Evaluation of Transcript-Guided Cell Segmentation on GBM-Derived Single Molecule Spatial Transcriptomics Data

Fatemah Alsolaiman1, Shamini Ayyadhury2,3 & Gary Bader2

[1]King Abdullah University of Science and Technology Academy, [2]The Donnelly Centre, University of Toronto, [3]Princess Margaret Cancer Centre

Introduction:

Glioblastoma (GBM) is a complex brain tumor characterized by its heterogeneity. This means tumor cells vary greatly in their genetic makeup and behavior. To better understand and treat GBM, it's crucial to analyze how these cells interact and organize within the tumor. To understand tumor organizational patterns better, 14 patient-derived GBM samples were analyzed using the single molecule spatial platform from 10xGenomics known as the Xenium analyzer. Using Xenium and a 414 gene panel, single molecule transcripts were localized across these tissues at subcellular resolution (<30nm) and subsequently analyzed using various gene and image feature profiling methods. However, accurate identification of individual tumor cells (segmentation) is key to this analysis. Cell segmentation, the process of identifying and isolating individual cells within an image, is essential for accurate disease diagnosis and classification.to ensure precise cell boundary prediction and transcript assignment. To determine the best approach, this project aims to compare two statistical segmentation methods, Baysor and Proseg, against the generic method (Xenium nuclear expansion). By evaluating these methods, we hope to improve spatial gene expression analysis in GBM tumors. Ultimately, this research could lead to better understanding of tumor behavior, development of targeted therapies, and improved patient outcomes.

Materials and Methodology

I contributed to this large-scale GBM project by focusing on a sub-project centered around cell segmentation methods as described above. I implemented and compared two statistical cell segmentation models, Baysor and Proseg, which require different parameter fine-tuning. Next, I compared the results from Baysor and Proseg with different parameters with the original xenium-derived segmentations. I aligned the resulting segmentation boundaries with corresponding DAPI and H&E image data.

Material:

1. **Datasets:**

Spatial transcriptomics data from sections of Human GBM FFPE samples were analyzed using the 10x Genomics Xenium. 14 patient-derived GBM samples were analyzed using the single molecule spatial platform from 10xGenomics known as the Xenium analyzer. Using Xenium and a 414 gene panel, single molecule transcripts were localized across these tissues at subcellular resolution (<30nm) and subsequently analyzed using various gene and image feature profiling methods. However, accurate identification of individual tumor cells (segmentation) is key to this analysis, and cell type annotation is applied, too.

2. **Image preprocessing:**

The above sections were stained using two methods post-xenium run: immunofluorescence antibody-based labeling (IF) and using Hematoxylin & Eosin (H&E). The Xenium in Situ workflow is non-destructive, which enables post-Xenium Analyzer staining and imaging on the same tissue section. Hematoxylin & Eosin (H&E) images can be imported and aligned to the original DAPI-stained image in Xenium Explorer v1.2 and later. In v1.3 and later, immunofluorescence (IF) images can be imported and aligned to the original DAPI-stained image (XE v2.0 and later supports importing images with >six channels and displays up to six channels at once).

Method:

1. **QuPath:**

QuPath is an open-source software designed for bioimage analysis, particularly for reviewing H&E-stained images. It enables users to annotate representative tumor regions across all H&E slides manually. Additionally, QuPath can enhance image clarity by removing artifacts, such as blood vessel leakage, providing clearer visuals for tumor cell analysis(Figure 1)

