Translation of Single-cell Gene Expression to Histopathological Images via Rectified Flow

Mengbo Wang*, Shourya Verma*, Aditya Malusare, Luopin Wang, Yiyang Lu, Mario Sola, Vaneet Aggarwal, Ananth Grama, Nadia Atallah Lanman

PURDUE UNIVERSITY® | Department of Computer Science

PURDUE UNIVERSITY® | Institute for Cancer Research

NEURAL INFORMATION PROCESSING SYSTEMS

Mapping Transcriptomes to H&E Images

- Spatial transcriptomics align transcriptomes with histopathology morphology and tissue structures, presenting opportunities for biomolecular discovery.

- We present a novel framework, **GeneFlow**, to map single and multi-cell gene expression onto paired cellular images. First to model the inverse problem.
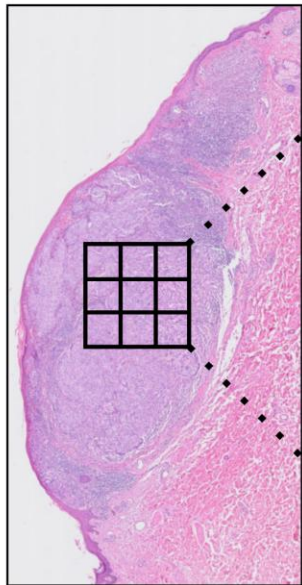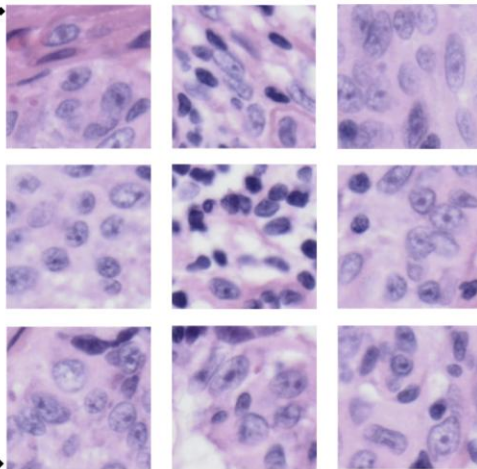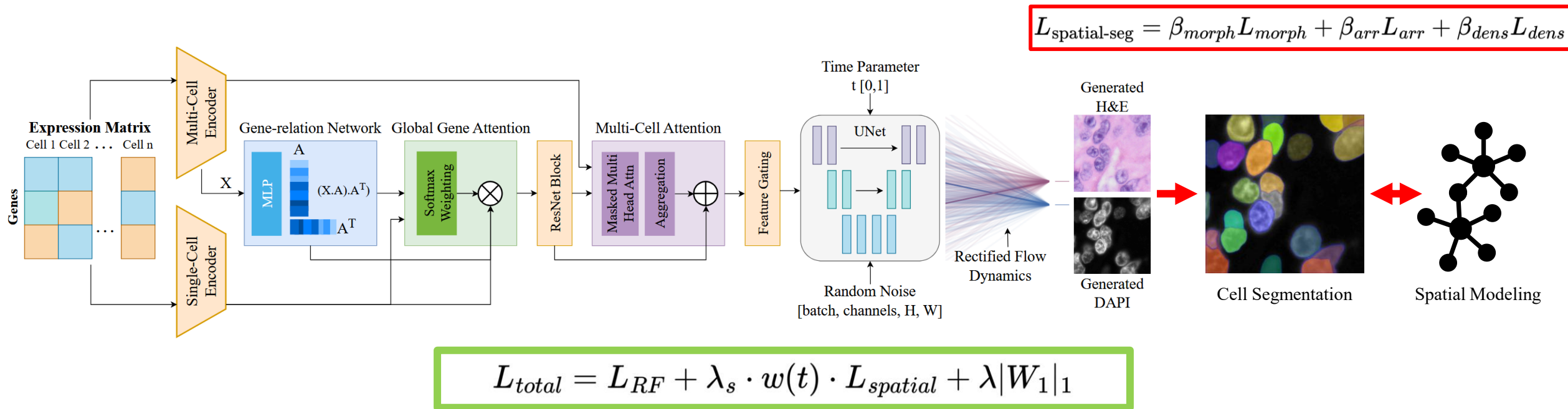
Method and Architecture

Input: Transcriptomics. **Output:** High-resolution H&E / DAPI images.

