E.T.S. de Ingeniería Industrial, Informática y de Telecomunicación

Image Analysis and Deep Learning for Urothelial Carcinoma Tumor Microenvironment Characterization

Máster Universitario en Ingeniería de Telecomunicación

Trabajo Fin de Máster

María Sangüesa Recalde

Director: Mikel Ariz Galilea

Pamplona, febrero 2024

upna

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Abstract

This project aims to develop image analysis and deep learning methods for the analysis of the tumor microenvironment in images of biopsies of urothelial carcinoma patients. Tumor biopsies were obtained from patients in the Clínica Universidad de Navarra, and 7-plex immunofluorescence images were acquired from tissue microarrays using the Vectra Polaris multispectral image acquisition device (Akoya Biosystems). The images are composed of seven channels corresponding to seven fluorochromes labelling tumor cells and different immune cell populations. The objectives of the project include the segmentation of cells based on nuclear and membrane markers, the classification of different cell phenotypes based on the expression of fluorescent markers, and the study of interactions between cell populations in relation with the patient clinical outcome. The nuclei segmentation deep learning-based method implemented and trained with our dataset outperforms other state-of-the-art methods in terms of segmentation accuracy. Whole-cell segmentation and cell phenotype classification have also been optimized to detect two cell populations in our patient cohort, due to their clinical interest: CD8+ cells and BATF3+ cells. Within the obtained results, it should be highlighted that the density of CD8+ cell population within the tumor microenvironment serves as an indicator of the response to Atezolizumab immunotherapy treatment in patients with urothelial carcinoma. Additionally, is has been found that spatial interactions between BATF3+ and CD8+ cell populations also have a direct impact on the patient outcomes. These conclusions significantly contribute to clinical research by identifying new relevant biomarkers within the immune system.

Key words: Deep learning, tumor microenvironment, urothelial carcinoma, multiplex imaging, spatial interactions.

Resumen

Este proyecto tiene como objetivo el desarrollo de métodos de análisis de imagen y aprendizaje profundo para el análisis del microambiente tumoral en imágenes de biopsias de pacientes con carcinoma urotelial. Las biopsias tumorales fueron obtenidas de pacientes en la Clínica Universidad de Navarra y las imágenes de inmunofluorescencia de siete canales se consiguieron utilizando el dispositivo de adquisición de imágenes multiespectrales Vectra Polaris de Akoya Biosystems. Las imágenes están compuestas por siete canales que corresponden a siete fluorocromos que etiquetan células tumorales y diferentes poblaciones de células inmunitarias. Los objetivos del proyecto incluyen la segmentación de células basada en marcadores nucleares y de membrana, la clasificación de diferentes fenotipos celulares basada en la expresión de marcadores fluorescentes, y el estudio de las interacciones entre las poblaciones celulares en relación con el resultado clínico del paciente. El método basado en Deep Learning utilizado para segmentación nuclear implementado y entrenado con nuestro conjunto de datos supera a otros métodos del estado del arte en términos de precisión de segmentación. La segmentación de célula complete y la clasificación de fenotipos celulares han sido también optimizados para detectar dos poblaciones celulares en nuestra cohorte de pacientes, dado su interés clínico: células CD8+ y células BATF3+.Entre los resultados obtenidos se debe destacar que la densidad de poblaciones celulares CD8+ dentro del microambiente tumoral sirve como un indicador de la respuesta mejor al tratamiento de inmunoterapia Atezolizumab en pacientes con carcinoma urotelial. Además, se ha descubierto que las interacciones entre las poblaciones celulares BATF3+ y CD8+ también tienen un impacto directo en la evolución de los pacientes. Estas conclusiones contribuyen significativamente a la investigación clínica al identificar nuevos biomarcadores relevantes dentro del sistema inmune.

Palabras clave: Aprendizaje profundo, microambiente tumoral, carcinoma urotelial, imágenes multiplexadas, interacciones espaciales.

Contents

List of Figures

Image Analysis and Deep Learning for Urothelial Carcinoma Tumor

List of Tables

Chapter 1. Introduction

This document has been divided into five chapters, starting with the Introduction, followed by Objectives, Materials, Methods, Results and concluding with Discussion and Conclusion and Future Work. Except for the Methods and Results chapters, further explained below, all other chapters are common to both the Final Master Project and the Complement to the Final Master Project courses.

The Methods chapter is separated into four sections: Nuclei Segmentation, Whole Cell Segmentation, Classification of Cell Phenotypes and Study of Spatial Interactions Between Cell . Nuclei Segmentation is part of the Complement to the Final Master Project, while the rest corresponds to the Final Master Project course.

Indeed, the same applies to the Results chapter, where only the Nuclei Segmentation subsection belongs to the Complement to the Final Master Project.

1. 1. Clinical Context

Cancer ranks as the second leading cause of death worldwide, surpassed only by cardiovascular diseases [1], [2]. Figure 1.1. illustrates the mortality trends for all types of cancer, excluding non-melanoma skin cancer, in Spain from 1943 to 2018. It can be observed, that the incidence of mortality is higher in males than females, but there is a decreasing trend in both genders referring to mortality rates.
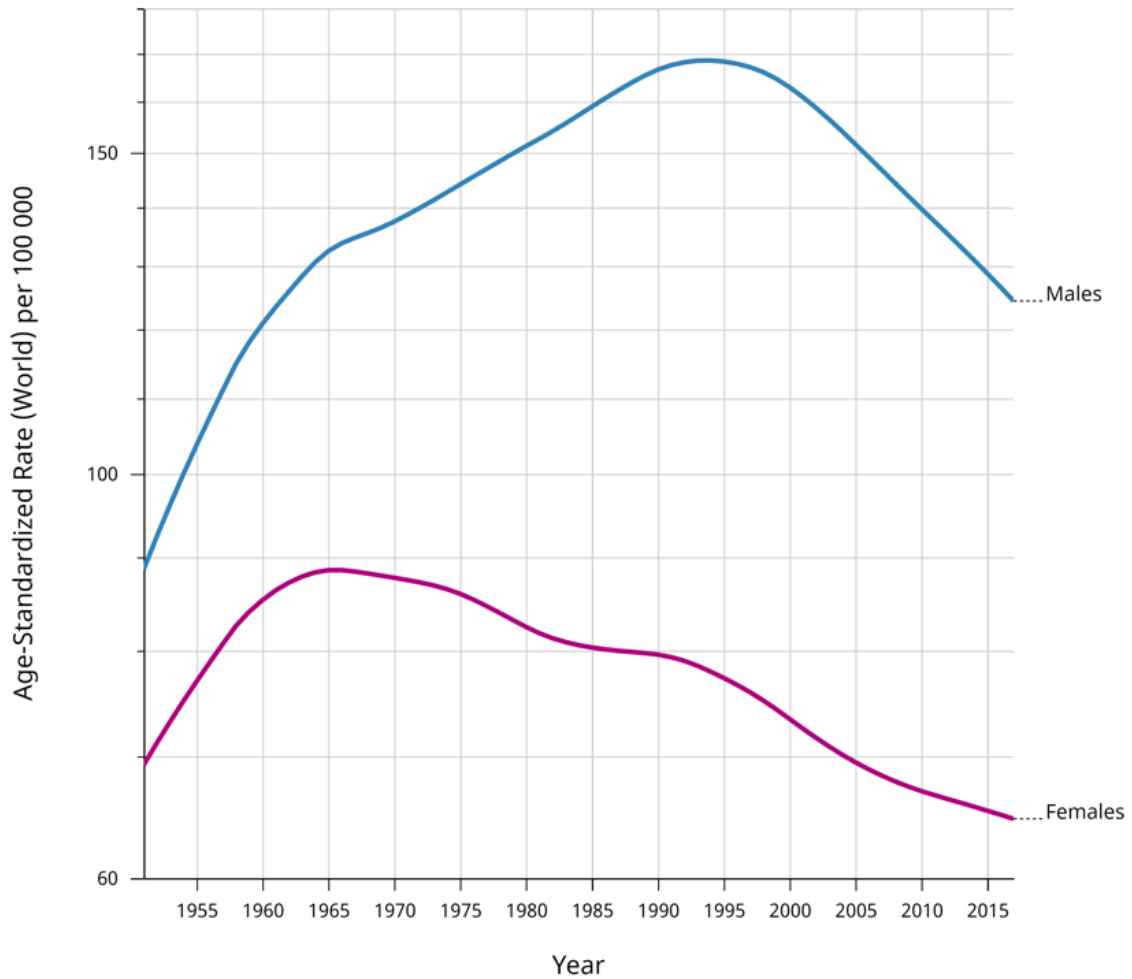
Figure 1.1: Cancer mortality statistics in Spain from 1943 to 2018 excluding non-melanoma skin cancer. Extracted from [3].

Urothelial carcinoma, which is predominantly seen in men [4], also known as transitional cell carcinoma, is a type of cancer that begins in urothelial cells. The term "transitional cells" is used for urothelial cells because they change shape; these cells present flexibility, stretching when the bladder is filled with urine and contracting when it is emptied. Urothelial cells cover various structures, including the urethra, bladder, ureters, renal pelvis, and certain other organs, see Figure 1.2.
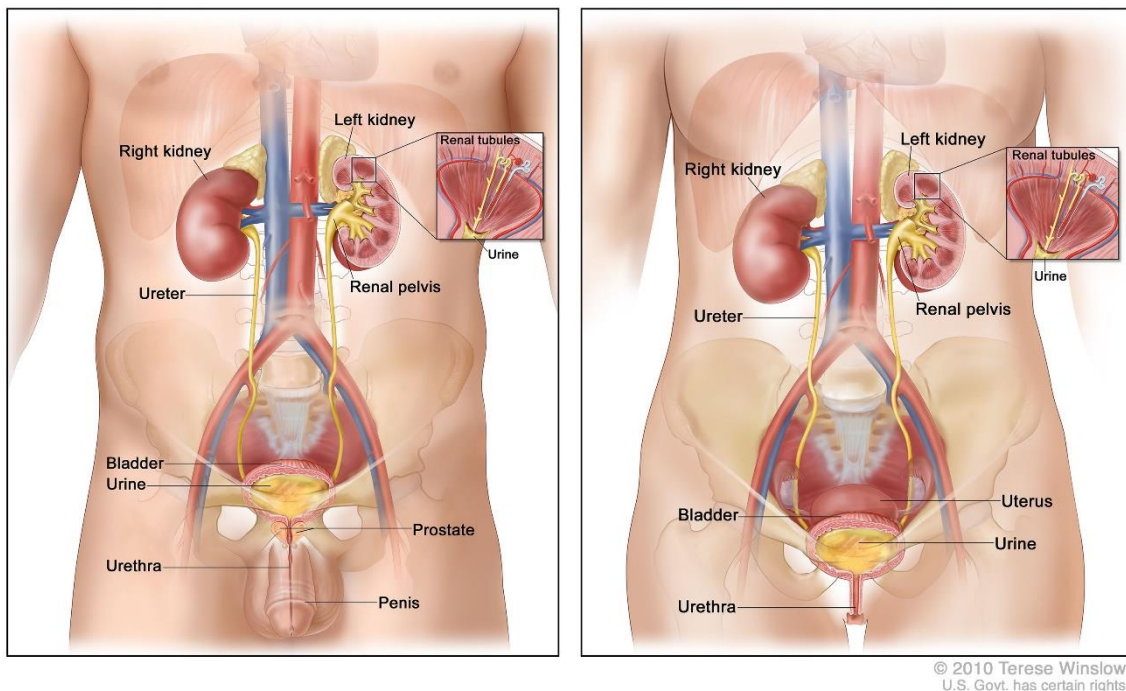
Figure 1.2: Anatomy of the male urinary system (left panel) and female urinary system (right panel), illustrating the kidneys, ureters, bladder, urethra, prostate, and penis (left panel), as well as the uterus (right panel). The internal view of the left kidney displays the renal pelvis. Extracted from [5].

Almost all bladder cancers fall under the category of urothelial carcinomas [6], making them the 10th most commonly diagnosed cancer and the 13th leading cause of cancer deaths worldwide [7], [8]. Bladder cancer is an aggressive malignancy with a five-year survival rate of about 5% in the metastatic cases [9]. Urothelial cancer is relatively infrequent in the upper urinary tract, specifically in the renal pelvis and ureter, constituting only 5-10% of cases. Its occurrence is even more uncommon in the urethra, comprising less than 1% [10].

The global cases and mortality of bladder cancer are expected to increase annually by 73% and 87%, respectively, by 2040 [11]. Therefore, there is an urgent need for investigation to develop treatments and facilitate rapid diagnosis due to its relevance on combating the disease [12].

Tumoral and immune cells coexist within the tumor microenvironment. The interactions between these cells, within the tumor itself or in its surrounding stroma, are crucial for patient diagnosis. Currently, several studies have been conducted to identify patterns in cellular interactions specific to each type of cancer, aiming to enhance the

understanding of the disease [13] [14]. Indeed, this approach enables the development of more individualized treatments for each type of cancer.

The tumour immune microenvironment has a direct impact on cancer progression, metastasis probability and response to cancer treatment, such as immunotherapy, being a pillar in the hallmarks of cancer [15][16][17][18]. The cancer grade is determined by the morphology of its cells and the tumour architecture, referred to as histopathology. At the same time, cancer biomarkers, which represent the phenotype, are used to categorize patients, predict prognosis, and adapt conventional or immune-based anticancer treatments [19].

The blockade of the PD-1/PD-L1 pathway is currently a standard treatment for advanced urothelial cancer [20]. However, reliable predictive biomarkers have not yet been defined. Therefore, one of the goals of this Final Master project is to identify cell populations, defined by the expression of specific biomarkers, that have a direct impact on the clinical response to treatment.

1. 2. Multiplexed Imaging

Cellular phenotypes and cellular interactions within the tumour microenvironment can be characterized thanks to the recent development of technologies such as multiplexed imaging technology [21], [22]. The analysis of multiple biomarkers present in each cell provide new biological insights that previous single-marker methods did not allow [23]. This, for instance, allows for establishing correlations between biomarker expression and treatment response, an ongoing task that remains both challenging and clinically relevant [24].

Multiplexed imaging currently allows the distinction of more than twenty of biomarkers, that reference cell phenotypes, within the tumor microenvironment. Analysing the interactions between these biomarkers enables a better characterization of the immunosuppressive mechanisms in the tumor microenvironment. Thus providing new insights into the prognosis of the disease [25].

Determining the expression and the spatial relations between the biomarkers remains a highly challenging task to humans due to the huge quantity of data that must be processed. For this reason, the development of machine learning-based methods to group cells with similar phenotypes into neighbourhoods where different cell phenotypes coexist, results in important advancements in clinical predictive tasks [19].

Lymphocytes, or T cells, are cells within the immune system that act as cell mediators with specific functions depending on their type. The most common types include CD8+ T cells, CD4+ T cells and T regulatory cells. In this project, the focus will be on CD8+ T cells, which are pathogens killers, because they play a vital role in anticancer treatment as key immune cells for killing cancer cells [26][27][28][29].

On the other hand, the expression of BATF3, basic leucine zipper transcription factor ATF-like 3, in conventional type-1 dendritic cells (cDC1) immediately before treatment is known to positively influence the response against tumors, despite the lack of studies on this topic. For a proper immune response by CD8 T lymphocytes, they need to encounter antigens presented by antigen-presenting cells such as type 1 dendritic cells detected by the presence of BATF3 expression. After interacting with antigen-presenting cells, CD8 T lymphocytes acquire cytotoxic functions. Subsequently, they migrate to the tumor site to generate an effector response. The BATF3 gene accomplishes this by regulating memory of the CD8+ T cells, while simultaneously enhancing their survival rates [30]. The presence of cDC1 cells is linked to the response to cancer treatments, as they present antigens to T cells within the tumor microenvironment [31].

1. 3. Deep Learning for Multiplexed Image Analysis

Histology images contain a wealth of valuable information due to the visualization of millions of different cells simultaneously. However, there is a significant challenge in optimizing the interpretation of these images, given the time-consuming nature of the process and its heavy

reliance on highly skilled professionals [32]. Recognizing the urgent need for more efficient, precise and cost-effective diagnostic methods for the analysis of histology images, the integration of Deep Learning techniques emerges as a crucial avenue.

Deep Learning, inspired by the function of the human brain, develops computational models capable of learning and performing complex tasks, mainly involving large datasets. It operates by the usage of artificial neural networks composed of interconnected layers of nodes. During the training process, the neural networks receive input data and adjust their internal parameters to learn patterns within the data to make predictions or classifications based on the input data. The complex features are extracted from the raw data as it passes through the layers of nodes.

The success of the proposed biological image analysis task in this project heavily relies on the accurate detection and segmentation of cells and nuclei from the microscopy images [33]. Innovative algorithms based on Deep Learning have made it possible to eliminate the laborious manual work involved. This is achieved by leveraging the algorithm's capacity to learn repetitive patterns within the cells of the tumor microenvironment, simplifying the analysis process and enhancing overall efficiency.

