Text-based Supervised Fine-tuning for Virtual RNA Inference Models

Pranay Kancharla, Ishan Ramrakhiani

Emerging Diagnostic and Investigative Technologies, Department of Pathology and Laboratory Medicine, Dartmouth Health

Dartmouth Health

ABSTRACT

Colon cancer remains the second leading cause of cancer-related deaths in the United States, with an estimated 52,900 annual deaths. To address the need for biologically coherent models of gene expression, we develop a proof-of-concept framework for predicting spatial gene expression directly from H&E-stained colon histology patches. Our baseline uses a frozen UNI2-h pathology foundation model paired with a minimal two-layer regression head (1536→1024→1000), which proved to be both stable and compute-efficient. We further fine-tune this baseline using MedVLM-R1–generated captions, applying a positive-only cosine alignment objective to encourage alignment between predicted gene activity and morphological features. Systematic ablations of model depth, width, dropout, and batch normalization yielded no consistent performance improvements, reinforcing the value of the simplest baseline. This study highlights the potential of incorporating text-based supervision to improve interpretability while maintaining robust predictive performance.

INTRODUCTION

Introduction

•Spatial transcriptomics links tissue morphology with gene expression, offering insight into tumor biology.

•Current VRI models show promise but often lack interpretability, limiting their clinical utility.

•Foundation models such as UNI2-h provide strong histology features but are typically used in a purely image-to-gene setting.

•Integrating language supervision can bridge morphology and biology, encouraging predictions that better reflect observable tissue features.

•Our study explores a lightweight framework combining frozen pathology features with text-based alignment to test this idea.

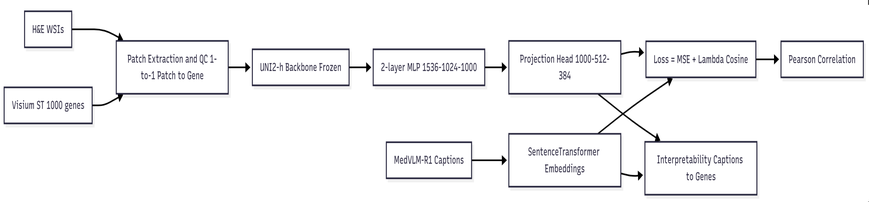

Figure 1: General workflow

METHODS

Data & Preprocessing

- Whole-slide H&E colon images paired with Visium spatial transcriptomics (1000 spatially variable genes).
- Patch extraction: One-to-one mapping between patches and gene vectors, aligned to Visium coordinates.
- Split: Train (29 samples), Validation (6), Test (5).

Model

- UNI2-h frozen backbone (681M params, ViT-H/14).
- Regression head: 2-layer MLP, ReLU, MSE loss.
- Optimizer: Adam (lr=1e-4), early stopping, gradient clipping.

Text Supervision

- Captions from MedVLM-R1 (≤25 words, concise morphology description).
- Embeddings via SentenceTransformer.
- Alignment head maps gene predictions → text embedding space (positive-only cosine alignment).
- Training loss: MSE + λ·alignment (λ=0.2, with warmup).

RESULTS

Seed	Final Mean Gene Correlation	Best Mean Gene Correlation	Best Epoch
40	0.419	0.423	2
41	0.435	0.445	2
42	0.414	0.414	1
43	0.403	0.403	1
44	0.463	0.464	2
45	0.452	0.452	1
51	0.457	0.461	3
54	0.451	0.451	1

Figure 2: Results of Training

RESULTS

Baseline

- Simplest 2-layer regression head outperformed deeper/wider/regularized heads.
- Robust across seeds (best correlations ≈0.42–0.47).

Fine-tuning with Captions

- Positive-only alignment → maintained baseline accuracy.
- Gains in interpretability; captions add morphological context to predictions.
- Example captions:
- “Features suggest adenocarcinoma with nuclear atypia and mitotic figures.”
- “Normal colon architecture with glandular patterns and smooth muscle cells.”

CONCLUSION

- **Future Directions:**
  - Experimenting with a further variety of VLM’s to understand which caption type would work based for text-based supervision.
  - Broadening text-based supervision across broader forms of cancer
  - Create a mobile app or web application built around our project
- **Limitations:**
  - Early experiments missed 2 samples; later recomputed but may have influenced initial architecture decisions.
  - Expanding cohort for DH with COBRE funding
- **Data and Code Availability:**
  - Directory to code: dartfs/rc/nosnapshots/V/VaickusL-nb/EDIT_Interns_2025/projects/TEXTGRAD_VRI

Acknowledgements:
Levy Lab, Vivek Pujara, Zarif Azher

edit.