- Encoder processes image patches containing multiple cells.

- Generating cell-specific embeddings utilizing gene relationships.

- Capturing intra/inter-cell interactions.

- Keeping parameters tractable through low rank factorization.

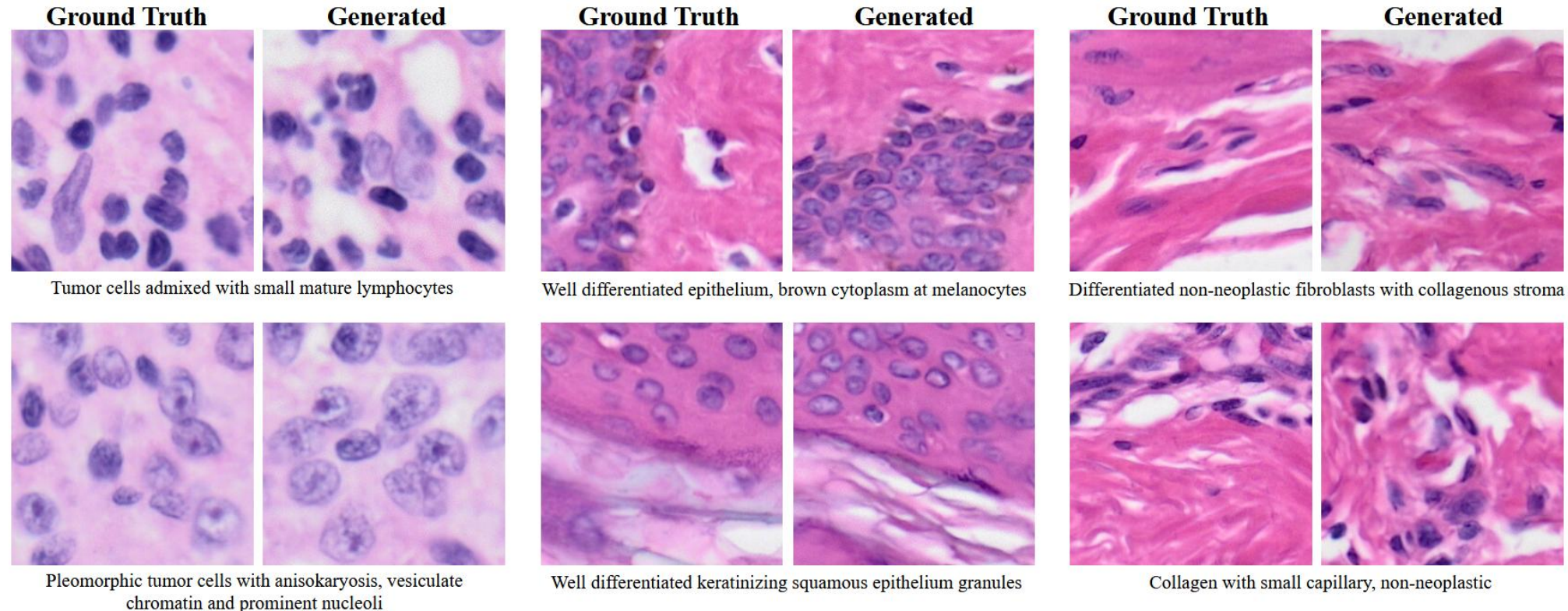$$L_{RF} = \mathbb{E}|v_\theta(x(t), t) - v^*(x(t), t)|^2$$
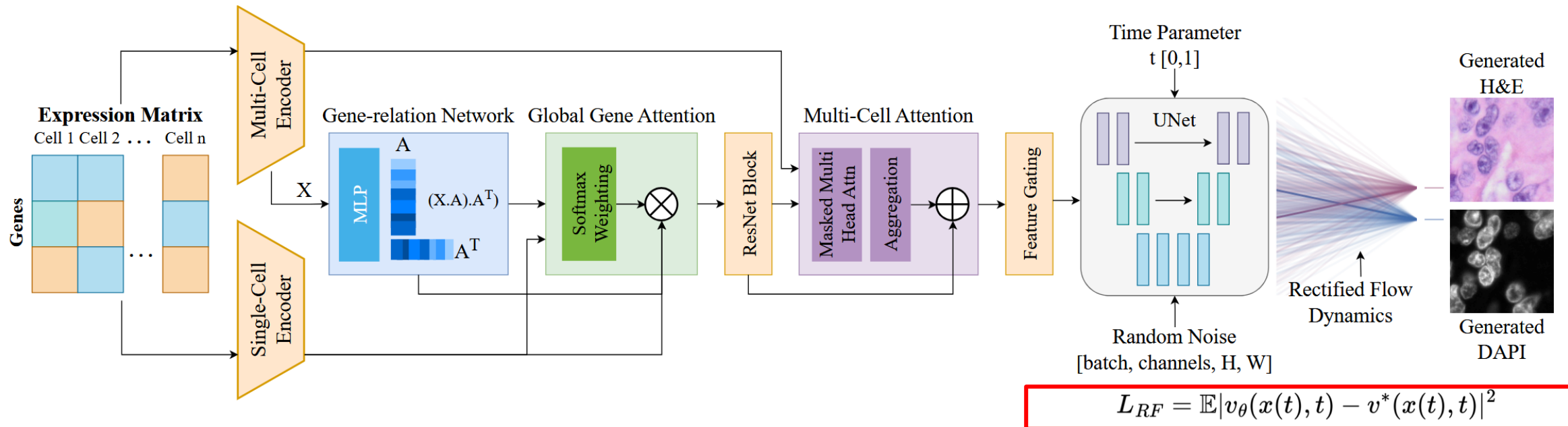
Spatial Modeling

