#5176

Efficient Large Scale Cell Classification and Analysis for MultiOmyx™ Assays: A Deep Learning Approach

Máté Levente Nagy • Arezoo Hanifi • Ahalya Tirupsur • Geoffrey Wong • Jun Fang • Chin-Chang Kuo • Nicholas Hoe • Qingyan Au • Raghav Krishna Padmanabhan

NeoGenomics Laboratories, Aliso Viejo, CA

Background

Comprehensive biomarker profiling from clinical samples with limited serial sections is not feasible using a traditional immunohistochemistry (IHC) assay. Standard clinical IHC assays stain one biomarker per slide, requiring multiple slides; due to IHC's cellular heterogeneity across multiple sections, co-expression analysis is difficult due to the presence of different cellular populations. MultiOmyx, a proprietary immunofluorescence (IF) multiplexing technology capable of staining up to 60 biomarkers from a single 4µm section, overcomes the limitations inherent in traditional IHC assays, and has been demonstrated in multiple studies **[1-6]**. As opposed to IHC's more qualitative pathologist scores, MultiOmyx's output contains quantitative profiling of tissues at a single cell level and data for millions of cells with billions of queryable data points. In order to detect and classify cells efficiently at this large scale, an image analysis framework using Deep Learning **[7]** was developed.

The Tumor Infiltrating Lymphocyte (TIL) Panel, listed in **Table 1**, can differentiate T cells, B cells, Macrophages, NK cells, and—through co-expression analysis—can further differentiate between various subsets such as T helper (CD3+CD4+) and T cytotoxic (CD3+CD8+) cells. Expanding the list of co-expressions to even more biomarkers allows cells to be classified into specific phenotypes (e.g. a cell that expresses CD3, CD8, and PD-1 is an anergic T cytotoxic cell). As such, the 12 TIL panel biomarkers can provide insight into the tumor microenvironment. One rewarding use-case is the analysis of clinical core needle biopsy samples before and after drug treatment to elucidate the interplay between tumor and immune cells.

With IF assays, tissue such as lung that exhibit high autofluorescence (AF) are difficult to analyze. MultiOmyx solves this by applying both biochemical and image analysis methods to remove AF and retain only true biomarker staining. Difficulties in classification also arise in tightly-packed regions of small-sized immune cells (e.g. lymphoma) belonging to different cell lineages. To classify these cells accurately, we exploit their biological taxonomy as seen in **Figure 2**.

The combination of these techniques (AF removal, incorporation of taxonomical exclusions and hierarchies, and Deep Learning) leads to a robust and reproducible MultiOmyx analysis pipeline required for high-throughput studies.

Biomarkers	Phenotype
CD3	T cell
CD4	T helper
CD8	T cytotoxic
CD20	B cell
CD68	Macrophage
CD56	NK cell
CD45RO	Memory T cell
PD-L1	PD1 ligand
PD-1	Inhibit T cell activation
CTLA-4	Inhibit T cell activation
FOXP3	T regulatory
PanCK	Epithelial cell

Table 1. TIL Panel biomarkers

Tissue Type	Reason for Difficulty
Lung	High autofluoresence
Liver	High autofluoresence
Dense Cellularity	Packed cells

Table 1. Difficult tissue types

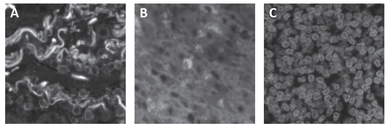


Figure 1. Difficult tissue: A: Lung, B: Liver, C: Lymph Node

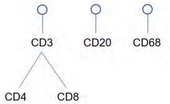
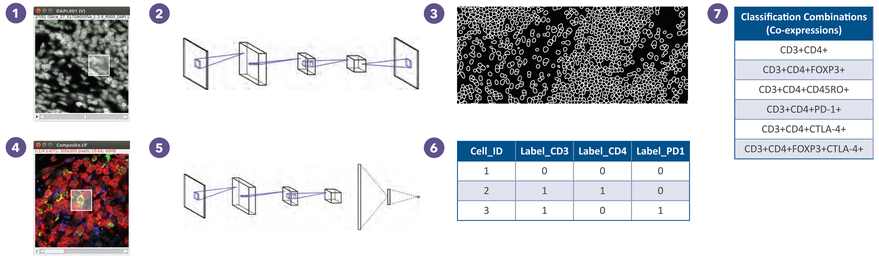


