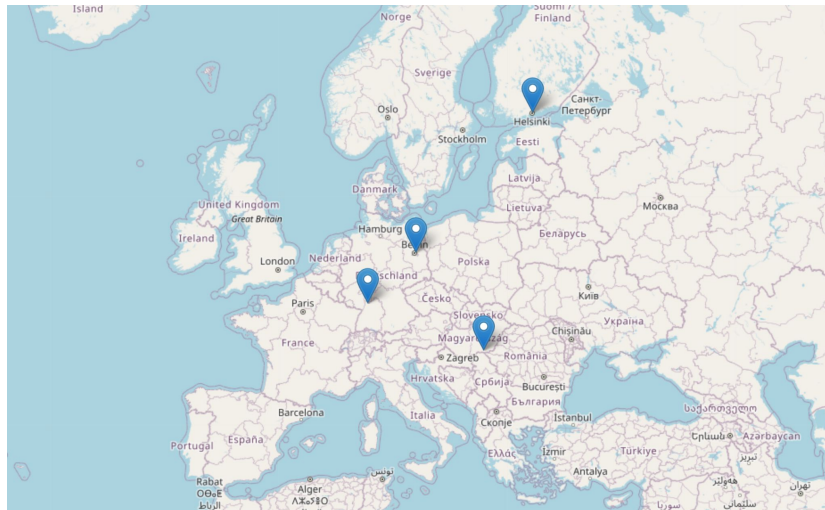


CytoData 2023 Hackathon

Group: BioMagicians

Divya Sitani, Loan Vulliard, Nikita Moshkov, Ziqi Kang



Task 1

**General exploration / profiling
of the data**

- Diverse cellular compositions across core locations.
- Data profiling performed at multiple levels (single cells, image crops, ROI locations, patients)

Single-cell embedding recapitulates cell lineage markers (SMA and panEpi)

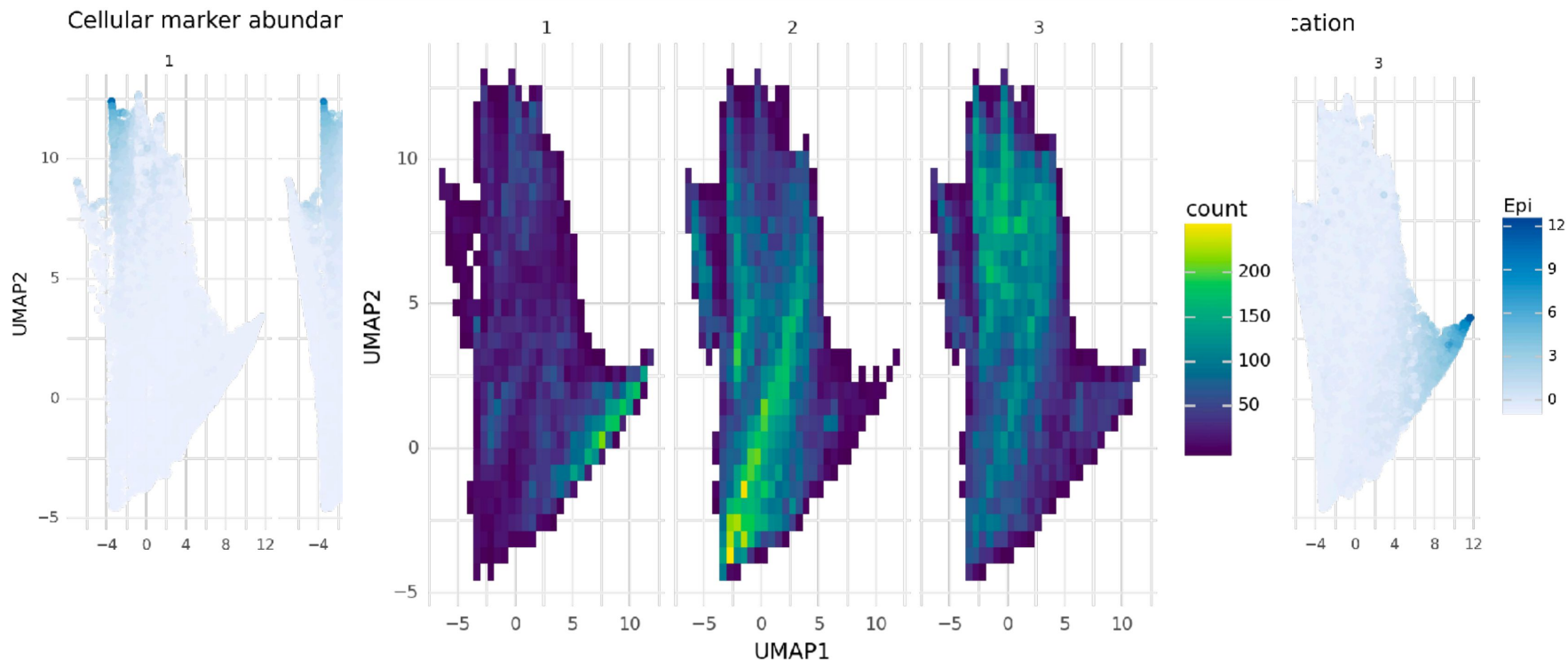
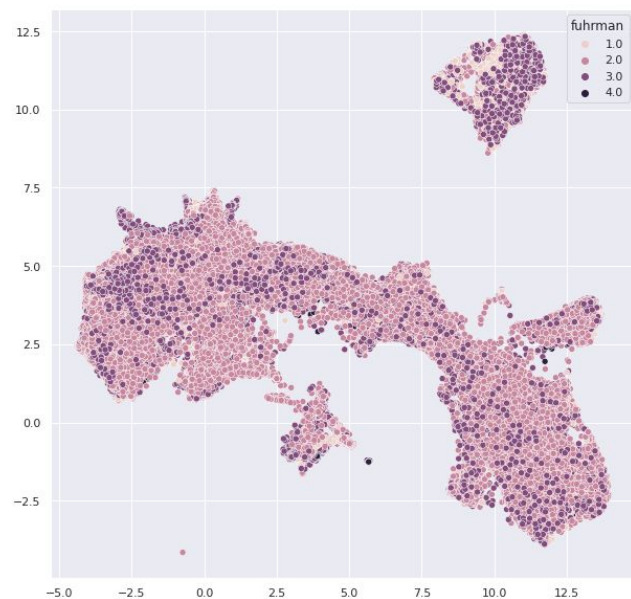


Image-based profiling on crop level

- Feature extraction per tile/channel with EfficientNet-B0 pre trained with ImageNet



Distinct multi-patient cluster visible at both ROI and crop level.