- Spatial Modeling to preserve local tissue organization during training.

- CellPose nuclei segmentation to model morphology, arrangement, density.

- Build kNN graphs over nuclear centroids to align structural relationships

- Enforce local texture & gradient similarity for efficient spatial regularization.

$$L_{\text{spatial-seg}} = \beta_{morph}L_{morph} + \beta_{arr}L_{arr} + \beta_{dens}L_{dens}$$

$$L_{total} = L_{RF} + \lambda_s \cdot w(t) \cdot L_{spatial} + \lambda|W_1|_1$$
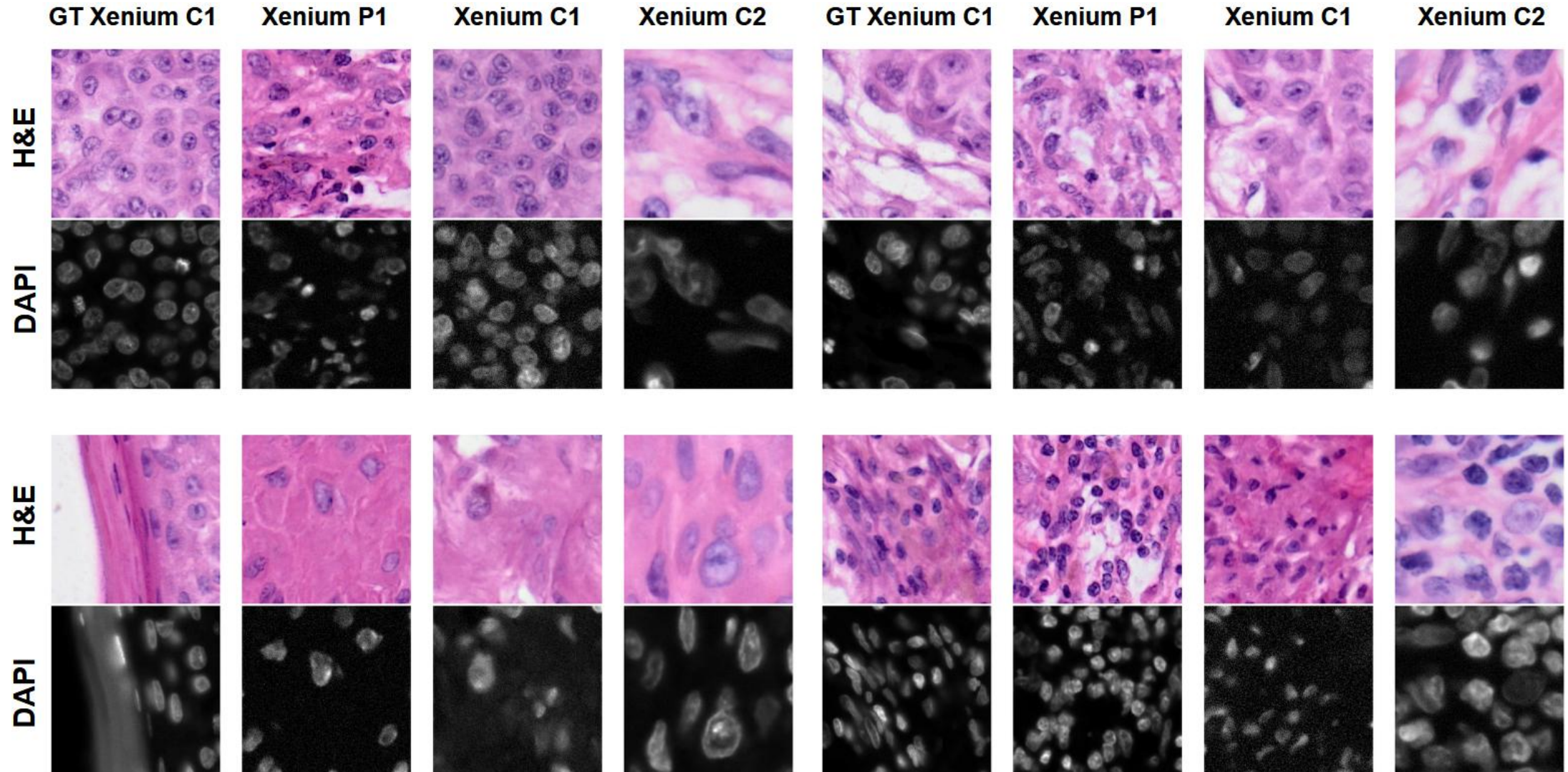
Coherent Tissue Morphologies

- ACVP board certified pathologist blindly diagnosed generated images.

