Using the HALO® High Dimensional Analysis Module for Advanced Phenotypic Characterization and Spatial Analysis of Highly Multiplexed Images

Introduction

From immuno-oncology to neuroscience, highly multiplexed tissue imaging has paved the way for deeper understanding of complex tissue microenvironments by enabling simultaneous visualization of dozens of biomarkers across entire tissue sections. However, the complexity of this data presents significant challenges, as potentially millions of cells per study, each described by numerous biomarker intensities and morphological features, can quickly outpace the ability of traditional tools to provide meaningful insights. With datasets growing more complex, researchers increasingly rely on high dimensional analysis techniques to extract patterns, identify novel phenotypes, and explore spatial context.

Two fundamental techniques frequently used to process high dimensional data are dimensionality reduction and unsupervised clustering. Dimensionality reduction techniques reduce the number of variables in a dataset while preserving the relationships between data points, allowing researchers to visualize complex patterns, such as cell populations, in lower-dimensional space. Unsupervised clustering groups objects based on similarity in their underlying features, such as biomarker positivity or cell shape, without prior knowledge of phenotypes. This technique enables the discovery of novel cell states or subpopulations as well as unbiased phenotyping of known populations.

To support researchers working with highly multiplexed data, Indica Labs developed the **High Dimensional Analysis Module** which launched alongside version 4.1 of the HALO® image analysis platform. This module integrates powerful dimensionality reduction and clustering techniques directly into the HALO interface, streamlining exploration of complex datasets and enabling interactive analysis of object-level data. In this application note, we walk through the steps and options in the module's workflow. In parallel, we demonstrate how the module enables identification of cell populations and their spatial distribution by analyzing a 30-plex lung adenocarcinoma sample imaged on the Cell Dive Multiplex Imaging Solution from Leica Microsystems, shown in **Figure 1**. The biomarkers analyzed included CD45, CD3-epsilon, CD8-alpha, CD4, CD20, CD79a, CD68, CD163, CD11b, CD56, FoxP3, Granzyme B, Ki67, PD-1, PD-L1, LAG3, TIM-3, CTLA-4, cytokeratins, EGFR, GFAP, NDRG1, SOX2, SOX9, survivin, CD31, aSMA, vimentin, ATP1A1, and GAPDH. Images were kindly provided by Leica Microsystems.

Generating Data and Selecting Features for High Dimensional Analysis

The High Dimensional Analysis module operates on object-level data generated by cell- and object-based HALO modules, including **Highplex FL**, **Multiplex IHC**, **FISH**, and **Object Colocalization** modules. Additionally, the output of any HALO AI algorithm(s) included in the module analysis, such as tissue classifiers or object phenotypers, can be ingested by the High Dimensional Analysis module.

After analysis has been completed using a compatible HALO module, analysis results from one or more images can be selected. Data from multiple images can be pooled for collective analysis as long

