Gene-Specific Cross-Attention vs Global Pooling for Virtual RNA Inference

Ishan Ramrakhiani

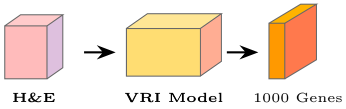
Q1: RESEARCH QUESTION & PURPOSE

Problem: VRI models lack optimal spatial reasoning and biological context integration.

> **Research Question:** *Can gene-specific cross-attention and text supervision improve VRI accuracy and interpretability?*

Significance: Colorectal cancer (107,320 cases/yr, 3rd most common). VRI enables gene expression prediction from routine H&E histology.

VRI Concept

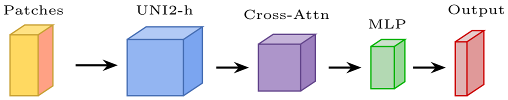


Q2: METHODOLOGY & APPROACH

Two Approaches:

1. Architectural: Gene-specific cross-attention + UNI2-h (681M params)
2. Biological: Text supervision via MedVLM-R1 pathology captions

Model Architecture



Data: H&E WSIs + Visium spatial transcriptomics (1000 genes)

Q3: RESULTS & KEY FINDINGS

Cross-Attention vs. Pooling Results (Mean Correlation)

Method	Mean	Best Seed
Global Pooling	0.442	0.472
Cross-Attention	**0.463**	**0.499**
Improvement	**+4.7%**	+5.7%

Key Findings:

- Consistent improvement across all 8 random seeds
- Gene-specific attention enables selective spatial reasoning
- Each gene learns to focus on relevant tissue regions
- Simple MLP heads sufficient with strong frozen features

Q4: DISCUSSION & IMPACT

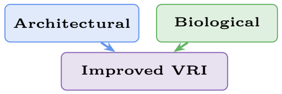
Conclusions:

1. Cross-attention (0.463) outperforms pooling (0.442)
2. Models learn which tissue regions matter per gene
3. Expression localizes to tissue compartments (crypts, stroma)

Impact: Scalable gene prediction from routine pathology. Enables precision medicine and therapeutic target identification.

Future: Richer text alignment, larger datasets, morphology-gene modeling.

Framework



Architectural
Biological
Improved VRI