Multiplex fluorescence imaging assays have great potential for translational research and clinical practice, especially when combined with new computational methods based on Deep Learning. These methods can identify patterns of cellular interaction in raw data, promising improvements in patient diagnosis. The union of advanced imaging techniques and state-of-the-art Deep Learning algorithms opens new horizons for a more efficient and insightful analysis of histology images, ultimately contributing to improved medical diagnostics and treatment development.

Deep Learning plays a crucial role in every step of highly multiplexed imaging analysis, covering from cell segmentation and classification to deriving biological insights from the resulting images. It enables the establishment of correlations between the high-level visual understanding of pathology and deeper levels of biological mechanisms. By exploiting Deep Learning techniques, researchers can effectively navigate through the complexity of vast image datasets, extracting

meaningful information that facilitates a deeper understanding of the cellular and disease behaviour. This integration of Deep Learning methodologies throughout the analysis pipeline not only enhances the accuracy and efficiency of image processing tasks but also empowers researchers to uncover previously unseen relationships and patterns within the data, ultimately advancing the comprehension of biological systems [32].

Chapter 2. Objectives

The main goal of this project is to conduct a comprehensive analysis of the complex tumour microenvironment within images obtained from biopsies of urothelial carcinoma patients who were treated with Atezolizumab [34]. All images were obtained before patients received the treatment, allowing for the identification of correlations between the response to immunotherapy and the image patterns present in the tumor microenvironment at the time of the treatment. This analysis employs cutting edge image analysis techniques and Deep Learning methodologies. Briefly, the goals of this study cover the following key aspects:

1. Cell segmentation: A prior aspect of the research involves precise nuclear segmentation and whole cell segmentation, achieved, respectively, through the utilization of nuclear and membrane and cytoplasmic markers. The aim of this process is the identification and delineation of individual cells and its nuclei within the tissue.

2. Phenotypic classification: Another crucial point of the study is the classification of diverse cell phenotypes. This classification will be based on the expression patterns of fluorescent markers, providing valuable insights into the heterogeneity of the cellular composition within the urothelial carcinoma microenvironment.

3. Population interactions: Beyond cellular characterization, this project aims to evaluate the presence of specific cell phenotypes and the spatial interactions among distinct cell phenotypes. Specifically, the presence of CD8+ and BATF3+ cells and the distances between them will be studied due to the existing of evidence of their impact on urothelial anticancer immunotherapy treatments. By correlating them with clinical outcomes, the study seeks to discover potential associations between the tumour microenvironment and patient response to the immunotherapy treatment.

Chapter 3. Materials

3. 1. Images

For the realization of this Final Master Project, two tissue microarrays (TMAs) were utilized. Tissue microarray technique organizes several diminutive biological samples of tissue in a single recipient, called microarray. This method introduced significant advantages, such as allowing to generate more data from a single sample, analysing a large number of data simultaneously, presenting experimental uniformity by treating each core in an identical manner or decreasing the volume, cost and time of the assay [35]. In this Final Master project, each TMA contains small circular pieces of tissue taken from 65 and 60 biopsy samples of urothelial carcinoma patients, respectively.

In Figure 3.1, the location of TMA1 images on the microarray is represented. This image offers an insight into the heterogeneity presented among the images of the TMA.
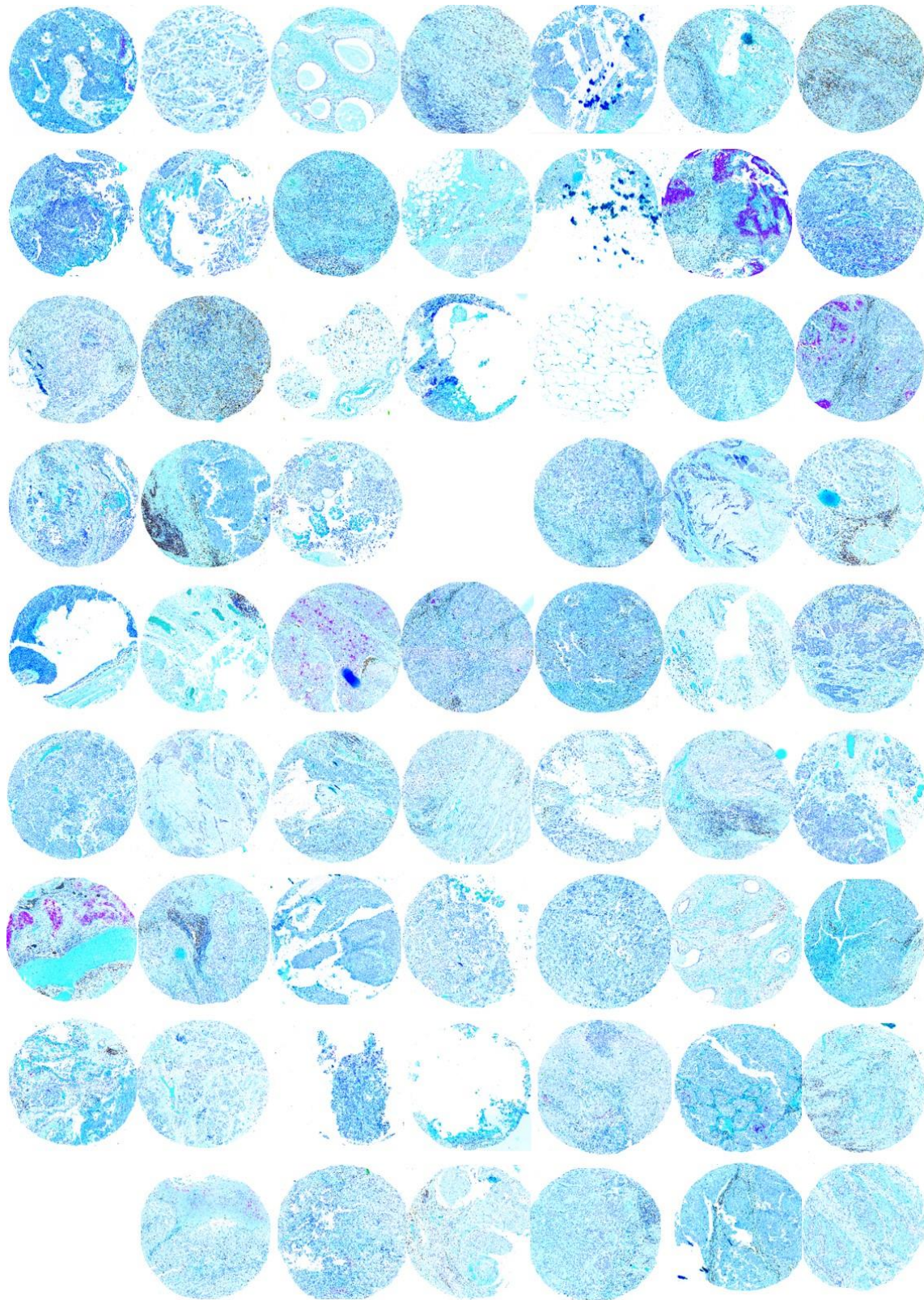
Figure 3.1: Distribution of TMA1 images on the slide.

Multiplexed immunofluorescence TMA slides were scanned using the Vectra-Polaris Automated Quantitative Pathology Imaging System (Akoya Biosciences). In brief, a spectral library was created, capturing the

spectral peaks emitted by each fluorophore from individually stained slides. This spectral library facilitated the spectral unmixing of the images, enabling color-based identification of the markers of interest. Each TMA core image was spectrally unmixed and exported as a TIF image (2656 × 2659 × 8 pixels) using Akoya Biosciences' Inform software.

Each image is composed of seven channels corresponding to seven fluorochromes labelling tumor cells and different immune cell populations. In this manner, the simultaneous detection of the relevant biomarkers for this study, DAPI, CD3 (T cells), CD8(CTLs), BATF3 and GranB (granzyme B), has been achieved.

Despite the original channel being composed of seven markers, during the spectral unmixing process, an estimation of the autofluorescence channel of the tissue is conducted. This channel is utilized to subtract the autofluorescence component from the rest of the channels, enabling a later work with a purer signal.

The eight channels are arranged as follows: DAPI, MelanA, CD3, CD8, BATF3, GranB, NK, and Autofluorescence. Each captured channel corresponds to a specific band in the visible spectrum, adjusted to the emission range of each fluorochrome. To visualize these multispectral images, an arbitrary colour is assigned to each channel for visualization purposes: blue, cyan, orange, red, green, yellow, magenta, and black, respectively, creating pseudo-colour image. An illustration of one of the images presenting all the channels is represented in Figure 3.2.
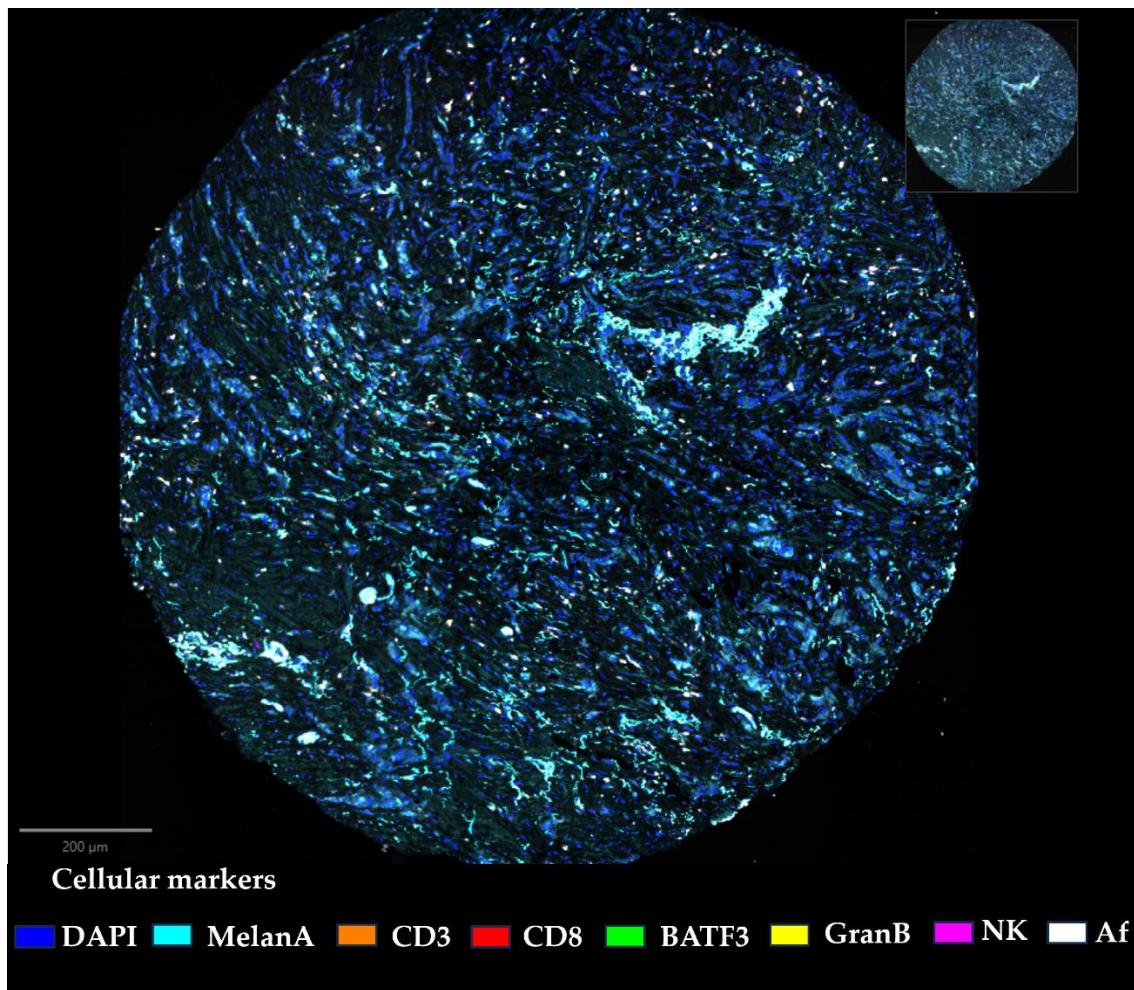
Figure 3.2: TMA2_1_5_C image visualization with the eight channels.

However, the MelanA (second channel), GranB (fifth channel) and NK (seventh channel) were excluded from the analysis due to their lack of relevance in the current study. Their presence is a result of previous studies which incorporated these biomarkers. The laborious process of staining and preparing the images obligates the reuse of the same images for multiple purposes, requiring the omission of channels with no significance in the current research work carried out.

To mitigate noise and artifacts of all images, a normalization process was implemented. This involved the application of various techniques, including a light Gaussian blur to reduce the noise level smoothing the image, gamma correction to adapt the saturation of each pixel, rolling ball to subtract background and brightness adjustment by saturating a percentage of pixels. Specific parameter values were adjusted for each channel of the images, see Table 3.1

Channel	Gaussian blur (sigma)	Gamma correction	Rollling ball	Pixels to saturate (%)
DAPI	0	1	0	0.1
MelanA	0	1	0	0.1
CD3	0.8	1.5	50	5
CD8	0.8	1.5	50	10
BATF3	1.5	3	50	50
GranB	0.8	1.5	50	5
NK	1.5	3	50	5
Autofluorescence	0	1	0	0.1

Table 3.1: Channel normalization parameters.

Figure 3.3 offers a visual representation of the advantages gained from the normalization procedure, particularly evident in the BATF3 (green colour) and CD8 (red colour) channels.
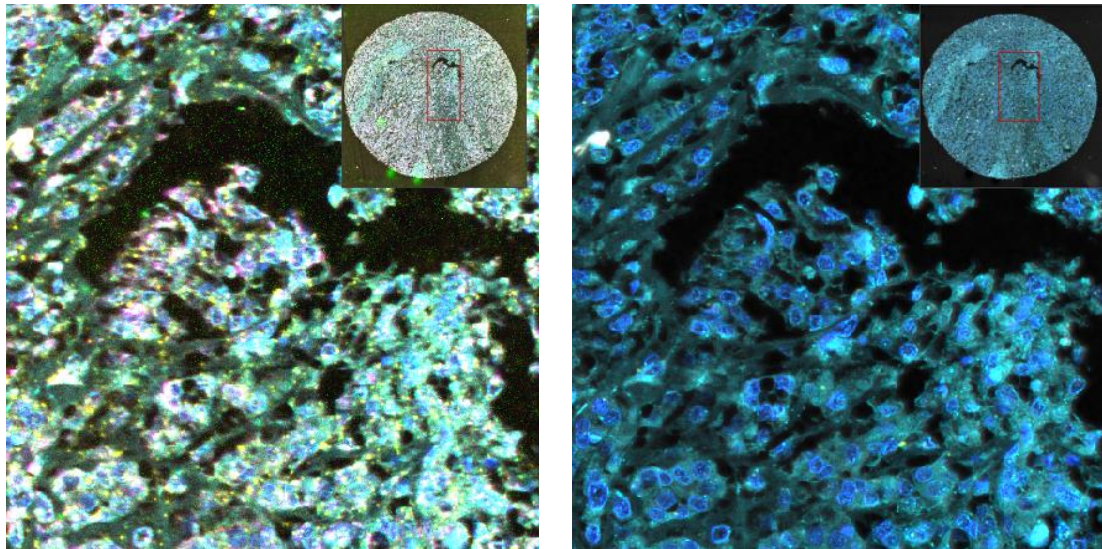
Figure 3.3: Comparison of images for TMA1_1_3_F without normalization (left) and with normalization (right).

3. 2. Datasets

Tumor samples were assessed from 61 urothelial carcinoma patients treated with Atezolizumab, from which 17 (27,87%) presented partial or complete response. The patient outcome is briefly represented in Table 3.2, specifying the response to immune-based treatment based on the gender and age of the patient at the time of biopsy.

	Complete Response	Partial Response	Not Evaluable	Progression Disease	Stable Disease
Sex (male/female)	7/3	6/1	4/0	15/6	17/2
Median age at time of biopsy (range)	66 (51-74)	76 (62-81)	75 (67-80)	65 (54-80)	69 (58-80)

Table 3.2: Patient summary demographics.

Based on the five types of responses to the immunotherapy treatment presented below, two main groups are defined: responders and non-responders. Complete and partial responses fall under the category of responders, while the group of non-responders is exclusively constituted by the progression disease response. The not evaluable group, as the name indicates, cannot be categorized as a positive or negative response. The same holds true for the stable disease response, given the uncertainty about the disease's future progression, as it may either improve or worsen over time after the biopsy acquisition.

Clínica Universidad de Navarra believes that there may be a correlation between the distance that separates BATF3+ and CD8+ cells and the response to Atezolizumab immunotherapy treatment in urothelial cancer patients. Due to the potential improvements in anticancer treatments that could be achieved and its clinical relevance, this Final Master Project will attempt to provide an answer to this hypothesis.

Chapter 4. Methods

4. 1. Nuclei Segmentation

Despite decades of progress, cell and nuclei segmentation for microscopic image analysis remains challenging [33]. Assigning a cell to every no-background object in the image is necessary for numerous biological tasks, including one of the project's objectives: the classification of cell phenotypes.