- Diagnosis on (non-)neoplastic cell morphology & spatial structure.

- Generated tissues accurately reproduced key diagnostic features.

Tumor cells admixed with small mature lymphocytes

Well differentiated epithelium, brown cytoplasm at melanocytes

Differentiated non-neoplastic fibroblasts with collagenous stroma

Pleomorphic tumor cells with anisokaryosis, vesiculate chromatin and prominent nucleoli

Well differentiated keratinizing squamous epithelium granules

Collagen with small capillary, non-neoplastic

Cross Dataset Generalizability
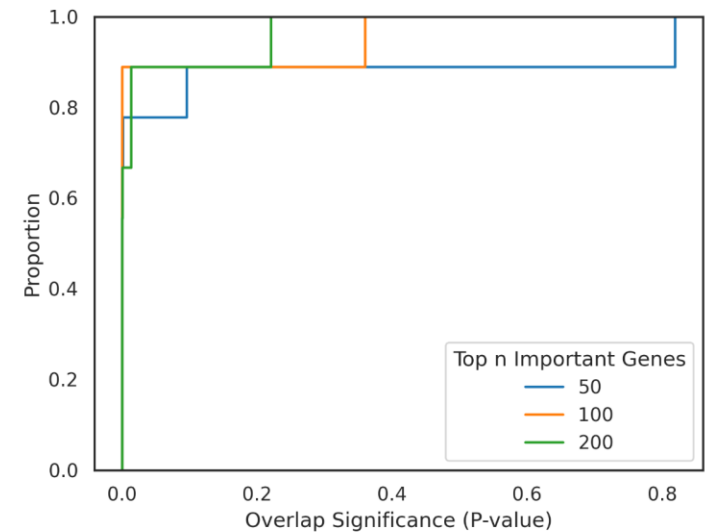
Gene Importance Analysis

- Measured overlap of top 50,100,200 important genes across splits/samples.

