An Automated Machine Learning Framework for Rapid Quantification and Analysis of Multiplexed Ion Beam Images (MIBI)

Máté Levente Nagy[1], Peng Si[1], Sweta Bajaj[1], Stanislaw Nowak[1], Monirath Hav[1], Raghav Padmanabhan[1]

[1]Ionpath Inc. Menlo Park, CA

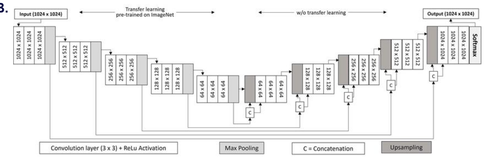
Background

Multiplexed Ion Beam Imaging (MIBI) offers high-parameter tissue imaging that is well suited for describing complex immuno-spatial features in tissues, including the enumeration of various cell phenotypes, expression of immune checkpoint proteins, and quantitative description of spatial distributions between different cell populations [1]. The high imaging resolution combined with the rich mass spectral information in each image allows for the quantification of up to 40 biomarkers in a single field of view (FOV), and enables immediate processing without the need for additional imaging rounds. Leveraging this, we outline our automated machine learning framework that enables rapid, deep phenotypic and spatial profiling of tissues at the single cell level.

A.

Category	Number
Patients	30
Field of View (FOV) per patient	6
Biomarkers	30
Cells per FOV	3000 - 6000
Cell Types / Phenotypes	30
Cell level classifications made per study	15 - 30 million

B.



https://www.researchgate.net/figure/ResNet50-UNet-architecture-for-automatic-detection-of-wheel-ruts-using-drone-imagery_fig3_361724164

Figure 1: A. Scale of Data: In a typical spatial proteomics study, millions of cell classifications need to be performed, highlighting the need for a scalable, automated workflow. **B. UNet Architecture:** Our models are based on the UNet architecture, with ResNet50 convolution blocks [2, 3].

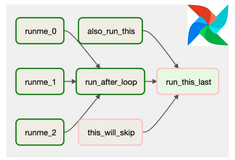
Methods: Workflow Building Blocks

Our automated machine learning framework consists of five steps, linked together using Apache Airflow with Kubernetes compute nodes [4, 5].

These five steps are as follows:

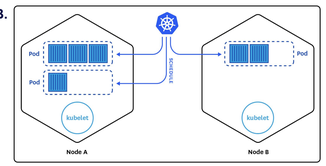
1. **Image Filtering** defines each channel in the staining panel by ensuring that there are no image artifacts that interfere with subsequent steps. Isobaric correction of known mass interferences between channels is also performed.
2. **Model Input Preparation** consists of combining multiple markers into multichannel images that are used by deep learning models in later steps.
3. **Segmentation** separates out individual cells and regions in the FOV. It is based on deep learning models that leverage the multiple biomarkers in the input image. Using multiple marker channels enables segmentation of challenging cells that lack dsDNA signal in the plane of imaging.
4. **Classification** uses deep learning models that leverage staining patterns alongside known phenotypic hierarchies to first define major cell lineages, then further divide them into specific subphenotypes.
5. **Quantification** calculates the expression levels for each checkpoint marker of interest and also computes counts and densities of each cell phenotype. Additionally, spatial analysis is performed, which allows for the quantitation of immune infiltration and various cell-to-cell and cell-to-region proximity features.

A.



https://github.com/apache/airflow

B.



https://coderise.io/kubernetes-communication-pods/

Figure 2: A. Airflow DAG Example: Airflow DAGs define the automated flow of data processing through nodes that allows us to track workflow status and run (or re-run) the whole pipeline or selected subsections of it. **B. Kubernetes Schematic:** Kubernetes works with a scheduler to run required processes in pods that get assigned to various compute nodes, thereby handing off resource management from manual setup and upkeep to an automated system.

