

Exploring Methods to Analyze and Process Spatial Proteomics Data

Masters Thesis in Bioinformatics (30 ECTS)

Tessa Jane Divita



AARHUS UNIVERSITY

Bioinformatics Research Center (BiRC)

Supervisor: Lars Dyrskjøt

Department of Molecular Medicine

Spring 2024

Abstract

This study used imaging mass cytometry (IMC) to profile the tumor microenvironment of 69 non-muscle invasive bladder cancer (NMIBC) tumors from 58 patients treated with Bacillus Calmette-Guérin (BCG). 199 regions of interest (ROIs) from tissue microarrays (TMA) were stained with 38 metal-conjugated antibodies and imaged. Ilastik pixel classification and DeepCell Mesmer were used for cell segmentation, and an anndata object was created with cell coordinates and marker expression levels.

Quality control filtering, batch correction using Harmony, PCA, UMAP dimension reduction, and Leiden clustering were applied to assign cell types. Differential expression analysis ranked markers for each cluster, and clusters were annotated with cell types like tumor, T cells, macrophages, etc. Neighborhood analysis using a k-nearest neighbors approach found spatially coherent neighborhoods enriched for different cell types.

The results provide a high-dimensional, spatially-resolved map of the NMIBC tumor microenvironment, revealing the complexity of immune and stromal cell interactions. The analytical pipeline integrates multiple machine learning techniques to robustly identify cell types and spatial patterns from IMC data. This approach can be applied to study the tumor microenvironment in other cancer types.

Abbreviations

Imaging Mass Cytometry (IMC)

Bacillus Calmette-Guérin (BCG)

Region of Interest (ROI)

Tissue Microarray (TMA)

Principle Component Analysis (PCA)

Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP)

Bladder Cancer (BC)

Muscle-invasive bladder cancer (MIBC)

Non-muscle-invasive bladder cancer (NMIBC)

Contents

1	Introduction	4
1.1	Non-muscle invasive bladder cancer	4
1.2	Tumor Microenvironment	6
1.3	Spatial Proteomics	8
1.4	Aim of Study	8
2	Methods	10
2.1	Dataset	10
2.2	Pre-processing	11
2.2.1	Pixel Classification	14
2.2.2	Cell Segmentation	15
2.2.3	Feature Extraction	16
2.3	Cell Typing	16
2.3.1	Quality Control	16
2.3.2	Signal Processing	18
2.3.3	Batch Correction	18
2.3.4	UMAP	19
2.3.5	Leiden Clustering	21
2.3.6	Ranking	22
2.4	Spatial Analysis	24
2.4.1	Cell-Specific Clustering	24
2.4.2	Cellular Neighborhood	25
3	Results	27
3.1	Single Cell Object Preparation	27
3.1.1	Quality Control	27
3.1.2	Signal Processing	28
3.1.3	Batch Correction	30
3.2	Towards Cell Typing	31
3.3	Spatial Analysis	35
4	Discussion	42
4.1	Image processing	42
4.2	Cell Typing	44
4.3	Spatial Analysis	45
5	Conclusion	53
6	Acknowledgements	54