- Evaluated overlap significance using a hypergeometric test.

- Comparisons showed strong agreement across splits.

- Models trained on all-cell-type vs melanoma-only data showed >80% overlap.

- No significant overlap between melanoma-only & non-melanoma.
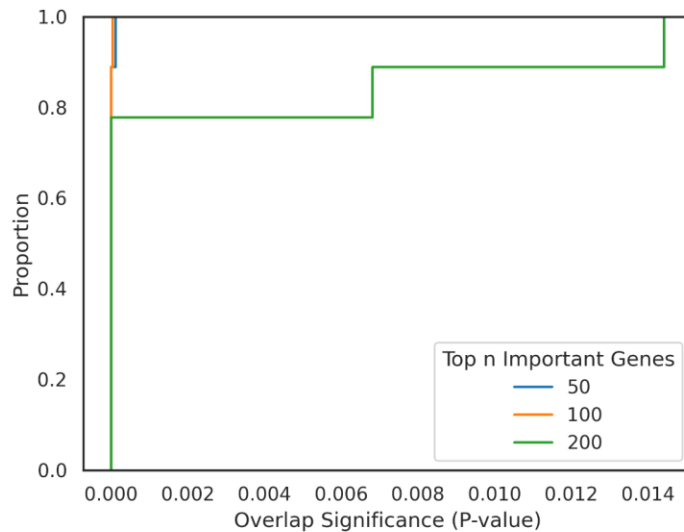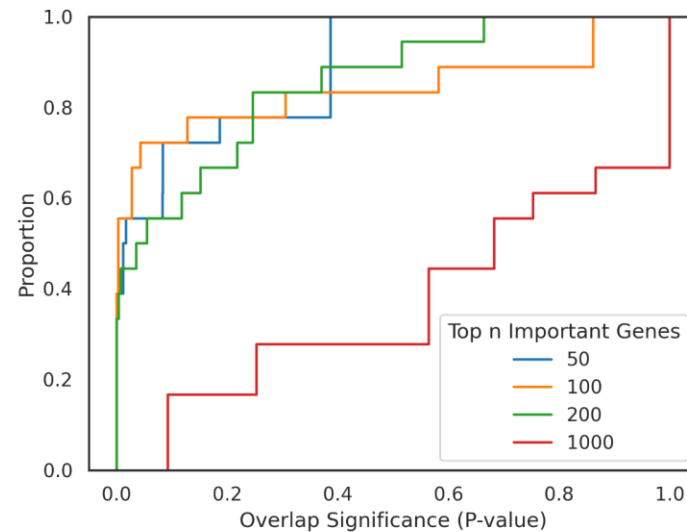

across splits   across samples   only melanoma cells

Quantitative Results

| Metric	UNet (MSE)	C1 Multi Diff.	C1 Multi Rect.	C1 Single Diff.	C1 Single Rect.
Image Quality Metrics ↓					
FID Overall UNI2-h	525.51	404.06	**87.96**	405.98	**39.27**
Inception Feat. Dist.	21.76±1.29	23.05±1.86	**16.72±1.64**	20.18±1.61	**14.50±2.52**
Biological Feature Similarity ↑					
UNI2-h Embedding Sim.	0.964±0.004	0.967±0.007	**0.979±0.004**	0.969±0.007	**0.983±0.003**
Nuclear Circularity Sim.	0.659±0.038	0.835±0.043	**0.844±0.037**	0.839±0.049	**0.874±0.028**
Nuclear Eccentricity Sim.	0.655±0.024	0.869±0.046	**0.954±0.015**	0.880±0.039	**0.964±0.012**
Nuclear Solidity Sim.	0.479±0.037	0.714±0.048	**0.888±0.036**	0.721±0.077	**0.867±0.028**
Spatial Feature Metrics ↑					
Spatial Energy Sim.	0.018±0.034	–	**0.283±0.083**	–	**0.678±0.176**
Spatial Complexity Sim.	0.099±0.060	0.167±0.138	**0.506±0.080**	0.130±0.111	**0.571±0.085**
Spatial Feat. Magnitude Sim.	0.112±0.061	0.167±0.143	**0.514±0.077**	0.135±0.119	**0.571±0.084**