The aim of this section is to provide a detailed explanation of the nuclei segmentation conducted in this project. The chosen method for nuclei detection was StarDist, an algorithm based on machine learning approaches, particularly the U-Net model. StarDist identifies cell nuclei through star-convex polygons [36]. This method was selected because its precise accuracy in detecting nuclei within images featuring crowded cells or irregular morphologies [32].

Additionally, the procedure for evaluating the segmentation performed will be explained. To begin, ground truth masks were necessary. These masks were generated through the manual annotation of every cell nucleus in the image.

4.1.1. Ground Truth Images Generation

Due to the laborious nature of this task, only eight images were annotated, as shown in Figure 4.1, each containing an average of over eight thousand nuclei. For instance, the annotation process for the image TMA2_1_1_E took nine and a half hours.

| TMA1_1_1_A | TMA1_1_1_B | TMA1_1_1_D | TMA1_1_6_E |
| TMA1_1_7_C | TMA1_1_8_F | TMA2_1_1_B | TMA2_1_1_E |
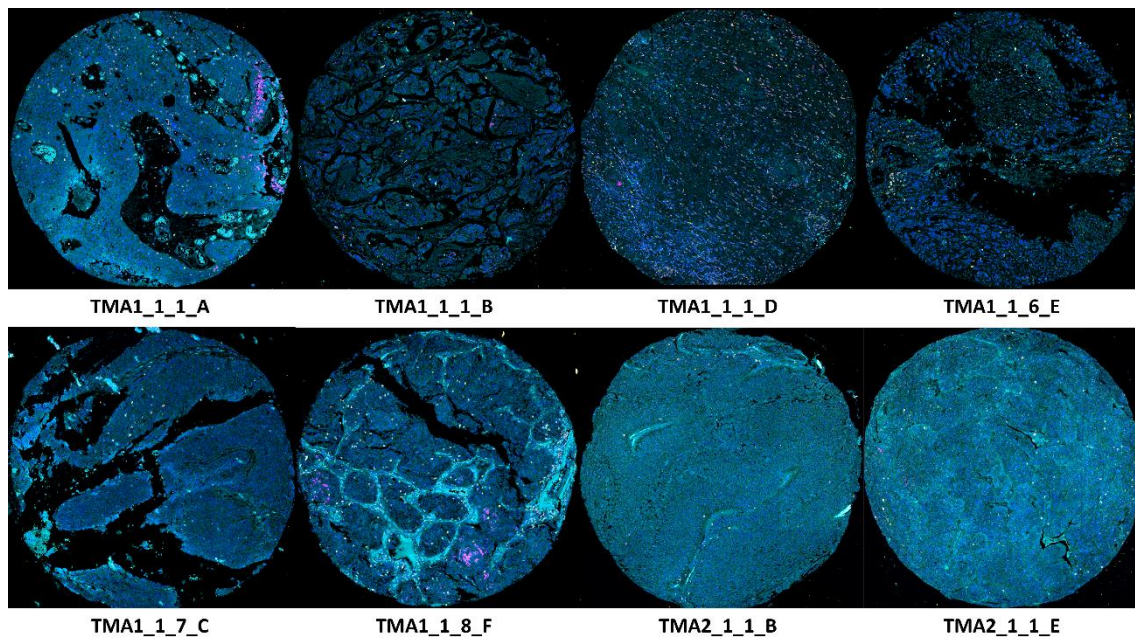
Figure 4.1: Images chosen for annotation.

All the annotation process has been supervised by a pathologist of Clínica Universidad de Navarra with a wide experience in multiplex imaging. He has provided histological indications and resolved specific doubts in complex annotation areas.

The ground truth images were generated using the bioimage analysis software *QuPath* [37] and the image processing package Fiji [38]. Briefly, the initial step involved the preliminary detection of nuclei contours in a selected image using the Cell Detection tool in *QuPath*. Each of these detections was then converted into an annotation using the *QuPath* scripting tool, facilitating the subsequent export to *Fiji*. Afterwards, each annotation generated by the Cell Detection tool underwent careful correction to ensure that the detected boundary of each nucleus accurately corresponded to its real shape, while also verifying that all nuclei were annotated, see Figure 4.2. Once all the annotations were ready, they could be exported to *Fiji* included in a .zip folder.
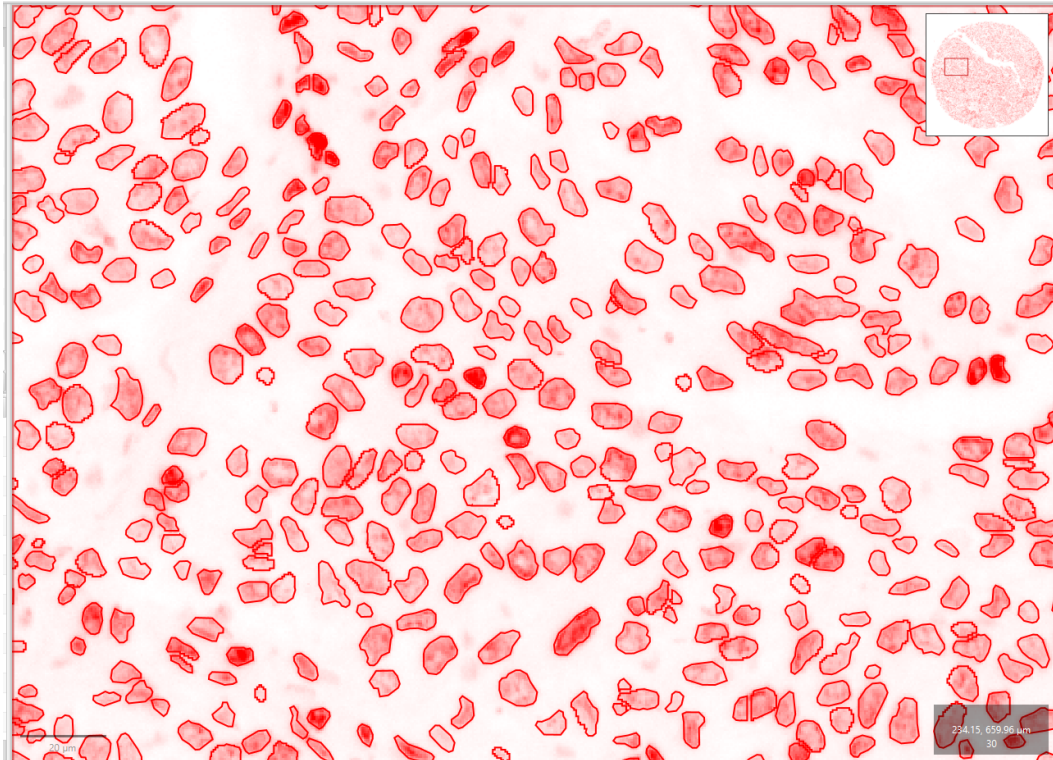
Figure 4.2: Nuclei annotations showed from QuPath. The DAPI signal is represented in red, with its boundaries outlined in the same colour. The background has been inverted to improve visualization.

In *Fiji*, all the annotations were treated as regions of interest, and each of them was filled with a unique grey level, thus creating an image with distinct nuclei; the ground truth image, as illustrated in Figure 4.3.
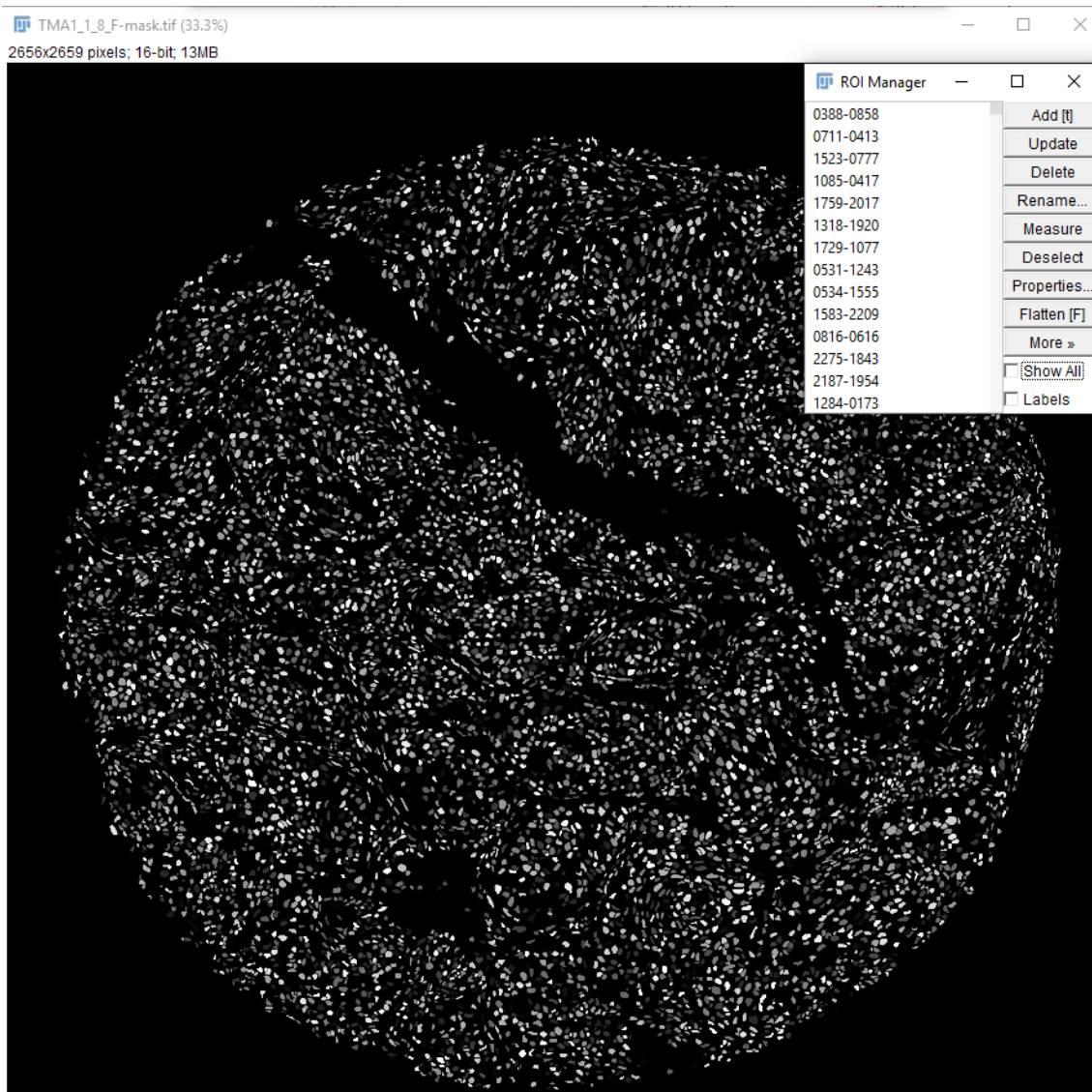
Figure 4.3: Ground truth image: TMA1_1_8_F. In the upper right corner, all regions of interest (ROIs) opened in the Fiji's ROI Manager.

4.1.2. Nuclear Segmentation Evaluation Methodology

As mentioned earlier, eight ground truth images are available to evaluate the trained model. Due to the limited dataset, the Leave One Out strategy was employed to estimate the accuracy of the model. In this procedure, the dataset is split into two subsets: training and test images. In each of the eight iterations, one image serves as the test set, while the remaining images constitute the training set, with 15% of them used as validation images. A new model is created in each iteration to prevent the algorithm from learning from previous training. The model is trained eight

times in total, allowing for evaluation on each image separately and providing an indication of the precision of the model.

During the leave one out strategy, input parameters such as the number of epochs, patch sizes and number of rays of the model must be set with the aim of achieving an optimal nuclear segmentation.

The number of epochs significantly affects the time required for the training process. Increasing the number of epochs leads to a longer computational time, but it also typically results in lower losses for both the validation and training sets. In the results chapter, it can be observed from the loss graphs a point at which the losses stabilize despite further increases in the number of epochs. At this point, a suitable fixed number of epochs can be selected for optimal training.

The selection of patch size between the values of (64,64), (128,128), and (256,256) will also be addressed. A larger patch size may provide more contextual information but can also increase computational load and memory consumption. Conversely, a smaller patch size may be more efficient but could sacrifice small details.

The effect of the number of rays on the precision of the model in segmenting nuclei will also be investigated. StarDist allows adjusting the number of rays to create the polygons that define the cellular nuclei. The number of rays determines the quantity of rays emanating from the centroid of each nucleus and extending to the nucleus's edge. Subsequently, these rays are connected to delineate the contour. Therefore, the analysis will explore all powers of two up to 128 to determine the optimal segmentation accuracy considering the computational efficiency.

Additionally, the nuclear segmentation model will be compared to different state of the art models, including DeepCell [39] and CellPose [40]. On the one hand, DeepCell offers two models for nuclei segmentation: the nuclear model and the Mesmer model. While Mesmer model enables an entire cell segmentation simultaneously, specifically its nuclear tool will be utilized for comparison in this context. On the other hand, CellPose exclusively provides one method with nuclear segmentation capabilities.

4. 2. Whole Cell Segmentation

Entire cell segmentation is one of the most challenging procedures in biomedical images analysis [41]. However, the lack of public datasets with annotations of entire cells and the difficulty in developing pre-trained models make progress in whole-cell segmentation challenging [42].

The deep learning library selected for whole cell segmentation was DeepCell, specifically its deep learning algorithm Mesmer, which was trained with the TissueNet dataset that contains more than one million annotated cells. This method has been chosen because of its robustness in outlining the cytoplasm of single cells [32].

The pipeline followed by Mesmer involves defining nuclei using a nuclear input image and outlining the whole cell using a cytoplasm or membrane input image. Once the boundaries and centroids of the entire cells and nuclei are defined, a watershed algorithm is applied to expand the objects detected and create the final labels.

Indeed, the parameters input and maxima threshold have been adjusted to achieve better results in this cohort of patients, visually observing how they modify the entire cell segmentation. The input threshold determines how conservative the model is in distinguishing between cells and background, while the maxima threshold parameter determines what the model defines as a distinct cell.

As the input nuclear image for the algorithm, the DAPI channel has been used due to its direct correspondence with the nuclei shape. However, despite advancements, there is currently no optimal biomarker to determine the cellular membrane or cytoplasm of every cell [43]. The variation in cytoplasmic or membrane fluorescent marker expression, which are GranB and CD8 in our cohort of patients, and the fact that many cells do not express either of these markers, make it challenging to define a complete cytoplasmic image. For this reason, it has been decided to create a cytoplasmic DeepCell input image as the sum of the CD8 and GranB channels, as shown in Figure 4.4. CD8 functions as both a membrane and cytoplasmic marker, whereas GranB acts as a membrane marker. This approach allows for a more comprehensive understanding of the shape of

each cell. Both input channels and its overlay, with nuclei and cytoplasm signals in green and blue respectively, can be seen in Figure 4.5.
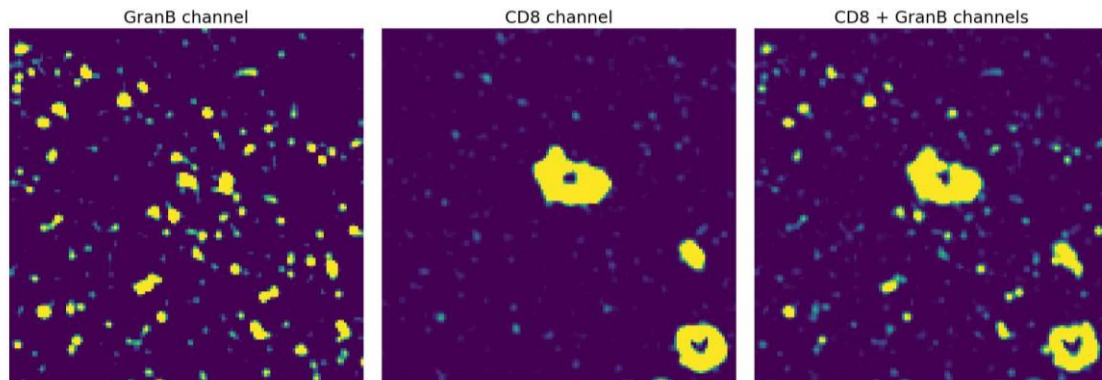
Figure 4.4: Combined visualization of GranB, CD8 and the sum of both channels.
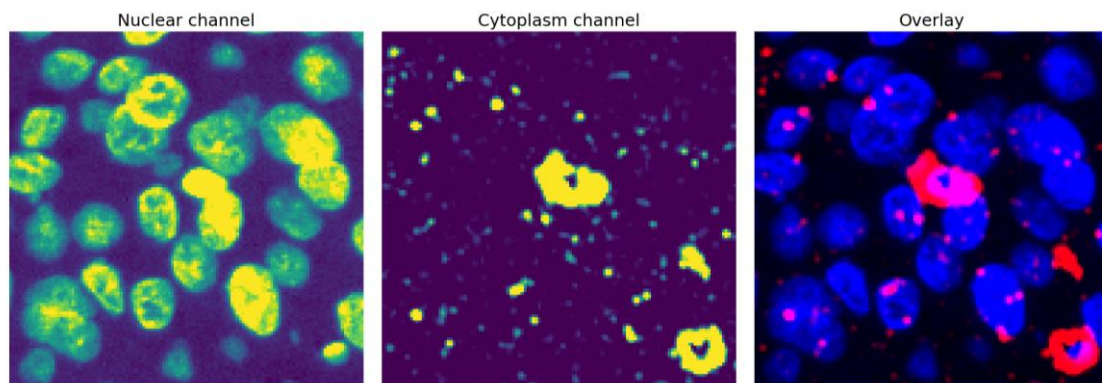
Figure 4.5: DeepCell Mesmer input images. In the overlay image the nuclei signal is represented in blue, while the cytoplasm signal is represented in red.