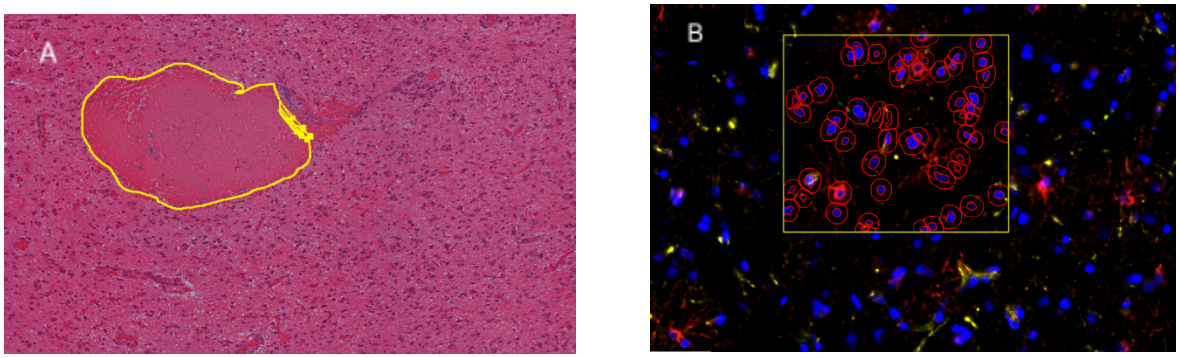


Figure 1: A. Annotating and removing unwanted areas from an image. B. Detecting cells within the selected region.

2. **Two Statistical Cell Segmentation Methods:**

Both Baysor and Proseg operate within a Python environment. The developers provided the base code for each tool, and additional code was written to explore different parameter settings. Statistical analysis was conducted on the outputs of both Baysor and Proseg, and image analysis was integrated with gene expression data in Python.

Baysor:

The most critical parameter is the scale, which specifies the expected cell radius and is the most sensitive aspect of the method. This parameter can be automatically inferred if cell centers are provided. Another important parameter is the min-molecules-per-cell, which determines the minimum number of molecules required for a cell to be considered valid. In this study, five scale settings (1, 3, 6, 9, 12), three min-molecules-per-cell settings (25, 50, 100), and random nine fields of view (FOV) were applied to 14 GBM samples. Each FOV provides a distinct result.

The key statistical analyses involve measuring the confidence level, which reflects the probability of a molecule being real. Additionally, the confidence in the assignment refers to the likelihood that a molecule has been correctly assigned to the appropriate cell.

Proseg:

Two key parameters are employed: schedule and scale. The schedule is a comma-separated list of numbers that dictates the sampling schedule, where the sampler runs for a given number of iterations, halves the voxel size, and then proceeds to the next set of iterations. The scale parameter controls the level of segmentation detail. In this study, three schedule measurements (0, 1, 2) and seven scales (0.3, 0.6, 0.9, 1.2, 1.5, 1.8) were applied.

Probabilistic segmentation analysis is used to investigate the transcripts within each segmented cell as well as background transcripts (those outside of cells). Adjusting the scale can reduce the probability of transcript contamination and alter the perceived size of a cell.

Results and Discussion:

Identifying the most effective segmentation method requires a combination of statistical evaluation and image analysis. In this study, we apply two different statistical models for segmentation and compare them against the original Xenium-derived segmentation. The evaluation of segmentation quality is based on three key criteria:

1. Visual assessment to ensure that segmentation boundaries accurately outline cells, using DAPI images as a reference;
2. Next, evaluating the confidence of these algorithms in transcript assignments; and finally,
3. comparing the new segmentations by analyzing their gene expression and cell-type labeling against single-cell data.