Figure 2. Immune cell taxonomy

Workflow Steps

The framework consists of seven major steps:



Cell_ID	Label_CD3	Label_CD4	Label_PD1
1	0	0	0
2	1	1	0
3	1	0	1

Classification Combinations (Co-expressions)
CD3+CD4+
CD3+CD4+FOXP3+
CD3+CD4+CD45RO+
CD3+CD4+PD-1+
CD3+CD4+CTLA-4+
CD3+CD4+FOXP3+CTLA-4+

Figure 3. Workflow:

1. Manual annotation of a small subset of the nuclear staining channel (DAPI)
2. Training of a fully convolutional neural network **[8]** on this annotation-set to generate a feature map identifying cell centers
3. Application of the trained network in (2) on the nuclear stain (DAPI) of the entire dataset to delineate individual cells
4. Manual annotation of a small subset of each biomarker channel
5. Training of convolutional neural networks on these annotation-sets for classification of biomarkers
6. Application of these convolutional neural networks to the entire dataset
7. Combination of the classification results to identify phenotypes of interest

The workflow output consists of both visual label maps and classification summary tables for individual and co-expressed biomarkers at the region of interest (ROI) level and the entire slide level. In addition, combining the phenotypes with their location information allows for the visualization of complex spatial relationships in the tissue.

Workflow Output

The primary way to understand the workflow output is via color overlays of the original IF stains and the corresponding classification label maps and tables. In addition to cell classification, tumor segmentation is also performed to aid in better understanding of the tumor micro-environment.

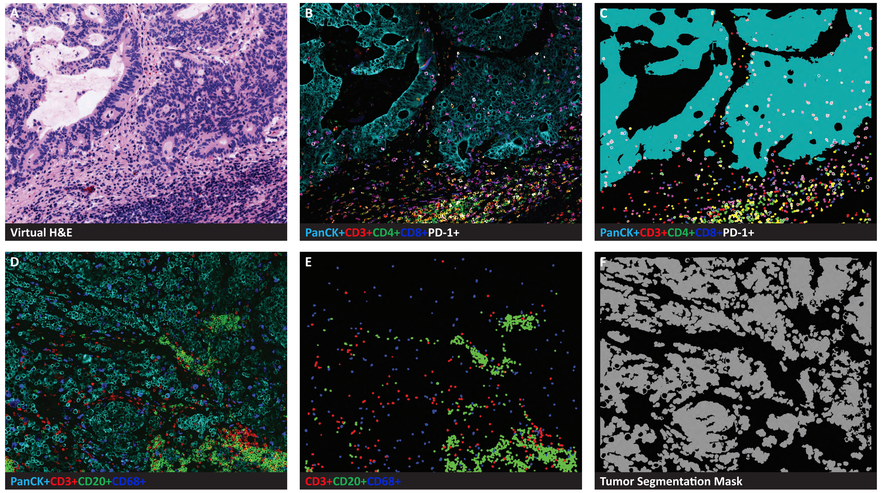


Figure 4. Multi-channel IF overlays (B, D) for various biomarker combinations, along with their corresponding classification label maps (C, E). Virtual H&E (A) and a tumor segmentation mask (F) are also shown. Top: S17050231004, R001 (colon); bottom: S17050231007, R001 (gastric).

A

Slide	ROI	Annotation	Initial_Cells	Roud_1_QC	Overall_QC	Overall_QC_Cells	QC_Pass
S17050231004	R001	Tumor plus immune cells	4069	99.78%	99.36%	4043	1
S17050231007	R001	Tumor plus immune cells	4246	99.86%	99.86%	4240	1

B

Cell_ID	Centroid_X	Centroid_Y	Area (px)	Round_1_QC	Overall_QC	TotalIntensity_CD3	Label_CD3	Within_Tumor
34	2370	71	382	1	1	115798	1	1
35	1049	74	313	1	1	28207	0	1
39	2211	78	227	1	1	36035	0	0
386	1294	274	344	1	1	86249	1	0

C

Slides	ROI	OutsideTumor-Area (mmsq)	TumorArea (mmsq)	TissueArea (mmsq)	ALLCELLS #	CD3+ #	CD20+ #	CD3+CD8+ #	CD3+CD8+PD-1+ #	ALLCELLS #per mmsq	CD3+ #per mmsq
S17050231004	R001	0.2302	0.2488	0.4790	4043	585	222	180	79	8439.80	1221.19
S17050231007	R001	0.2529	0.2758	0.5287	4240	181	732	55	5	8019.81	342.35