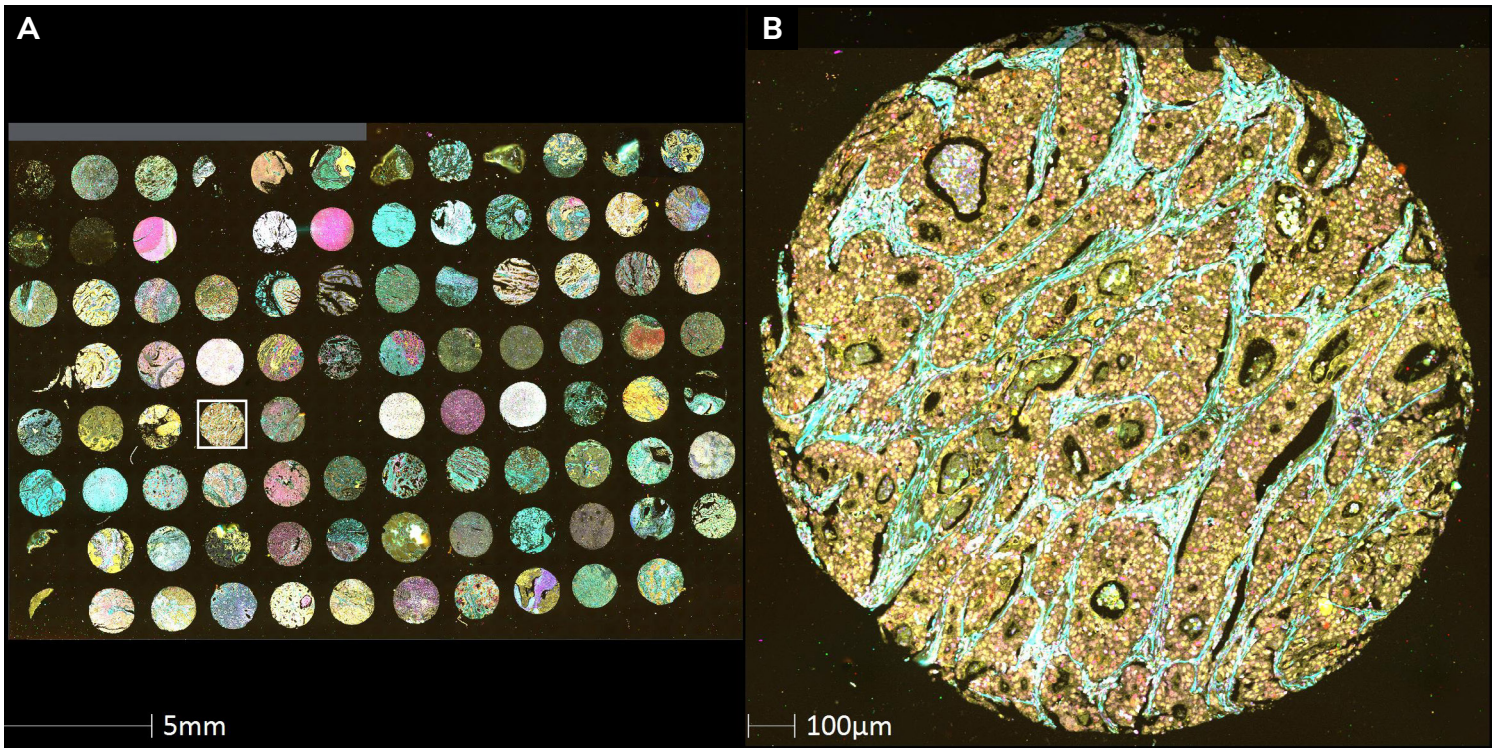


Figure 1. Tissue microarray (TMA) and analyzed lung adenocarcinoma core. **A:** A TMA containing 40 tumor tissues from 27 different anatomic sites was stained and imaged for 30 protein targets. **B:** Core of a stage three lung adenocarcinoma that was subsequently analyzed using the High Dimensional Analysis module.

as the same analysis settings were used across all analysis jobs. After selecting the analysis result(s) of interest and launching the High Dimensional Analysis workflow, users are presented with a list of available features present in the analysis. Depending on the module and analysis settings, features could include marker cell and compartment intensity, marker positivity classification, cell or object area, nuclear shape, and others.

Group	Feature
Cell Intensity	
☑	CD45 Cell Intensity
☐	DAPI Cell Intensity
☑	SOX9 Cell Intensity
☑	CD31 Cell Intensity
☑	FoxP3 Cell Intensity
☑	CD8a Cell Intensity
☑	GAPDH Cell Intensity
☑	PD-L1 Cell Intensity

Figure 2. Feature selection in the High Dimensional Analysis module. Users can select individual features or add entire feature groups with a single click, enabling fine tuning in a streamlined workflow.

Researchers can include any combination of features in their high dimensional analysis, selecting them either individually or as a feature group. Choosing appropriate features is critical to meaningful analysis, with redundant, omitted, or irrelevant features potentially obscuring true biological trends. Reviewing the clustering of known cell populations in a sample can help with optimizing the selection of features for analysis.