1. Comparing segmentation outputs over original DAPI images

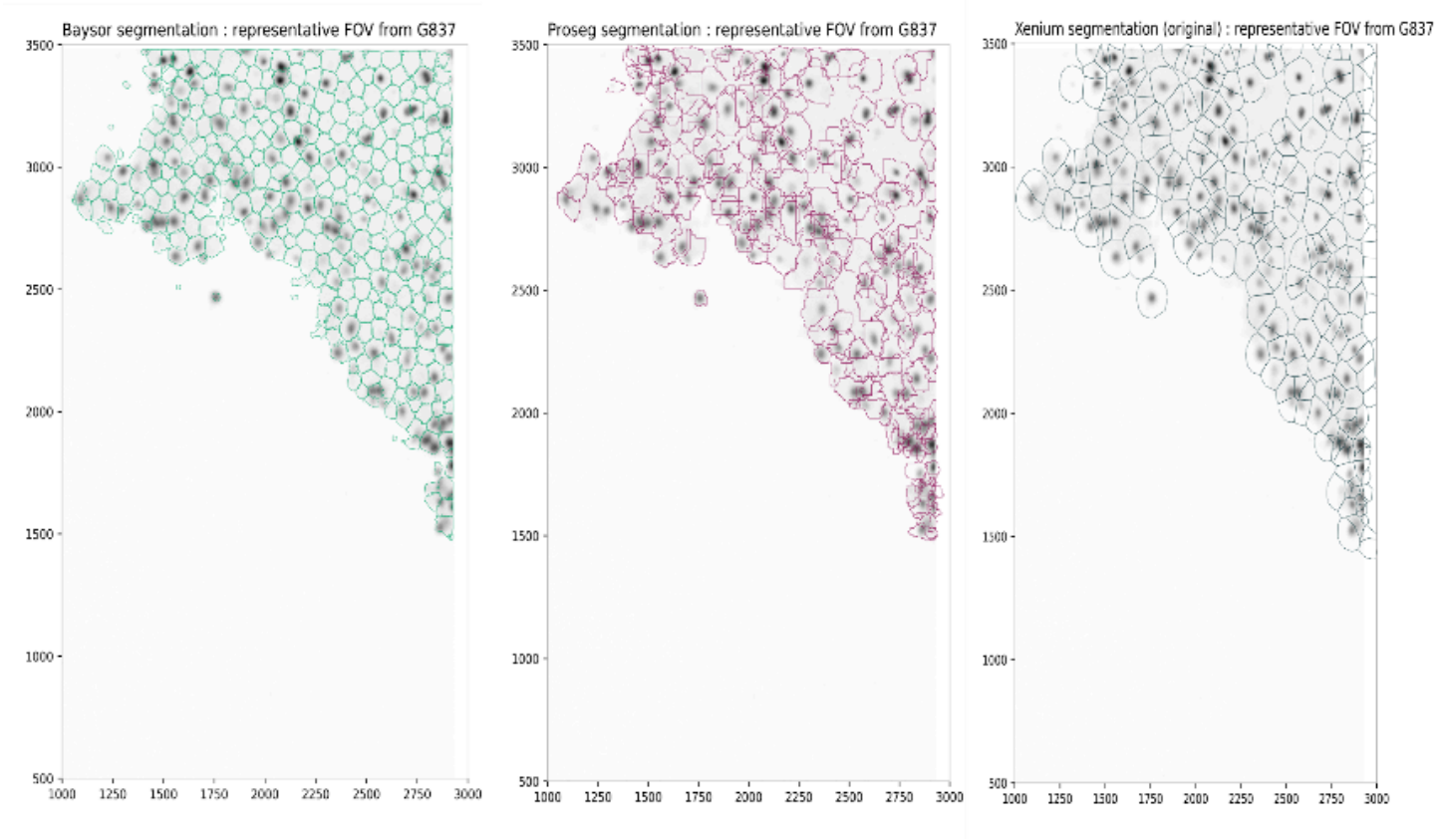
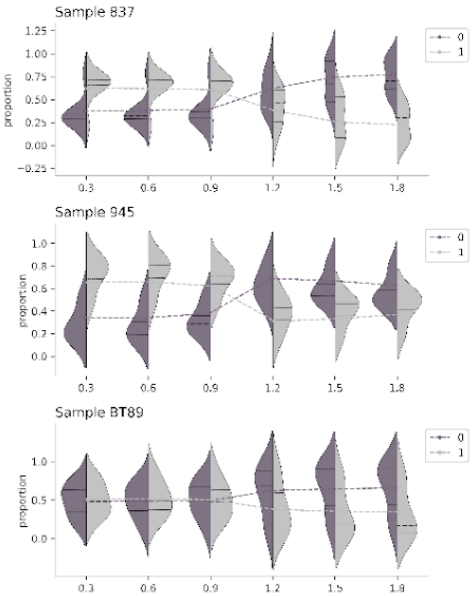


Figure2. Baysor, Proseg, and Xenium (original) segmentation. Using G837 – a patient GBM sample, we show here the differences in segmentation obtained from three different methods: Baysor (statistical model), Proseg (Cell-potts model), and the original nuclear expansion (deep-learning model from the Xenium).

