An unsupervised learning clustering workflow for automated cell phenotyping of hyperplex immunofluorescence images reveals cell types and neighbourhoods

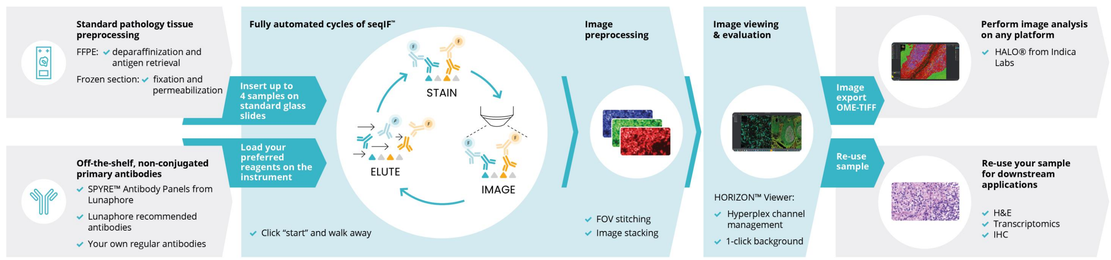
W. Edelman[1], I. Nearchou[1], J. Kowal[2], F. Rivest[2], A. Ortiz[1], D. Bowman[1], J. Dzubay[1], B. Pelz[2], and D. Dupouy[2]

[1]Indica Labs Inc., Pharmaceutical Services, Albuquerque, New Mexico, United States.
[2]Lunaphore Technologies SA, Tolochenaz, Switzerland.

BACKGROUND

The emergence of multiplex immunofluorescence approaches and multiomics technologies allowed the simultaneous detection of dozens of biomarkers in a single tissue slide[1]. Manually defining phenotypes can be a bottleneck for analysis of whole tissue images. Previously, we described the phenotyping of cells across different tumour types using HALO® and 20 biomarkers. In this work, we use the automated sequential immunofluorescence (seqIF™) protocol on the COMET™ platform[2], HALO® to segment and organize a tissue microarray (TMA), and HALO AI to segment and delineate regions of interest. Here, we leverage open-source tools for the unsupervised clustering (Seurat)[3] of cells present in the TMA and perform neighbourhood analysis[4] to gain insight into changes in population structure between groups of tumour, non-tumour, and tumour types. We highlight changes in neighbourhood representation between groups and the workflow necessary to accomplish this level of analysis. The workflow presented here provides an easy-to-adopt solution to extract quantitative information from hyperplex datasets and to transform pixels into biologically relevant insights.

COMET™ FULLY-AUTOMATED WORKFLOW



COMET™ workflow applies sequential immunofluorescence (seqIF™) protocol, where tissues are undergoing cycles of staining, imaging and antibody elution in automated manner without the need of user intervention. COMET™ provides the possibility to perform up to 20 cycles per automated run, what yields in 40 plex immunostaining image. Final images are delivered ready to be analyzed and there is no need of data postprocessing from the user. seqIF™ protocol was executed in fully automated fashion on COMET™ system. FFPE slide was preprocessed with the use of PT module with dedicated reagents at pH9 (Epredia). Afterwards, slides were stored in Multistaining Buffer (Lunaphore) prior to loading on the system. The dedicated protocol was created with COMET™ Control Software. Off-the-shelf antibodies were prepared offline and loaded on the system together with proprietary buffers enabling all the steps of seqIF™ protocol: washing, imaging, quenching and elution buffer (Lunaphore). At each cycle signals from 2 markers and DAPI were acquired by the integrated fluorescent microscope. As final results, single multichannel OME.TIFF image is assembled.

Instrument

- COMET™ - for staining and imaging

Accessories

- PT Module
  Dewaxing & antigen retrieval
- Compressor

Software

- COMET™ Control Software
- HORIZON™ Viewer for image quality control
- Panel Builder app: create seqIF™ protocol *in silico*

