#8029

A Multi-Omics Cell Segmentation Pipeline Using RNA Spots and High-Plex Protein Neuro Images

Aster Wardhani[1], Winnie Leung[1], Lidan Wu[1], Joseph Phan[1], David Ross[1], Mithra Korukonda[1], Ashley Heck[1], Shilah Bonnett[1], Alyssa Rosenbloom[1], Joseph Beechem[1]

[1]Nanostring Technologies, Seattle WA 98109

nanoString

NanoString Technologies
530 Fairview Avenue North, Seattle, WA 98109

Abstract

The most common brain cell segmentation method relies on nuclei information followed by fixed dilation to capture cytoplasmic areas. This arbitrary dilation disregards the cell shape and structure, potentially leading to erroneous inclusion of projections from neighboring cells. Our multi-omics profiling using CosMx™ Spatial Molecular Imaging, proteomics data is acquired through iterative cycles of re-staining and clearing, yielding an extensive source of morphological information that is virtually unlimited and can be harnessed to bolster cell boundary delineation. Furthermore, in cases where membrane staining is absent, we harness transcriptomics data to infer cell locations. Our multi-omics cell segmentation framework amalgamates transcriptomics insights and high-dimensional protein images into a suite of input channels that feed into our custom machine learning segmentation model.

Overview of multi-omics cell segmentation pipeline

Image acquisition system

- High Dynamic Range (HDR) mode used in protein image acquisition to improve dynamic range and avoid saturation

Registration & alignment

- Transcriptomic cycling data is aligned with proteomic cycling data using template matching across each acquisition for each field of view (FOV)

Image Corrections

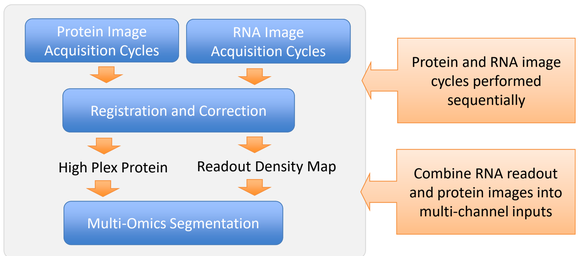
- Intensity based corrections for illumination non-uniformity and bleaching
- Spatially corrections for lateral chromatic aberration and radial distortion

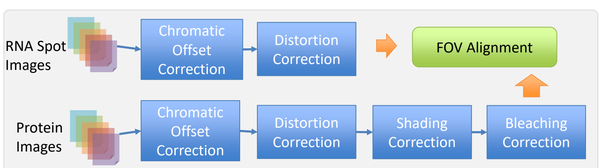
Sample & tissue challenges

- Variable staining and tissue quality
- High cell density
- High auto-fluorescence
- Complex connection of neurons

Multi-modal cell segmentation

- Using custom algorithm leveraging publicly available pre-trained ML based segmentation and auto thresholding to generate initial ground truth
- Train a new multi-channel neural net model using curated ground truth
- Leverage Nvidia RTX A4000 GPU to gain 5X inference performance





Generating Spot Density Heat Map

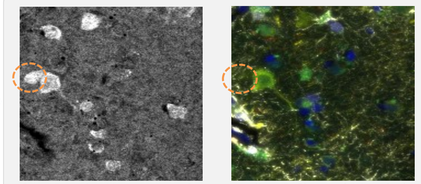
RNA reporter detection

- During a series of encoding cycles, each RNA spot is detected using Laplacian of Gaussian (LoG) filter matched to the expected spot morphology
- Localized with sub-pixel precision using a paraboloid surface fitting function

Spot Density Heatmap

- Each FOV yields over 10 million potential target locations across the entire encoding space
- These locations are used to populate a 2D spatial histogram which serves as a spot density map that highlight objects of interest in and around the cell

Providing cell boundary in instances where membrane staining is missing



Spot Density Heatmap *Corresponding 5-Channel Morphology*

Multi-Omics Segmentation Input

- 34 protein images were selected + RNA Spot Density heatmap
- Split into six distinct protein groups, characterized by shared morphological features:

Immune group
- gH2AX
- MHCII
- SOX10
- CD11b
- CD45
- CD68
- DAP12
- Iba1
- Itgax
- P2ry12
- SORL1
- TMEM119

