Decomposing Predictive Uncertainty in Virtual Spatial Transcriptomics of Colorectal Cancer Using Bayesian Deep Learning

Thanosan Prathifkumar
California Institute of Technology
Pasadena, CA 91125, USA
Email: tprathif@caltech.edu

Nehan Mohamed
Rutgers University, New Brunswick
New Brunswick, NJ 08901, USA
Email: nehanmohamed1@gmail.com

Minh-Khang Le
Cedars Sinai Medical Center
Los Angeles, CA 90048, USA
Email: khang.le@cshs.org

Zarif Azher
California Institute of Technology
Pasadena, CA 91125, USA
Email: zazher@caltech.edu

Anvith Kakkera, Fred Kolling IV, Laurent Perreard
Dartmouth College Geisel School of Medicine
Hanover, NH 03755, USA
Email: kakkeraanvith@gmail.com, fred.w.kolling.iv@dartmouth.edu, laurent.perreard@dartmouth.edu

Xiaoying Liu, Scott Palisoul, Louis J. Vaickus
Department of Pathology and Laboratory, Dartmouth-Hitchcock Medical Center
Lebanon, NH 03756, USA
Email: xiaoying.liu@hitchcock.org, scott.m.palisoul@hitchcock.org, louis.j.vaickus@hitchcock.org

Mohamed Omar, Arkadiusz Gertych, Maha Guindi,
Keluo Yao, Jane C. Figueiredo, Kyle Coleman, Joshua Levy[1]
Cedars Sinai Health Sciences University
Los Angeles, CA 90048, USA
Email: mohamed.omar@cshs.org, Arkadiusz.Gertych@csmc.edu, Maha.Guindi@cshs.org, keluo.yao@cshs.org, jane.figueiredo@csmc.edu, kyle.coleman@csmc.edu, joshua.levy@csmc.edu

The increasing emphasis on the tumor microenvironment (TME) for colorectal cancer prognosis has accelerated adoption of high-plex spatial profiling technologies such as spatial transcriptomics (ST). While ST enables transcriptome-wide characterization within intact tissue architecture, its cost and technical complexity limit application to the large cohorts needed for biomarker discovery. Consequently, deep-learning approaches have emerged to infer spatial gene expression directly from routine hematoxylin and eosin (H&E) whole-slide images, effectively expanding the scale of spatial analyses where transcript abundance is reflected in tissue morphology. However, these methods typically produce point estimates without indicating whether prediction error arises from

[1] To whom correspondence should be addressed.

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

intrinsic biological and assay variability or from limitations of the predictive model itself. Distinguishing these sources is important for determining which genes are fundamentally recoverable from morphology and which predictions may improve through larger datasets, improved annotations, or more expressive models. Here, we present an initial characterization of predictive uncertainty in virtual spatial transcriptomics using Bayesian deep learning. A heteroscedastic ResNet-34 trained on 45 Visium colorectal cancer tissue sections with matched H&E images predicted expression for 991 spatially variable genes, while Monte Carlo dropout (MC dropout) decomposed predictive uncertainty into epistemic (model) and aleatoric (intrinsic) components. The model achieved gene-wise predictive performance that varied substantially in magnitude (Spearman ρ range: 0.071 – 0.596; mean ± SD = 0.297 ± 0.079). Under the specified MC dropout configuration, aleatoric uncertainty explained most of the reductions in predictive accuracy, whereas epistemic uncertainty showed only weak associations with performance. The most accurately predicted genes were epithelial differentiation and adhesion markers (ELF3, EPCAM, CD24), whereas immediate-early response genes (JUN, JUND, IER2) were consistently difficult to infer. Predictive uncertainty was further structured by tissue architecture: neoplastic epithelium exhibited the lowest total and aleatoric uncertainty, while pathway-level analyses demonstrated distinct uncertainty profiles across metabolic, biosynthetic, proliferative, and stress-response programs. These findings provide a proof of principle that predictive uncertainty can help disentangle intrinsic biological and measurement-related variability from uncertainty attributable to the predictive model, offering a framework for identifying morphology–transcriptome relationships that are reliably recoverable from H&E versus those requiring additional data, improved modeling, or direct molecular profiling. Such uncertainty-aware virtual spatial transcriptomics may ultimately guide biomarker discovery, cohort design, and confidence-aware deployment of virtual RNA inference in precision oncology.

Keywords: spatial transcriptomics, deep learning, uncertainty quantification, colon cancer, histology

1. Introduction

Colorectal cancer (CRC) remains a leading cause of cancer-related mortality, with more than 150,000 new diagnoses and over 50,000 deaths projected annually in the United States [1,2]. Particularly concerning is the rising incidence of early-onset CRC [3], which is increasing rapidly among younger adults and often presents with more aggressive clinicopathologic features. Despite advances in diagnosis and treatment, metastatic progression and disease recurrence remain the primary causes of CRC-related mortality. While tumor-specific molecular alterations and clonal evolution contribute substantially to these outcomes, increasing attention has focused on the tumor microenvironment (TME) – an ecosystem of malignant, stromal, immune, and vascular cells together with extracellular matrix and signaling molecules whose spatial organization governs cell-cell interactions within the tumor. The composition and architecture of the TME play critical roles in tumor progression, immune evasion, therapeutic response, and metastatic dissemination [4]. Accordingly, TME-derived molecular signatures have emerged as important biomarkers for predicting recurrence, metastatic potential, and response to therapy [5]. These advances have made comprehensive characterization of the TME a central goal of precision oncology.

Spatial transcriptomics (ST) platforms, such as the 10x Genomics Visium platform, enable mapping of gene expression within a tissue while preserving its spatial architecture [6,7]. Coupled with imaging of histopathologic tissue slides, ST allows molecular heterogeneity within the TME to be mapped directly onto tissue morphology, revealing how gene expression varies across distinct histologic contexts [8]. However, despite this promise, applying spatial transcriptomics at a population scale remains difficult. Current platforms are limited by high cost, low throughput, and batch-

dependent reproducibility, among other challenges (e.g., RNA degradation) that complicate reproducibility and routine clinical use in large cohorts. This gap motivates the need to infer ST from more affordable sources, such as from routine tissue stains, such as hematoxylin and eosin-stained whole slide images (WSI).

A growing body of work has demonstrated that spatial gene expression can be inferred directly from routine H&E whole-slide images, particularly for genes and pathways whose activity is reflected in tissue morphology. HER2RNA predicted bulk RNA-seq expression from whole slides [9], while ST-Net predicted spot-level spatial expression from paired Visium H&E images using a neural network with a convolutional backbone [10]. HisToGene uses a visual transformer to incorporate more spatial context [11], while Hist2ST combines convolutional, transformer, and graph neural networks to recover ST-level information [12]. More recently, large publicly available datasets have accelerated development of vision foundation models capable of virtual spatial transcriptomic inference across diverse tissues [13–16]. While these resources promise broad applicability, further work is needed to understand how well they reflect clinical-grade specimens and real-world pathology workflows. Regardless of model architecture, however, nearly all existing approaches report only point estimates of gene expression, providing little indication of the confidence or reliability of individual predictions. Such predictions fail to account for the range of plausible transcriptomic states consistent with a given histologic pattern or to distinguish uncertainty arising from intrinsic biological and measurement variability from uncertainty attributable to limited training data or model capacity.