4. 3. Classification of Cell Phenotypes

Once all the cells and their nuclei are properly segmented, it is possible to characterize each cell individually. The aim of this study is to study the correlation of the patient's response to treatment with the abundance of CD8+ and Batf3+ cells, as well as with the spatial interactions between both cell phenotypes. To achieve it, we first need to classify cells into three groups: CD8+ cells, BATF3+ cells, and other cells. The code developed for this task has been entirely written in Fiji's macro language.

4.3.1.Areas to Exclude from the Analysis

Neither whole-cell segmentation nor nuclear segmentation consider areas of the images that should be excluded from the analysis, such as regions with broken tissue or image artifacts. Therefore, regions to be eliminated from the analysis have been manually annotated, after a visual inspection of all the images. This process has been supervised by a pathologist with a vast experience in the field. An example of annotations for the excluded areas within the TMA1_1_2_A can be seen in Figure 4.6. In the image on the right, a zoomed snapshot of one of the regions to be excluded reveals nuclei in blue, whose shapes are distorted possibly due to a scratch in the tissue sample. Besides, areas with presence of CD8 marker, colorized in red, which do not correspond to positive cells for that marker are also observed. These are erythrocytes, highly auto-fluorescent cells that are present when bleeding has occurred in part of the tissue. This approach prevents the classifier from detecting false positives due to undesired artifacts in the images under analysis.
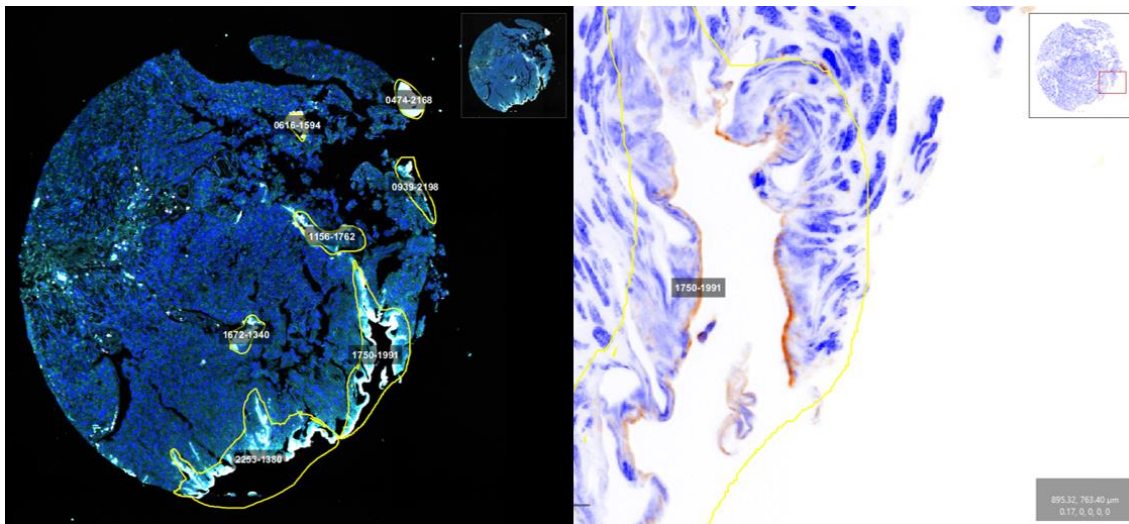
Figure 4.6: Annotations of zones to be excluded in image TMA1_1_2_A. The left image displays the original fluorescence image, while the right image presents the same image with its background inverted, enhancing visualization of details.

4.3.2. Tissue Area Localization

For a later assessment of the classification results, it will be essential to determine the total area of the analysed tissue due to its variability in each image. This way, realistic comparisons can be carried out by normalizing the quantified cell populations by the total tissue area.

Tissue areas have been calculated excluding from the analysis the regions detailed previously. The segmentation procedure followed to obtain a total tissue mask has been as follows. First, a Gaussian Blur filter with a sigma of 4 to reduce the noise level was applied to all the images, which were previously transformed from 32 bits to 8 bits for ease of the analysis. Then, a threshold value was selected, thus eliminating pixels with intensities below 15. Despite automatic thresholding was considered, there were several cases of failure. For this reason, a fixed value has been experimentally chosen, by observing instances when the algorithm failed to detect tissue due to the use of excessively high thresholds. Given the prior normalization of the images, this approach has been shown to function effectively. Finally, particles with an area below 5000 pixels were removed. It has been determined that tissue zones with an area less than 5000 pixels, equivalent to 1250 µm² of tissue area, are categorized as non-analysable because they could correspond to disregarded tissue and may result in artifacts. This value was also experimentally set to ensure that

tissue areas are not left undetected while also avoiding the detection of non-tissue area. Following this image processing workflow, the remaining available tissue area was measured. Examples of the whole tissue segmentation are illustrated in Figure 4.7.
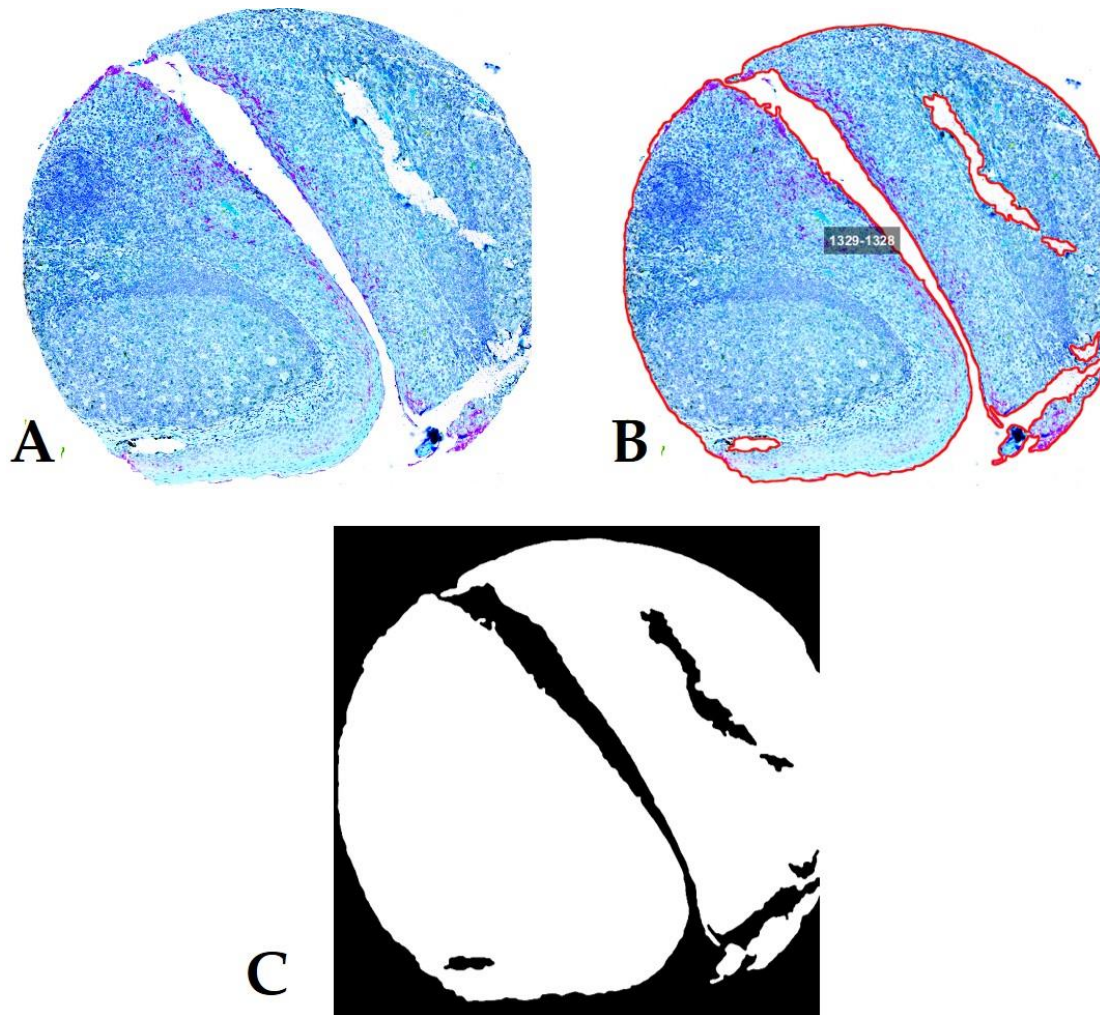
Figure 4.7: Detected tissue of TMA1_1_11_F image. A. Original image: background has been inverted to improve visualization. B. Detected tissue boundaries outlined in red. C. Whole tissue mask.

4.3.3. Cell Classification Strategy

Once the tissue area is identified, the nuclei and cytoplasm of each cell, segmented as explained in Nuclei Segmentation and Whole Cell Segmentation, are analysed within it to detect BATF3+ and CD8+ cells, respectively. One region of interest (ROI) is created for each nucleus during the BATF classification, and for each cytoplasm in the case of CD8, due to the specific locations of these proteins within the cells. The method for

designing these ROIs involves analysing the intensity of pixels in the image. Both segmentations assign a different level of intensity to each label. Therefore, pixels with a same intensity level correspond to a single ROI, where all the measurements are conducted.

To evaluate the cell phenotype classification, eleven ground truth images were generated by annotating CD8+ cells to evaluate CD8+ classification, as CD8 is a cytoplasmatic marker. Additionally, fifteen ground truth images were created by annotating nuclei with positive BATF3 expression to assess BATF3 classification, since BATF3 corresponds to a nuclei marker. All assesments of the classifier have been acquired using an IoU threshold of 30%, as it evaluates the precision of the classifier to detect positive cells for each population, rather than precisely outlining their boundaries.

The criteria selected to consider a cell positive for CD8 or BATF3 biomarkers consist of pixels in the corresponding channel presenting an intensity above a threshold and a minimum positive area percentage within the nucleus, in the case of BATF3, or the membrane and cytoplasm, in the case of CD8.Thus, an intensity threshold has been applied to the CD8 and BATF3 channels, considering that pixels with an intensity value below the chosen threshold do not correspond to actual CD8 or BATF3 expressions but to background noise. Due to the significant differences between the expression of both fluorochromes, two different threshold values have been determined, one for each marker.

When masks for the channels of interest were created, the area of non-zero pixels within each of ROI, representing either the nuclei or cytoplasm shapes, was calculated. If the ratio of pixels to CD8 and BATF3, relative to the cytoplasm and nucleus areas, respectively, exceeds the specified minimum percentage, the cell will be categorized as positive for the respective marker. As in the threshold case, due to the differences between both cell phenotypes, two different minimum area percentage values have been determined, one for each marker.

Optimal parameters have been calculated separately, and the methods used will be detailed in the following sections: Optimal Threshold Selection for Cell Phenotype Classification, Optimal Minimum Area Percentage Selection for Cell Phenotype Classification.

It must be considered that the DeepCell Mesmer model occasionally encounters challenges in detecting cells without nuclei due to the presence of signal in the cytoplasmic input image where DAPI does not show any nuclear signal. This failure can be seen in Figure 4.8, which illustrates how the model detects a cell despite the absence of any nuclear signal. To address this issue, an image post-processing step has been implemented; if the CD8+ cells detected by the classifiers lack a nucleus, the cell will be excluded from the analysis.
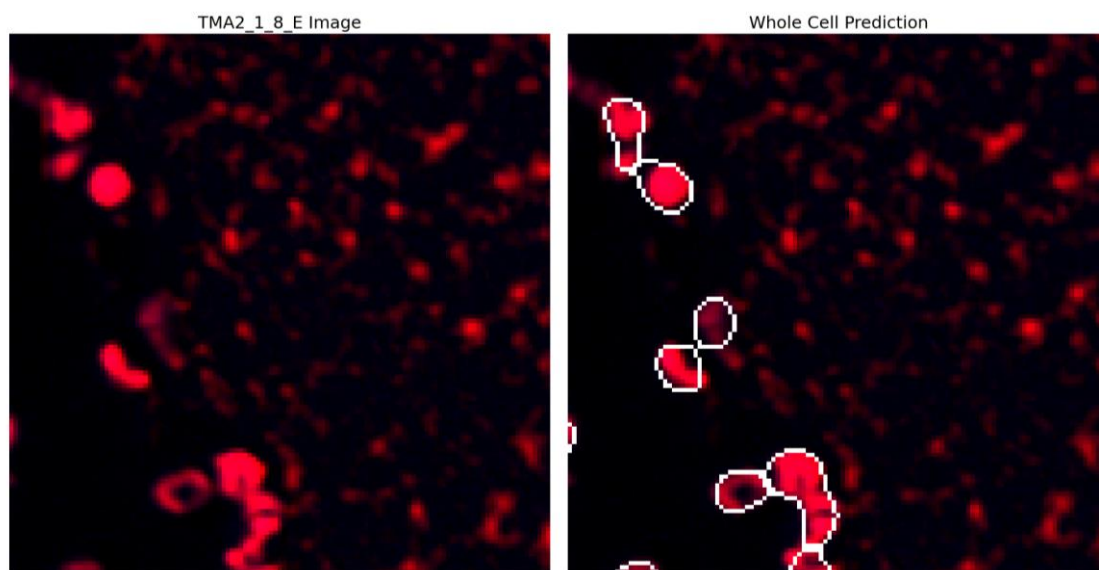
Figure 4.8: DeepCell failure in detecting cells without a nucleus. Cytoplasmic signal is represented in red, while the entire cell boundary is outlined in white. The nuclear signal would be presented in blue color.

4.3.4. Optimal Threshold Selection for Cell Phenotype Classification

Firstly, the optimal threshold value for phenotype classification has been sought. Given the variations in intensity levels observed in the CD8+ and BATF3+ cell phenotypes, the methods detailed in this section will be applied separately to both phenotypes.

The Otsu automatic thresholding method performance will be firstly analyzed for this purpose [44]. Furthermore, various correction factors will be applied to the Otsu value to assess the results. This approach allows to explore how adjustments to the automatic threshold impact the quality of

phenotype classification results, facilitating a more refined optimization of the segmentation and classification process.

Since the histograms of both studied cell phenotypes are unimodal, the automatic threshold method Triangle has been implemented, which is reported to perform better in these scenarios [45].

Both automatic approaches and their respective correction factors will be compared in terms of cell phenotype classification ability.

4.3.5. Optimal Minimum Area Percentage Selection for Cell Phenotype Classification

The minimum area percentage to classify CD8 T cells and BATF3 as positive has been optimized. To achieve this, evaluation metrics have been analysed for a range of minimum area percentage values. As previously mentioned, these metrics are obtained by assessing the matching between the ground truth images created with only CD8+ cells and BATF3+ cells and the predictions made by the classifier.

F1 score will be examined to select the optimal parameters because it represents the harmonic mean of the recall and precision metrics. This approach is considered a good compromise, since it allows to establish a balance between not leaving positives unidentified, as indicated by precision metric, while also avoiding the detection of more positives than the actual number present, as indicated by recall metric.

4. 4. Study of Spatial Interactions Between Cell Phenotypes

Based on the cellular classification detailed in the previous section, the CD8+ and BATF3+ cell populations have been detected in each image of the dataset. Their abundance, normalized per square micron of tissue area, as well as their spatial localization, have been identified. For the evaluation of their spatial interactions, the distance between each CD8+ cell and its nearest BATF3+ cell within the tumor microenvironment of each patient have been measured. To calculate these variables, a simple code written in Python programming language has been used.

With the classification information and the knowledge of the clinical response of 61 urothelial carcinoma patients to Atezolizumab immunotherapy, correlations between these quantified image parameters and the clinical response to treatment can be examined for this patient cohort.

On the one hand, the average density of CD8 and BATF3 cell populations for the two types of clinical response have been calculated. On the other hand, it has been explored whether the measured distances between CD8 and BATF3 cell populations have an effect in the clinical response to Atezolizumab immunotherapy. In both studies, a t-test has been used to assess the statistical significance of differences between the two groups.