In order to reach high levels of automation, we built our workflow on top of Apache Airflow, a workflow management platform for data engineering pipelines that streamlines the connection of multiple complex data processing tasks through directed acyclic graphs (DAGs), with definable dependency structures for a seamless data processing flow. Many of our processing tasks require high compute resources that would require large compute servers. Processing a high volume of MIBI image analysis tasks would quickly lead to infrastructure bottlenecks. In order to prevent these bottlenecks, we use Kubernetes to define compute nodes for each task that get spun up to allow high parallelization without the concern of reaching resource limits.

MIBI Images

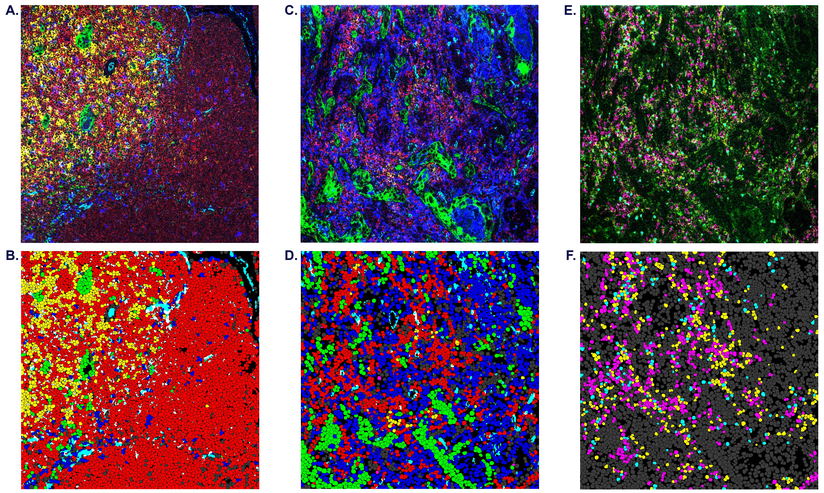


Figure 3: A-B. Thymus Control: A. Lineage Markers (CD3, Keratin, CD68 & CD11b & CD163 & CD14, CD31, CD20, CD56), B. Workflow Output of Lineage Cell Classes (T Cells, Epithelial Cells, Myeloid Cells, Vessels, B Cells, NK Cells); **C-F. Head and Neck Squamous Cell Carcinoma:** C. Lineage Markers (colors same as in A), D. Workflow Output of Lineage Cell Classes (colors same as in B), E. T Cell Markers (CD3, CD4, FoxP3, CD8), F. Workflow Output of T Cell Subtype Classes (Helper Ts, Cytotoxic Ts, Regulatory Ts).

Workflow: Example Subset from Ionpath Checkpoint Panel

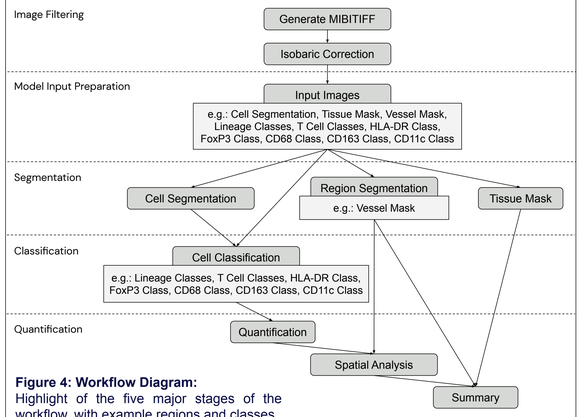


Figure 4: Workflow Diagram: Highlight of the five major stages of the workflow, with example regions and classes.

Hierarchical Classification

Lineage Model
T Cells
B Cells
Myeloid Cells
Tumor Cells
NK Cells

T Cell Subset Model
T Helper Cells
T Cytotoxic Cells

Individual Marker Models
CD11c
CD163
CD68
FoxP3
HLA-DR

Figure 5: Hierarchical Classification: Cell classification uses a set of hierarchical deep learning models where the domain knowledge of the disease of interest is incorporated into defining mutually-exclusive cell populations.