Genes whose expression is linked to tissue architecture are expected to manifest, either directly or indirectly, through histologic morphology. Conversely, many molecular programs, including transient signaling events, secreted factors, and stress-response pathways, may exhibit weak or inconsistent morphologic correlates, as suggested by prior studies [17]. Nevertheless, many current virtual ST methods routinely report predictions across the transcriptome with limited consideration of whether individual genes can be reliably inferred from morphology. Consequently, deterministic point estimates alone cannot convey the confidence or reliability of individual predictions, nor distinguish uncertainty arising from intrinsic biological and assay variability from uncertainty attributable to limited data or the predictive model itself. These limitations have motivated growing interest in uncertainty-aware approaches for virtual ST.

However, predictive uncertainty has previously been explored in related spatial transcriptomic tasks. Gene expression imputation methods infer missing transcriptomic measurements from neighboring spatial transcriptomic observations, either alone or in combination with histologic information. TransImpute and stImpute estimate per-gene reliability scores to identify which imputed genes can be interpreted with confidence [18,19]. TISSUE instead provides a model-agnostic uncertainty framework by applying conformal inference to generate calibrated prediction intervals around inferred expression values [20].

Uncertainty quantification is also common in probabilistic cell-type deconvolution. Methods including Stereoscope, cell2location, and DestVI model spot-level gene counts using hierarchical Bayesian generative models based on negative binomial or zero-inflated negative binomial (ZINB) likelihoods, enabling cell-type abundances and gene expression to be estimated jointly with posterior uncertainty [21–24]. More closely related to the present work are uncertainty-aware histology-based predictors. StImage parameterizes a negative binomial expression distribution for each gene from an H&E-derived latent representation and combines multiple stochastic predictions through

ensembling to estimate predictive variability [25]. CHRep, on the other hand, uses a histology encoder and then applies a post-hoc calibration step at inference to improve cross-slide reliability [26]. UTOPIA, developed for confocal image-based molecular prediction, similarly calibrates confidence estimates across multiple spatial scales [27].

Collectively, these approaches quantify overall predictive uncertainty using confidence scores, prediction intervals, calibrated variances, or posterior distributions. However, relatively little attention has been given to interpreting the origin of that uncertainty. It remains unclear whether poor prediction reflects intrinsic biological and measurement variability that cannot be recovered from morphology or uncertainty arising from limited data and model capacity that may be reduced through improved modeling. Distinguishing these sources has important implications for biomarker discovery, model development, and clinical deployment of virtual spatial transcriptomics.

Decomposition of predictive uncertainty into aleatoric and epistemic components may help address such questions. Aleatoric uncertainty represents variability intrinsic to the data-generation process, including biological heterogeneity and measurement noise, whereas epistemic uncertainty reflects uncertainty in the learned model and is potentially reducible through additional data or improved modeling [28,29].

Bayesian neural networks (BNNs) provide a principled probabilistic framework for jointly estimating aleatoric and epistemic uncertainty [30]. More broadly, Bayesian inference permits incorporation of prior information to regularize parameter estimation, potentially improving robustness and reducing overfitting in limited-data settings while naturally propagating uncertainty into the predictive distribution [31]. In this work, we employ an approximate Bayesian deep learning approach based on Monte Carlo dropout, which offers a computationally scalable means of estimating predictive uncertainty in large-scale virtual spatial transcriptomic models while retaining many of the practical advantages of Bayesian inference. Gal and Ghahramani showed that dropout at inference approximates variational Bayesian inference over network weights [32,33]. Kendall and Gal formalized how to obtain epistemic and aleatoric uncertainty within a single model [34].

Building on this prior work, we present a proof-of-principle workflow for uncertainty-aware virtual spatial transcriptomics in colorectal cancer. Rather than simply quantifying overall predictive uncertainty, we decompose uncertainty into epistemic and aleatoric components and characterize their spatial, histologic, and gene-specific distributions. We hypothesize that this decomposition can distinguish transcriptomic programs that may benefit from additional data collection or improved modeling from those whose morphology-transcriptome relationships are intrinsically uncertain to guide model development for future biomarker discovery.

2. Materials and Methods

2.1. *Data Collection*

This study utilized a previously described in-house colorectal cancer (CRC) spatial transcriptomic cohort from Dartmouth-Hitchcock Medical Center [1]. Briefly, the cohort comprised 41 patients with stage T3 colorectal adenocarcinoma (mean age, 70.5 ± 15.4 years), including 22 metastatic cases (53.7%), with tumors arising from the ascending (56.1%), transverse (12.2%), and descending (31.7%) colon. The cohort was predominantly low grade (54%) and included 15 MLH1-deficient tumors (36.6%) by immunohistochemistry. ST profiling was performed on 21 paired 10x Genomics Visium capture areas (1-2 tissue sections per capture area) with matched H&E whole-slide images,

yielding 213,036 tissue-overlapping Visium spots following quality control. H&E-stained slides were digitized using a Leica GT450 whole-slide scanner prior to spatial registration and analysis. Twenty-one capture areas were generated using the standard 10x Genomics Visium/CytAssist workflow and processed with the Space Ranger pipeline. Board-certified gastrointestinal pathologists manually annotated the standard Visium capture areas into nine major histologic architectures–benign epithelium, inflammation, cancer, interface, submucosa, muscularis, serosa, adipose, stroma– stored in GeoJSON format for downstream architecture-specific analyses.

