Reading the Residual: Understanding Uncertainty in Virtual RNA Inference from H&E in Colorectal Cancer

Alice Qian, Joshua Levy PhD,
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

ABSTRACT

- Spatial transcriptomics measures gene expression in tissue but is far less accessible than routine H&E histology. Virtual RNA inference (VRI) could extend molecular information to H&E-only slides, but its usefulness also depends on whether the model can recognize when its predictions are unreliable.
- We developed an uncertainty-aware, single-cell VRI pipeline using paired Xenium and H&E from colorectal cancer. The model captured several gene programs well across patients (best ρ = 0.71), yet prediction accuracy and uncertainty reliability were largely independent across 20 programs (ρ = −0.22)
- Conformal calibration restored near-nominal coverage across all pathways, while epistemic uncertainty failed to increase for a new patient despite patient identity being detectable in the frozen feature space. **Accuracy alone is insufficient for evaluating trustworthy VRI and that important sources of error may lie outside the model's own uncertainty estimates.**

INTRODUCTION

- **H&E histology:** routine and widely archived, but shows tissue morphology rather than molecular activity.
- **Spatial transcriptomics:** measures gene expression while preserving cell location, but it is far more costly and less widely available.
- **Virtual RNA Inference (VRI):** deep learning approach that predicts molecular state directly from H&E, potentially extending transcriptomic information to H&E-only tissue archives.
- Existing VRI studies focused mainly on prediction accuracy, often at spot-level resolution where multiple cells are pooled together.
- H&E-only slides: no transcriptomic ground truth
- Goal: Build and evaluate a single-cell VRI that provides useful molecular predictions and uncertainty estimates, test whether that uncertainty is actually reliable + characterise where and why it fails.

METHODS

Study design

- Paired same-section Xenium + H&E colorectal cancer: 2 patients, 8 sections, 3.23M cells, 322 genes
- Train Patient A (60,000 cells) → evaluate on Patient B (300,344)
- **Four-Part Error Framework**

Image encoding + Prediction targets

- Extracted 40 μm H&E crops centered on individual cells
- Frozen UNI pathology foundation model converted each crop into a 1024-D feature vector
- Final targets: 20 gene programs (8 curated + 12 Enrichr)

Model

- **Small MLP: 1024 → 512 → 256**
- Outputs for each pathway: Predicted pathway score + Cell-specific variance (aleatoric uncertainty)
- Trained with heteroscedastic Gaussian loss

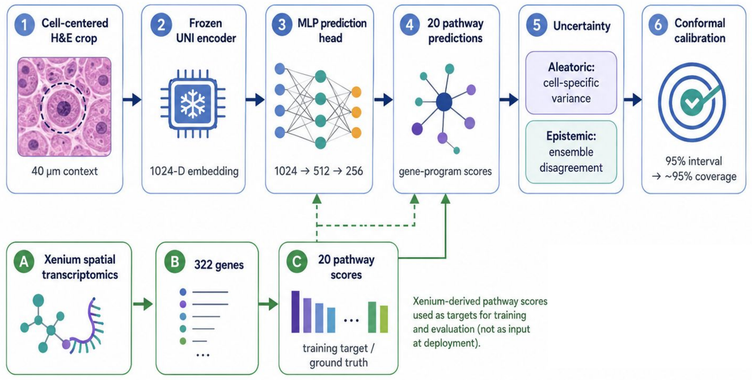
METHODS

Uncertainty & calibration

- Compared 5 uncertainty methods: MC dropout, 2 Bayesian NNs, packed ensemble, deep ensemble
- **Deep ensemble (10 independent MLP heads)** used as primary model (Model disagreement → epistemic uncertainty)
- Conformal calibration: held-out errors to adjust prediction intervals

Evaluation

- **Evaluated externally**: setup/batch + unexplained residual
- Prediction accuracy: Spearman ρ between predicted and Xenium pathway scores
- Uncertainty reliability measured by whether uncertainty increased with actual prediction error



RESULTS

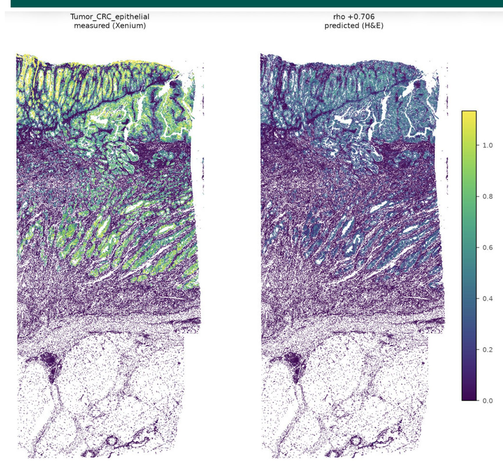
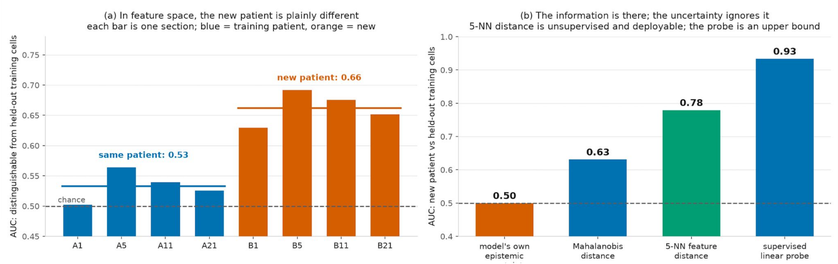


