

# Machine Learning Based Prediction of Proteomic Signatures using H&E Histology and Pathomic Features

Julio Alfredo Maragall | April 19, 2024 | (HuBMAP) HIVE grant OT2 OD033753



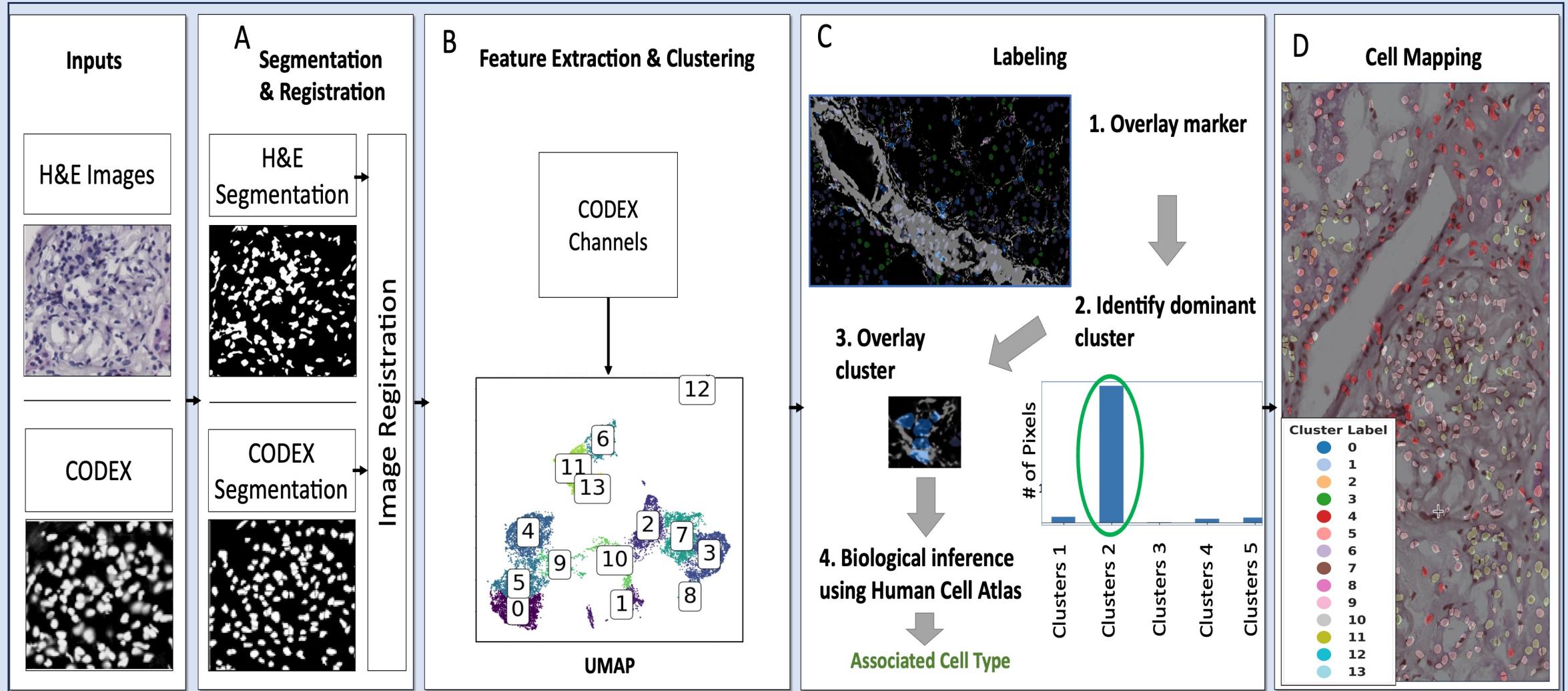
Me and Dr. Naglah are developing a system with the aim of submitting towards a high-impact journal.

We need your feedback!

# Project Objective:

- **segment** individual cells in both images,
- **extract** features from each
- **cluster** feature manifold handle
- **label** the cells
- **map** these labeled clusters back onto the original tissue image

# High-level Overview



# Now The Details:

(Thank you Nicole)

# Positive Updates/Summary

- **Project Overview:**

- **Aim:** To combine CODEX with H&E staining for detailed analysis of renal tissues.
- Enhances diagnostic precision by revealing cellular neighborhoods.
- Bridges detailed molecular insights with conventional histopathology.
- Transforms standard practices by making molecular insights accessible.

- **Positive Updates Since Last Presentation:**

- Advanced integration of CODEX and H&E data, integrating nephrological neighborhood analysis.
- Integrated deep learning models for enhanced biological inference.
- Working with Dr. Ahmed me has taught me how to develop projects systematically.