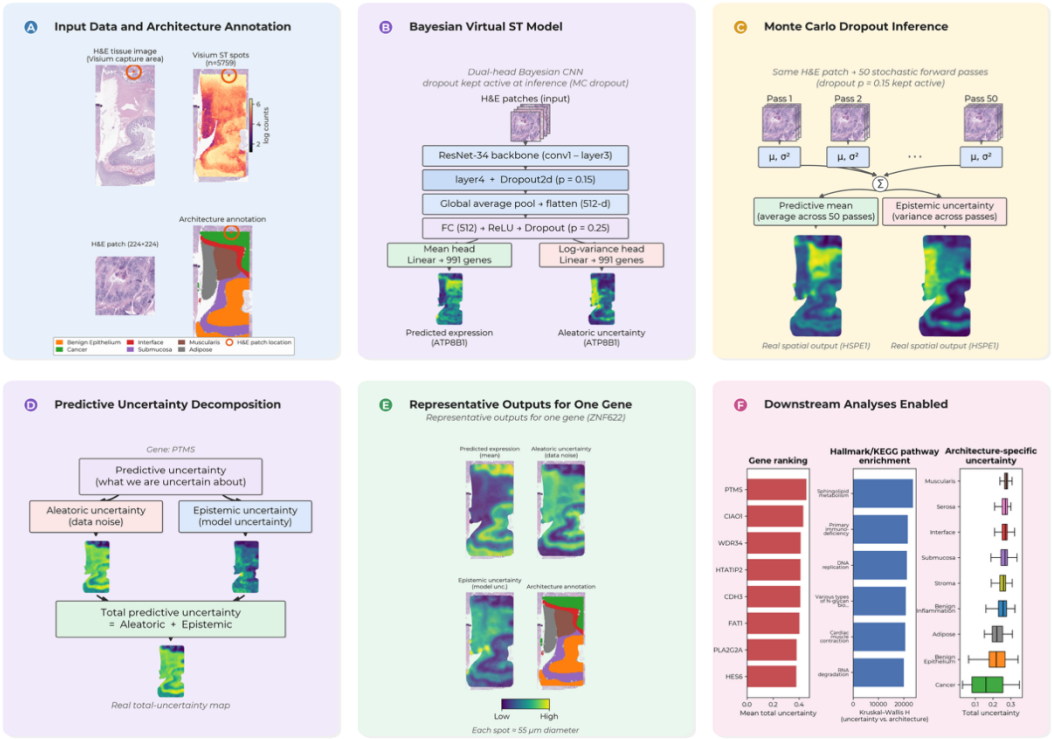


Figure 1: Overview of uncertainty-aware virtual spatial transcriptomics workflow. **A)** Input data and annotation; H&E whole-slide image with the Visium capture area outlined, spot-level expression across the capture area colored by log-counts, a 224x224 patch supplied to the model, and pathologist annotation of histologic architecture, with circle indicating the location of the depicted patch. **B)** Heteroscedastic Bayesian regression network; a ResNet-34 backbone followed with 2D dropout, global average pooling, and a fully connected layer with ReLU and dropout, branching into mean and log-variance heads. **C)** Monte Carlo dropout inference; dropout is retained at test time, and each patch is passed through 50 times. **D)** Decomposition of predictive uncertainty into aleatoric and epistemic components, which sum to total predictive uncertainty. **E)** Representative outputs for a single gene (NOC2L); predictive expression, aleatoric uncertainty, epistemic uncertainty, and the corresponding architectural annotation. **F)** Downstream analyses enabled by the framework, such as ranking genes by total uncertainty, pathway enrichment of uncertainty-associated gene sets, and comparison of total uncertainty across annotated architectures.

2.2. *Data Preprocessing and Patch Extraction*

The prior candidate panel of spatially variable genes identified by Srinivasan et al. were used [1]. The panel of genes was identified using SpatialDE [35]. Spots were restricted to annotated tissue regions, and then quality filtered to ensure there is a minimum of 500 counts per spot. Genes that were detected in fewer than 3 spots were removed. Spot-level expression was normalized by library-size, pseudo-log transformed and reduced to the top 5000 highly variable genes prior to running SpatialDE. SpatialDE was then run independent of each tissue region, producing a q-value reflecting section-specific spatial variation for each gene. For each gene, q-values were averaged across the

tissue sections. The 991 genes with the lowest average q-value across the slides were retained as the candidate SVG panel.

For each Visium spot, a 512×512 pixel image patch centered on the spot location was extracted from the full-resolution H&E whole-slide image, resulting in 213,036 paired image–gene expression samples for downstream model development.

2.3. *Model Training and Validation*

Data were partitioned into training (n=14 capture areas), validation (n=3) and test (n=4) sets, grouped by capture area with validation driving early stopping. Training patches were augmented with random horizontal and vertical flips, brightness/contrast jitter, and shift/scale/rotation. All patches were resized to 224×224 and channel normalized.

Gene expression was predicted using a heteroscedastic regression network developed using a ResNet-34 backbone. The ResNet-34 model was pre-trained from ImageNet, with the final classification layer removed, and a 2D dropout at p = 0.15 added to the final residual block in order to balance backbone capacity with non-degenerate MC sampling. Global average pooled backbone features were passed through a fully connected layer before branching into two parallel output heads, each containing 991 neurons: 1) a mean prediction head (μ_i) representing point estimates, and 2) a log-variance prediction head ($\log \sigma_i^2$). Using the Adam optimizer, the model was trained using the Beta-Gaussian negative log-likelihood (NLL) loss, similar to Mean Squared Error (MSE) but accounting for predictive variance for each data point. The formulation of the heteroscedastic loss and subsequent uncertainty decomposition are described in Section **2.4**.

During model training and validation, a coarse hyperparameter scan demonstrated optimal validation performance with a 1 x 10^{-4} learning rate, a weight decay of 1 x 10^{-5}, a cosine-annealing learning-rate schedule with warm restarts, gradient clipping (max norm = 1.0), mixed precision training, and batch size of 512. Models were trained for 1,000 epochs with early stopping on minimum validation loss.

We intentionally selected a conventional CNN architecture rather than a pathology foundation model to emphasize the effects of uncertainty decomposition. Foundation models leverage extensive pretraining on external histopathology datasets that may alter predictive uncertainty, making it difficult to distinguish the effects of richer image representations from those attributable to the proposed uncertainty framework.

2.4. *Uncertainty Decomposition During Training and Inference*

Let θ represent the set of neural network parameters/weights and the conditional mean, or the prediction for given spot, is $\mu_\theta(x_i)$. The sample-specific variance is given by $\sigma_\theta^2(x_i)$, reflecting heteroscedastic noise, used later to calculate aleatoric uncertainty. The model was trained using the following heteroscedastic Gaussian NLL function for pseudo-log expression y over N spots:

$$L(\theta, X) = \frac{1}{N}\sum_{i=1}^{N}\left(\sigma_\theta^2(x_i)\right)^{\beta}\left[\frac{1}{2\sigma_\theta^2(x_i)}\|y_i - \widehat{\mu_\theta}(x_i)\|^2 + \frac{1}{2}\log\sigma_\theta^2(x_i)\right] + \lambda\|\theta\|^2 \quad (1)$$

The first term penalizes prediction error while weighting observations according to their predicted variance, whereas the logarithmic regularization term (2nd term) discourages unrealistically large variance estimates. The Beta-NLL modification to the objective was originally proposed by Seitzer et al., which reweights the heteroscedastic Gaussian NLL by detaching the variance term, weighting

the exponent with β=0.5 as recommended by the original authors [36]. Under the variational interpretation of Monte Carlo dropout, the final weight regularization term approximates the Kullback–Leibler (KL) divergence component of the evidence lower bound (ELBO), thereby encouraging the approximate posterior over network weights to remain close to the prior. The regularization coefficient λ is related to the dropout probability p and prior variance σ_0^2 as:

$$\lambda = \frac{1 - \mathrm{p}}{2\mathrm{N}\sigma_0^2} \tag{2}$$

Epistemic and aleatoric uncertainty were disentangled at inference using Monte Carlo Dropout, within the aforementioned heteroscedastic regression framework outlined by Kendall and Gal, with spot level uncertainty decomposition illustrated below:

$$\text{Aleatoric Uncertainty} = \mathrm{E}_\theta[\sigma_\theta^2(x_i)] \approx \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} \sigma_{\theta_t}^2(\mathrm{x_i}) \tag{3}$$

$$\text{Epistemic Uncertainty} = \mathrm{Var}_\theta[\widehat{\mu_\theta}(\mathrm{x_i})] \approx \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} \left(\widehat{\mu_{\theta_t}}(\mathrm{x_i}) - \bar{\mu}_\mathrm{i}(\mathrm{x_i})\right)^2 \tag{4}$$