C1 Single-Cell model baseline comparison
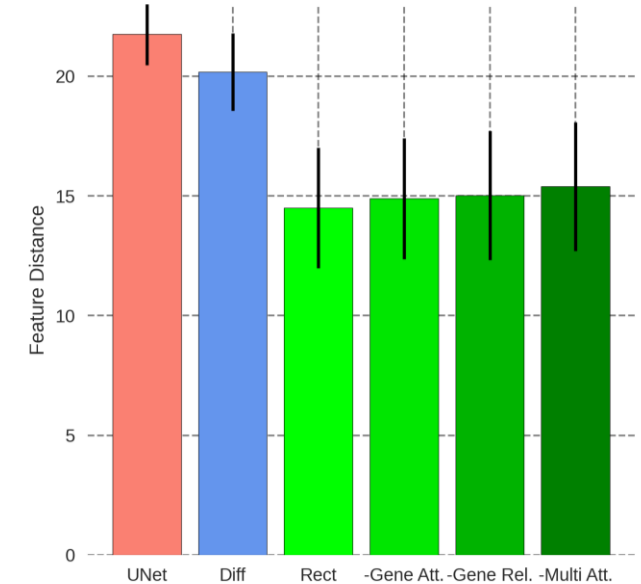
Ablation Study

Term	P-value	Adj. P-value	Odds Ratio	Combined Score	Genes
Epithelial Mesenchymal Transition	2.09E-05	5.02E-04	17.48	188.37	POSTN;LUM;MMP2;IGFBP2;MGP
KRAS Signaling Dn	3.56E-04	4.27E-03	13.45	106.80	TFAP2B;KRT15;IGFBP2;KRT5
Angiogenesis	1.70E-03	1.36E-02	36.64	233.56	POSTN;LUM
Apoptosis	2.55E-03	1.53E-02	12.13	72.44	CCND1;LUM;MMP2
Myogenesis	4.69E-03	2.25E-02	9.71	52.07	ACHE;APOD;CRYAB
Coagulation	2.30E-02	9.18E-02	9.11	34.39	C1QA;MMP2

Gene set enrichment analysis of the most influential genes. Top 10 enriched pathways ranked by adjusted p-value.