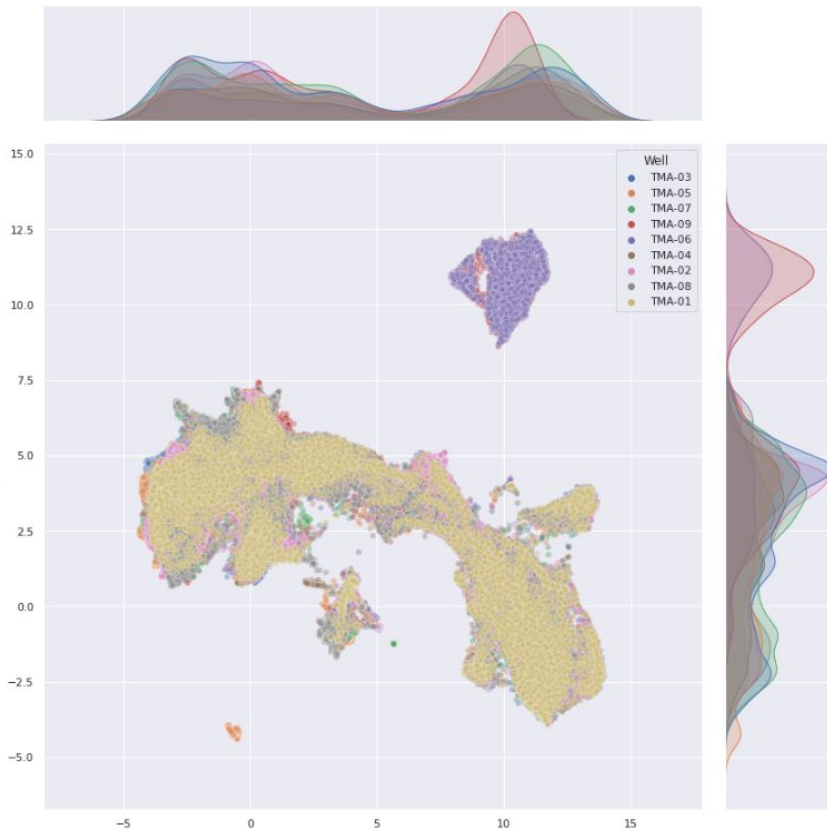
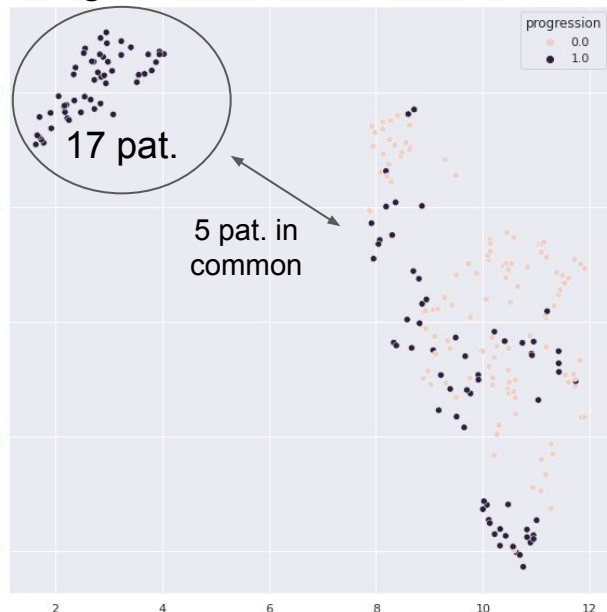
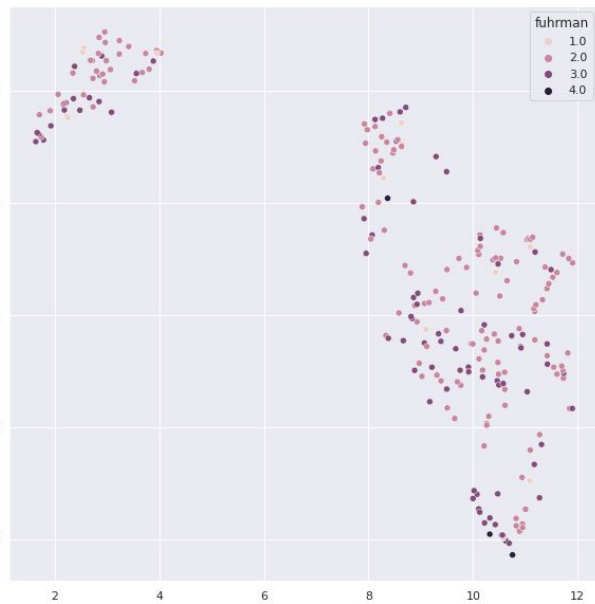
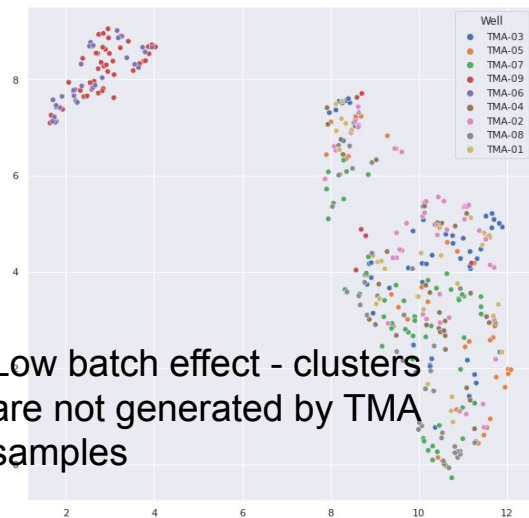