HALO®, HALO AI IMAGE ANALYSIS and DATA ANALYSIS WORKFLOW

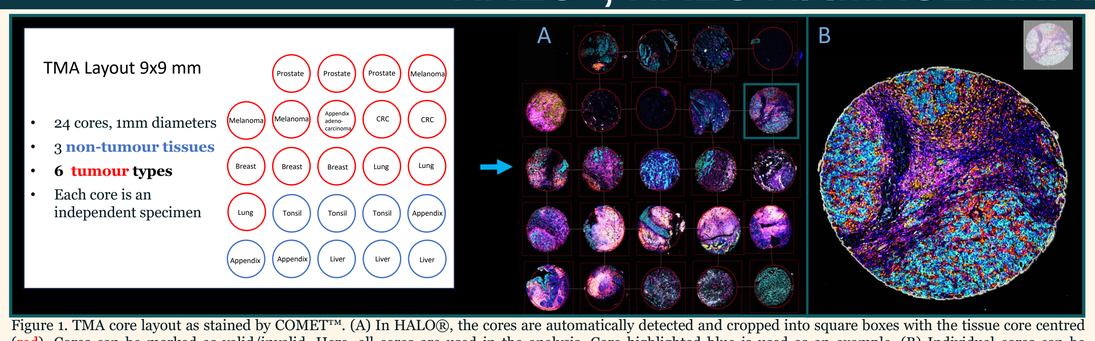


Figure 1. TMA core layout as stained by COMET™. (A) In HALO®, the cores are automatically detected and cropped into square boxes with the tissue core centred (red). Cores can be marked as valid/invalid. Here, all cores are used in the analysis. Core highlighted blue is used as an example. (B) Individual cores can be automatically annotated using a tissue detection classifier, the flood tool, or with manual edits.

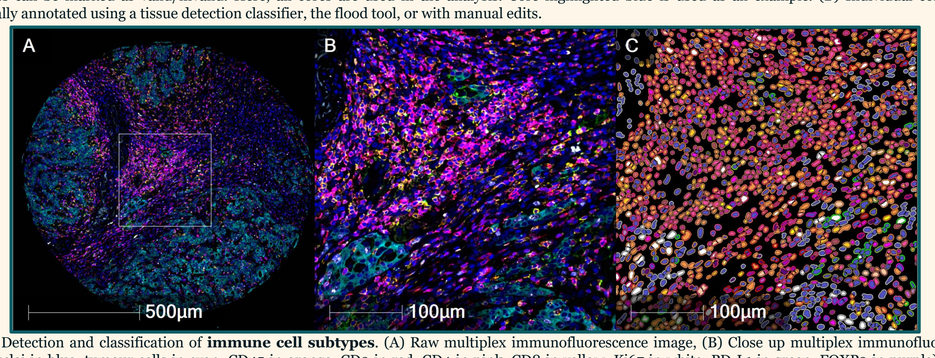


Figure 2. Detection and classification of **immune cell subtypes**. (A) Raw multiplex immunofluorescence image, (B) Close up multiplex immunofluorescence image (nuclei in blue, tumour cells in cyan, CD45 in orange, CD3 in red, CD4 in pink, CD8 in yellow, Ki67 in white, PD-L1 in green, FOXP3 in purple), (C) cell classification/analysis mark-up (colocalization) using the Highplex FL module.

HALO AI

Tissue classification

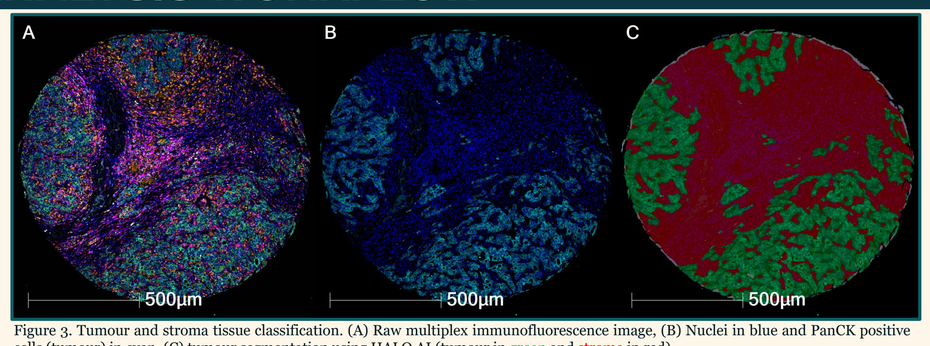


Figure 3. Tumour and stroma tissue classification. (A) Raw multiplex immunofluorescence image, (B) Nuclei in blue and PanCK positive cells (tumour) in cyan, (C) tumour segmentation using HALO AI (tumour in green and stroma in red).