Qualitative Results
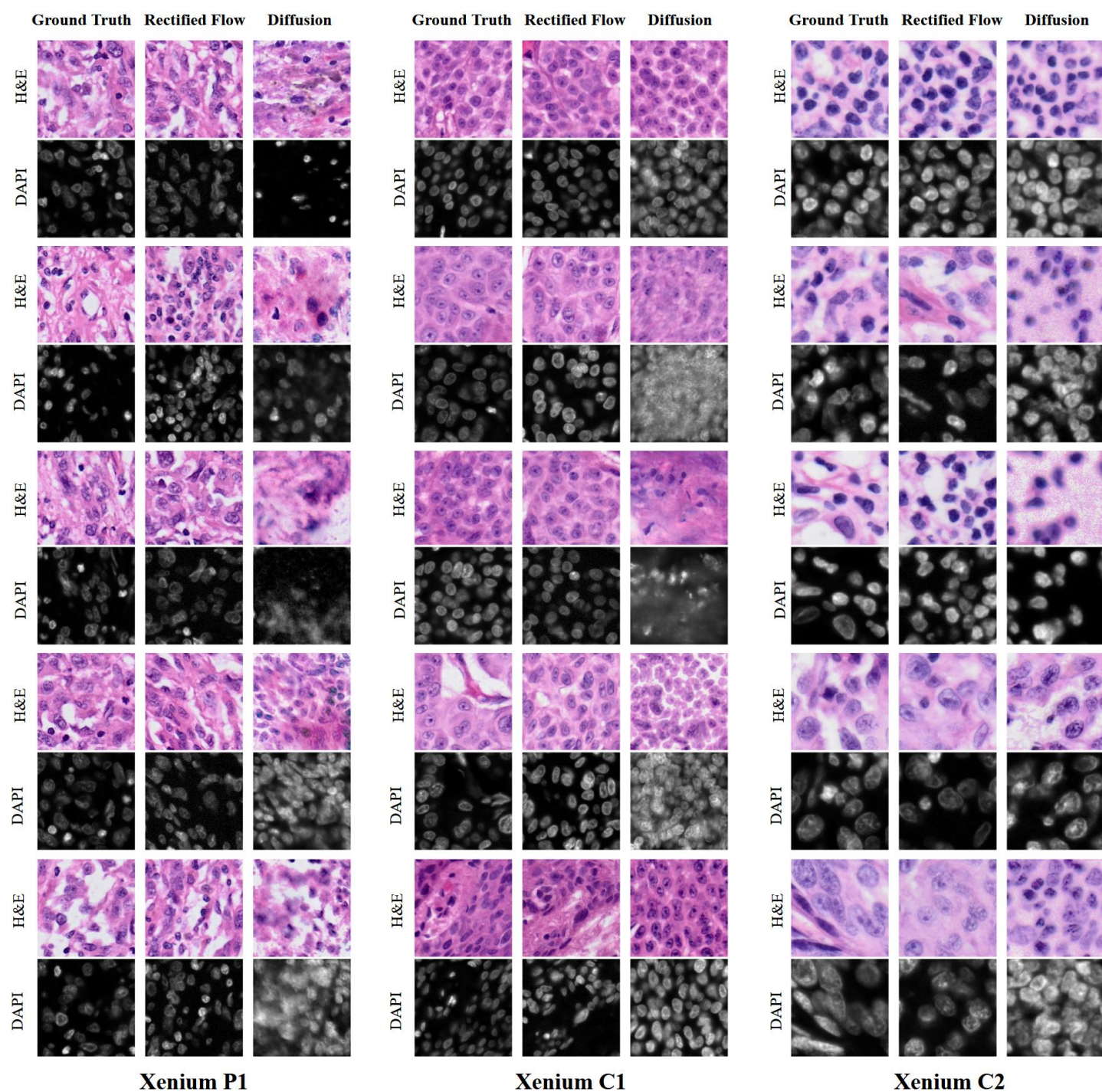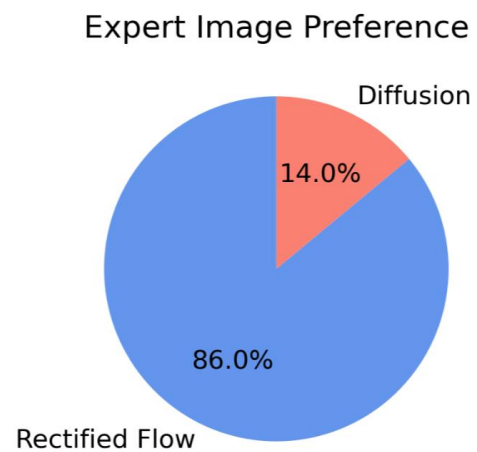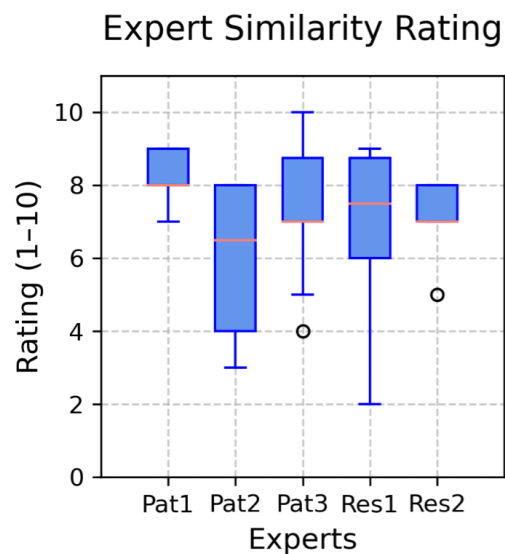
Conclusion

- GeneFlow translates transcriptomes to realistic H&E images, linking gene-expression with cellular morphology.

- Operates on 256×256 tiles & depends on Xenium single-cell data.

- Accurately generates coherent tissue morphologies.