Table 3: A: Number of cells passing QC for each staining round (two stains per round), B: Cell-level Quantification Table, C: ROI-level Summary Table

Tissue Quality Mask (12 Biomarkers)	Cell Segmentation*	Single-marker Classification	Multi-marker Classification (3 Biomarkers)	Tumor Segmentation Mask	Summary Table (31 Co-expressions)	Total Run Time (12 Biomarkers)**
27.2s	288s	8.4s	14.6s	1.6s	3.8s	402.4s (6.7 min)

Table 4: Average run times per ROI for major workflow steps

* Cell segmentation contains: cell segmentation, identification of which cells pass each QC in each round, and the calculation of intensities for each biomarker (12 biomarkers).

** Total run time assumes one multi-marker classification (for three biomarkers) and eight single-marker classifications, with the 12th biomarker used for tumor segmentation.

Data Visualization, Spatial Analytics, and Complex Queries

Numerous visualizations for initial data exploration and for high-level messaging about the outputs are provided.

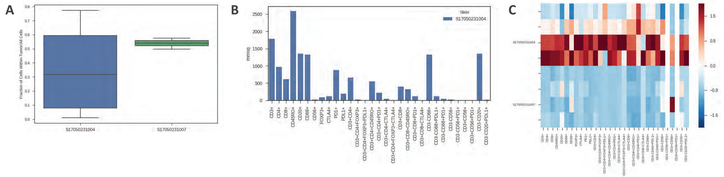


Figure 5. A: Ratio of cells within tumor region to all cells, B: Biomarker and co-expression densities, C: Biomarker and co-expression heat map

This technology allows for advanced spatial analytics and is capable of answering complex queries such as:

- What is the extent of immune infiltration into the tumor region?
- Are certain immune biomarkers excluded from the tumor region?
- What is the spatial distribution of immune activators (OX40, ICOS, GITR) and suppressors (PD-1, PD-L1, CTLA-4) in context to the tumor region

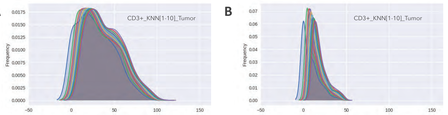


Figure 6. Average distance to the K-nearest tumor-region neighbors (Ks of 1 to 10) of CD3 cells. A: Many CD3 cells further from the tumor region, B: Most CD3 cells very close to the tumor region

Summary

The benefits of using this Deep Learning framework are greatly felt through increased time efficiency without a loss in accuracy, when compared to more traditional computer vision methods requiring high levels of parameter fine-tuning that were employed in the past. Given trained models, the current workflow is able to produce results in the form of both images and tables in under seven minutes per ROI for numerous biomarkers and co-expressions. This results in great overall study-level speedup, with a threefold decrease in average per-study analysis time. A fully automated approach that removes much of the remaining manual touch points is currently in development, which will take MultiOmyx a step closer to real-time analysis.

The output consists of:

- Cell segmentation label maps
- Cell classification label maps for every biomarker within a study panel
- Tumor segmentation masks
- Overall QC masks
- Quantification files listing cell morphological features, intensity measurements, and classification results
- Summary table listing cell counts for various biomarkers and co-expressions
- Visualization of data to ease understanding of the complex output

References

1. Ribas, Antoni, et al. "Oncolytic virotherapy promotes intratumoral T cell infiltration and improves anti-PD-1 immunotherapy." *Cell* 170.6 (2017): 1109-1119.
2. Santamaria-Pang, Alberto, et al. "Robust single cell quantification of immune cell subtypes in histological samples." Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on. *IEEE*, 2017.
3. Sood, Anup, et al. "Multiplexed immunofluorescence delineates proteomic cancer cell states associated with metabolism." *JCI insight* 1.6 (2016).
4. Hollman-Hewgley, Denise, et al. "A single slide multiplex assay for the evaluation of classical Hodgkin lymphoma." *The American journal of surgical pathology* 38.9 (2014): 1193-1202.
5. Gerdes, Michael J., et al. "Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue." *Proceedings of the National Academy of Sciences* 110.29 (2013): 11982-11987.
6. Nelson, Deirdre A., et al. "Quantitative single cell analysis of cell population dynamics during submandibular salivary gland development and differentiation." *Biology* open 2.5 (2013): 439-447.
7. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436.
8. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.