Image-based profiling in image (ROI) level shows

Fuhrman & Progression

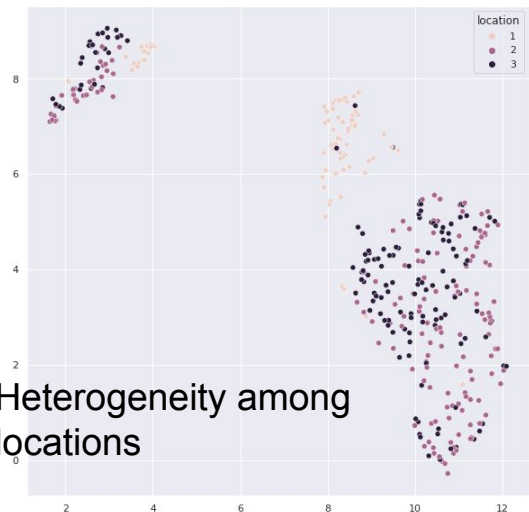


Low batch effect - clusters are not generated by TMA samples

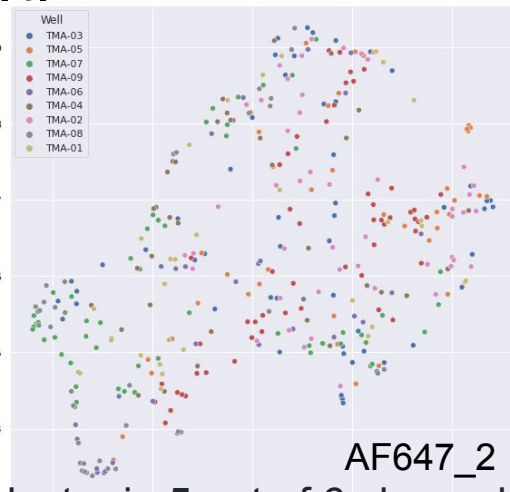
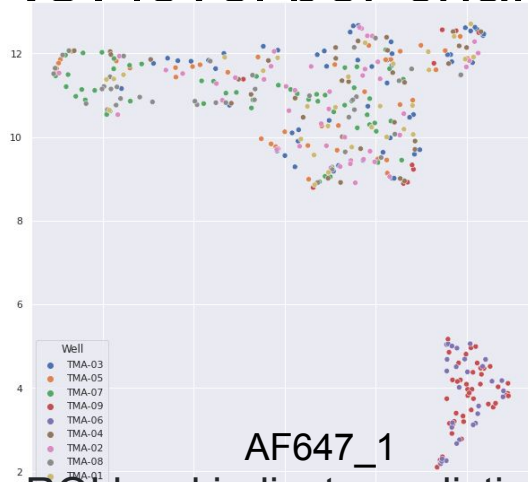
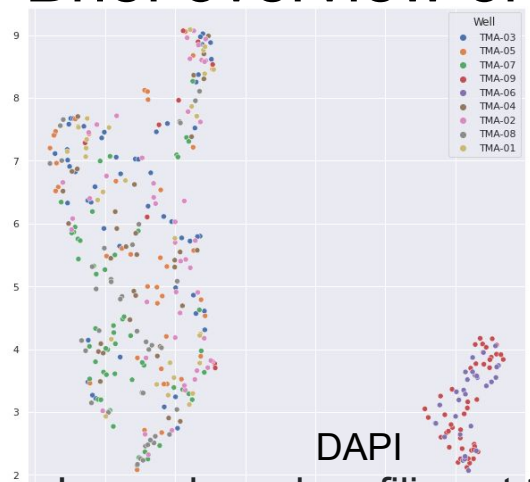


- This cluster consists of 12 unique patients, with an overlap of 5 patients shared with the primary cluster.

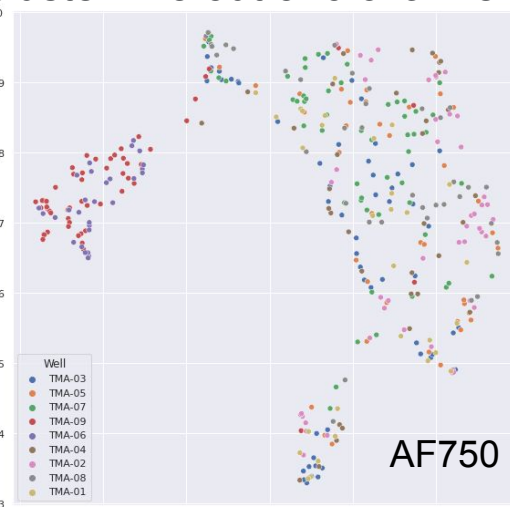
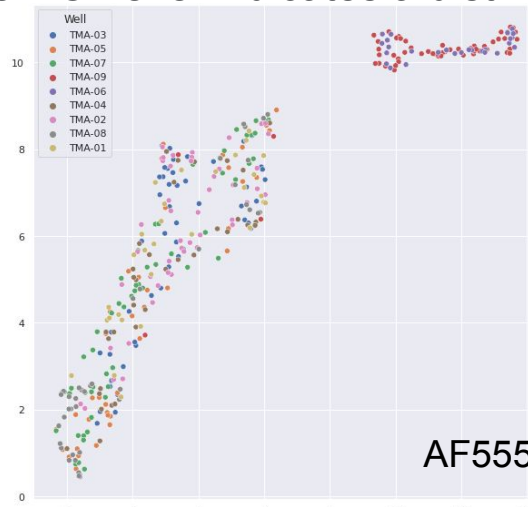
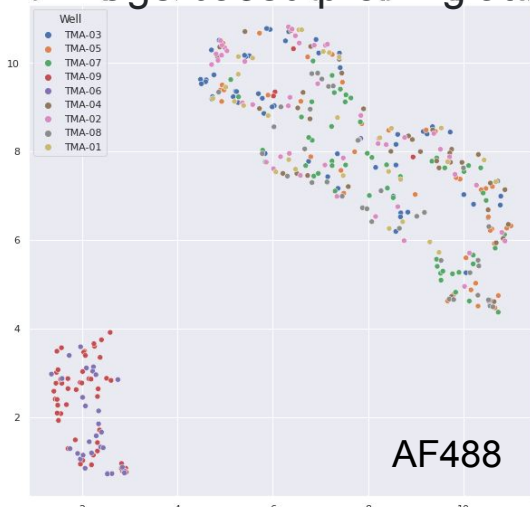
Heterogeneity among locations