# IU Data Summary

MAIN STUDY DATA

CODEX (n=21), 32-bit | H&E (n=29), 8-bit

Categories:

Histology

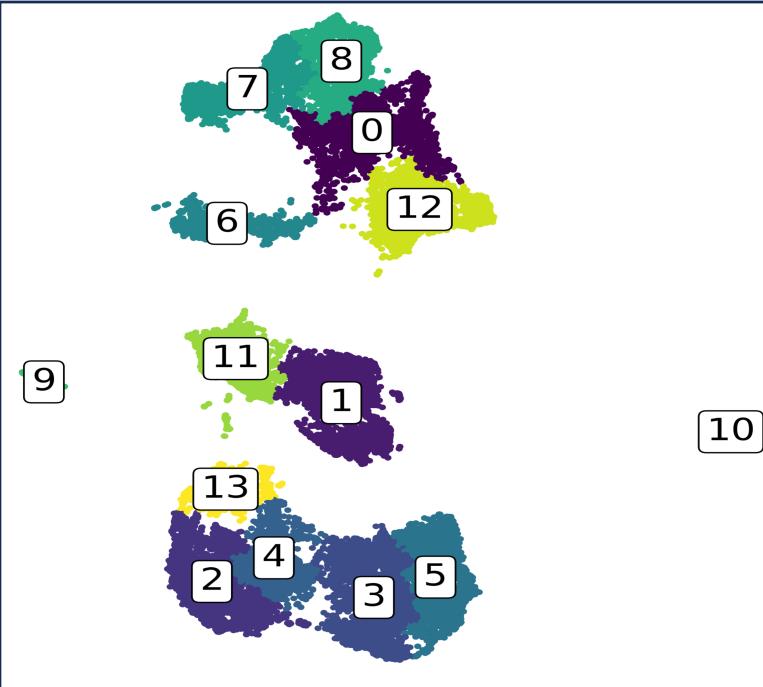
Molecular: Pixel Intensity

# Research Strategy

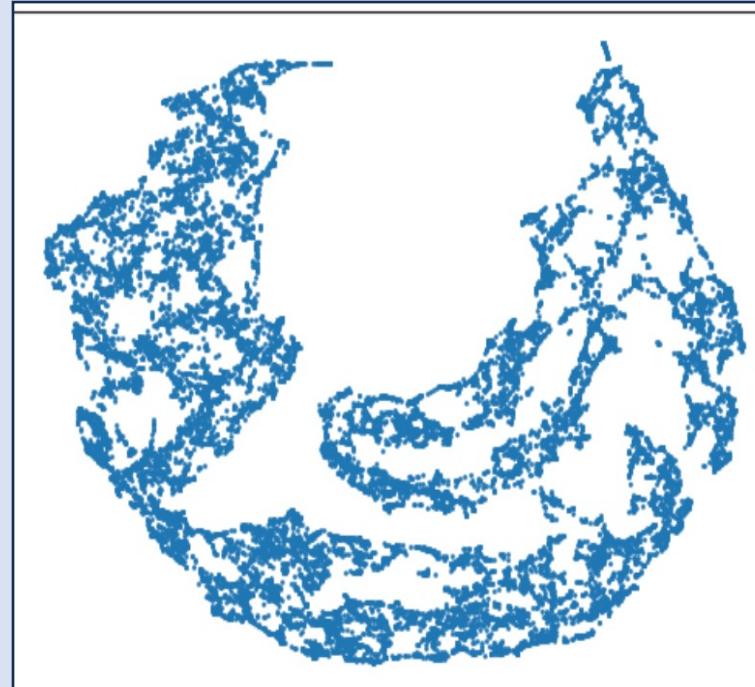
- Direction 1:
  - HE nuclei-level is noisy.
  - Exploring whether we can **aggregate** CODEX data on additional levels of abstraction (e.g., an FTU with glom and 1<sup>st</sup> order neighbor tubules.)
  - We want to **find** signal between HE and CODEX.
- Direction 2:
  - We want to study methodology for **fusing** the histology and proteomic imaging data.
  - Independently from nuclei, glom, FTU's, and neighborhoods, we want to integrate features from histology onto CODEX optimally.
  - We want to **see** how the H&E modality affects CODEX UMAP clusters.