2. Evaluating the best parameter combination that results in the highest confidence in transcript assignment.

Proseg



Baysor

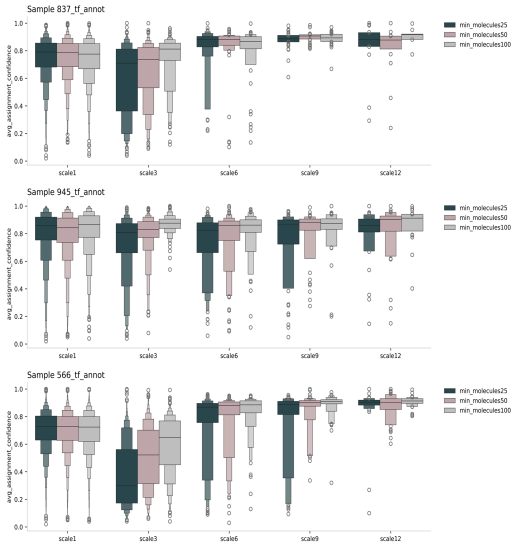


Figure3. Evaluation of model performance. A. Representative probabilities from proseg segmentation of three samples (‘1’ means background, ‘0’ means cells transcript). B. Representative assignment confidence of three samples – two parameters: 1) scale and 2) min-molecules.

Conclusion:

Our research shows that integrating gene and protein expression data with geometric patterning offers a comprehensive view of GBM (glioblastoma). This combined approach, using hierarchical pattern analysis, improves understanding of GBM and could be valuable in clinical profiling and studying tumor growth and resistance. Lastly, comparing the new segmentations by analyzing their gene expression and cell-type labeling against single-cell data is future development.

Code availability:

The code developed for this sub-project is currently on the Bader lab virtual clusters as well as on a private GitHub repo. As the code is being prepared as part of a manuscript that is currently being written, we can only share our codes upon manuscript publication, which we estimate to be towards the end of 2024 or early 2025.

If you would like to make any inquiries regarding the code development used by Fatemah Alsolaiman, please do not hesitate to reach out to me at shamini.ayyadhury@utoronto.ca

Acknowledgments:

My appreciation goes to Bader Lab at Terrence Donnelly Centre for Cellular & Biomolecular Research University of Toronto [Dr. Gary Bader, PhD (gary.bader@utoronto.ca) and Dr. Shamini Ayyadhury, PhD (shamini.ayyadhury@utoronto.ca)] and Data Science Institute, University of Toronto for their support during this study. KAUST Academy financially supports Fatemah Alsolaiman.

References:

1. Ma, J., Xie, R., Ayyadhury, S., Ge, C., Gupta, A., Gupta, R., Gu, S., Zhang, Y., Lee, G., Kim, J., Lou, W., Li, H., Upschulte, E., Dickscheid,

T., de Almeida, J. G., Wang, Y., Han, L., Yang, X., Labagnara, M., … Wang, B. (2024, March 26). The multimodality cell segmentation

challenge: Toward universal solutions. Nature News. https://www.nature.com/articles/s41592-024-02233-6

2. Jones, D. C., Elz, A. E., Hadadianpour, A., Ryu, H., Glass, D. R., & Newell, E. W. (2024b, January 1). Cell simulation as cell segmentation.

bioRxiv. https://www.biorxiv.org/content/10.1101/2024.04.25.591218v1

3. Petukhov, V., Soldatov, R. A., Khodosevich, K., & Kharchenko, P. V. (2020, January 1). Bayesian segmentation of spatially resolved transcriptomics data. bioRxiv. https://www.biorxiv.org/content/10.1101/2020.10.05.326777v1.abstract
4. QuPath GitHub Repository. (n.b). From https://github.com/qupath/qupath
5. Proseg GitHub Repository. (n.d.).From https://github.com/dcjones/proseg
6. Baysor GitHub Repository. (n.b). From https://github.com/kharchenkolab/Baysor