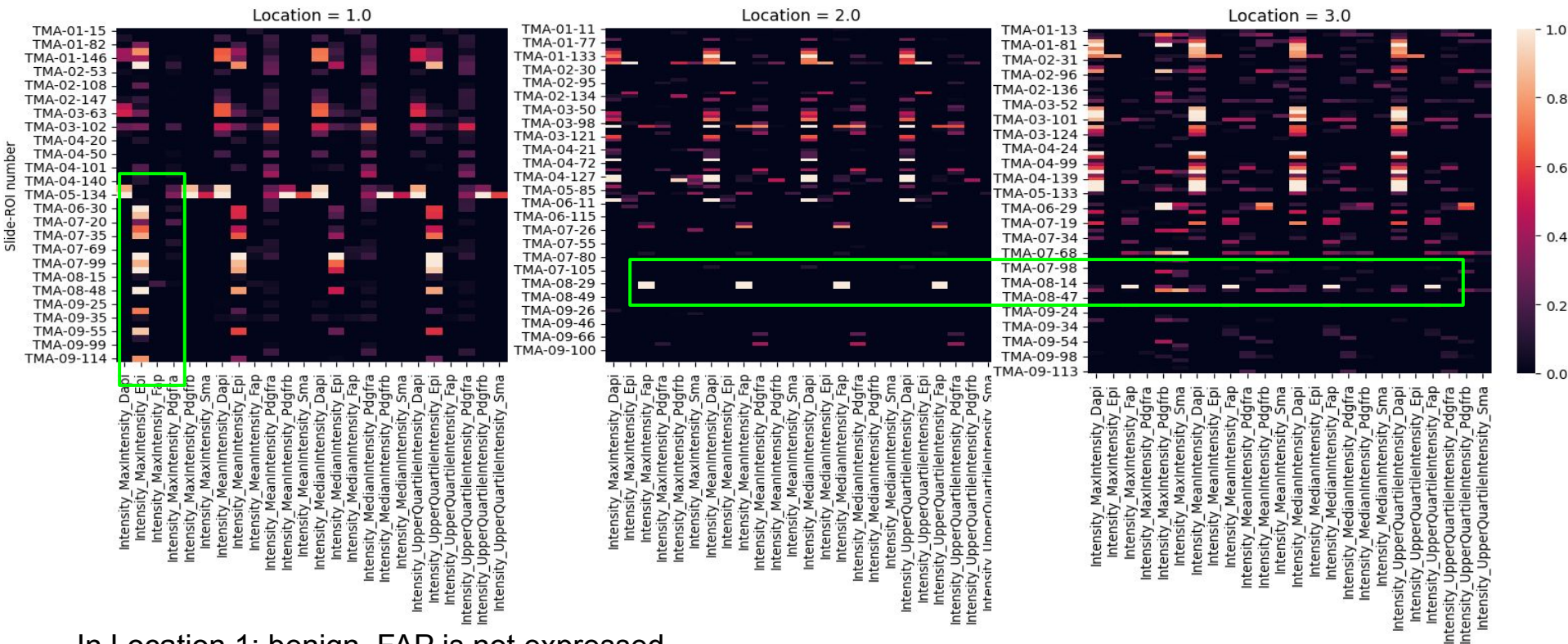
Brief overview on ROI-level per channel



● Image-based profiling at the ROI level indicates a distinct cluster in 5 out of 6 channels.



Slide-ROI Number vs. Features Across Locations



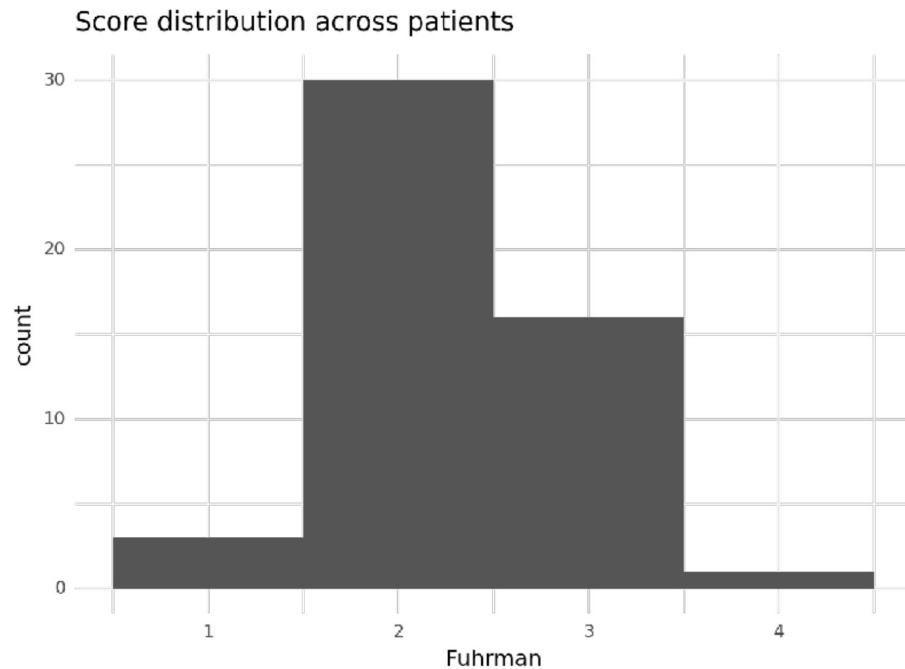
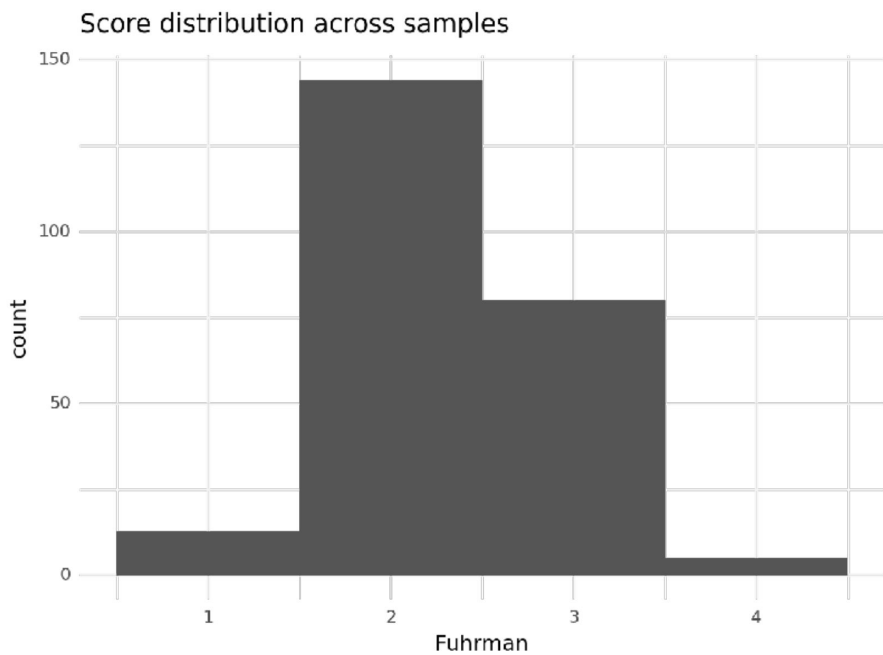
In Location 1: benign, FAP is not expressed.