Time Profile of Workflow (per FOV)

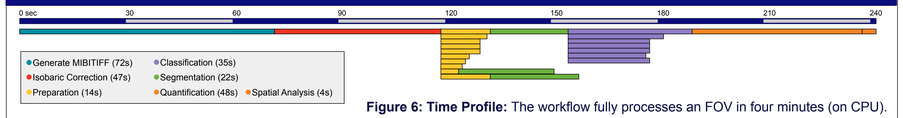


Figure 6: Time Profile: The workflow fully processes an FOV in four minutes (on CPU).

Workflow Outputs and Example Analysis Results

The workflow provides end-to-end analysis of multiple tissue types with diverse morphology and tissue architectures, adapting to a wide range of tissue background and noise, achieving a high degree of accuracy in segmentation and classification. This obviates the need for multiple iterations and parameter tuning to optimize algorithm performance.

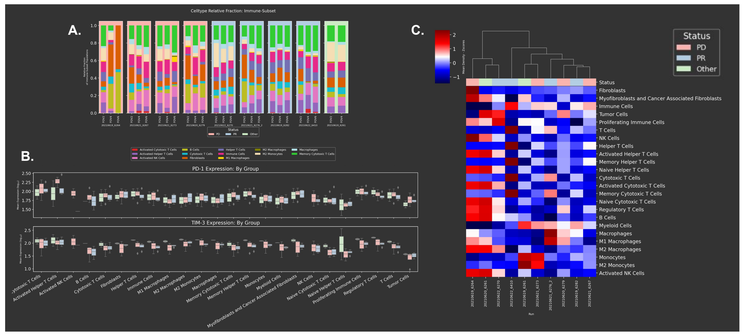
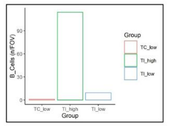


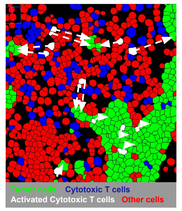
Figure 7: A. Relative fraction of cell phenotypes for each FOV and sample. **B. Box plots** of expressions for two checkpoint markers on multiple cell phenotypes. **C. Heatmap** showing clustering of patient samples based on cell phenotype densities for progressive disease (PD) and partial responders (PR).

Case Studies on TNBC and HNSCC Data

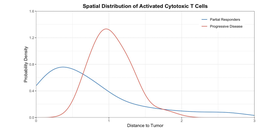
The high quality results generated with our workflow have been used to reveal biological insights and discover spatial patterns of immune cells in the tumor microenvironment and other diseases.

A study at the University of Minnesota (UMN) has found a potentially novel interaction between RHAMM expression and B-cell infiltration in TNBC [6].





At the Medical University of South Carolina (MUSC), a statistically significant increase in interactions between Activated Cytotoxic T Cells and Tumor Cells was seen in partial responders treated with a neoadjuvant PD-1 inhibitor compared to patients with progressive disease [7].



Conclusions

We introduce an automated machine learning framework for deep tissue profiling from MIBI images. The combination of pre-trained deep learning models connected through Airflow's directed acyclic graphs on a Kubernetes cluster leads to a rapid and scalable bioinformatics solution for MIBI images that can be used to uncover novel biology.

References

1. Keren L et al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell. 2018 Sep 6;174(6):1373-1387.e19. doi: 10.1016/j.cell.2018.08.039. PMID: 30193111; PMCID: PMC6132072.
2. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
3. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
4. Apache Airflow can be found at https://airflow.apache.org.
5. Kubernetes can be found at https://kubernetes.io.
6. He, Yuyu et al. "Multiplexed ion beam imaging identifies B-cell enrichment in the RHAMM-high invasive niche of breast cancer". SITC Conference 2022.
7. Raghav Padmanabhan, Hannah M. Knochelmann, Monirath Hav, et al. "Multiplexed Ion Beam Imaging (MIBI) uncovers adaptive immune responses associated with clinical outcomes in oral cavity squamous cell carcinoma treated with neoadjuvant PD-1 inhibitor". SITC Conference 2022.