# H&E is Noisy

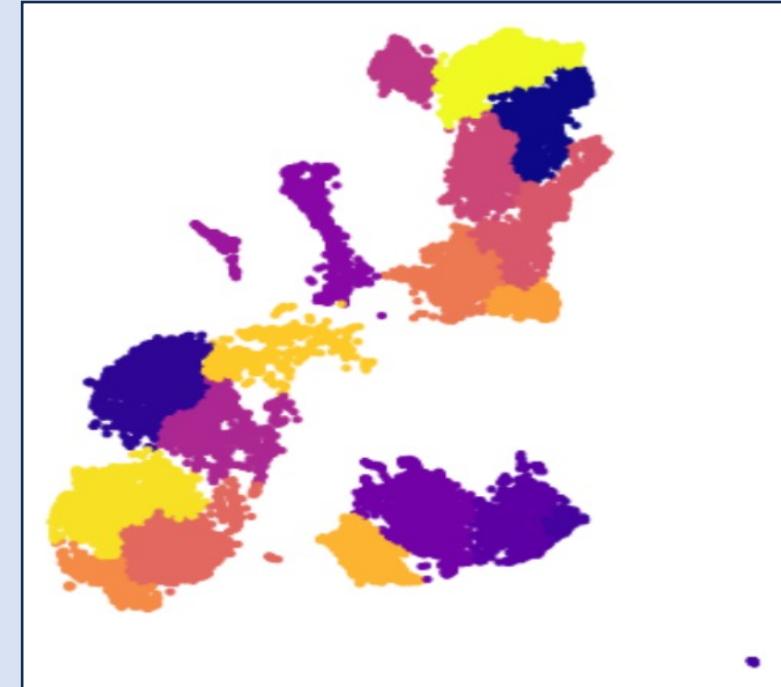
CODEX, Handcrafted



HE, Handcrafted

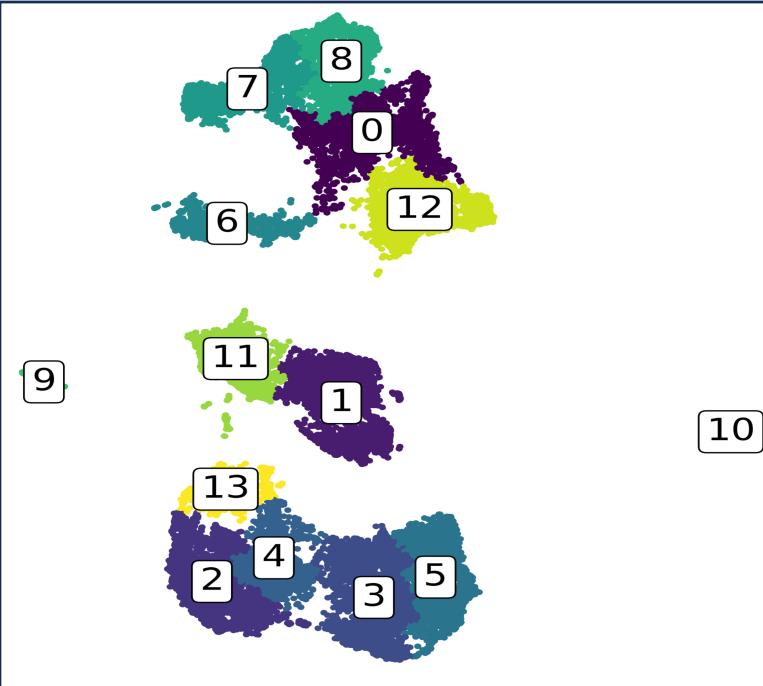


CODEX + HE

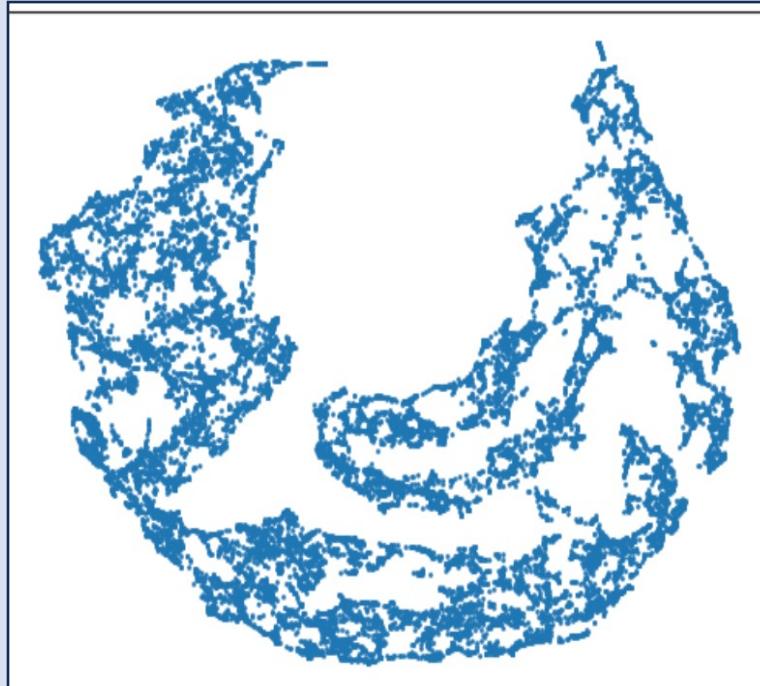


# H&E is Noisy

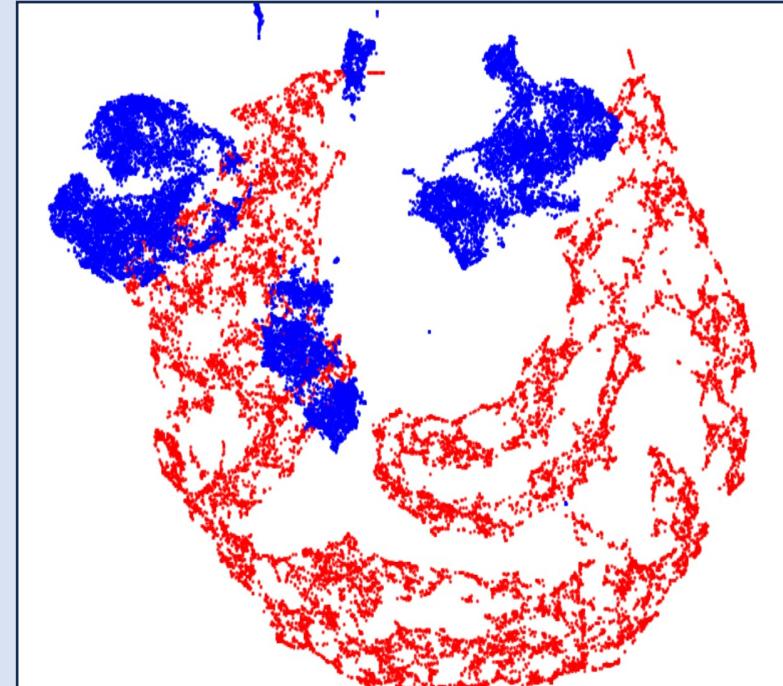