Task 2

**Fuhrman grading prediction on
test data**

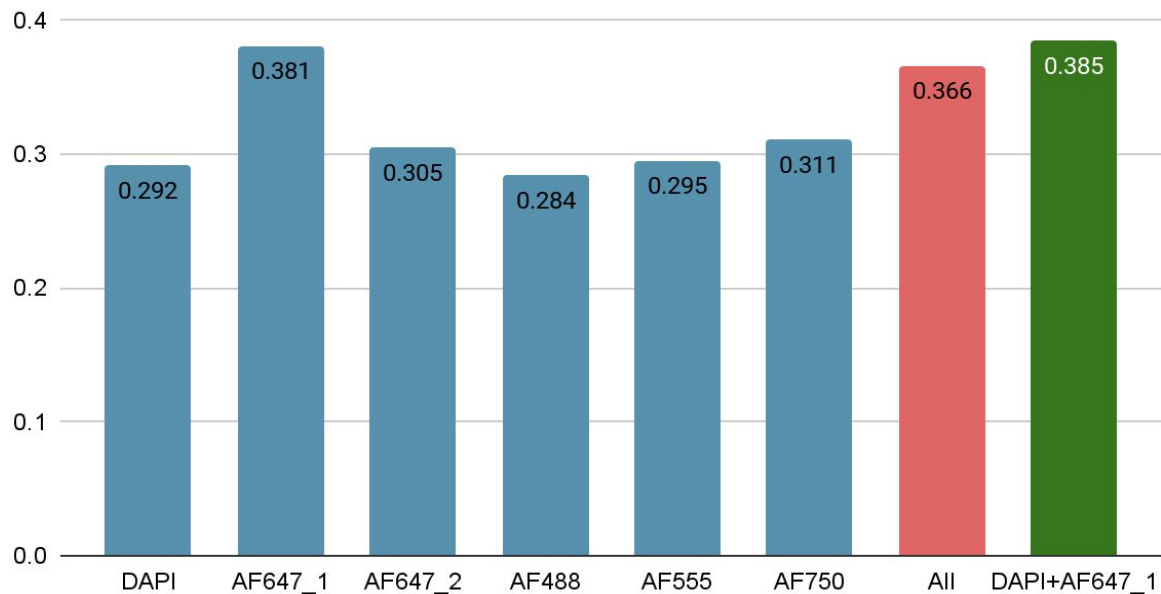
- Explore feature selection: DAPI and AF647_1 (SMA) channels combined are sufficient to make final Fuhrman grading predictions.

Imbalanced Fuhrman grading labels in training set



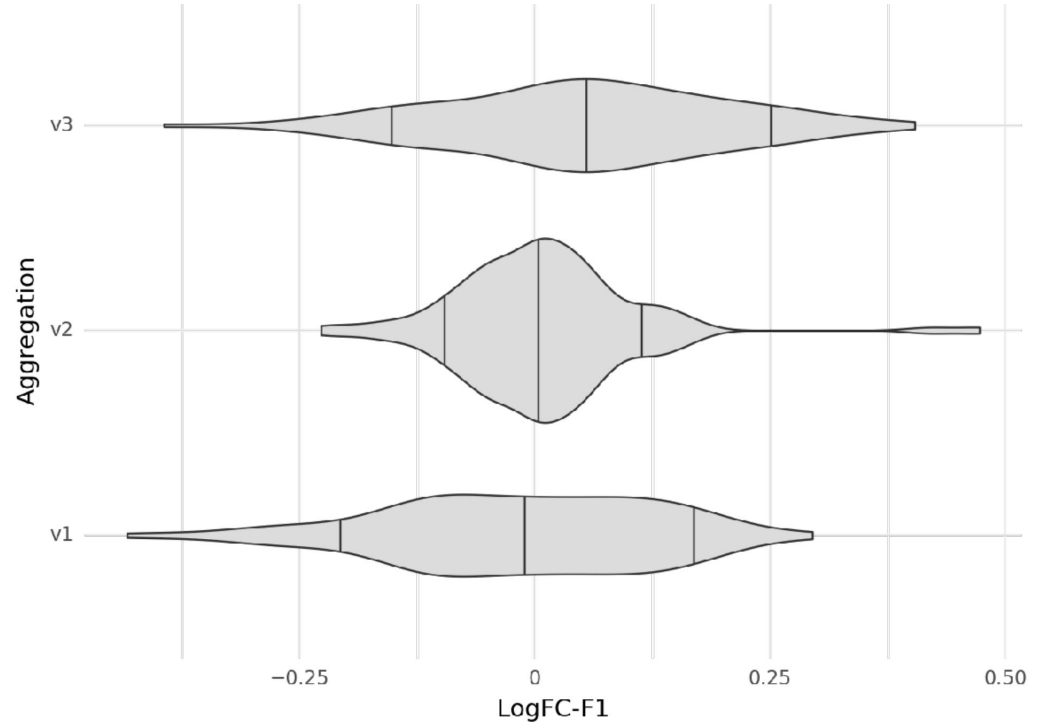
Fuhrman prediction with roi-level features

F1-macro, 5-fold CV on training set



Aggregating labels per patient

- V3: Taking the highest label of all samples per patient
- V2: Taking the most common label if found for >50% of samples
- V1: Taking the most common label per patient