Chapter 5. Results

5. 1. Analysis of TMA heterogeneity

Firstly, the mean intensity values of the images for each TMA have been compared in Figure 5.1. This analysis aids in determining the possibility of using the same segmentation tools for both TMAs. As illustrated, the variations in grey levels between the TMAs are negligible for every marker. Therefore, there is no need to treat them differently during image processing.
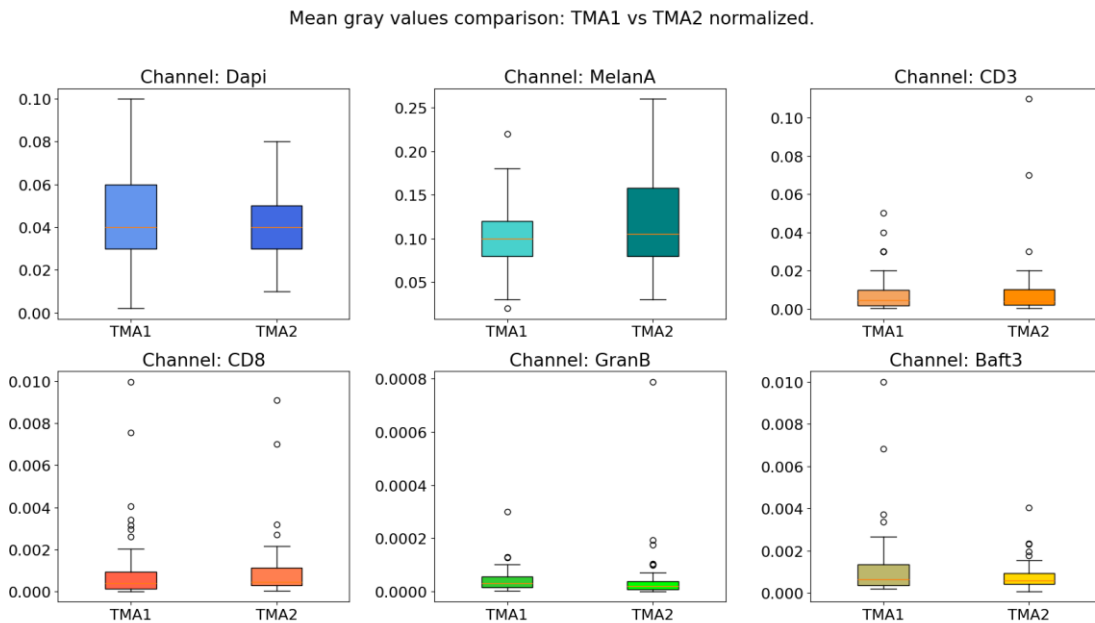
Figure 5.1: Comparison of mean grey values between the two normalized Tissue

5. 2. Nuclear Segmentation

The Leave One Out procedure enables the evaluation of nuclear segmentation in the most robust possible manner, given the limited availability of ground truth images. Using this method, the input parameters number of epochs, patch size and number of rays of the StarDist model are adjusted to this cohort.

In this section, the performance of the trained model will be compared to that of the pretrained StarDist model, the pretrained DeepCell Mesmer model, and the pretrained CellPose model, using various IoU thresholds and evaluation metrics.

5.2.1. Selection of Number of Epochs for the StarDist Model

To mitigate the risk of overfitting [46] and to reduce the computational time, the number of training epochs have been limited to 200. This choice has been made after observing the behaviour of each test image, whereby the number of epochs has been selected when the losses stabilized.

An example illustrating the behaviour of the losses for both the validation and training datasets out of the number of epochs, represented for the test image TMA1_1_1_A, noting that the remaining images are similar, is represented in Figure 5.2. As mentioned earlier, the effect of this parameter has been analysed during the leave one out procedure.
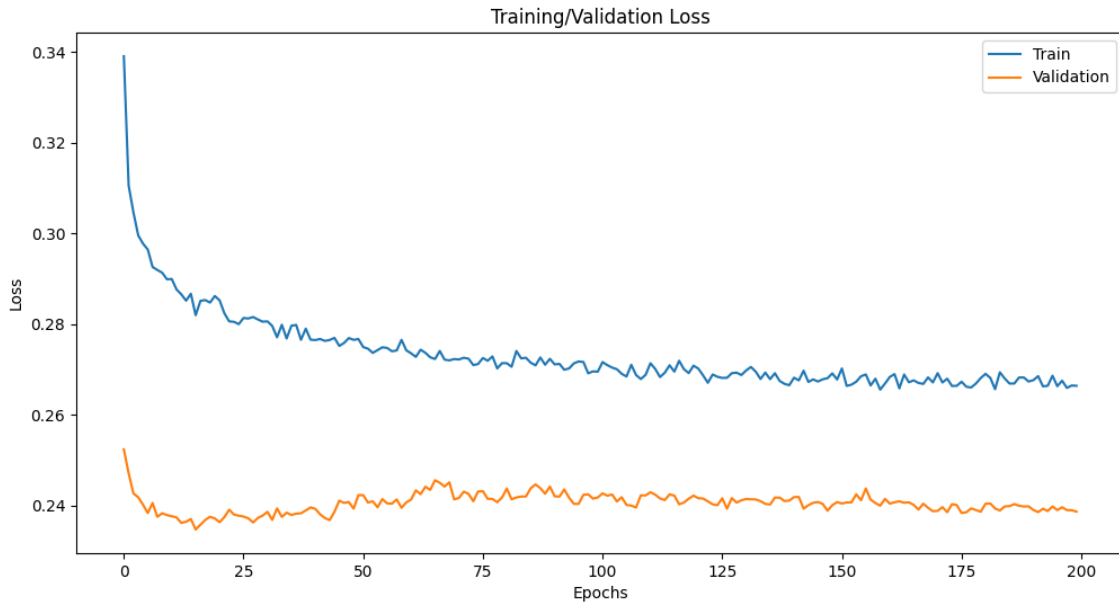
Figure 5.2: Losses vs number of training epochs for training and validation sets for the test image TMA1_1_1_A.

5.2.2. Selection of the patch size for the StarDist model

The adjustment of the training model's patch size was also considered. Patch size refers to the smaller sections in which an image is divided during the training process. To assess the impact, training and validation losses were compared across different patch sizes: (64,64), (128,128) and (256,256) for the image TMA1_1_1_A. The results are illustrated in Figure 5.3. A noticeable difference was observed between the losses for the first two patch sizes. However, no significant disparities in losses were found between (128,128) and (256,256). Nevertheless, there was a considerable increase in computational time.

A

B

C
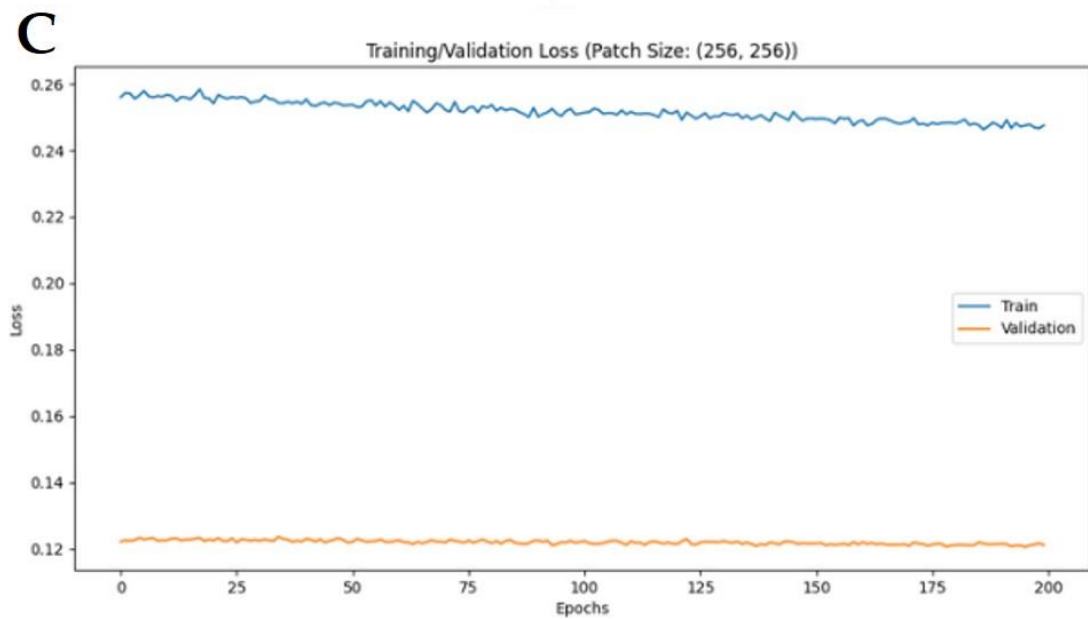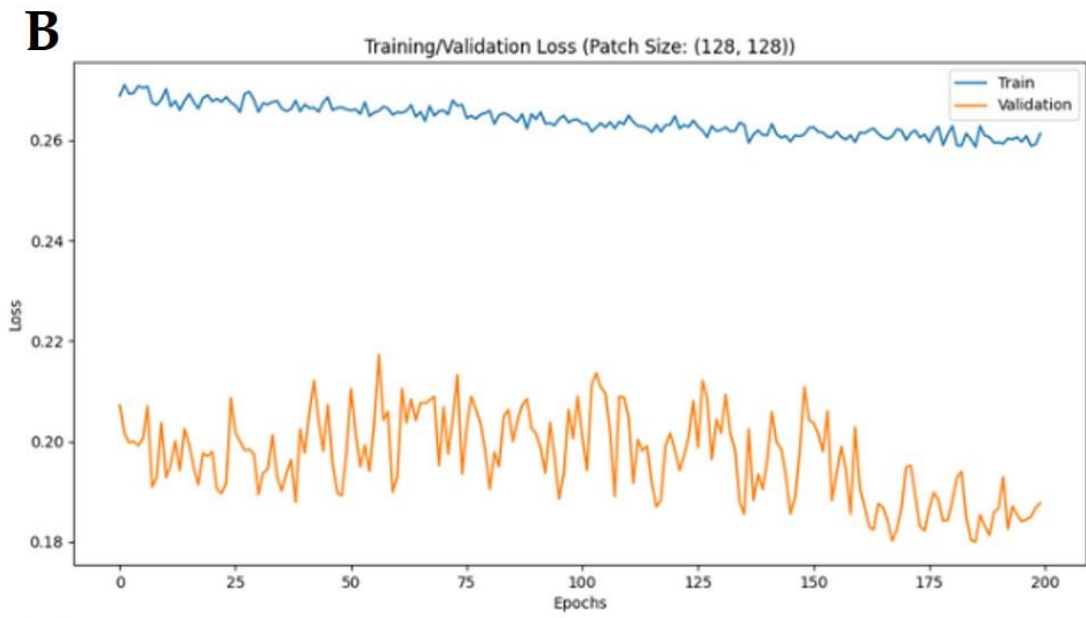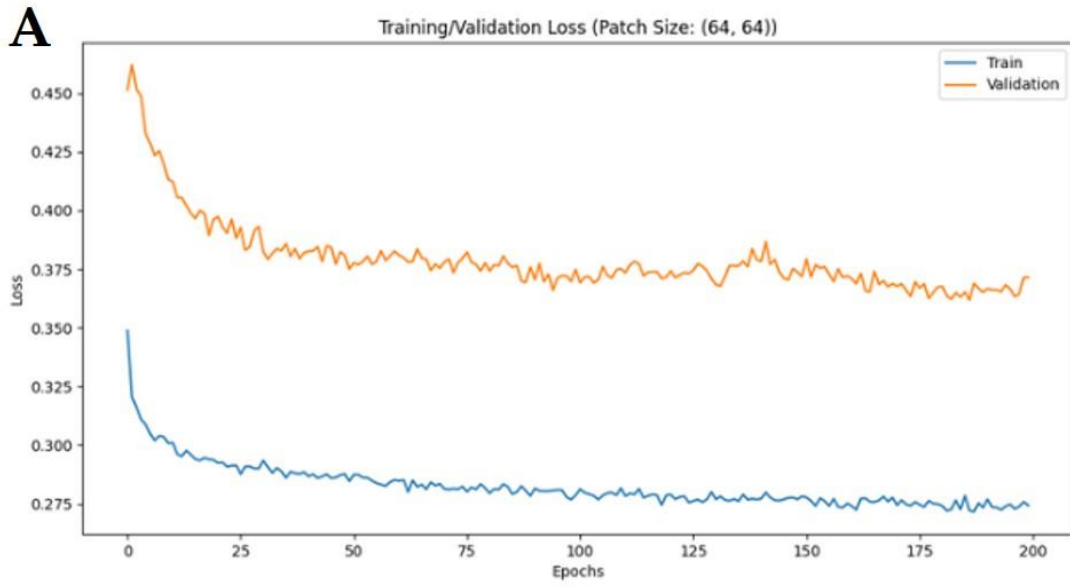
Figure 5.3: Losses out of number of epochs for a fixed patch size of: A. (64,64) B. (128,128) C. (256,256)

For the same image, the computational times for the various patch sizes are represented in Table 5.1. Balancing time efficiency and precision, the (128,128) patch size was chosen for training the model.

Patch size	Comptutational time
(64,64)	4h 10min
(128,128)	6h 20min
(256,256)	9h 30 min

Table 5.1: Computational times represented for each patch size.

5.2.3. Selection of the number of rays for the StarDist model

StarDist outlines the nuclei boundary using a star-convex polygon with a number of rays selected by the programmer. To determine this parameter, its impact on prediction accuracy was measured, as shown in Figure 5.4. The visual effect is further represented in Figure 5.5. A total of 32 rays were selected as there were no significant precision differences with the next power of two, 64. Moreover, increasing the number of rays to 64 led to a noticeable computational time increment. The accuracy of the ground truth reconstruction, based on the number of rays for star-convex polygons in StarDist's trained model, is obtained through the evaluation of the eight images with ground truth. The *Matching_dataset* function of StarDist allows for the simultaneous evaluation of the segmentation of a group of images, returning the mean value of various evaluation metrics.
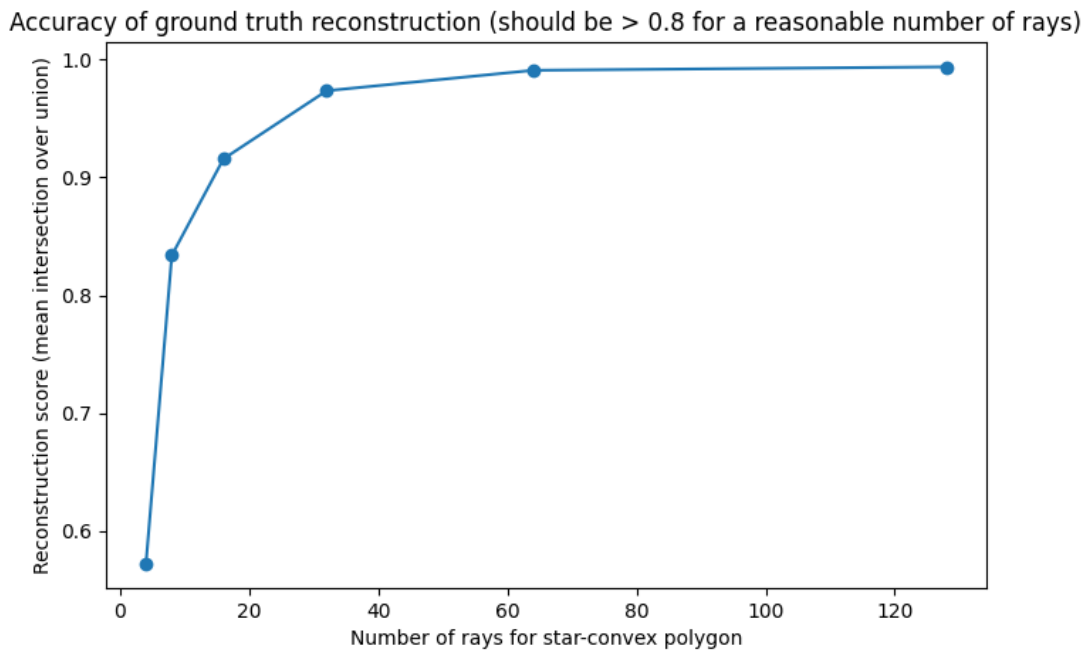
Figure 5.4: Accuracy of ground truth reconstruction based on number of rays for star-convex polygon in StarDist's trained model. Obtained from the evaluation of the eight images with ground truth.
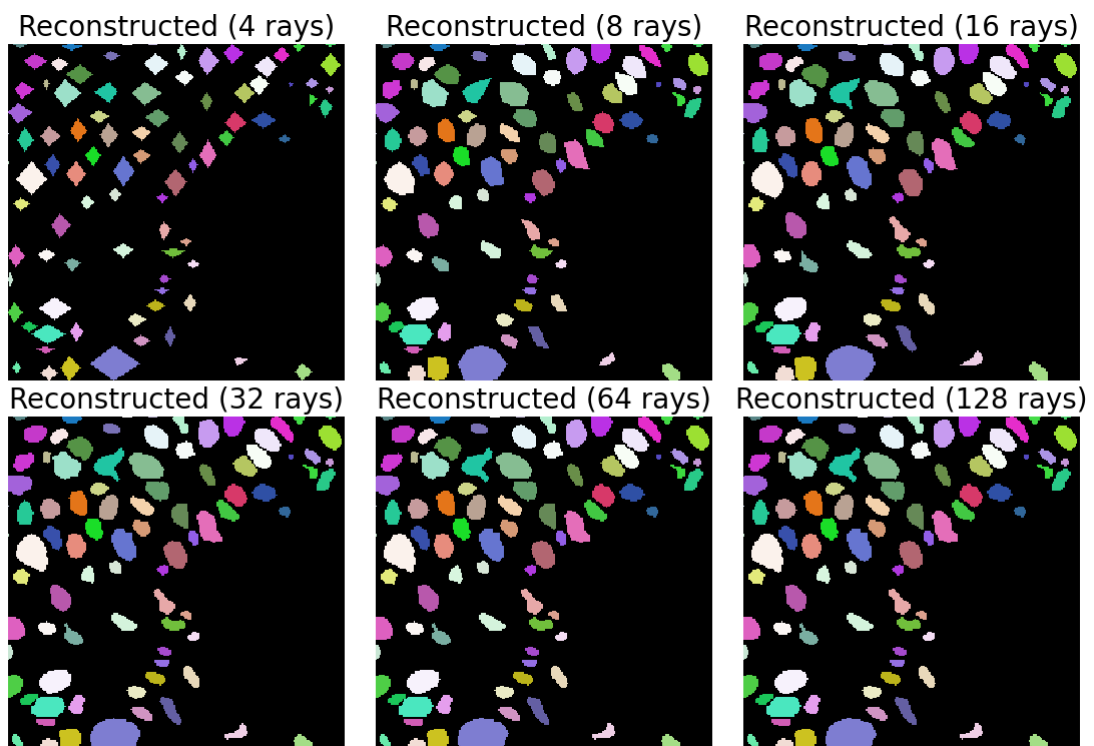
Figure 5.5: Nuclear segmentation performance based on the number of rays for star-convex polygon in StarDist model, showing predicted labels for the image TMA1_1_1_A.