- Outperforms diffusion-based models through rectified flow.

- Validated by quantitative and expert assessments.

- Spatial modeling creates realistic cellular micro-environment.

- **Limitation:**
  - Scalability to larger slides & diverse datasets remains a challenge.

- **Future Work:**
  - Expand to whole-slide synthesis, integrate sc-foundational models, and refine cell-context modeling for complex tissue reconstruction.
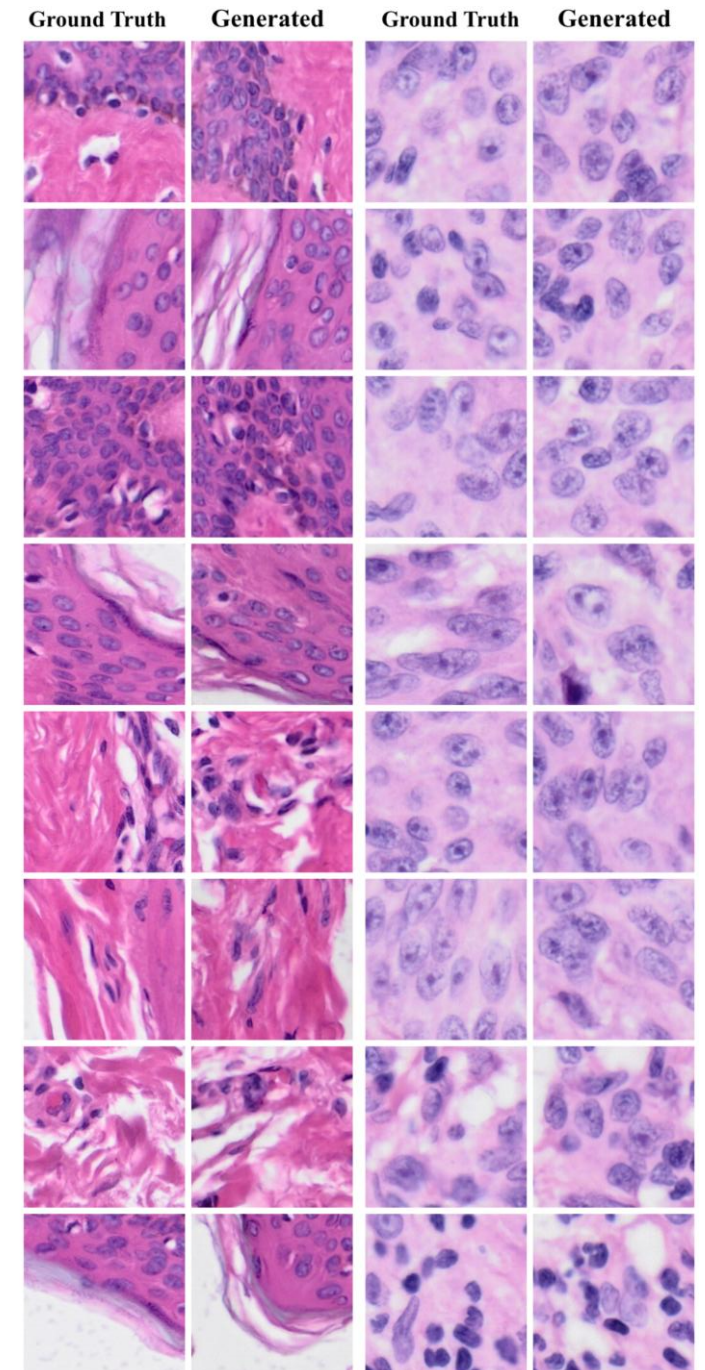
Thank you!

@inproceedings{wanggeneflow,
        title={GeneFlow: Translation of Single-cell Gene Expression to Histopathological Images via Rectified Flow},
        author={Wang, Mengbo and Verma, Shourya and Malusare, Aditya and Wang, Luopin and Lu, Yiyang and
                        Aggarwal, Vaneet and Sola, Mario and Grama, Ananth and Lanman, Nadia Atallah},
        booktitle={The Thirty-ninth Annual Conference on Neural Information Processing Systems},
        year={2025}
}