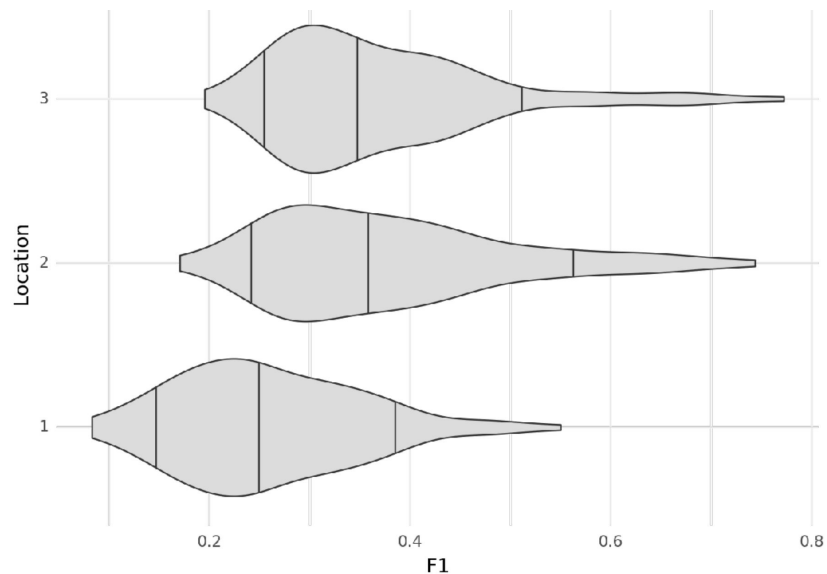


Task 3

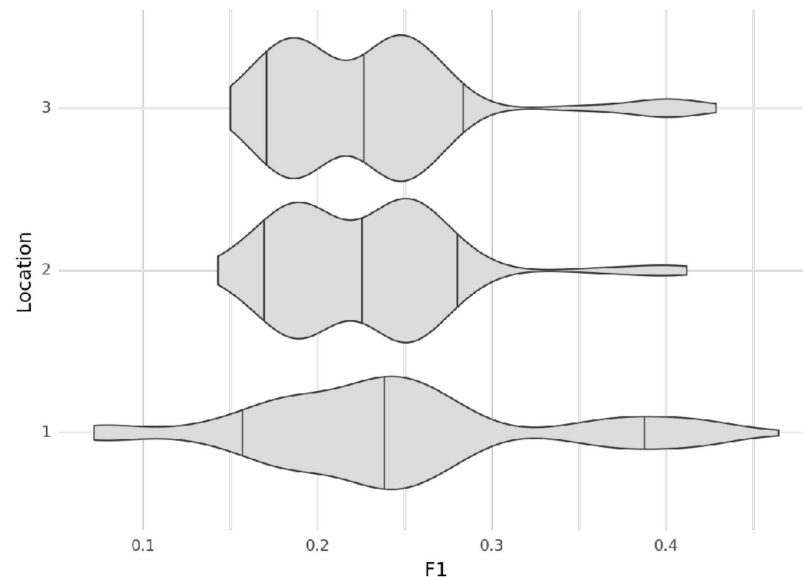
**Study the effect of spot location
for the prediction**

- Quantified the importance of spot location & single-cell spatial information
- Explore on tissue architecture

Insight into location role from predictive performance

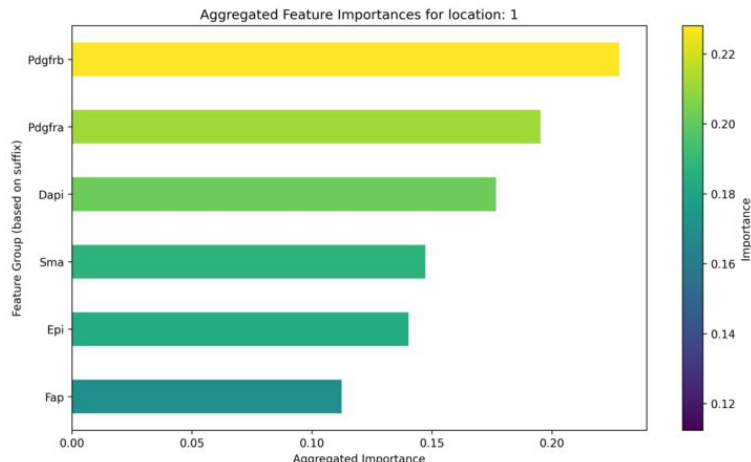


Core locations have various predicting powers.

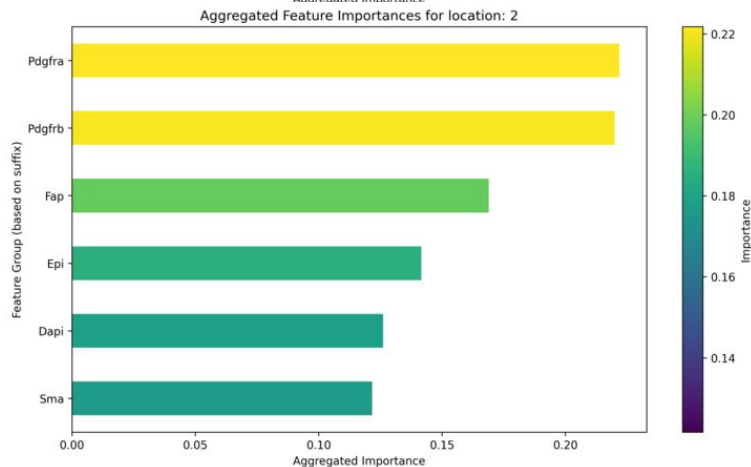


Predictions using only discrete cell type abundance loses information in cancer cores.

Aggregated feature importance varies among Core Locations



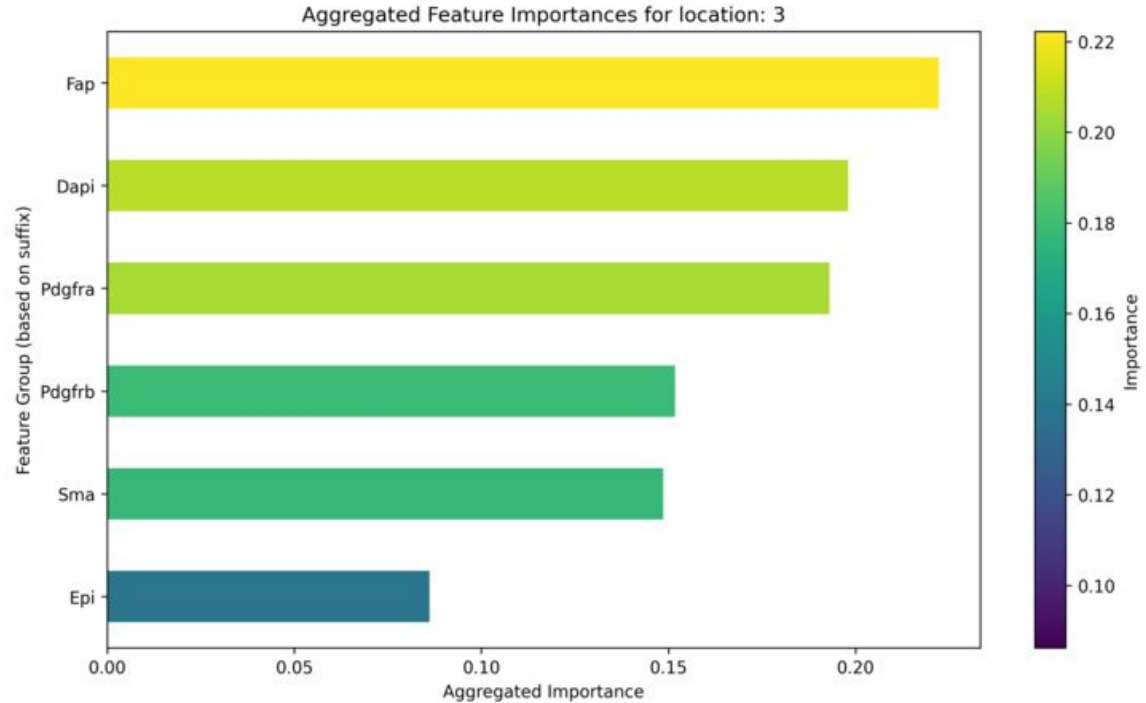
- Highest feature importance of PDGFA/B (Platelet-derived growth factors A and B) in Location 1: Benign and Location 2: Tumor center.