Where $\bar{\mu}_\iota(x_i)$ represents average prediction across all T stochastic forward passes.

$$\bar{\mu}_\mathrm{i}(\mathrm{x_i}) = \mathrm{E}[\widehat{\mu_\theta}(\mathrm{x_i})] = \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} \widehat{\mu_{\theta_t}}(\mathrm{x_i}) \tag{5}$$

During inference, Monte Carlo dropout was enabled, and each image patch was passed through the network 50 times, producing 50 stochastic predictions of the mean gene expression vector and corresponding log-variance vector for each spot. Epistemic uncertainty was estimated as the variance of the predicted mean expression across the 50 stochastic forward passes. Aleatoric uncertainty was estimated as the mean predicted variance across the 50 forward passes. Total predictive uncertainty was computed as the sum of the epistemic and aleatoric uncertainty, while the final gene expression prediction was obtained by averaging the predicted means across all stochastic forward passes. Unless otherwise noted, subsequent analyses focused on aleatoric uncertainty and the epistemic uncertainty fraction, defined as the proportion of total predictive uncertainty attributable to epistemic uncertainty (epistemic uncertainty / total uncertainty). Uncertainty decomposition was performed on the held-out test set at each training epoch to evaluate uncertainty dynamics during model optimization. Mean Spearman correlation across genes was correlated with the mean, 25th percentile, median, 75th percentile, 90th percentile, and maximum aleatoric uncertainty and epistemic uncertainty fraction across genes, under the expectation that epistemic uncertainty would decrease with training while aleatoric uncertainty remained stable.

2.5. *Model Evaluation and Interpretation*

Model performance was evaluated on the held-out test set using the per-gene Spearman correlation between predicted and ground-truth log-normalized expression across all Visium spots. For each gene, epistemic, aleatoric, and total uncertainty were averaged across all spots, yielding ranked gene lists describing genes with the highest and lowest predictive performance and uncertainty. The top-50 (highest) and bottom-50 (lowest) genes ranked by aleatoric uncertainty and epistemic uncertainty fraction underwent over-representation analysis using Enrichr with the MSigDB Hallmark 2020 gene set collection [37]. The relationship between predictive performance and uncertainty was evaluated using linear regression with quadratic effects, modeling Fisher-transformed gene-wise

Spearman correlations as a function of log-transformed aleatoric uncertainty or epistemic uncertainty fraction (U): $\text{atanh}(\rho_{\text{Spearman}}) \sim U + U^2$.

Architecture-specific uncertainty was quantified by averaging epistemic, aleatoric, and total uncertainty across all spots within each histologic architecture. To characterize biological processes associated with uncertainty, gene-specific uncertainty scores were aggregated at the pathway level using KEGG-predefined gene sets, and pathway-level uncertainty was subsequently averaged within each tissue architecture to generate architecture-by-pathway uncertainty maps which were hierarchically clustered for visualization.

3. Results

3.1. *Association between Uncertainty Estimates and Predictive Performance*

Across 991 genes, per-gene Spearman correlations ranged from 0.071 to 0.596 (mean ± SD = 0.297 ± 0.079). Decomposition of predictive uncertainty revealed a marked asymmetry between uncertainty sources. Aleatoric uncertainty accounted for 98.1% of total predictive uncertainty (mean aleatoric uncertainty = 0.861) and exhibited substantial variation across genes (median = 0.861, IQR = 0.677–1.037). In contrast, the mean epistemic uncertainty fraction was only 1.77% at convergence (median = 1.48%, IQR = 0.13–2.00%, maximum = 22.77%), indicating that model uncertainty contributed only a small proportion of predictive uncertainty after training. Aleatoric uncertainty was negatively associated with performance (β = -0.063, p = 1.1×10^{-6}).

The relationship between aleatoric uncertainty and per-gene predictive performance was nonlinear (**Figure 2**), with increasingly negative effects observed at higher levels of aleatoric uncertainty (25th percentile: $\beta = -0.090 \pm 0.019$, $P = 1.1 \times 10^{-6}$; median: $\beta = -0.202 \pm 0.017$, p-value = 7.6×10^{-32}; 90th percentile: $\beta = -0.362 \pm 0.035$, p-value = 2.2×10^{-25}). In contrast, the relationship between epistemic uncertainty fraction and predictive performance remained consistent across its observed range, with modest positive effects (β = 0.058–0.077, p-value ≤ 0.025).

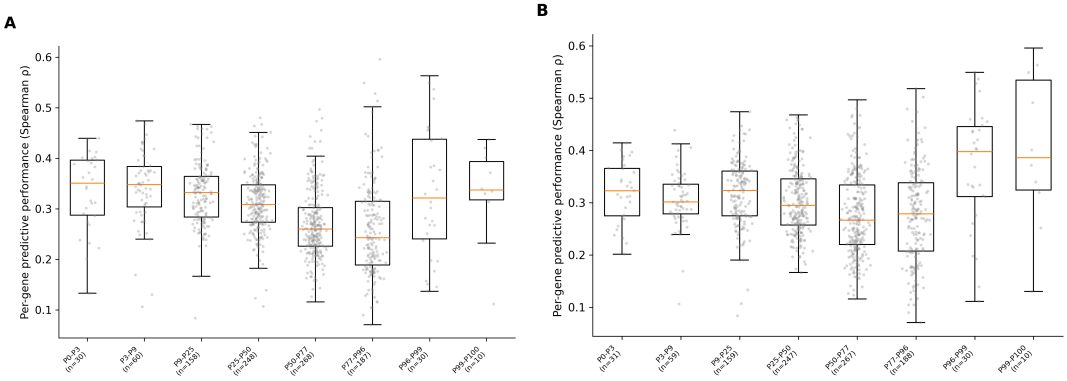


Figure 2: Per-gene predictive uncertainty on the held-out test set. Each data point represents a gene. Genes were binned by percentile of **A)** mean aleatoric uncertainty (across test-set spots) and **B)** mean epistemic uncertainty fraction. Boxplots summarize the distribution of the underlying gene-level data within each uncertainty percentile bin to facilitate visualization of overall trends; boxes denote the median and interquartile range with individual genes overlaid, and n denotes the number of genes per bin. The y-axis represents gene-wise Spearman ρ (range 0–1). Quadratic regression was fit to the original gene-level data (log-transformed uncertainty) rather than the binned boxplot summaries.