HALO®

Object data, per cell export, .csv (.fcs export feature also in HALO®).

SEURAT[3]- R toolkit for single cell genomics v4.3.0→ neighbourhood analysis [4]

1. Import csv into R as a data.table, all keeping all "Cell intensity" columns and "Nucleus Area (µm²)": **CD8, CD3, CD4, CD45, Ki67, PD-L1, FoxP3, CD16, CD68, HLA-DR, PanCK**
2. Convert HALO .csv object data to *Seurat object*

```
CreateSeuratObject(dataframe)
```

3. Scale data

```
ScaleData(dataframe)
```

4. Data Dimension reduction

```
RunPCA()
RunUMAP() #n.neighbors = 25
FindNeighbors #k.param = 25
FindClusters()#original Louvain algorithm
```

5. Plotting cluster onto UMAP
6. Export .csv with clusters, cell coordinates, metadata and run Neighbourhood analysis[4]. Example code available through paper, and current workflow available up on request.

HALO® DATA ANALYSIS RESULTS

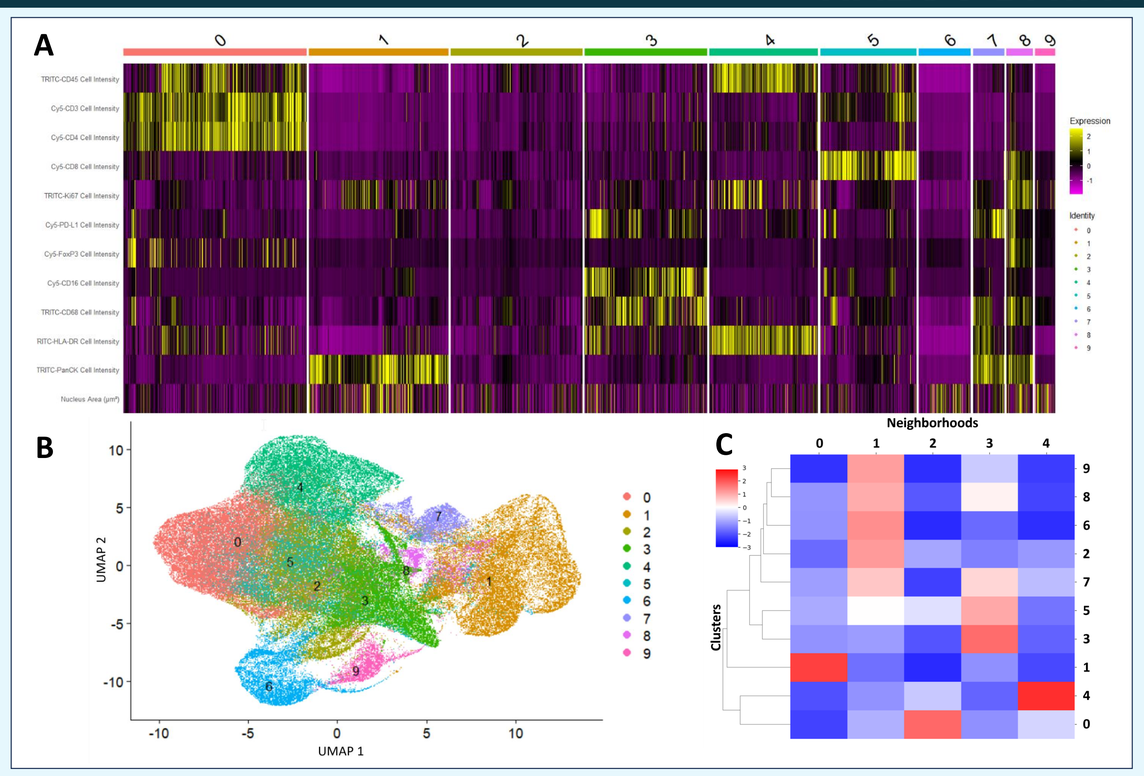


Figure 4. Identification of potential cell types from cells segmented by HALO. (A) Unsupervised clustering in Seurat[3] using the original Louvain algorithm, k.param = 25, resulted in 9 clustered populations. Their expression level and frequency of each marker is shown. (B) UMAP dimension reduction plot generated potential groups of cells, which are assigned a colour based on clustering. (C) Cellular neighbourhood identification (window = 5 nearest cells, by Euclidean distance) heatmap coloured by the neighbourhood enrichment score (frequencies), showing abundance of clusters 0-9 (y-axis) for a given cellular neighbourhood 0-4 (x-axis).