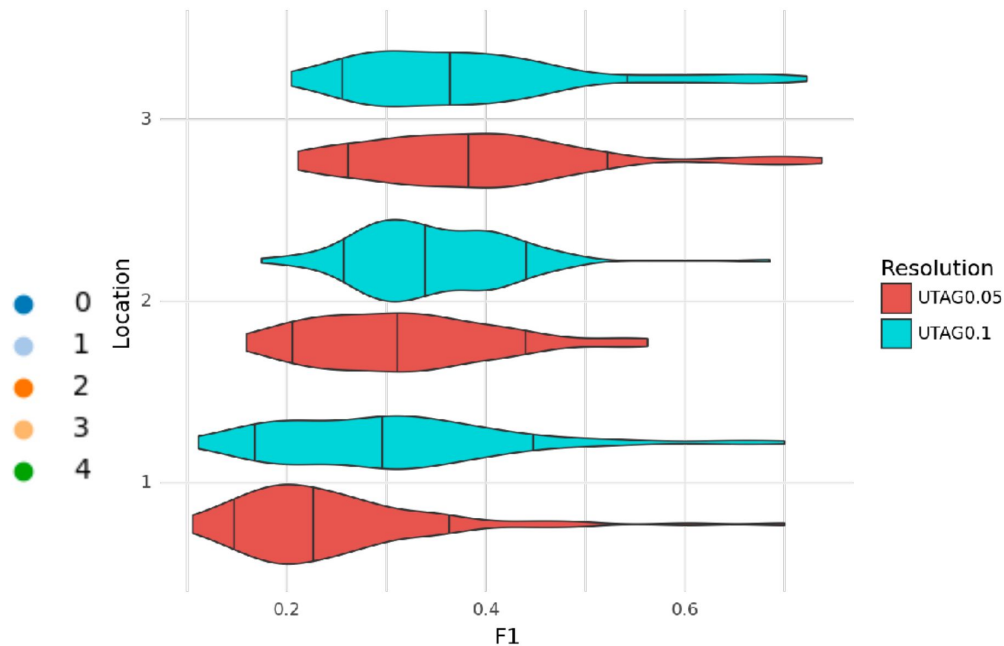
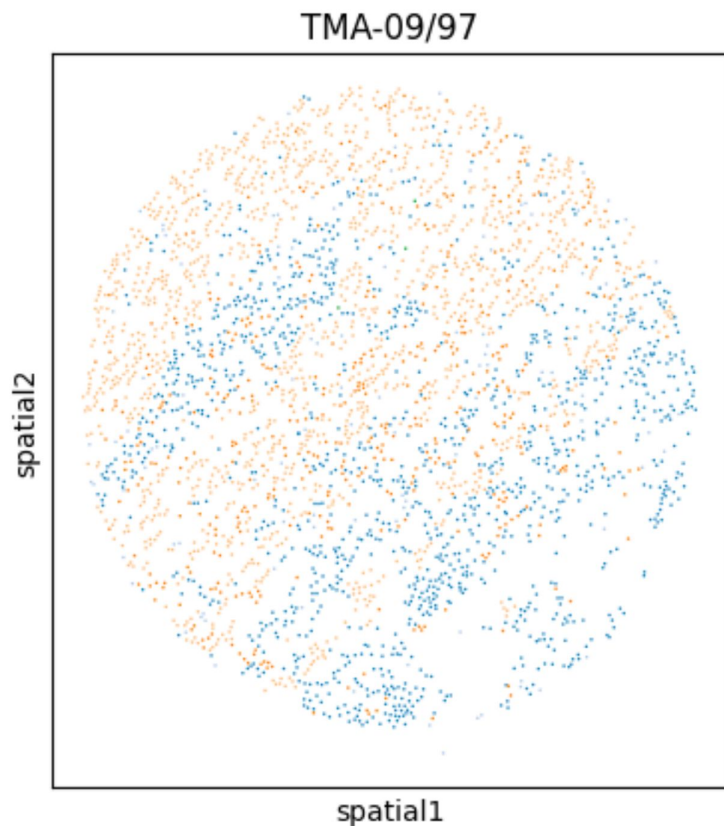
- The high importance in the benign and tumor center regions indicates their involvement in tumor initiation or growth in these regions.

Aggregated feature importance varies among Core Locations

- Highest feature importance of FAP (Fibroblast Activation Protein) in Location 3: tumor border.
- Its high importance in the tumor border indicates its role in tumor invasion, metastasis, or interactions with surrounding tissues.



Unsupervised discovery of tissue architecture with graphs could highlight spatial organization



Kim *et al.* Unsupervised discovery of tissue architecture in multiplexed imaging. *Nat Methods* (2022)

Thanks for listening!