3.2. *Highest and Lowest Uncertain Genes by Pathway Enrichment*

Epistemic and aleatoric uncertainty were strongly positively correlated across genes (Spearman ρ = 0.84, p-value < 0.001) (**Table 1**). Genes with high aleatoric and high epistemic uncertainty showed substantial overlap (18/50 genes), including CEACAM5, EPCAM, FCGBP, PHGR1, PIGR, SPARC, TGFBI, THBS1, and TSPAN8. This shared set was enriched for epithelial differentiation and secretory markers, extracellular matrix remodeling genes, and mesenchymal-associated transcripts, suggesting that both uncertainty sources preferentially identified genes with complex, spatially heterogeneous expression patterns. In contrast, genes with low aleatoric and low epistemic uncertainty exhibited more limited overlap (12/50 genes), including DPM2, GAREM1, NDUFAF6, SINHCAF, SLCO5A1, SPRTN, THNSL1, and ZNF215, consistent with a shared subset of constitutively expressed, biologically stable genes (**Figure 3**). Consistent with this observation, genes with high aleatoric and high epistemic uncertainty exhibited similar Hallmark enrichment profiles, including epithelial–mesenchymal transition and TNF-α signaling via NF-κB, plus myogenesis (**Table 2**). In contrast, genes with low uncertainty showed substantially less overlap (12/50 genes) and minimal Hallmark enrichment, with the exception of low-aleatoric genes, which were enriched for G2/M checkpoint, mitotic spindle pathways, and E2F targets.

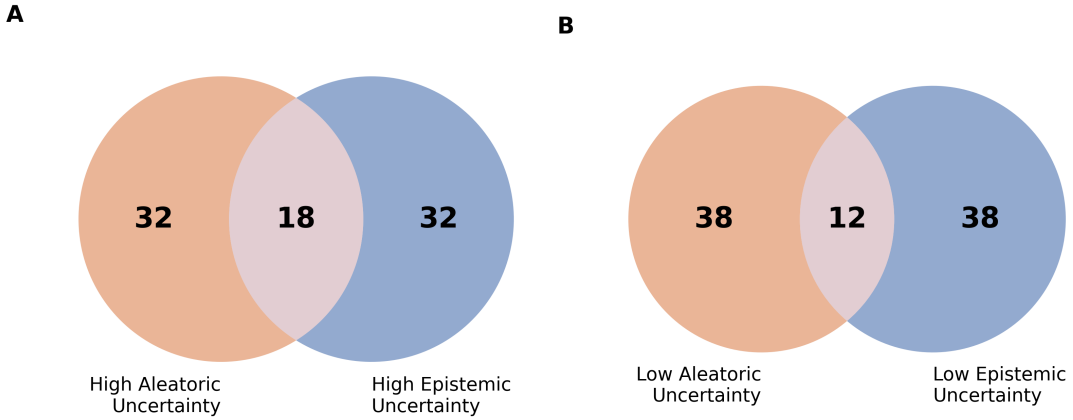


Figure 3: Overlap between genes ranked by aleatoric and epistemic uncertainty. **A)** Intersection of the 50 genes with the highest mean aleatoric uncertainty and the 50 genes with the highest epistemic uncertainty fraction. **B)** Intersection of the corresponding lowest-ranked gene sets. Substantial overlap among high-uncertainty genes contrasts with the limited overlap among low-uncertainty genes.

Table 1: Top-20 and bottom-20 genes by predictive performance and uncertainty decomposition.

Top 20 genes					Bottom 20 genes				
Rank	Gene	Spearman ρ	Aleatoric uncertainty	Epistemic fraction	Rank	Gene	Spearman ρ	Aleatoric uncertainty	Epistemic fraction
1	VIM	0.596	1.068	9.51%	972	XBP1	0.137	1.323	2.53%
2	SPARC	0.563	1.362	7.96%	973	OASL	0.133	0.359	1.59%
3	LGALS1	0.549	1.256	5.76%	974	ETNK1	0.133	1.085	1.93%
4	IGFBP7	0.549	1.353	5.98%	975	MAT2A	0.131	1.113	2.83%
5	PHGR1	0.537	1.339	5.28%	976	TMSB4X	0.13	0.481	22.77%
6	CTSB	0.528	1.05	6.18%	977	TMEM176B	0.128	1.222	3.09%
7	HIST1H1B	0.518	1.407	2.38%	978	VPS28	0.126	1.034	1.61%
8	IGFBP4	0.513	1.097	5.89%	979	CTSA	0.125	1.082	2.14%
9	GPX2	0.502	1.263	3.41%	980	ATP2A3	0.123	0.815	2.00%
10	WNK2	0.497	0.869	2.41%	981	H1FX	0.117	1.087	3.39%
11	CD74	0.491	1.303	5.96%	982	CLSTN1	0.116	1.035	1.61%
12	MKI67	0.48	0.798	2.28%	983	PILRB	0.115	1.106	2.09%
13	MCM7	0.479	1.037	2.20%	984	PLA2G2A	0.111	2.27	2.35%
14	CENPF	0.474	0.437	1.70%	985	MGMT	0.107	0.817	0.97%
15	SRPK1	0.473	0.959	2.22%	986	CAP2	0.106	0.517	0.96%
16	AGMAT	0.468	0.451	2.10%	987	CYBA	0.105	1.195	2.94%

17	MCM4	0.467	0.776	1.95%	988	TGM2	0.104	1.188	2.30%
18	UHRF1	0.467	0.596	2.05%	989	NPDC1	0.09	1.233	2.28%
19	TPX2	0.464	0.688	2.04%	990	CCL14	0.084	0.625	1.14%
20	KRTCAP3	0.463	0.641	1.73%	991	IER2	0.071	1.298	2.47%

Table 2: Hallmark gene set over-representation analysis for the top-50 and bottom-50 genes ranked by aleatoric uncertainty and epistemic uncertainty fraction, reporting gene set overlap, FDR-adjusted p-value, and overlapping genes; no significantly enriched pathways were identified among genes with the lowest epistemic uncertainty fraction.

Uncertainty Group	Hallmark pathway	Overlap	FDR-adjusted P-value	Overlapping genes
Epistemic High	Epithelial Mesenchymal Transition	11/200	6.25e-11	JUN;SPARC;LGALS1;IGFBP4;FLNA;VIM;TGFBI;THBS1;SAT1;CD44;RHOB
	Apoptosis	7/161	2.22e-06	JUN;CCND2;LMNA;TXNIP;SAT1;CD44;RHOB
	Myogenesis	6/200	7.92e-05	SMTN;KLF5;SPARC;IGFBP7;SORBS1;CNN3
	p53 Pathway	6/200	7.92e-05	GPX2;JUN;CCND2;TXNIP;SAT1;ZFP36L1
	TNF-alpha Signaling via NF-kB	5/200	8.67e-04	NFKBIA;SOC3;NR4A1;ZFP36;MYC;BHLHEP40;PLAUR;FOS;PHLDA1;KLF2;IER2
Aleatoric High	TNF-alpha Signaling via NF-kB	11/200	7.262e-11	CFD;SPARC;SCD;BHLHE40;IGFBP7;CRYAB
	Myogenesis	6/200	1.84e-04	CFD;SERPINA1;SPARC;PECAM1;THBS1
	Coagulation	5/138	2.88e-04	XBP1;MYC;BHLHE40;EMP1;PHLDA1
	IL-2/STAT5 Signaling	5/199	7.19e-04	XBP1;MYC;BHLHE40;EMP1;PHLDA1
	Hypoxia	5/200	7.19e-04	ZFP36;BHLHE40;PLAUR;TGFBI;FOS
Aleatoric Low	G2-M Checkpoint	6/200	1.43e-04	POLQ;CENPF;EXO1;KIF23;KIF2C;KIF11
	Mitotic Spindle	4/199	7.28e-03	CENPF;KIF23;KIF2C;KIF11
	E2F Targets	4/200	7.28e-03	UBE2T;KIF2C;CIT;SPC25