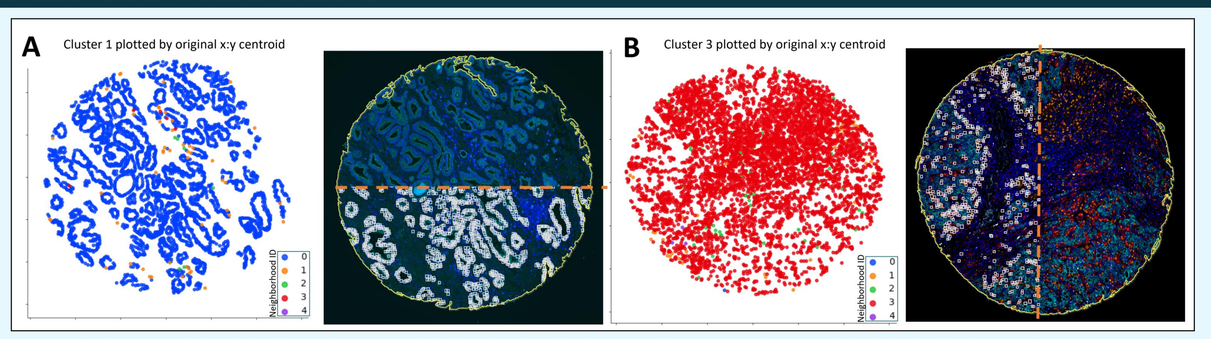


Figure 5. Confirmation of identified cell types by unsupervised clustering in original X:Y coordinate space reflects morphological colocalization. (A) Core 3, prostate acinar adenocarcinoma. Cluster ID "1", abundant for PanCK+ cells, is plotted, whilst neighbourhood ID label colours the cells (left). Core image in HALO shows original staining pattern in tumour cells, and PanCK+ cells are highlighted in bounding boxes (right below dotted line). (B) Core 9, Colon adenocarcinoma. Cluster ID "3" is plotted showing an abundance for $CD68^{+}CD16^{+}PD\text{-}L1^{+}$ cells. In rightmost image, HALO phenotyped CD68+CD16+ manually threshold cells are highlighted in bounding boxes.

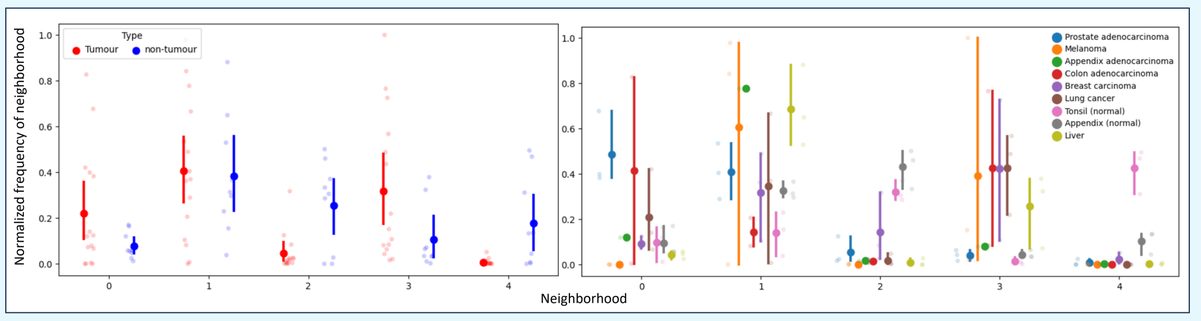


Figure 6. Neighbourhood analysis plot. Percent of total cells allocated to each neighbourhood between Tumour and non-tumour cores (left plot). As expected from the neighbourhood heatmap, neighbourhood "0" with abundant panCK, shows ~2-fold change between Tumour and non-tumour cores. Relative frequencies of neighbourhood between indications grouped by tissue type. In prostate cancer, we observed higher relative frequencies of neighbourhood "0" which was driven by high pancytokeratin-expressing cell clusters (1) as compared to the lowest representation in melanoma. This pattern is between prostate cancer and melanoma is a known contrasting feature[5]