:)

[Draft] Tasks & Solutions

1. General exploration / profiling of the data. Are you able to find anything interesting?
 - Heatmap
 - UMAP to show heterogeneity between among locations?
2. Fuhrman grading prediction on test data
 - Worth to notice: Baseline. The number of Fuhrman grading labels are not balanced
 - We have a table to include feature selection & performance
3. Study the effect of spot location for the prediction
 - Feature Importance
 - UTAG (cell location) still running
 - misty

Background & Tasks

80 patients (clear-cell renal cell carcinoma)

384 cores (images)

- Core location 1: Benign
- Core location 2: Tumor center
- Core location 3: Tumor border

1. General exploration / profiling of the data. Are you able to find anything interesting?

- Heatmap & UMAP to show heterogeneity among locations

2. Fuhrman grading prediction on test data

- Worth to notice: Baseline. The number of Fuhrman grading labels are not balanced
- Feature selection & performance

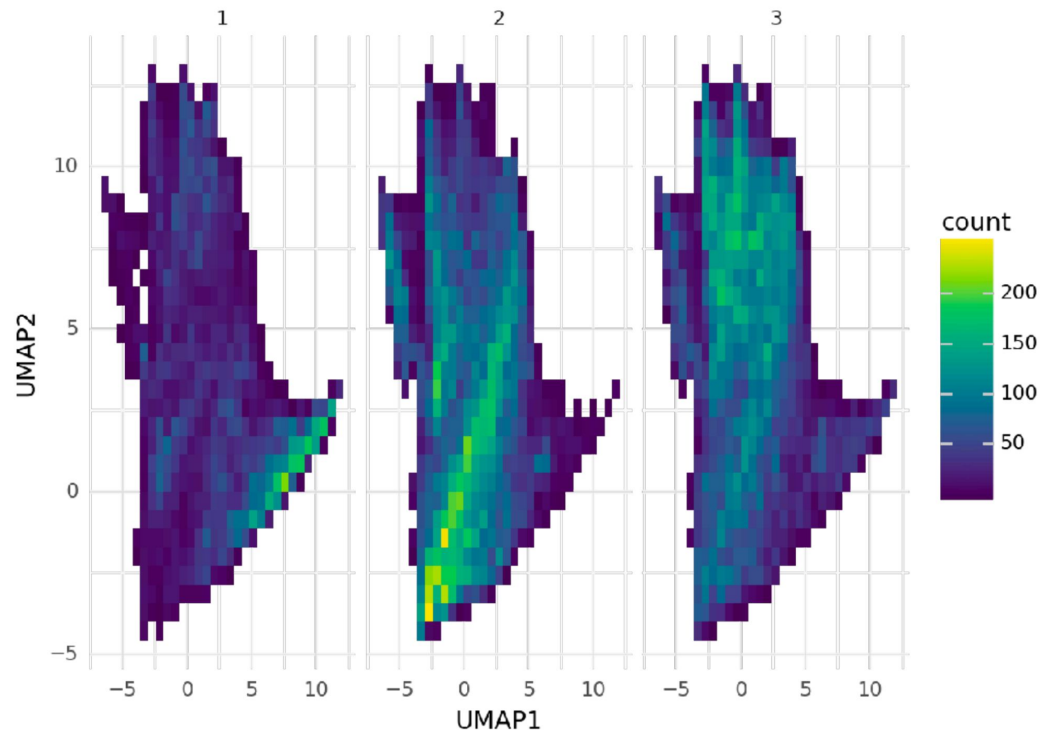
3. Study the effect of spot location for the prediction

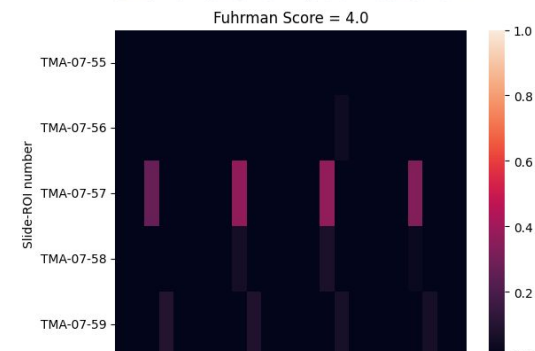
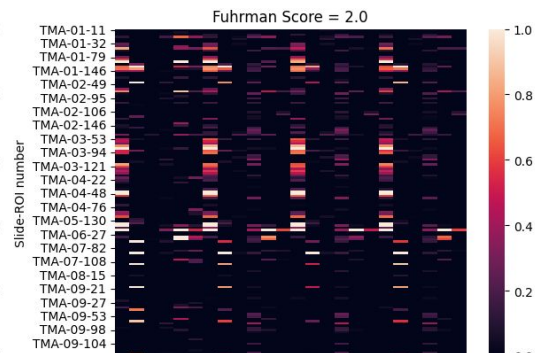
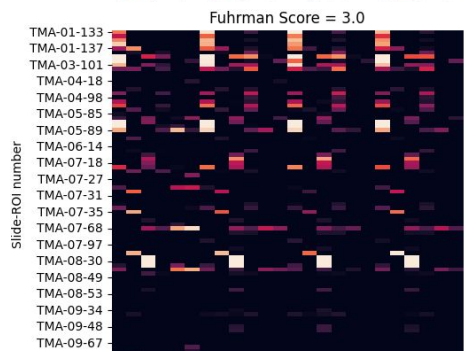
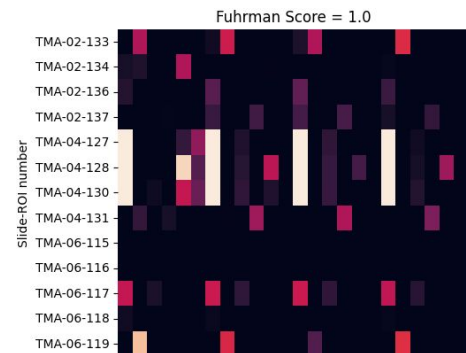
- Feature Importance
- UTAG (cell location)

Findings from aggregated feature importance plots:

1. Highest feature importance of PDGFA/B (Platelet-derived growth factors A and B) in Location 1: begin and Location 2: tumor center.
2. Their high importance in the benign and tumor center regions indicates their involvement in tumor initiation or growth in these regions.
3. Highest feature importance of FAP (Fibroblast Activation Protein) in Location 3: tumor border.
4. Its high importance in the tumor border indicates its role in tumor invasion, metastasis, or interactions with surrounding tissues.

Cellular composition differs among core locations





[Draft] Data info

80 patients (clear-cell renal cell carcinoma)

384 cores (images)

- Core location 1 (Lime): Benign
- Core location 2 (Red): Tumor center
- Core location 3 (Blue): Tumor border