5.2.4. Nuclei Segmentation Evaluation

The IoU threshold is used to assess the nuclei detection by evaluating the overlap between the detected object and the real object. An Intersection over Union of 50% indicates that the detected object corresponds to the real one at a 50% overlap. All the segmentation metrics between ground truth label masks and predictions were computed by using the *stardist.matching* submodule, a tool that works not exclusively with StarDist predictions.

The results obtained for Intersection over Union (IoU) evaluation metrics of 50% and 75% for both pretrained and trained by us models are respectively shown in Table 5.2.

Model	IoU Th	Precision	Recall	Accuracy	F1
Pretrained StarDist	50%	0.80±0.06	0.84±0.08	0.70±0.08	0.81±0.07
Pretrained StarDist	75%	0.41±0.10	0.42±0.12	0.25±0.09	0.41±0.11
Trained StarDist	50%	0.90±0.04	0.89±0.07	0.81±0.08	0.89±0.04
Trained StarDist	75%	0.75±0.09	0.75±0.13	0.61±0.12	0.75±0.10

Table 5.2: Evaluation results for pretrained and trained by us StarDist models at IoU threshold of 50% and 75%. The scores for each evaluation metrics represent the mean value, with the standard deviation indicated as ±.

Visually, the comparison between the trained and the pretrained model is represented in the Figure 5.6, where different IoU thresholds are compared for the two models, showcasing the accuracy, precision, F1 and recall metrics.
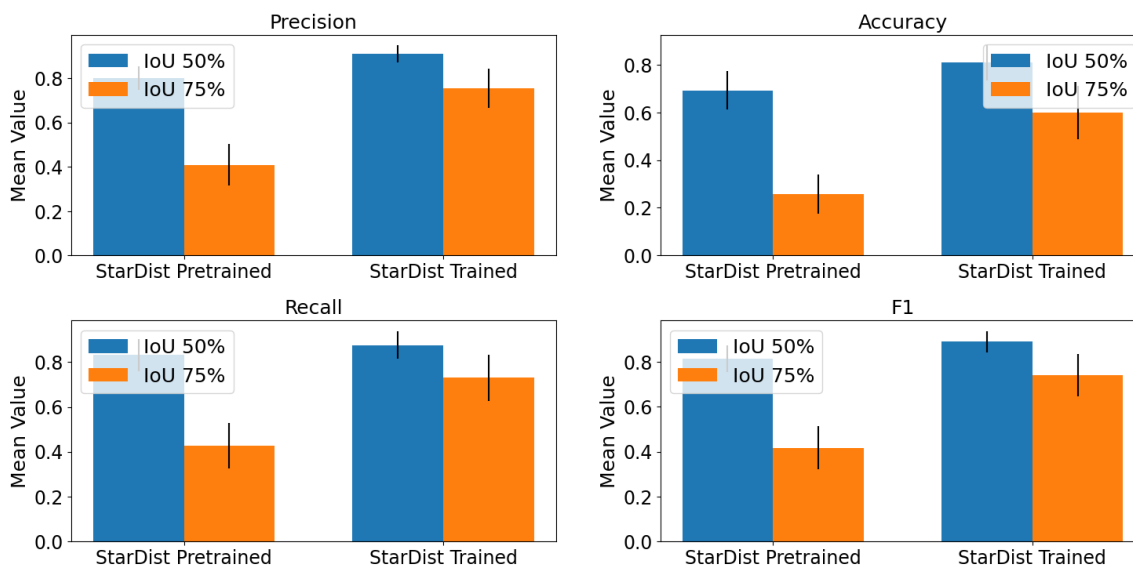
Figure 5.6: Comparison of mean values between the pretrained and trained models for StarDist at IoU thresholds of 50% and 75%.

The improvement achieved after training StarDist is notably superior, especially when comparing the models with an IoU of 75%, as quantitively illustrated in Table 5.3.

	IoU 50%	**IoU 75%**
Accuracy	13.79%	91.96%
Precision	17.17%	148.98%
F1 score	5.59%	77.45%
Recall	9.57%	84.48%

Table 5.3: Improvement in % of the trained StarDist model compared to the pretrained model, for evaluation metrics: accuracy, precision, F1, and recall, and IoU thresholds of 50% and 75%.

Moreover, Figure 5.7 visually represents this substantial enhancement. Based on this information, it can be concluded that training StarDist on our images improves the performance of the pre-trained model in terms of nucleus detection and segmentation. The significant enhancement, particularly notable at an IoU of 75%, signifies that the trained model not only detects nuclei more effectively, but, above all, exhibits an improvement in outlining the contours of the nuclei.
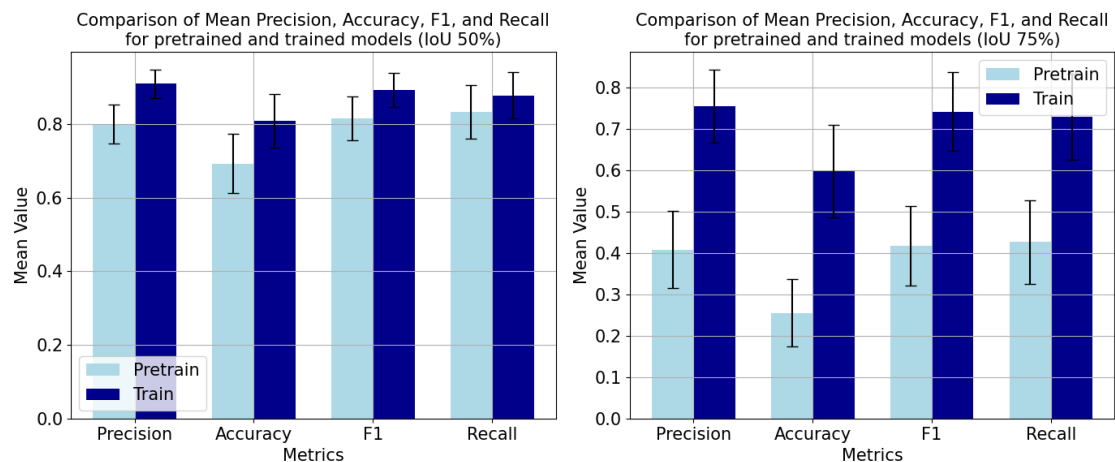
Figure 5.7: StarDist evaluation: comparison between precision, accuracy, F1 and recall for the trained and pretrained models at an IoU of 50% (left panel) and 75% (right panel).

Figure 5.8 illustrates an example of the enhanced predictive capabilities of the model trained by us in contrast to the pretrained one. The comparison has been made across the original image, the ground truth image and the predictions from both the pretrained and trained StarDist models. This comparison demonstrates how the pretrained model encounters challenges, such as erroneously splitting a nucleus into two distinct entities, marked with the blue circle. Additionally, the pretrained model fails by incorrectly joining two different nuclei, misidentifying them as a singular label, as can be seen within the grey circle.
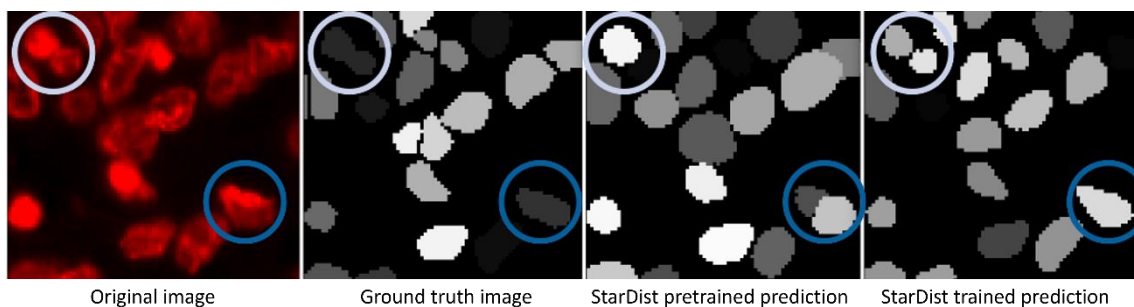
Figure 5.8: Visual comparison example of StarDist's predictions and the ground truth labels.

Figure 5.9 represents how the evaluation of the StarDist trained model performance depends on the IoU threshold for each of the test images used during training. The figures illustrate the model's decline in precision with an increase in the Intersection over Union value.
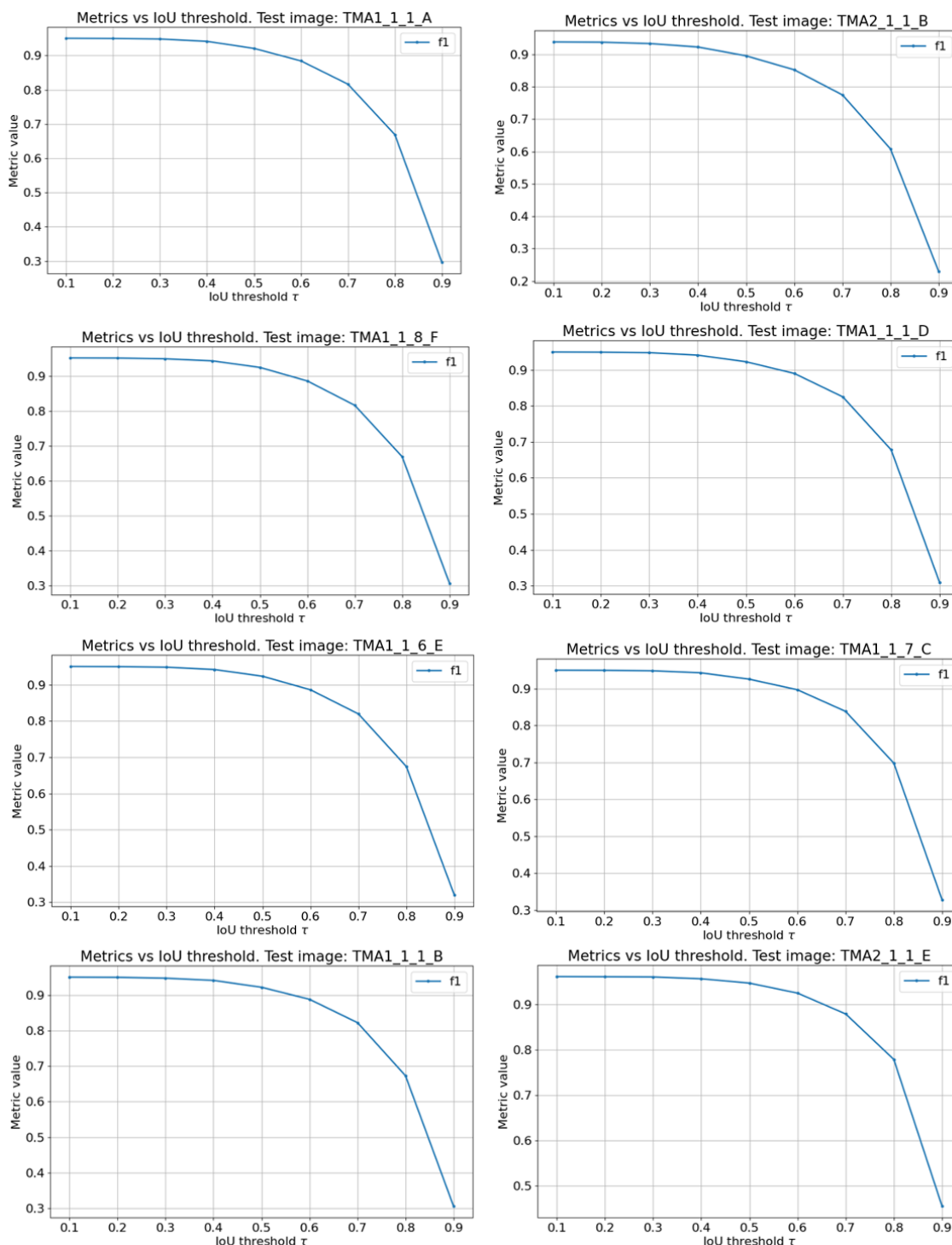
Figure 5.9: F1 value with respect to the IoU threshold for various test images.

5.2.5. Nuclear Segmentation Models Comparison

The Stardist nuclei segmentation model has been compared to other state-of-the-art cell segmentation models such as CellPose and DeepCell. As mentioned in the methods chapter, CellPose, like StarDist, offers a

single model for nuclear segmentation, while DeepCell provides two alternatives: its nuclear model and the Mesmer model, both represented in Figure 5.10. As evident from the results, the DeepCell's nuclear model tends to predict nuclei bigger than their actual size.
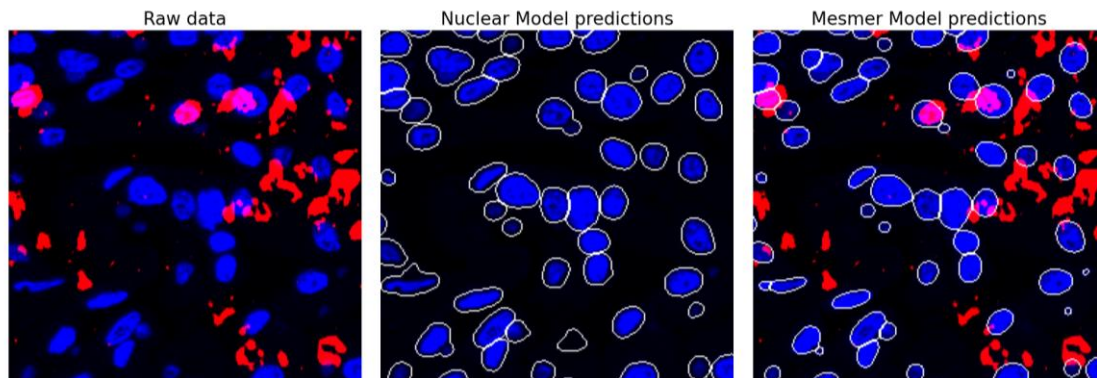
Figure 5.10: Comparison of DeepCell's Mesmer and Nuclear models. The nuclear signal is represented in blue, while the cytoplasmic image, required for the Mesmer model, is displayed in red. Entire cell boundaries are delineated by a thin white line.