In the current study, one 30-plex lung adenocarcinoma core was analyzed using the Highplex FL module with pretrained AI nuclear and membrane segmentation algorithms. Proper cell segmentation was verified by visual inspection and marker positivity thresholds were not defined. Analysis was performed on the entire tissue and per-object data was saved. Features selected for the subsequent high dimensional analysis included cell intensity values for all markers except for DAPI, with a subset of these features shown in **Figure 2.**

Algorithm	Parameter	Higher Value	Lower Value
K-means	Number of clusters	More clusters	Fewer clusters
PhenoGraph	K-nearest neighbors	Fewer clusters	More clusters

Table 1. Exposed clustering algorithm parameters. Adjustments made to the exposed parameters for K-means and PhenoGraph have opposite effects on the number of clusters detected.

Selecting and Optimizing Parameters for Dimensionality Reduction and Clustering

Once features are chosen, users select and configure dimensionality reduction and clustering algorithms. The High Dimensional Analysis module supports three algorithms for these steps: UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction and K-means and PhenoGraph for unsupervised clustering.

UMAP was included in the High Dimensional Analysis module due to its balance of speed, scalability, and preservation of meaningful relationships within data. In the module, UMAP projects high-dimensional data into two dimensions while maintaining both the global and finer local details of cellular or object relationships. The module exposes one UMAP parameter, the neighbor count, which allows fine tuning of the balance between global and local structure preservation. Lower values emphasize local relationships and are better for identifying specific subpopulations of cells, while higher values capture broader relationships and trends across cell types.

Among the two clustering methods, K-means is a fast, readily interpretable algorithm that requires the user to specify the number of clusters in advance. It performs well when the number and identity of phenotypic groups are already known or hypothesized. With its adjustable number of clusters parameter, the algorithm also enables a "first look" at dimensionally reduced data, with users able to set the number of clusters to one to review the data for apparent trends.

PhenoGraph is a data-driven clustering algorithm that is more computationally intensive than K-means but does not require users to define the number of clusters beforehand. Instead, it constructs a graph based on nearest neighbors and uses community detection to find clusters. This makes PhenoGraph particularly well suited for exploratory analyses where rare or subtle populations may be present. PhenoGraph exposes a single parameter, k-nearest neighbors, that influences graph density and the sensitivity of cluster detection. **Table 1** summarizes how the parameters exposed by PhenoGraph and K-means impact the number of detected clusters. No matter which clustering algorithm is used, the identified cluster information is saved back to the object data table that was generated during analysis with the initial cell- or object-based HALO module.

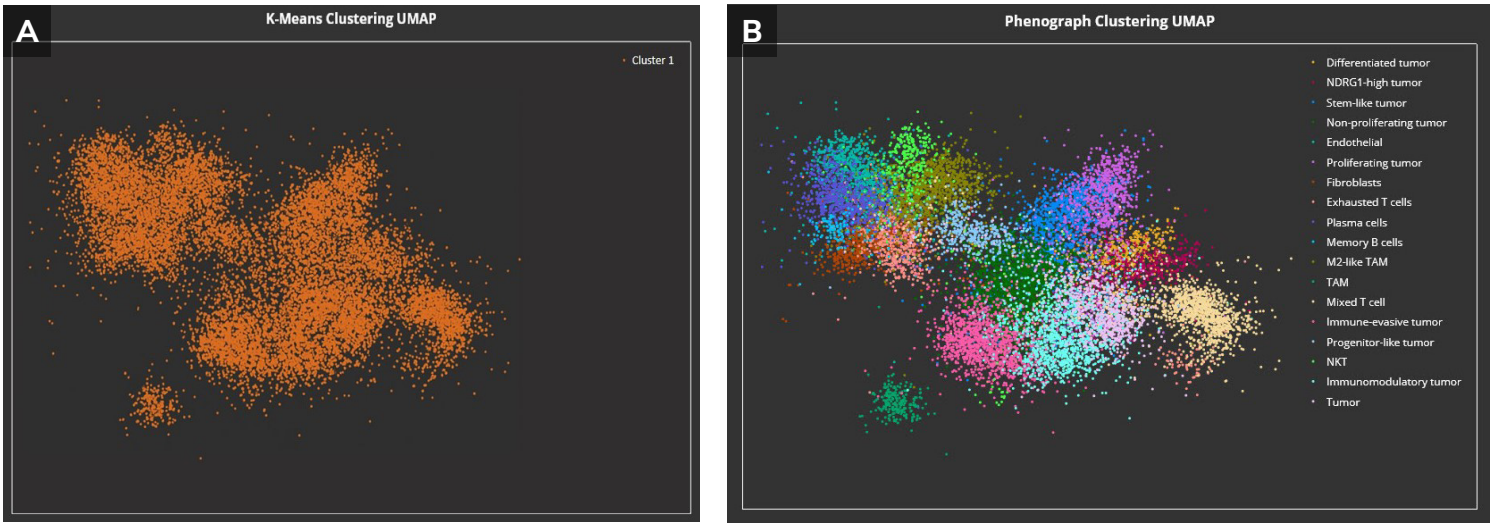
In this study, the UMAP neighbor count parameter was maintained at the default value of 15 and an initial exploratory analysis was performed using K-means clustering with the number of clusters set at one. The generated UMAP plot is shown in **Figure 3A**. The data in this "first look" analysis showed satisfactory separation with several apparent clusters, supporting the use of this UMAP neighbor count value in subsequent analysis. A PhenoGraph algorithm was used in the final analysis to increase the probability of identifying

