).

[29] Li, R. et al. Accurate imputation of pathway-specific gene expression in spatial transcriptomics with PASTA. *Nature Communications* **17**, 726 (2026).

Supplementary Files

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

- Supplementaryinformation.pdf