CODEX, Handcrafted



HE, Handcrafted



CODEX + HE

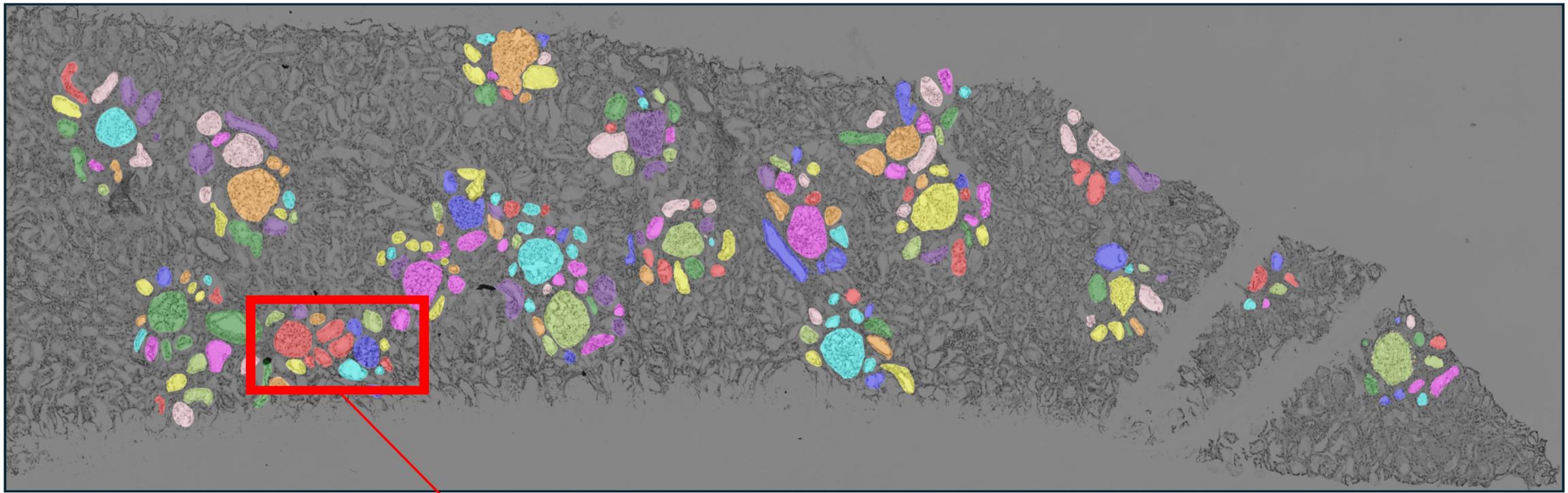


# Direction 1

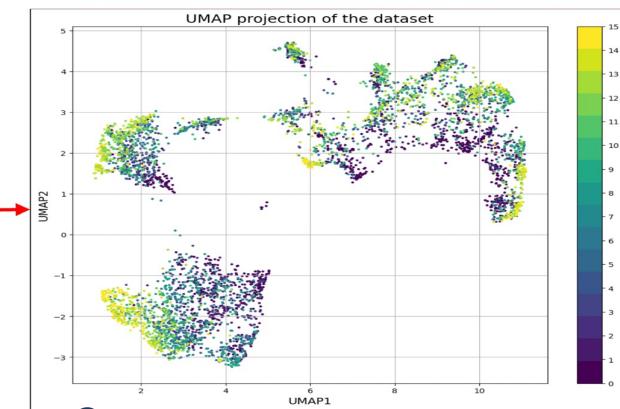
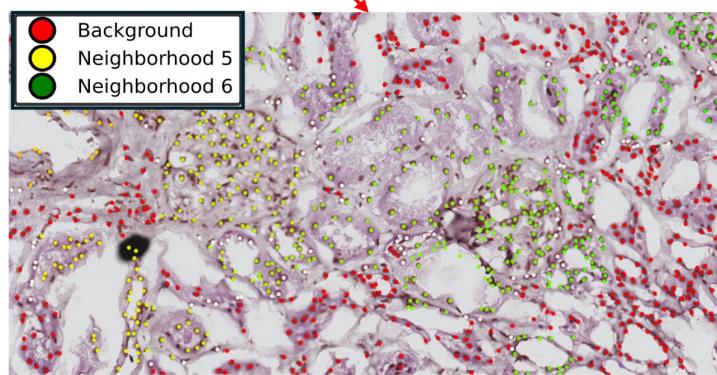
CODEX UMAP Color-coded by nephron neighborhood