rare populations. k-nearest neighbors was set to 50 to decrease the sensitivity of cluster detection after observing some suboptimal cluster differentiation in the initial results with the default k-nearest neighbors value of 20. The algorithm identified 18 clusters with these final settings, with results shown in **Figure 3B** and putative cluster names listed in **Figure 3C**.

Reviewing and Interacting with Analysis Results

Upon completing the analysis, the High Dimensional Analysis module presents results in an interactive interface. The UMAP plot displays each cell as a point, colored by cluster assignment, and users can zoom in and pan for a closer look at their data and rename clusters directly within the interface. In **Figure 3B**, the 18 clusters identified in analysis of the lung adenocarcinoma core were renamed after the initial review of the underlying phenotypes.

A key feature of the plot is its integration with the original slide image(s). The centroids of cells in the image are automatically marked with a colored dot corresponding to their assigned cluster, as shown in **Figure 4A**. When data points are selected in the UMAP plot, the corresponding cells are highlighted in the slide image, as demonstrated in **Figure 4B** with the cluster identified as endothelial cells. This live link assists researchers in validating phenotypes or exploring novel findings by examining cellular morphology, marker combinations, spatial distribution, and neighboring



C

Differentiated tumor	NDRG1-high tumor	Stem-like tumor
Non-proliferating tumor	Endothelial	Proliferating tumor
Fibroblasts	Exhausted T cells	Plasma cells
Memory B cells	M2-like TAM	TAM
Mixed T cell	Immune-evasive tumor	Progenitor-like tumor
NKT	Immunomodulatory tumor	Tumor

Figure 3. UMAP plot with initial and final clustering. **A:** Using the K-means algorithm set to a single cluster enabled review of UMAP performance while minimizing clustering compute time. **B:** The final unsupervised clustering using a Phenograph algorithm with K-nearest neighbors set to 50 identified 18 clusters that were later renamed based on the combination of marker expressions. **C:** A list of the putative clusters.