3.3. *Uncertainty varies by Tissue Architecture*

Differences in uncertainty decomposition were observed across histologic architectures by comparing gene-averaged uncertainty, computed by averaging uncertainty across all 991 genes for each spot and subsequently summarized within each tissue architecture. Total uncertainty differed significantly across the nine annotated architecture categories (Kruskal-Wallis $H = 22{,}713$, $p < 0.0001$), as did its aleatoric ($H = 22{,}608$, $p < 0.0001$) and epistemic ($H = 16{,}329$, $p < 0.0001$) components individually. Neoplastic (cancer) epithelium had the lowest mean aleatoric uncertainty of the nine (0.622), while slightly higher for benign epithelium (aleatoric uncertainty=0.978). Epistemic uncertainty was heightened across regions of inflammation and benign epithelium (**Figure 4**).

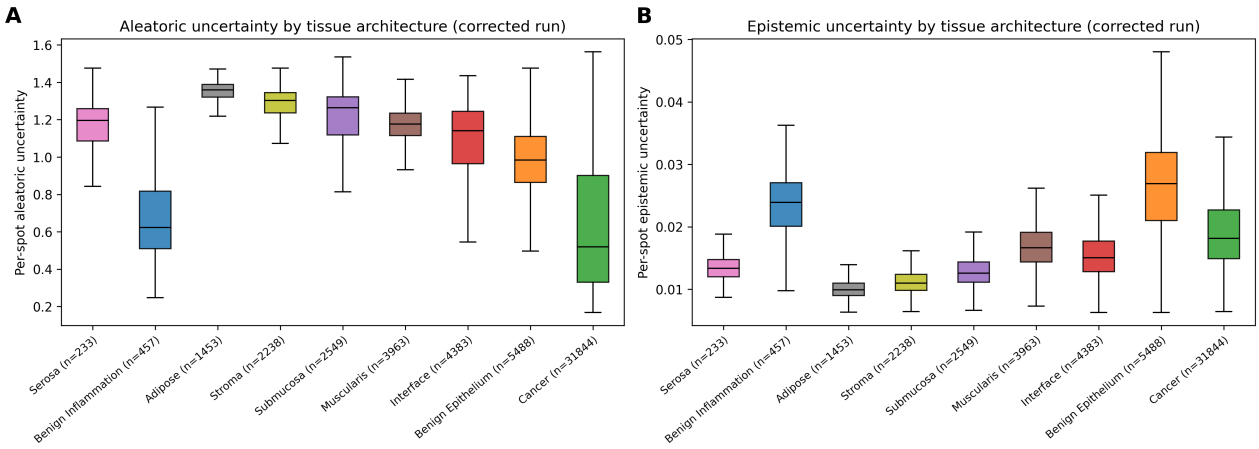


Figure 4: Per-spot predictive uncertainty across annotated histologic architectures, averaged over all 991 genes. **A)** Aleatoric and **B)** epistemic uncertainty for each of the nine architecture categories in the held-out test set, ordered from rarest to most common.

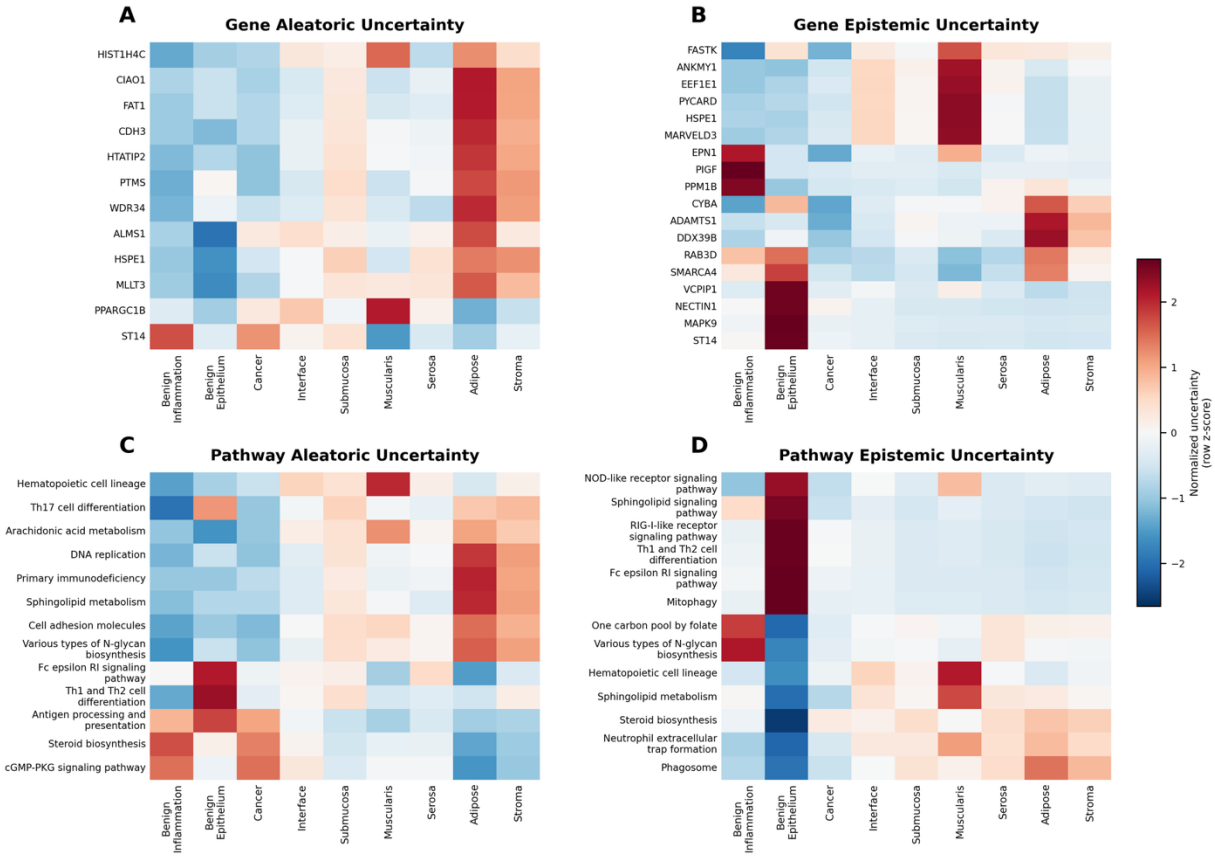


Figure 5: Hierarchical clustering of uncertainty by tissue histology architecture. Mean **A)** aleatoric and **B)** epistemic uncertainty for representative genes, and mean **C)** aleatoric and **D)** epistemic uncertainty aggregated over KEGG gene sets, shown across the nine annotated architecture categories. Red indicates relatively higher uncertainty, while blue indicates relatively lower uncertainty. Uncertainties were normalized prior to plotting, with blue indicating less uncertainty and red indicating greatest uncertainty. Ordering of rows was determined by consensus hierarchical clustering across the four panels to find architectures most related by their pathway uncertainty profiles.