# Highlights:

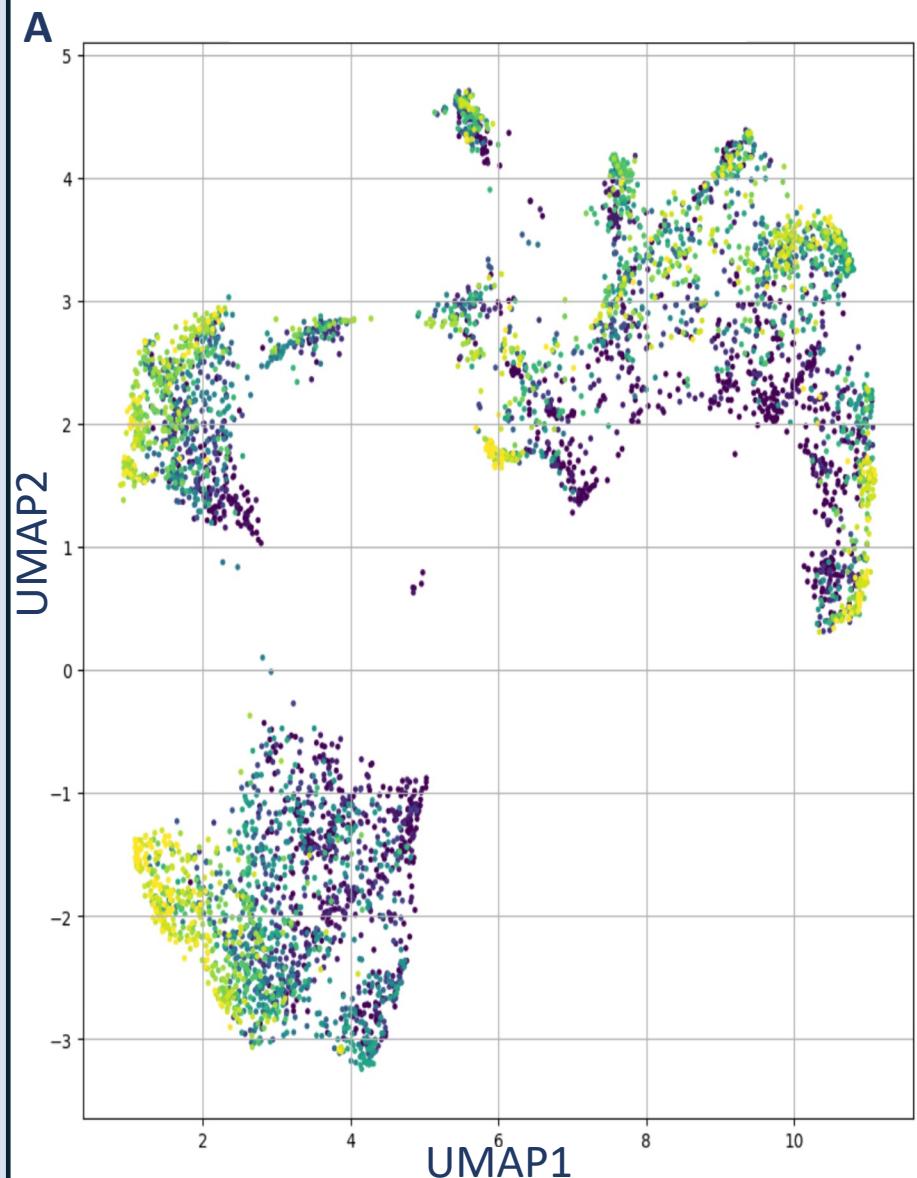
- Developed a pipeline that:
  - Associates nephron membership with CODEX feature manifold.
  - Explains cluster heterogeneity based on nephron membership.
- Filtered dataframe by overlaying spatial locations of nuclei onto a hand-annotated binary mask.
- Associated filtered nuclei with glomeruli.
- Defined glomerular neighborhood as tubules closest to the glomerular object.
- Color coded the UMAP with glomerular neighborhood after association.