Nuclei group
- DNA
- NeuN
- TDP-43

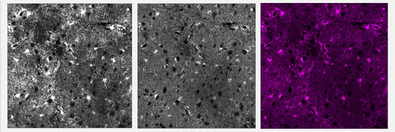
Soma group
- Spot density map
- Ubiquitin
- rRNA
- APP-Hu-Mm

Astrocytes group
- GFAP
- MX1
- NCAM
- S100B

Neurons group
- ChAT
- MAP2
- NRGN
- 5-HTT
- Neurofilament light
- TUBB3

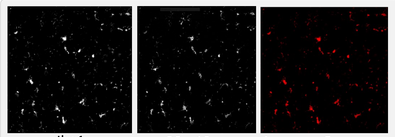
Endo Epi group
- APOE
- Doublecortin
- Laminin
- Nestin
- PDGFRB
- VIM

Normalized to ensure uniform ranges



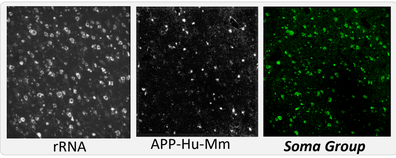
GFAP S100B ***Astrocytes Group***

Combined into groups using **maximum intensity projection**

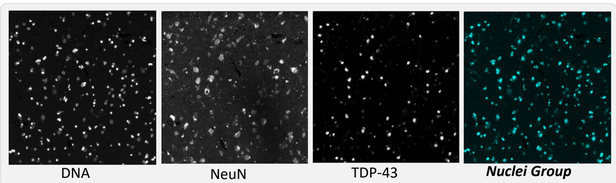


Iba1 P2ry12 ***Immune Group***

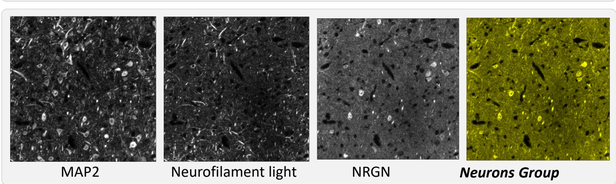
Each group is fed into ML segmentation module



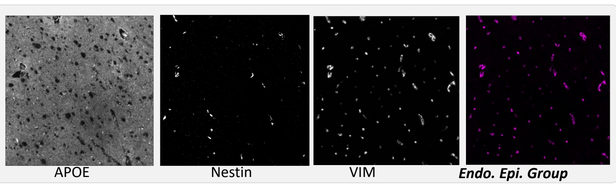
rRNA APP-Hu-Mm ***Soma Group***



DNA NeuN TDP-43 ***Nuclei Group***



MAP2 Neurofilament light NRGN ***Neurons Group***

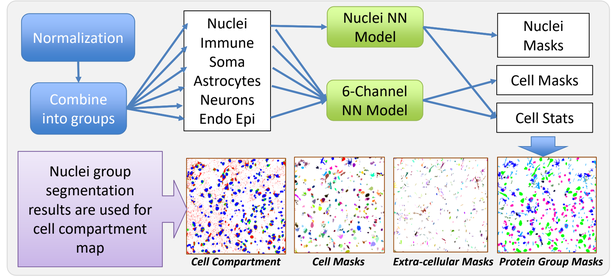


APOE Nestin VIM ***Endo. Epi. Group***

Segmentation Architecture

ML Based Architecture

- Each group is fed into a 6-channel Neural Net (NN) model based on Cellpose [1]
- Using transfer learning, the new model is initialized with weights from pretrained models



Training Datasets and Custom Algorithms

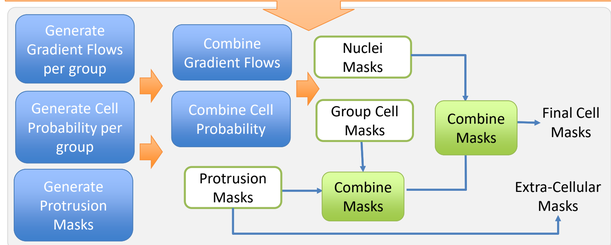
- To train the NN model, we need training datasets
- Using a hybrid segmentation approach to generate the initial ground truth:

➢ Protein groups are segmented using Cellpose CP and Cyto2 models [2] producing gradient flows and cell probability matrices

➢ The matrices are combined to create the combined flow and cell probability, which are then converted into **cell masks** using gradient tracking