3.4. *Hierarchical Clustering of Pathway-Architecture Uncertainty*

Hierarchical clustering of pathway-level uncertainty demonstrated that the biological programs exhibiting the greatest predictive uncertainty varied substantially across histologic architectures (**Figure 5**). Distinct architecture-specific patterns were observed, broadly separating metabolic and biosynthetic pathways from innate and adaptive immune signaling programs. Metabolic and biosynthetic pathways, including sphingolipid metabolism, N-glycan biosynthesis, and DNA replication, exhibited elevated aleatoric uncertainty within stromal and adipose regions. In contrast, immune-related pathways, including Th1/Th2 cell differentiation, antigen processing and presentation, and mitophagy, demonstrated comparatively greater epistemic uncertainty, concentrated in benign epithelium and, secondarily, inflammation. Overall, these findings suggest that uncertainty is both pathway- and tissue-specific, with metabolically associated programs predominately characterized by irreducible (aleatoric) uncertainty, whereas immune-associated programs exhibited proportionally greater reducible (epistemic) uncertainty.

4. Discussion

While predictive uncertainty has become an increasingly important component of virtual spatial transcriptomics, most efforts have focused on quantifying or reducing overall uncertainty rather than

understanding its underlying sources. Here, we present a proof-of-principle application of an approximate Bayesian deep learning framework based on Monte Carlo dropout and a heteroscedastic ResNet-34 to decompose predictive uncertainty into its epistemic and aleatoric components. Although preliminary, our findings suggest that uncertainty decomposition provides a practical framework for distinguishing uncertainty that may be reduced through additional data collection or improved modeling from uncertainty that likely reflects intrinsic limitations of morphology-based gene expression inference.

An initial observation from this study was that predictive uncertainty was overwhelmingly dominated by the aleatoric component at convergence. A training size ablation (25-100% of the cohort) found that for ~97% of genes ($p < 1e\text{-}230$), epistemic uncertainty decreased as additional training spots were included. Aleatoric uncertainty showed no trend. More broadly, these findings align with the growing interest in data valuation and active data selection strategies, speculating that prioritizing training examples with the greatest expected contribution to model performance may also preferentially reduce epistemic uncertainty while leaving aleatoric uncertainty largely unchanged [38,39]. Such approaches, when used in tandem, could provide a principled strategy for identifying which additional tissue samples, architectures or biological states would most effectively improve virtual spatial transcriptomic models.

Gene-level uncertainty patterns further support this conclusion. Highly predictable genes (e.g., EPCAM, CD24, ELF3, GRHL2) were enriched for epithelial differentiation and adhesion markers, whose expression may be more consistently reflected in histologic architecture, whereas poorly predicted genes were enriched for immediate-early genes (IEGs) / transcription factors involved in transient signaling responses which may not relate to morphologic changes [17]. Notably, both groups exhibited relatively low epistemic uncertainty, suggesting that the poor performance of IEGs is unlikely to be resolved through additional training data. Instead, these findings support the hypothesis that certain programs are weakly linked to tissue morphology, representing intrinsic, biologic limits of morphology-based gene expression inference rather than deficiencies in model training. However, it is important to note that these findings were derived from the top 991 spatially variable genes more likely to have expression tied to tissue architecture. Extending uncertainty decomposition to broader transcriptome-wide gene sets may reveal substantially different uncertainty profiles and presents further opportunity to appreciate the limitations of morphology-based inference.

The elevated aleatoric uncertainty observed for metabolic and biosynthetic pathways within stromal and adipose regions may reflect heterogeneous cellular composition, reduced cellularity, and lower transcript counts per Visium spot. Conversely, the concentration of epistemic uncertainty within immune signaling pathways, particularly in inflamed and epithelial regions, likely reflects limited representation of these comparatively rare morphologic and immunologic states in the training cohort, suggesting that targeted acquisition of underrepresented tissue architectures or immune microenvironments could preferentially improve prediction of these biological programs. Tumor regions exhibited the lowest median aleatoric uncertainty but the greatest variability, likely reflecting strong morphology-transcript correspondence for highly expressed epithelial genes alongside substantial intratumoral heterogeneity, including differences in tumor differentiation across patients and variable stromal and immune admixture within annotated tumor regions. Finally, architecture size may influence uncertainty through increased sampling, although larger regions

such as stroma and submucosa may also exhibit greater intrinsic uncertainty due to lower transcript abundance.

Several methodological considerations should be acknowledged. This proof-of-principle study employed an approximate Bayesian deep learning framework based on Monte Carlo dropout rather than a fully Bayesian neural network with explicit prior distributions over model parameters. Although Monte Carlo dropout provides an efficient and scalable approximation to Bayesian inference, predictive uncertainty estimates are influenced by architectural choices, including dropout placement, dropout probability, and the implicit regularization imposed through the L2 weight penalty associated with the variational approximation. In the present study, dropout was introduced only within the final residual blocks to balance predictive performance; however, alternative dropout configurations may yield different uncertainty estimates. Likewise, we intentionally used a conventional ResNet-34 backbone instead of pathology foundation model representations to isolate the effects of uncertainty decomposition without the additional influence of large-scale transfer learning. More broadly, exhaustive comparisons across alternative uncertainty-aware architectures and Bayesian inference strategies, including but not limited to Hamiltonian Monte Carlo, stochastic variational inference, Laplace approximations, fully Bayesian neural networks, spatially informed Gaussian processes, graph neural networks, and physics-based diffusion models, were beyond the scope of this initial study but represent important directions for future work [40–45]. A calibration assessment has not been performed, and as such, variances are relative rankings. Heteroscedasticity can also be incorporated into a ZINB likelihood. Rather than identifying an optimal Bayesian framework, the objective of this work was to further interrogate uncertainty decomposition as supportive interpretation of virtual spatial transcriptomics that can be extended to more sophisticated probabilistic models to guide targeted data acquisition, model refinement, and experimental design.

Looking ahead, uncertainty decomposition has implications extending well beyond improving virtual spatial transcriptomic models. Predicted gene expression is increasingly used for downstream analyses, including differential expression, pathway enrichment, and spatial colocalization analyses, where uncertainty in the initial predictions may propagate through subsequent statistical analyses. Incorporating predictive uncertainty into these downstream workflows could improve the interpretability of biological conclusions. Likewise, uncertainty-aware predictions have the potential to enhance clinician trust by explicitly identifying transcriptomic programs or tissue regions where predictions are less reliable, facilitating more informed interpretation and decision-making [46]. Integrating uncertainty decomposition with causal machine learning frameworks may further distinguish biological limitations from systematic sources of predictive uncertainty, including institution-specific effects, scanner characteristics, tissue processing, collection time, RNA degradation, and other measured (patient/tumor characteristics) or unmeasured confounders [47], while enabling principled propagation of uncertainty from individual spatial locations to patient-level biomarkers, prognostic models, and clinical decision support technologies [48]. Together, these future directions could strengthen the role of uncertainty decomposition in trustworthy and interpretable computational pathology workflows.