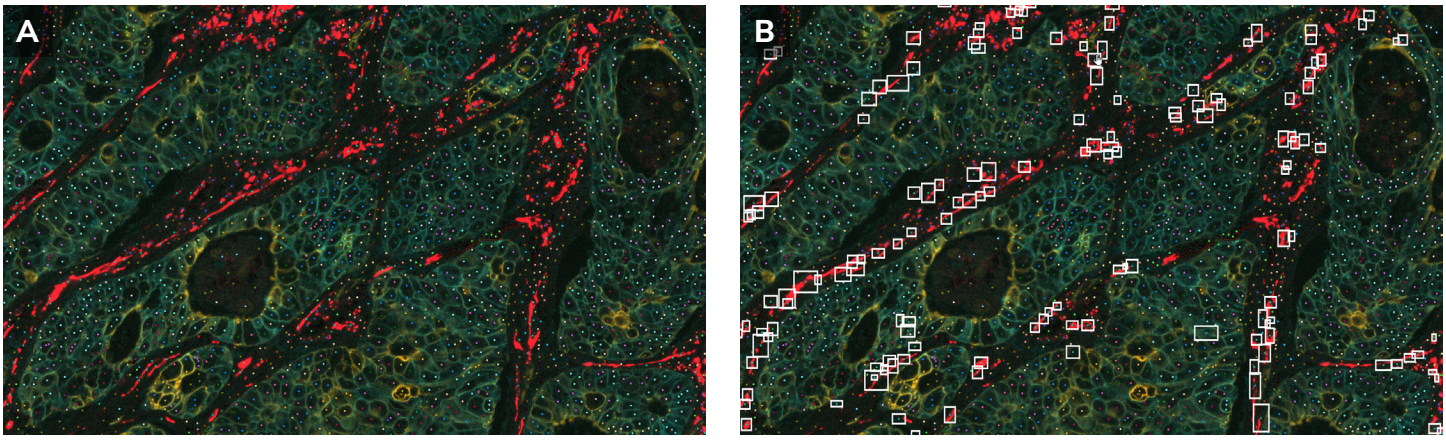


Figure 4. Image overlaid with cluster assignments. **A:** Image showing the cell centroids marked with colored dots corresponding to cell clusters. Only selected protein markers are shown (CD31, EGFR, and cytokeratins) to help visualize the cell clusters. **B:** Image with the cells assigned to the endothelial cell cluster selected.