The comparison between the models has been carried out in terms of evaluation metrics, as shown in Figure 5.12 (all calculated for an IoU of 50%). In the studied image dataset, the nuclei model of CellPose demonstrates good performance. However, the behaviour of the nuclei model of DeepCell is noticeably inferior. Despite the Mesmer model providing better results than DeepCell's nuclear model, its performance is still worse than that of StarDist. Nevertheless, the StarDist model trained on our images exhibits the best performance among the four models for this dataset.
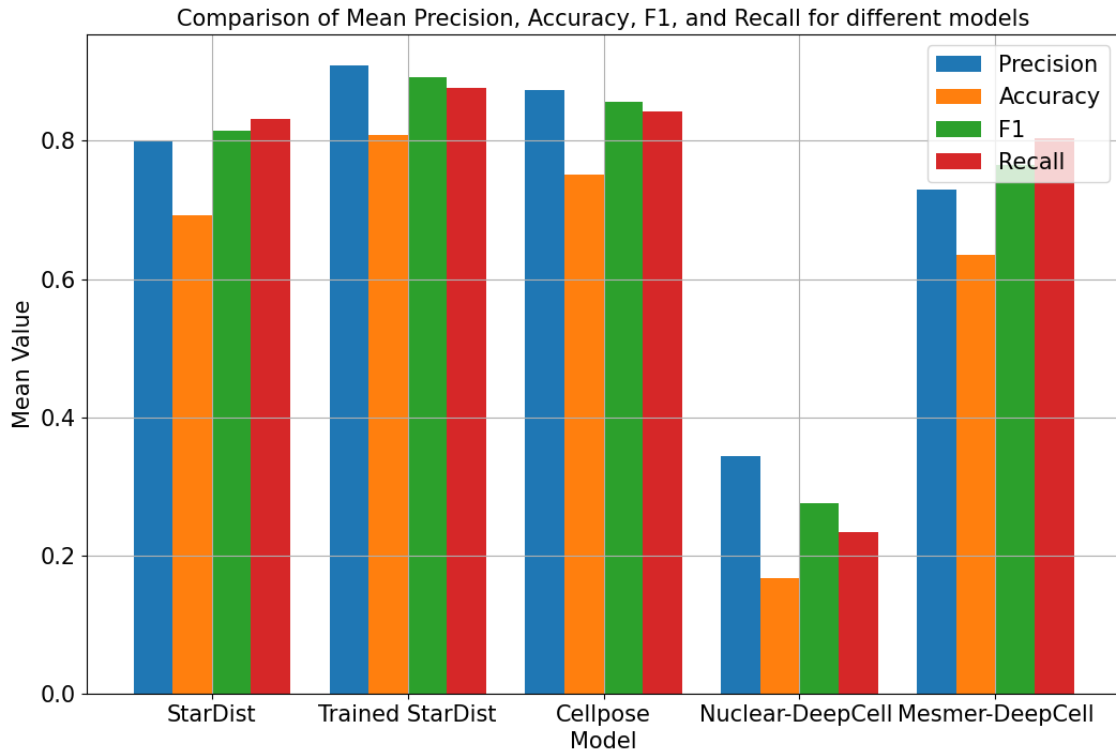
Figure 5.11: Comparison of mean values for different model metrics for IoU of 50%.

Indeed, for a more restrictive IoU threshold of 75%, the trained model of StarDist on our images continues to exhibit the best performance for all the evaluated metrics, as represented in Figure 5.12. In this case, the differences between the five compared models are more pronounced, highlighting the clear benefits of our training process.
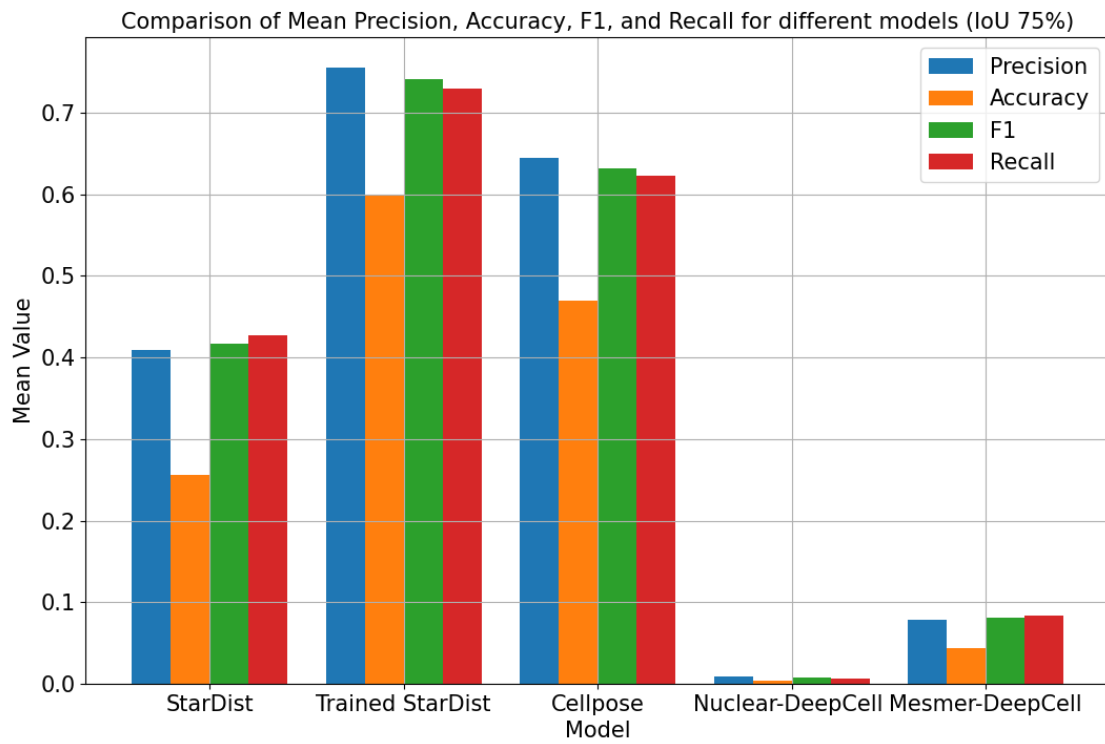
Figure 5.12: Comparison of mean values for different model metrics for IoU of 75%.

5. 3. Whole Cell Segmentation

Mesmer, the DeepCell model selected for whole-cell segmentation, has two adjustable input parameters to determine the membrane of each cell. These parameters are the interior threshold and the maxima threshold.

The interior threshold defines how conservative the model is when estimating if a pixel corresponds to background or cell. Low values result in larger cells, while high values result in smaller cells. Its effect on whole cell segmentation has been demonstrated for two different images with different cell densities in Figure 5.13. The default interior threshold, 0.2, has been selected for entire cell segmentation as it can been seen that it provides better results. When increasing the parameter, the cells are detected with a smaller size, potentially excluding parts of their nuclei from the segmentation.
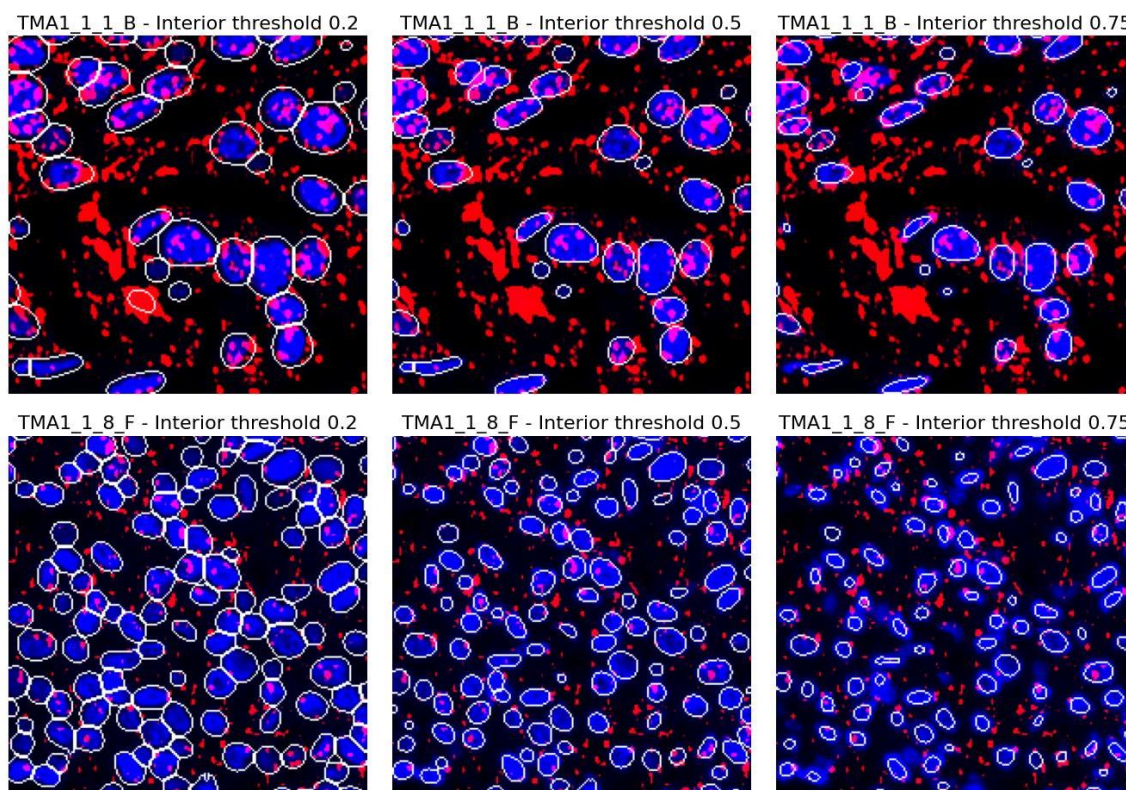
Figure 5.13: Comparison of various interior threshold values for the Mesmer model comparison in two different images. The nuclear images are represented in

blue, while the cytoplasmic images are displayed in red. Entire cell boundaries are delineated by a thin white line.

On the other hand, the repercussion of the maxima threshold on Mesmer's precision has been analysed. This parameter controls what the model considers as a single cell. Low values result in more separate cells, while high values result in fewer cells, as can be seen in Figure 5.14. In the figure, the zones where the effect of the maxima threshold is observed are marked with a red circle. Again, the default value has been selected because it provides the best performance in separating cells. As there are several zones of aggregated cells in this cohort of patients, it is required to properly separate the cells.
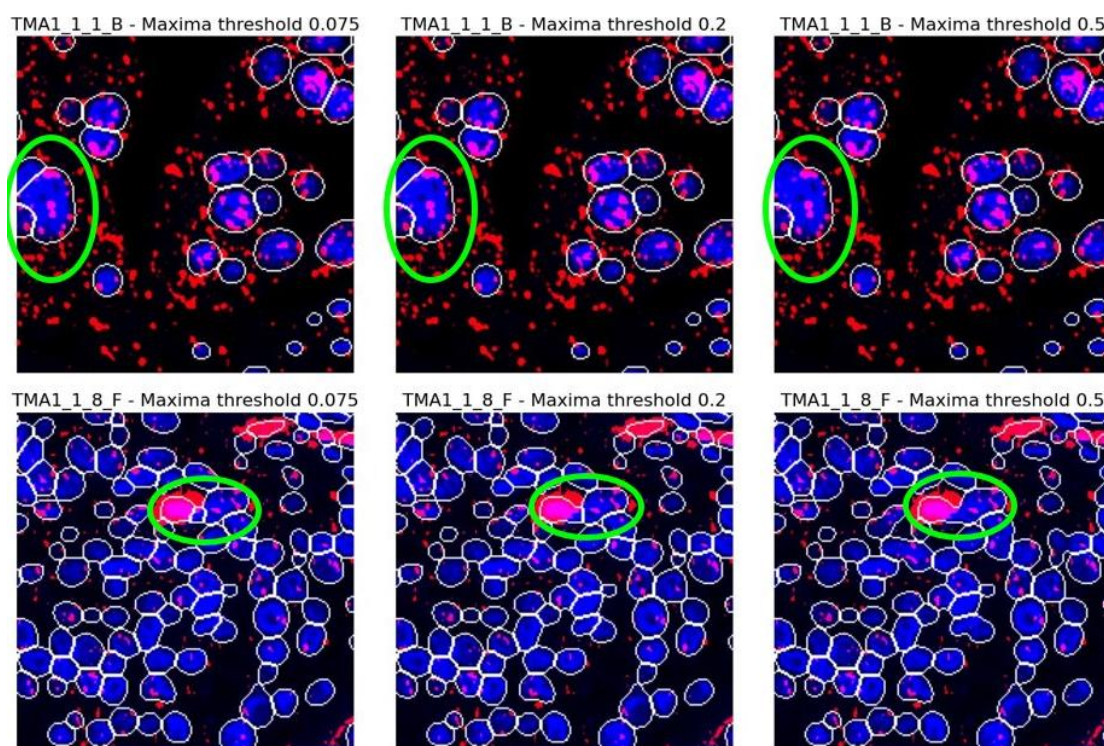
Figure 5.14: Comparison of various maxima threshold values for the Mesmer model in two different images. Nuclear images are represented in blue, while cytoplasmatic images are displayed in green. Entire cell boundaries are delineated by a thin white line. Green circles indicate the zones where the effects of the maxima threshold are evident.

As mentioned in the Methods chapter, the input cytoplasmic image that has been chosen for Mesmer model segmentation resulted from the sum of both CD8 and GranB channels. An illustration of how the model

predicts depending on the input cytoplasmic image can be seen in Figure 5.15. The comparison includes the CD8 channel, GranB channel, and the sum of both as input images. The predictions made by DeepCell for each input demonstrate how the model fails, for example, detecting a unique cell as two when using the GranB single channel, as observed within the green circle.
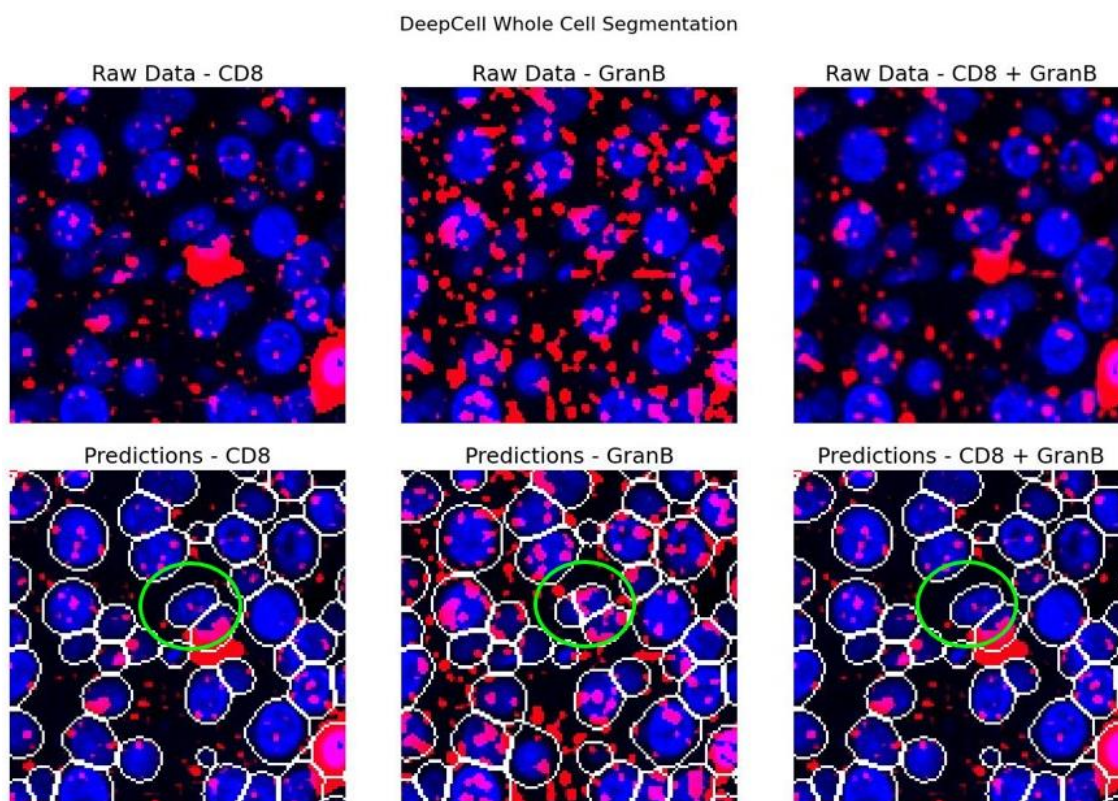
Figure 5.15: Whole Cell outline prediction with DeepCell using different cytoplasmic input images. The nuclear image is represented in blue, while the various cytoplasmatic images are displayed in red.