CONCLUSIONS

Tumour microenvironment (TME) composition is implicated as an important regulator of the response to treatment and patient prognosis[6]. Significant efforts are continuously carried out to better understand how tumours affect their surrounding by recruiting healthy cells to their proximity and interfering with their functional states. Hyperplex immunofluorescence enables interrogating TME in the manner, where complex phenotypes can be identified within the spatial context of the tissue, empowering researchers to understand better intercellular interactions and tissue-intrinsic biology[7]. Current state-of-art methodologies allowing questioning proteomic composition of TME at large scale are costly in terms of time and resources to execute both experiments and data analyses, limiting the adoption and day-to-day use of hyperplex immunofluorescence. The workflow presented here highlights the accessible of adoption of seqIF™ protocol and an unsupervised image analysis pipeline. COMET™ platform ensures single-cell resolution and the simultaneous detection capability of multiple protein biomarkers with high reproducibility. COMET™ automates all steps of protocol execution and limits inter-experiments variability, while using standard reagents and delivering 12 40-plex images in 1 week. We demonstrate further how the combination of COMET™ hyperplex images with the HALO® and HALO AI image analysis platform from Indica Labs, results in an easy and straightforward workflow for interrogating heterogeneous TME and depicting tissue architecture on the single-cell level. HALO® and HALO AI enable guided and automated data extraction from hyperplex images with flexible workflow design. Our analysis of a 20-plex immuno-oncology panel on a TMA with both normal and tumour tissues, yielded several insights into the cellular phenotypes (Figure 4B) and varying neighbourhood compositions between indications (Figure 6). We were able to use a previously described method for colorectal cancer neighbourhood analysis[4] and apply it to a hyperplex TMA workflow from HALO®, with relative ease. We highlight two publicly available open-source algorithms, and show they analyse HALO outputs which include cell coordinates, intensities, nuclear size, and metadata (e.g. tumour vs. non tumour core, core location, and indications). While this workflow demonstrates the integration of large amounts of data, TMAs require special considerations. Under sampling of relevant cell types is a potential risk. Secondly, careful curation input data is necessary to ensure that each cell is labelled with its proper image name + TMA coordinate, such that the nearest spatial neighbours (Figure 4C) are only calculated within each respective core. COMET™ images, together with HALO® and HALO AI™, can be directly used for quantitative analysis of the TME, enabling researchers the identification of the biomarkers across different tumours and at a single-cell level. The high throughput of COMET together with HALO® and HALO AI workflows stand out as tools that allow bringing the hyperplex immunofluorescence to every laboratory and streamlines the study of TME across basic and translational research.

REFERENCES

(1) Bressan, Dario et al. "The dawn of spatial omics." *Science (New York, N.Y.)* vol. 381,6657 (2023): eabq4964. doi:10.1126/science.abq4964
(2) Rivest, François et al. "Fully automated sequential immunofluorescence (seqIF) for hyperplex spatial proteomics." *Scientific reports* vol. 13,1 16994. 9 Oct. 2023, doi:10.1038/s41598-023-43435-w
(3) Hao, Yuhan et al. "Integrated analysis of multimodal single-cell data." *Cell* vol. 184,13 (2021): 3573-3587.e29. doi:10.1016/j.cell.2021.04.048
(4) Schürch, Christian M et al. "Coordinated Cellular neighbourhoods Orchestrate Antitumoural Immunity at the Colorectal Cancer Invasive Front." *Cell* vol. 182,5 (2020): 1341-1359.e19. doi:10.1016/j.cell.2020.07.005
(5) Menz, Anne et al. "Pan-keratin Immunostaining in Human tumours: A Tissue Microarray Study of 15,940 tumours." *International Journal of Surgical Pathology*. 31,6 (2023): 927-938. doi:10.1177/10668969221117243
(6) Hanahan D and Weinberg RA. Hallmarks of Cancer: The Next generation. Cell 5 :144 (2011). https://doi.org/10.1016/j.cell.2011.02.013
(7) Lewis, S.M., Asselin-Labat, ML., Nguyen, Q. et al. Spatial omics and multiplexed imaging to explore cancer biology. Nat Methods 18, 997–1012 (2021). https://doi.org/10.1038/s41592-021-01203-6

Correspondence: bedelman@indicalab.com, joanna.kowal@bio-techne.com