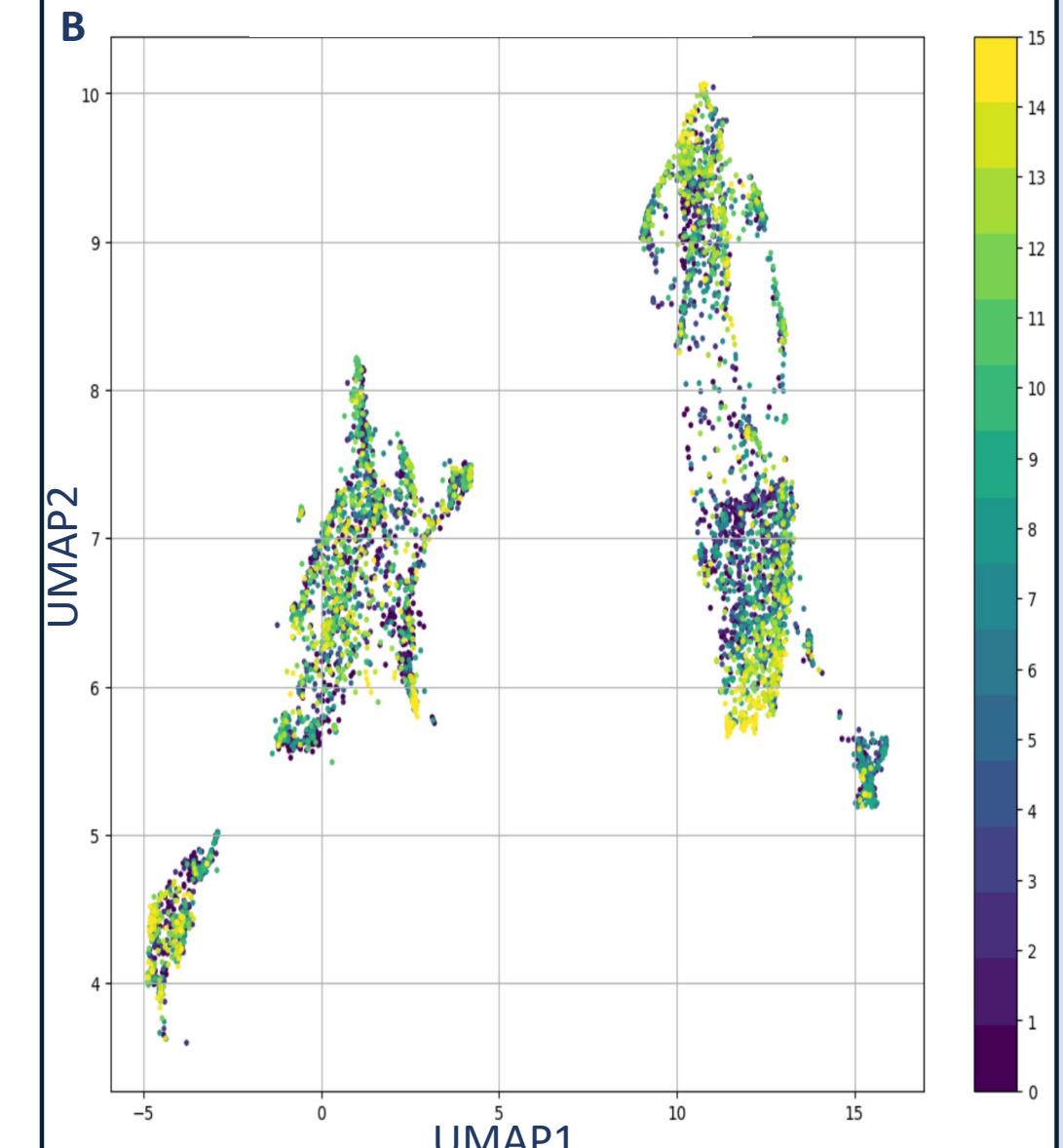
**A**



UMAP projection of the dataset  
Minkowsky distance metric



UMAP projection of the dataset  
Chebyshev distance metric



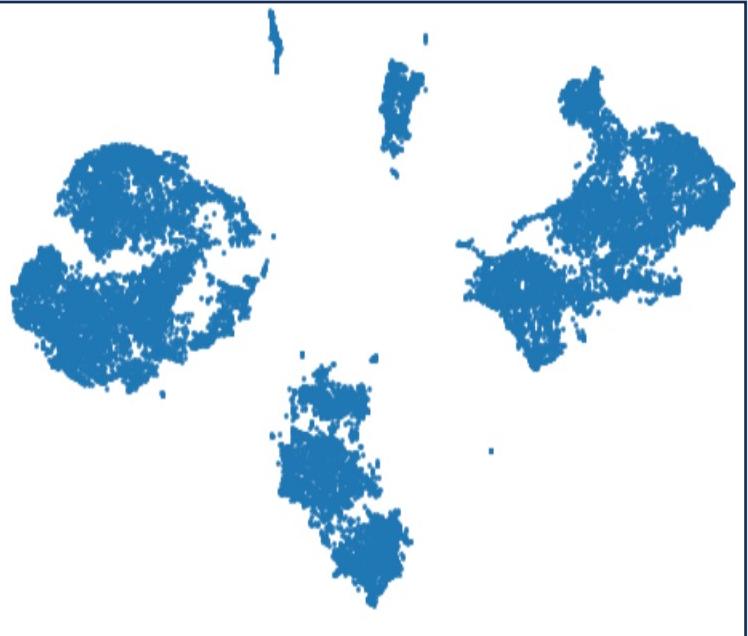
# Direction 2

Weighted HE + CODEX UMAP

## Highlights:

- Weighted HE features (contrast, smoothness, intensity variance, locally linear embedded autoencoder vector) onto CODEX manifold.
- Evaluated mutual information between weighted and unweighted CODEX using element-wise multiplication with varying alpha.
- Graphed CODEX and CODEX+HE similarity (alpha vs. cosine similarity)
- Maximum similarity: smoothness at certain alpha peaks, consistently by contrast.
- Average cosine similarity: 0.6723 for variance and autoencoder vector, 0.8432 for others.
- Some glomerular neighborhoods show homogeneity; 3, 6, and 7 have higher closeness scores.