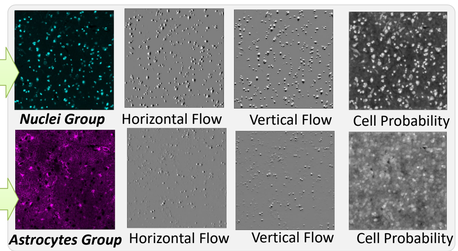
➢ Using **Moment auto thresholding** to segment protrusions [3]

➢ Combine protrusion and protein group cell masks using **intersection of union** analysis



The horizontal and vertical gradient flow indicate the location of cell centers

Cell probability applies weights between real cell signals and noise

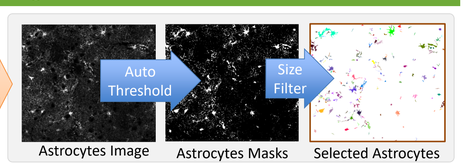


Nuclei Group Horizontal Flow Vertical Flow Cell Probability

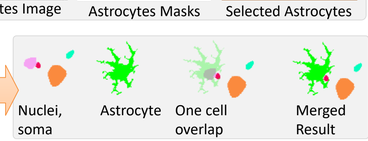
Astrocytes Group Horizontal Flow Vertical Flow Cell Probability

Segmentation of Cell Protrusions

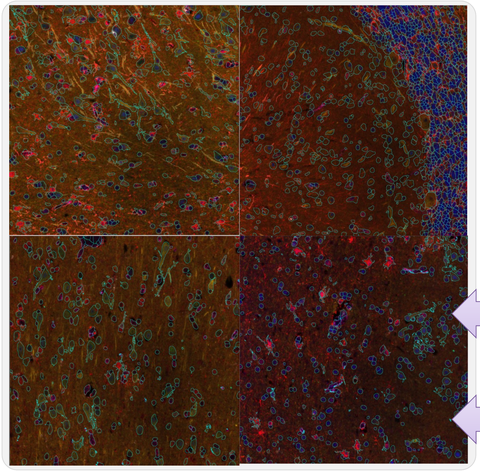
Protrusion mask is obtained by applying intensity threshold calculated by preserving the moments



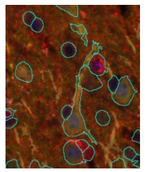
Protrusion needs to be combined with cell masks. The rules of merging/splitting is calculated using the amount of intersection over union and area



Hybrid Segmentation Results



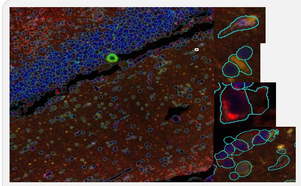
Astrocytes: **Red**
Nuclei: **Blue**
Immune: Gray
Neurons: **Yellow**
Endo. Epi.: **Magenta**
Soma: Green



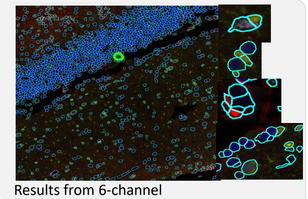
Protrusions that are not attached to soma are marked extra-cellular regions

Hybrid segmentation results are used for training

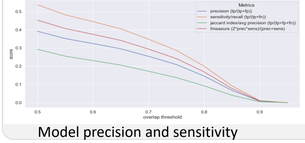
Results from 6-channel Model



Hybrid Segmentation Result

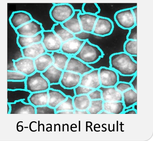


Results from 6-channel



Model precision and sensitivity

Hybrid Result



6-Channel Result

Training parameters: 500 epochs and a learning rate of 0.2. The training dataset comprises 130 images, each with dimensions of 4256x4256 pixels subsampled to 2128x2128.

Conclusions

- Multi-omics segmentation significantly enhances segmentation performance both using the hybrid method and the 6-Channel NN model. This combination of modalities allows for the recovery of cells with weak positive staining that were initially not segmented in the original framework.
- To further improve the 6-channel NN model, the initial ground truth datasets need to be curated.

References

[1] Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nature methods, 18*(1), 100-106.

[2] Pachitariu, M., Stringer, C. Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634–1641 (2022). https://doi.org/10.1038/s41592-022-01663-4

[3] W. Tsai, "Moment-preserving thresholding: a new approach," Computer Vision, Graphics, and Image Processing, vol. 29, pp. 377-393, 1985.

Scan here to download or learn more