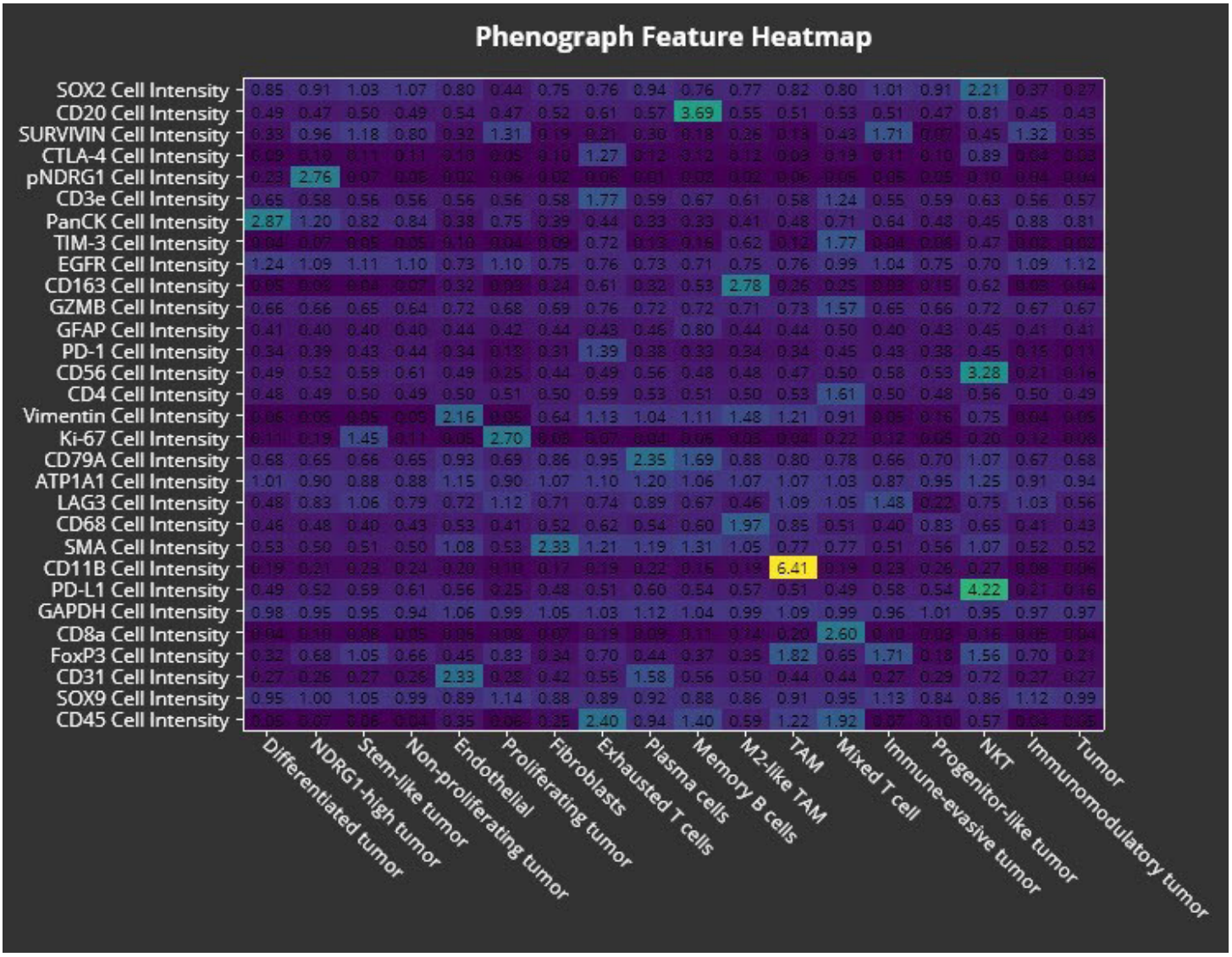


Figure 5. Feature heatmap. The feature heatmap for the lung adenocarcinoma core displays all the features selected for analysis on the y-axis and the renamed clusters on the x-axis.

context. Researchers can quickly perform quality control of clustering and assess whether clusters correspond to expected or potentially novel populations.

The module also generates a feature heatmap that complements the UMAP plot by displaying average values for all included features across each cluster, as shown in **Figure 5**. This visual summary supports phenotype interpretation by showing which biomarkers or morphological attributes differentiate clusters. The heatmap can also assist with verifying biologically significant separation of clusters while initially optimizing clustering algorithm parameters.

At any point during analysis, users can quickly iterate by modifying the features included, adjusting UMAP or clustering parameters, and re-running the analysis. Adjusting features or UMAP neighbor counts necessitates rerunning the full analysis, while changes to clustering algorithms are faster, requiring just computation of the adjusted clusters. Once satisfied, researchers can save the results to HALO, export data for additional review, or continue downstream analysis within the HALO Spatial Analysis module as the clustering results are saved to the object data.

After iterating our algorithm settings as outlined previously, our analysis of a lung adenocarcinoma core putatively identified diverse tumor, immune, and stromal phenotypes. Investigation of the spatial distribution of well-defined populations, including endothelial (**Figure 4B**) and fibroblast (not shown) cells, revealed characteristic patterns, supporting their proper clustering.

Conclusions

With highly multiplexed imaging, the promise of investigating phenotypic diversity, uncovering novel cell populations, and probing deeper into complex spatial relationships comes with the challenge of interpreting complex datasets with dozens or hundreds of dimensions. Here, we walked through the High Dimensional Analysis module workflow and demonstrated how it enables researchers to interactively explore these datasets and extract biologically meaningful insights. Combined with HALO's robust segmentation, quantification, and spatial analysis capabilities, the High Dimensional Analysis module provides a powerful end-to-end solution for highly multiplexed image analysis.

US Headquarters

Indica Labs, Inc
8700 Education Pl NW, Bldg B
Albuquerque, NM 87114 USA
+1 (505) 492-0979
info@indicalab.com

UK and Europe

+44 (0) 333 090 1113
emea@indicalab.com

Japan

+81 (0)3 4400 0466
japan@indicalab.com

China

+86 13761896143
china@indicalab.com

Tech Support: support@indicalab.com

Website: https://indicalab.com

indica labs @IndicaLabs.HALO