CODEX, Handcrafted

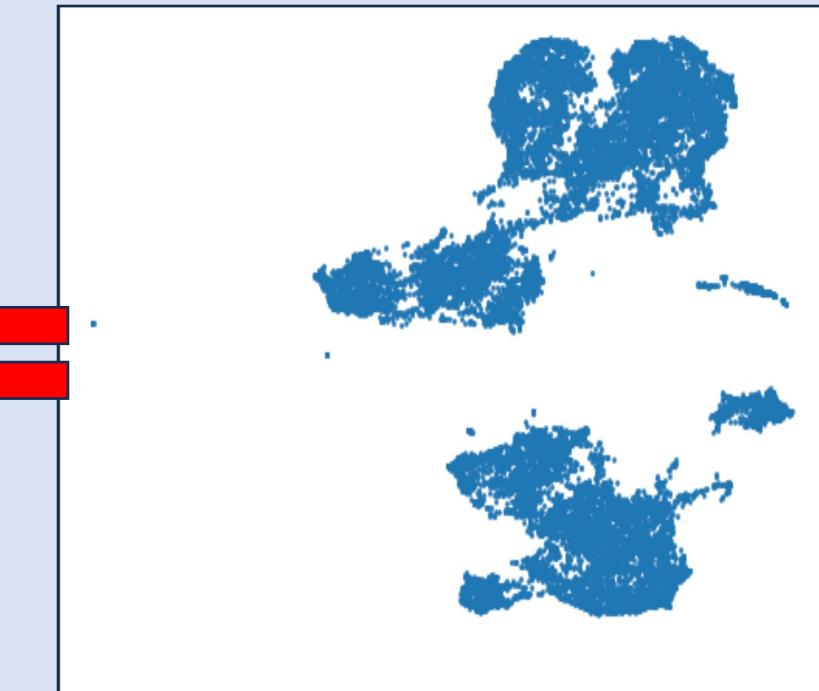


$\alpha$

HE\_autoencoder

AE_Val
0 0.742504
1 0.684280
2 0.734316
3 0.699771
4 0.766316
...
22275 0.660592
22276 0.708865
22277 0.734787
22278 0.718412
22279 0.741837

22280 rows x 1 columns



CODEX, Handcrafted



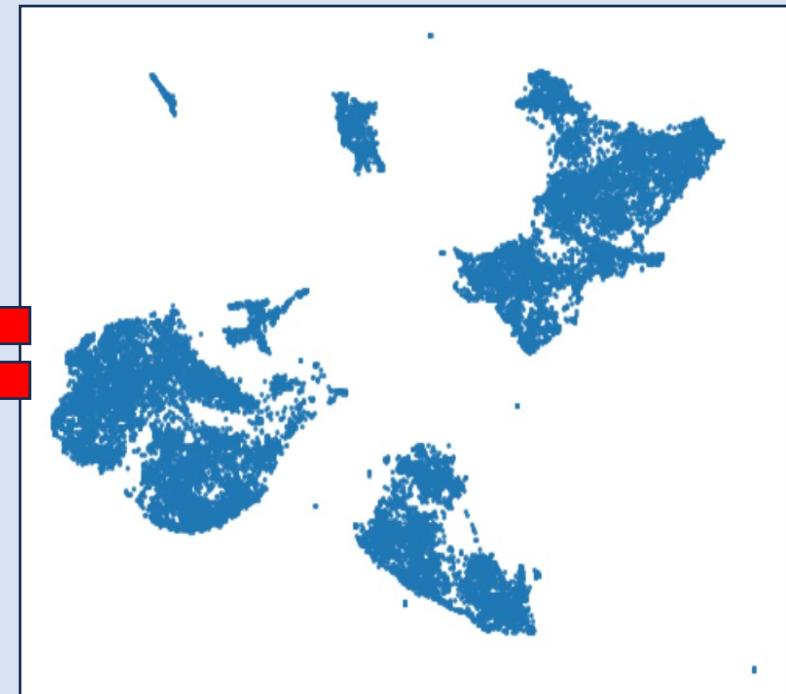
$\alpha$

HE\_Smoothness

Smoothness
0.998823
0.999038
0.999682
0.999386
0.999282
...



22280 X 1



CODEX, Handcrafted



$\alpha$

HE\_Intensity\_Var

Variance_Intensity
848.774722
1038.386537
3147.484627
1627.074481
1390.951334
...
903.165235
4665.596808
814.832070
1230.769759
1891.959180

22280 X 1



CODEX, Handcrafted



$\alpha$

HE\_Contrast

**Contrast**

132

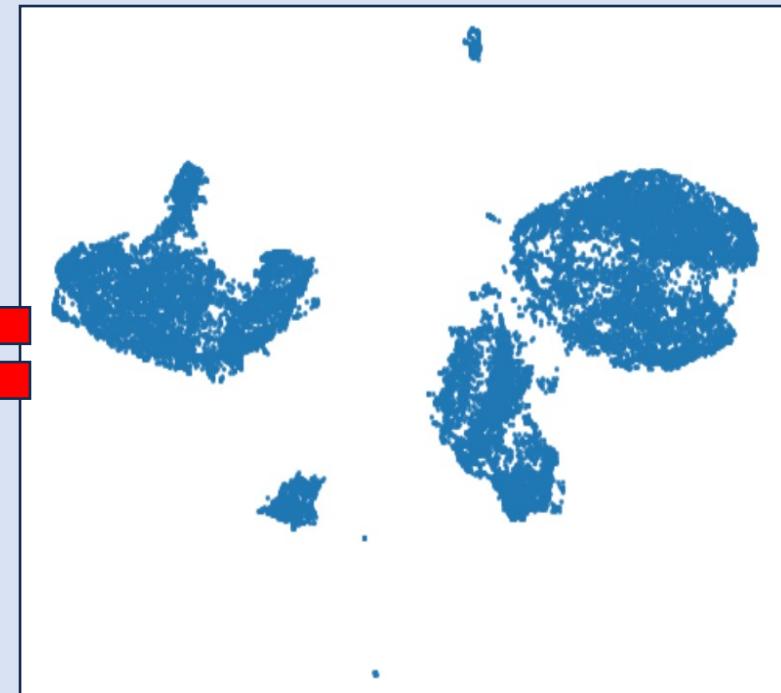
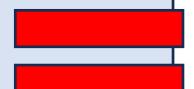
145