5. Acknowledgments

The authors acknowledge support from the National Institutes of Health (R37CA299857 and P20GM130454), Cedars-Sinai Medical Center, the Cedars Cancer Developmental Funds (Team Science and Molecular Twins Awards), the Dartmouth Cancer Center (5P30CA023108-47), philanthropic support from Dartmouth Health Pathology, and the EDIT AI Program. The authors also acknowledge travel support provided TP to present preliminary findings from this work at the American Association for Cancer Research (AACR) Annual Meeting, 2026. Large language models (ChatGPT, OpenAI) were used to assist with language editing, manuscript organization, and code development. All scientific content, analyses, interpretations, and final manuscript revisions were reviewed and verified by the authors, who take full responsibility for the accuracy and integrity of the work.

References

1. Srinivasan, G. *et al.* Histology-Based Virtual RNA Inference Identifies Pathways Associated With Metastasis Risk in Colorectal Cancer. *Modern Pathology* **38**, 100866 (2025).
2. Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA: A Cancer Journal for Clinicians* **73**, 233–254 (2023).
3. Akimoto, N. *et al.* Rising incidence of early-onset colorectal cancer — a call to action. *Nat Rev Clin Oncol* **18**, 230–243 (2021).
4. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31–46 (2022).
5. Galon, J. *et al.* Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* **313**, 1960–1964 (2006).
6. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat Methods* **19**, 534–546 (2022).
7. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
8. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**, 425–436 (2021).
9. Schmauch, B. *et al.* A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nature Communications* **11**, 3877 (2020).
10. He, B. *et al.* Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* **4**, 827–834 (2020).
11. Pang, M., Su, K. & Li, M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. 2021.11.28.470212 Preprint at https://doi.org/10.1101/2021.11.28.470212 (2021).
12. Zeng, Y. *et al.* Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. *Briefings in Bioinformatics* **23**, bbac297 (2022).
13. Xu, H. *et al.* A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**, 181–188 (2024).
14. Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nat Med* **30**, 850–862 (2024).
15. Chen, J. *et al.* STimage-1K4M: A histopathology image-gene expression dataset for spatial transcriptomics. Preprint at https://doi.org/10.48550/arXiv.2406.06393 (2024).
16. Jaume, G. *et al.* HEST-1k: A Dataset for Spatial Transcriptomics and Histology Image Analysis. Preprint at https://doi.org/10.48550/arXiv.2406.16192 (2024).
17. Bahrami, S. & Drabløs, F. Gene regulation in the immediate-early response process. *Advances in Biological Regulation* **62**, 37–49 (2016).
18. Qiao, C. & Huang, Y. Reliable imputation of spatial transcriptomes with uncertainty estimation and spatial regularization. *Patterns (N Y)* **5**, 101021 (2024).
19. Zeng, Y. *et al.* Imputing spatial transcriptomics through gene network constructed from protein language model. *Commun Biol* **7**, 1271 (2024).
20. Sun, E. D., Ma, R., Navarro Negredo, P., Brunet, A. & Zou, J. TISSUE: uncertainty-calibrated prediction of single-cell spatial transcriptomics improves downstream analyses. *Nat Methods* **21**, 444–454 (2024).
21. Andersson, A. *et al.* Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun Biol* **3**, 565 (2020).
22. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* **40**, 661–671 (2022).
23. Li, H. *et al.* A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nat Commun* **14**, 1548 (2023).
24. Lopez, R. *et al.* DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol* **40**, 1360–1369 (2022).

25. Tan, X. *et al.* Robust and interpretable prediction of gene markers and cell types from spatial transcriptomics data. *Nat Commun* **17**, 1781 (2026).
26. Wang, C. *et al.* CHRep: Cross-modal Histology Representation and Post-hoc Calibration for Spatial Gene Expression Prediction. Preprint at https://doi.org/10.48550/arXiv.2604.21573 (2026).
27. Jin, K. *et al.* Multiscale confidence quantification for virtual spatial transcriptomics with UTOPIA. *bioRxiv* 2026.03.01.708850 (2026) doi:10.64898/2026.03.01.708850.
28. Chakraborti, T. *et al.* Personalized uncertainty quantification in artificial intelligence. *Nat Mach Intell* **7**, 522–530 (2025).
29. Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach Learn* **110**, 457–506 (2021).
30. Neal, R. M. *Bayesian Learning for Neural Networks*. (Springer Science & Business Media, 2012).
31. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. (CRC press, 2020).
32. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *arXiv:1506.02142 [cs, stat]* http://arxiv.org/abs/1506.02142 (2016).
33. Rączkowski, Ł., Możejko, M., Zambonelli, J. & Szczurek, E. ARA: accurate, reliable and active histopathological image classification framework with Bayesian deep learning. *Sci Rep* **9**, 14347 (2019).
34. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).
35. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat Methods* **15**, 343–346 (2018).
36. Seitzer, M., Tavakoli, A., Antic, D. & Martius, G. On the Pitfalls of Heteroscedastic Uncertainty Estimation with Probabilistic Neural Networks. in (2021).
37. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *cels* **1**, 417–425 (2015).
38. Fatemi, M. Y. *et al.* An initial game-theoretic assessment of enhanced tissue preparation and imaging protocols for improved deep learning inference of spatial transcriptomics from tissue morphology. *Briefings in Bioinformatics* **25**, bbae476 (2024).
39. Tang, S. *et al.* Data valuation for medical imaging using Shapley value and application to a large-scale chest X-ray dataset. *Sci Rep* **11**, 8366 (2021).
40. Cobb, A. D. & Jalaian, B. Scaling Hamiltonian Monte Carlo inference for Bayesian neural networks with symmetric splitting. in *Uncertainty in Artificial Intelligence* 675–685 (PMLR, 2021).
41. Izmailov, P., Vikram, S., Hoffman, M. D. & Wilson, A. G. G. What are Bayesian neural network posteriors really like? in *International conference on machine learning* 4629–4640 (PMLR, 2021).
42. Weyrauch, A. *et al.* torch_blue: A Flexible Python Package for Bayesian Neural Networks in PyTorch. *JOSS* **11**, 9415 (2026).
43. Krishnan, R., Esposito, P. & Subedar, M. Bayesian-torch: Bayesian neural network layers for uncertainty estimation. *Zenodo* https://ui.adsabs.harvard.edu/abs/2022zndo...5908307K/abstract (2022).
44. Lafage, A., Laurent, O., Gabetni, F. & Franchi, G. Torch-uncertainty: Deep learning uncertainty quantification. *Advances in Neural Information Processing Systems* **38**, (2026).

45. Gardner, J., Pleiss, G., Weinberger, K. Q., Bindel, D. & Wilson, A. G. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. *Advances in neural information processing systems* **31**, (2018).
46. Kompa, B., Snoek, J. & Beam, A. L. Second opinion needed: communicating uncertainty in medical machine learning. *npj Digit. Med.* **4**, 4 (2021).
47. Howard, F. M. *et al.* The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nat Commun* **12**, 4423 (2021).
48. Chernozhukov, V. *et al.* Double/debiased machine learning for treatment and structural parameters. (2018).