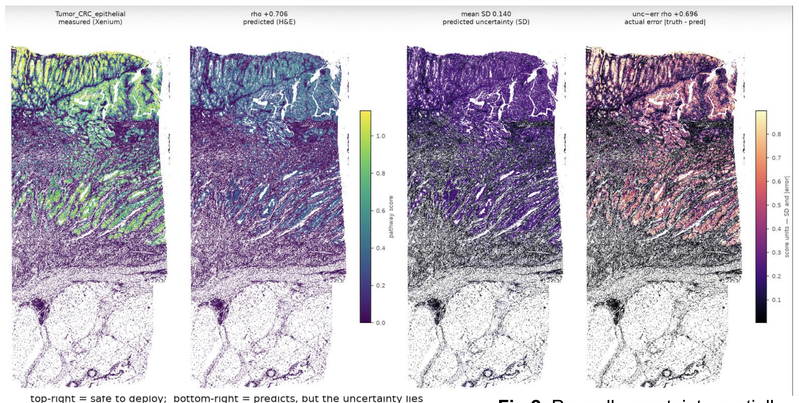
Fig 1: Cross-patient prediction: mean ρ = **0.296** (20 pathways); best ρ = **0.706**.

method	accuracy ρ	uncertainty vs error
Packed ensemble	+0.212 ± 0.006	+0.477 ± 0.020
Deep ensemble (10 models)	+0.211 ± 0.004	+0.502 ± 0.008
Variational BNN	+0.202 ± 0.008	+0.388 ± 0.020
torchbayesian BNN	+0.191 ± 0.010	+0.256 ± 0.015
MC dropout	+0.190 ± 0.011	+0.413 ± 0.013

Table 1: Five uncertainty methods performed similarly



RESULTS



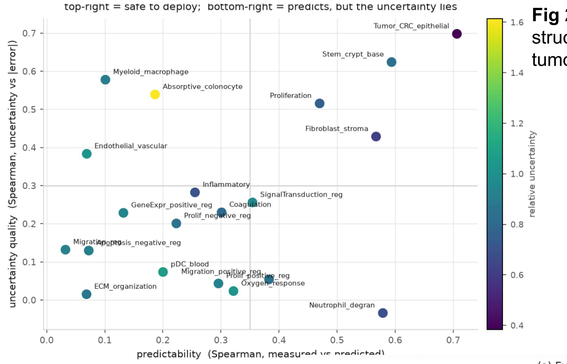
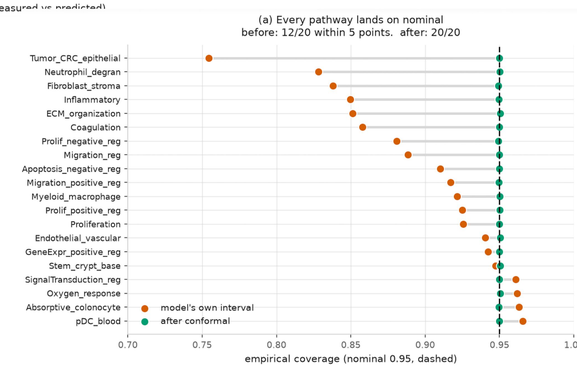


Fig 2. Per-cell uncertainty spatially structured+aligned with actual error for tumor epithelium (ρ = 0.70)

Fig 3. Accuracy vs. uncertainty reliability: ρ = −0.22

Fig 4. Nominal 95% coverage: 12/20 → 20/20 pathways after conformal calibration

Best-predicted pathway: 74.5% → ~95% coverage (Intervals: ~36% wider)



CONCLUSION

- Single-cell VRI can recover meaningful molecular signal from H&E, but performance varies substantially by pathway.
- Prediction accuracy and uncertainty reliability are distinct, an accurate prediction can still have uninformative confidence.
- Conformal calibration can restore reliable coverage even when the model's own uncertainty is miscalibrated.
- In our frozen-UNI setup, ensemble epistemic uncertainty missed an unfamiliar patient, despite patient novelty being present in the feature representation.
- Model-reported uncertainty captures only part of the failure landscape; patient/setup effects and unexplained residuals require external evaluation.

edit.