203

173

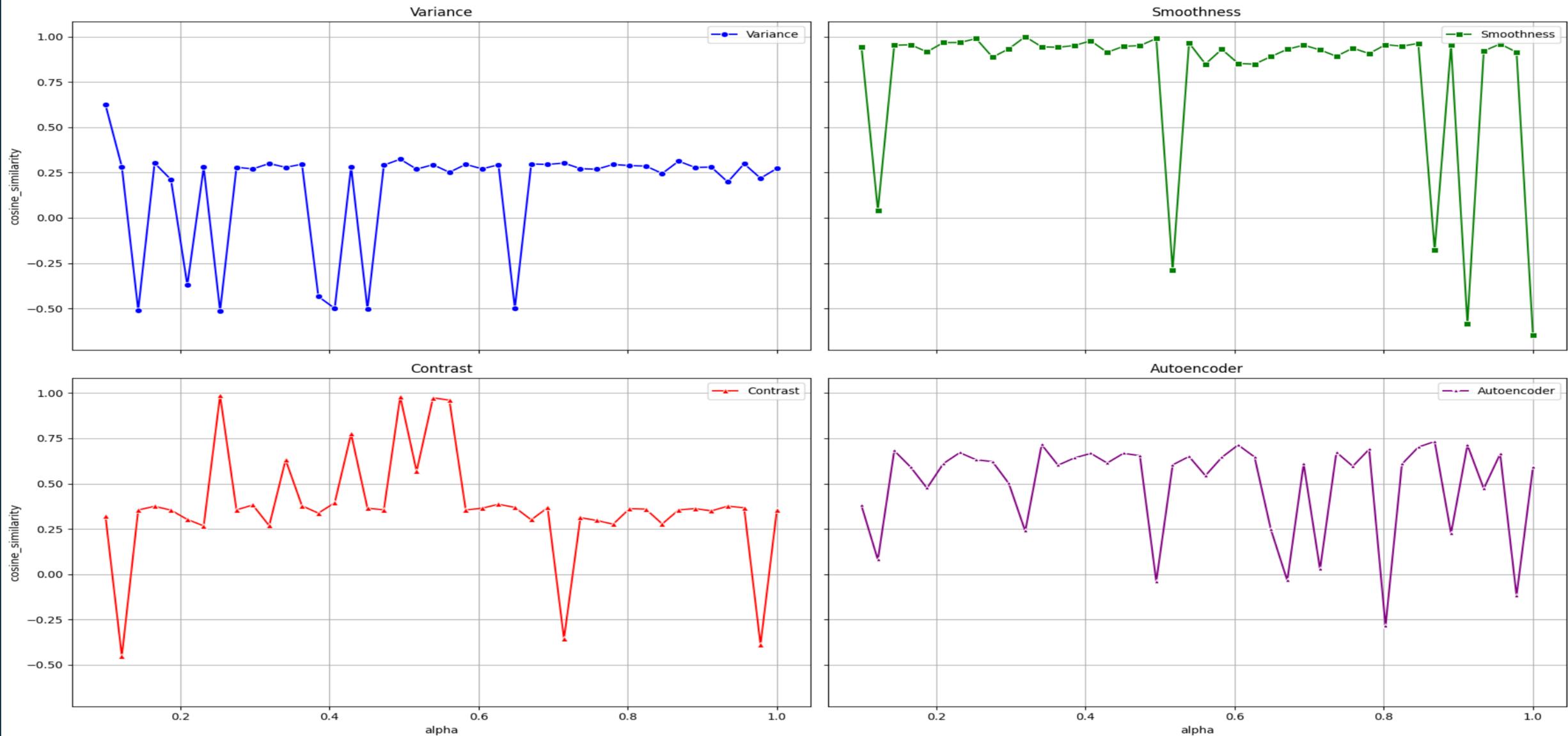
166

...



22280 X 1

Alpha vs. Cosine Similarity for Various DF Types



# Pain Points

- **Data Processing Challenges:**
  - Handling and aligning massive CODEX and noisy H&E datasets requires substantial computational power and can lead to inaccuracies in cell mapping.
- **Feature Extraction Limitations:**
  - The introduction of noise through extensive feature extraction (626 features) complicate accurate analysis.
- **Clustering and Interpretability Concerns:**
  - The unsupervised nature of UMAP clustering and the interpretation of overlapping cell type clusters present significant challenges, impacting the biological relevance of results.
- **Need for Advanced Analysis Tools:**
  - Addressing the gap between traditional histology and spatial proteomics necessitates the development of more sophisticated computational models and machine learning techniques.

# Next Steps

- **Innovative Computational Techniques:** Advance our computational models with sound algorithms for more precise data integration and analysis.
- **Interdisciplinary Collaborations:** Strengthen partnerships with clinical experts to ensure biological and clinical relevance of our findings.
- **Data Complexity:** Develop sophisticated alignment algorithms for precise CODEX and H&E image integration.
- **Resource Efficiency:** Optimize our deep learning models for computational efficiency, ensuring robust analysis with extensive datasets.
- **Clinical Interpretation:** Work closely with nephrologists for insightful interpretation and validation of our research outcomes.
- Submit our findings and future plans **to Scientific Reports**, aiming to contribute significantly to the fields of nephrology and computational pathology.

Thank You  
**for** Your Attention **in** this system!