DeepCell, in its cell segmentation process, assigns a distinct grey value to each cell, producing 16-bit images. Consequently, all pixels within a cell share the same intensity value. Figure 5.16 illustrates a colorized example of this unique grey value assignment. This approach ensures a clear differentiation of individual cells in the analysis.
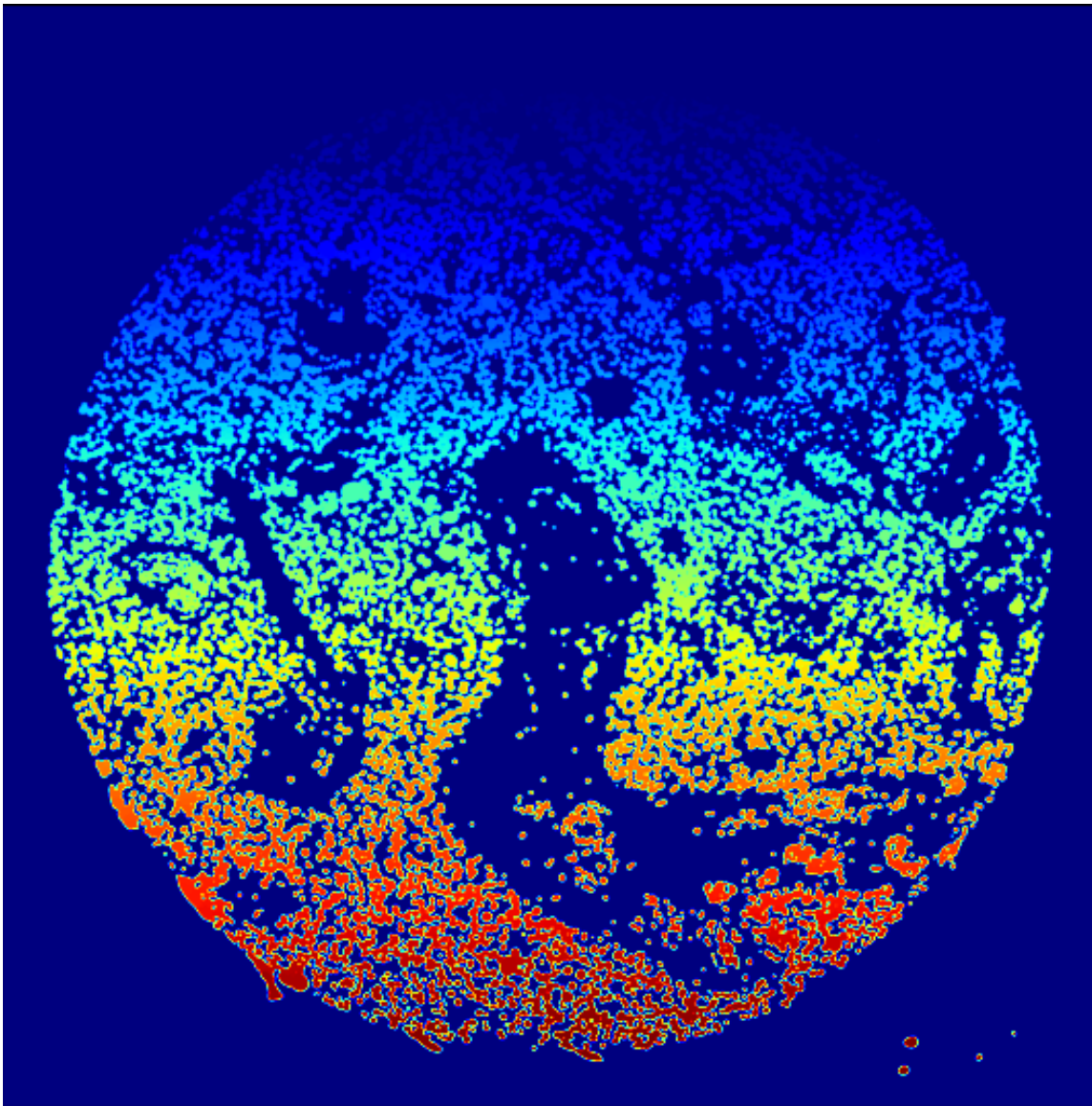
Figure 5.16: Colorized Mesmer predictions for the image TMA1_1_1_A.

In contrast to validating nuclei segmentation, generating ground truth images for whole cell evaluation is not trivial due to the difficulty in delineating cell boundaries without a universal cell membrane fluorescent marker. If the analysed cell lacks positive signals for any of the cytoplasm or membrane biomarkers, annotating the entire cell boundary becomes unfeasible. Unlike nuclei segmentation, where StarDist, using the DAPI expression, outlines the nuclei directly, the Mesmer model's outlining of the entire cells does not exactly correspond to the signal in the images.

In this case, images are obtained from the expression of the DAPI biomarker as the nuclear image and the sum of CD8 and GranB expressions as the cytoplasmatic image. However, Mesmer achieves this

by using the results of applying a marker-controlled watershed algorithm, with the two images as inputs. Consequently, the cells expand by a few pixels based on the signal level in the image, considering the proximity to other cells.

This complexity makes evaluating the model's performance in this patient cohort challenging. As mentioned earlier, ground truth images would be essential for a comprehensive evaluation. As will be described later, we have chosen to directly evaluate cell classification using the CD8 biomarker because this marker is expressed in the cellular cytoplasm.

5. 4. Classification of Cell Phenotypes

In this section the obtained results for the CD8+ and BATF3+ cell phenotypes will be detailed. Starting with the optimal parameters selected to consider a cell positive for one of the phenotypes and concluding with the results presentation. Noting that the evaluation in each analysis has been conducted using the ground truth of CD8+ and BATF3+ cells described in the corresponding section of the methods.

5.4.1. Optimal Threshold Selection for Cell Phenotype Classification

The initial intention was to use the Otsu automatic threshold. However, despite the precision metric indicating the classifier was not detecting many false positives, the results obtained for the recall metric suggested that there were numerous positve cells that the classifier was not detecting. This occurred in CD8+ and BATF3+ cellular phenotypes as can be seen respectively in Figure 5.17 and Figure 5.18.
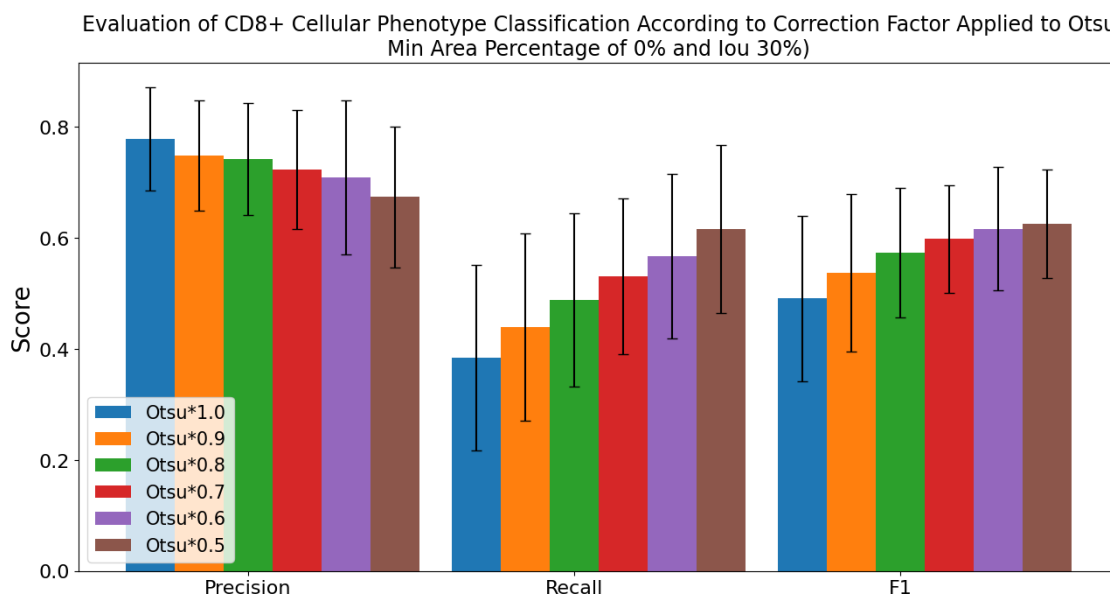
Figure 5.17: Evaluation metrics scores for different intensity thresholds for CD8+ phenotype classification.
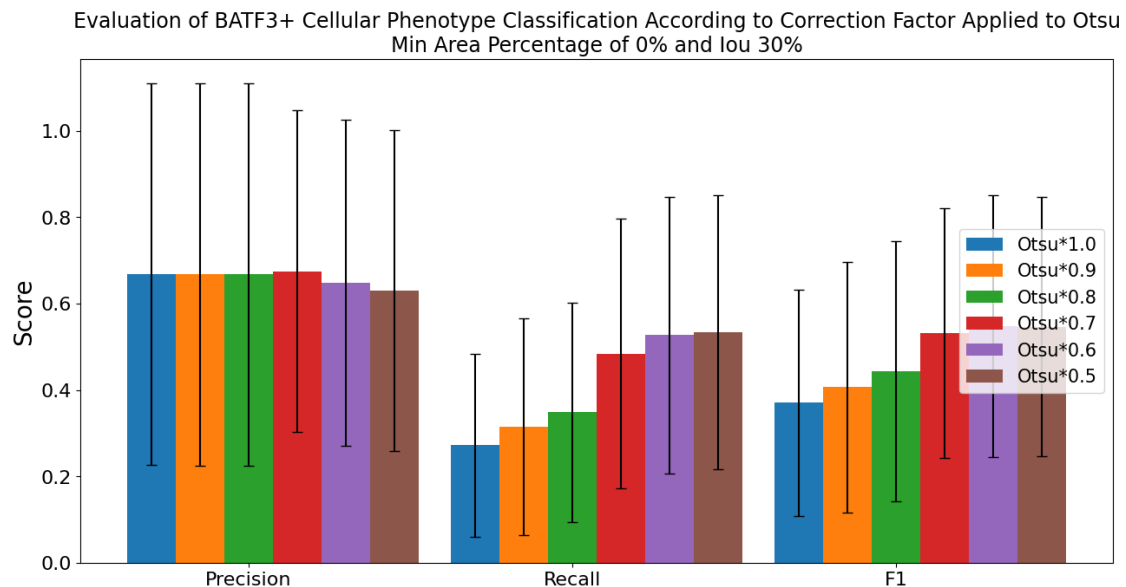
Figure 5.18: Evaluation metrics scores for different intensity thresholds for the BATF3+ phenotype classification.

Observing both previous figures, as the threshold value decreases, achieved by reducing the multiplying factor for Otsu, there is an observable improvement in the F1 score, which represents the harmonic mean of precision and recall metrics. This improvement suggests that a lower threshold should be used to detect the majority of positive expressions of CD8+ and BATF3+ phenotypes.

We conclude that the Otsu's method is not a suitable approach for the analysed images as it establishes excessively high thresholds and therefore too many false negatives, which can be inferred from the low recall values. Both the CD8 and BATF3 channel images present a unimodal histogram with a dominant peak corresponding to background pixels. The histogram of both channels for the image TMA1_1_1_A has been represented as an example in Figure 5.19.
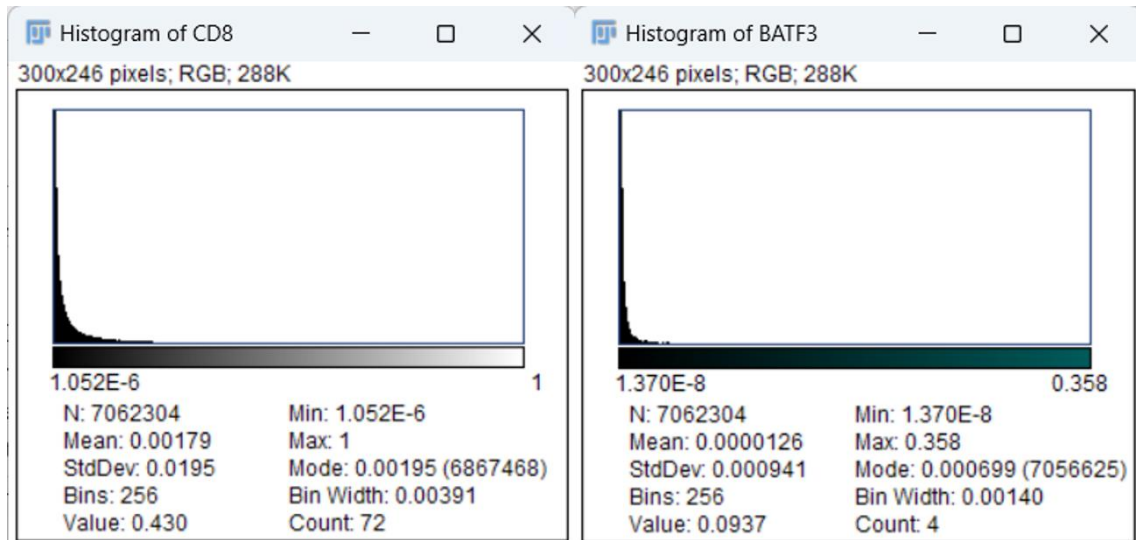
Figure 5.19: Histograms of the CD8 (left image) and BATF3 (right image) channels for the TMA1_1_1_A image.

Due to the unimodal histogram, the Triangle method approximation results in better recall metric compared to the one previously obtained with Otsu, as illustrated in Figure 5.20 and Figure 5.21. The observed results demonstrate a clear enhancement in recall metric, signifying that over 80% of positive cells are detected.
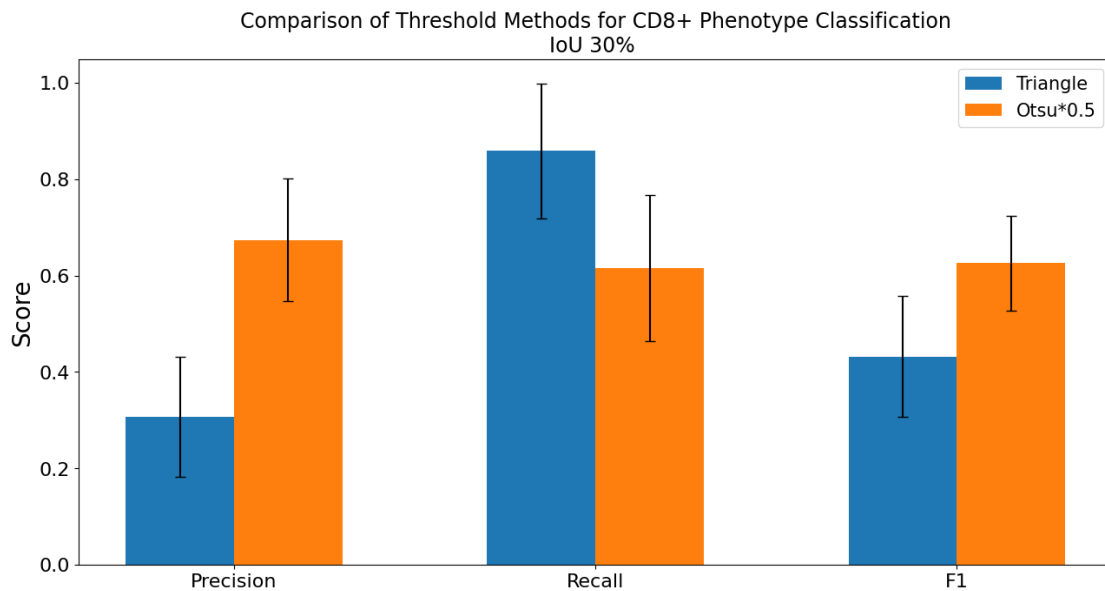
Figure 5.20: Triangle and Otsu thresholding methods comparison for the CD8+ phenotype classification.
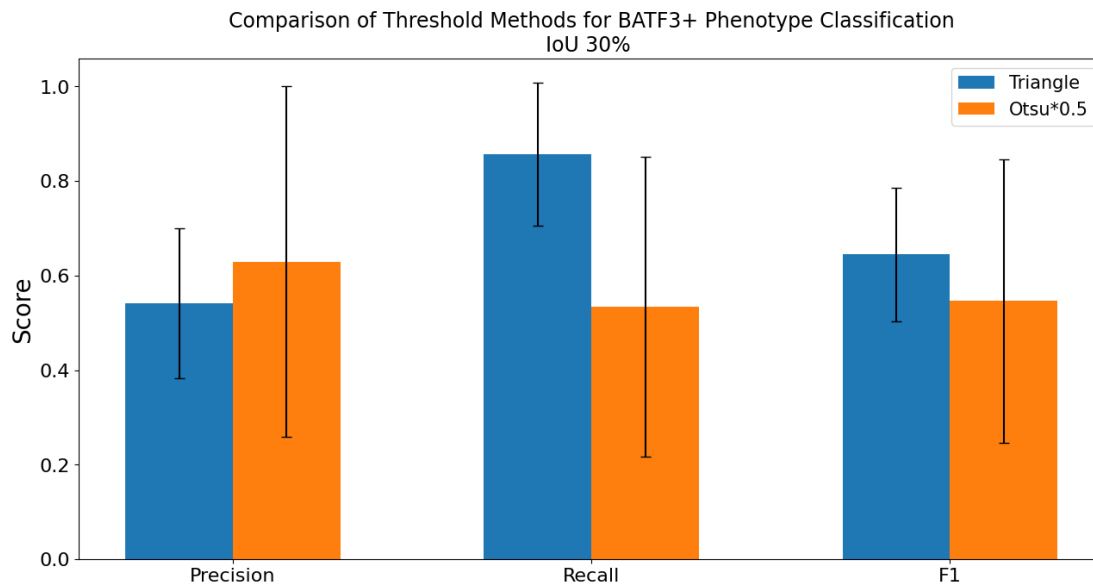
Figure 5.21: Triangle and Otsu thresholding methods comparison for the
BATF3+ phenotype classification.

However, there is a decline in precision results, as seen in the previous figure, indicating that some cells are identified as positive when they are, in fact, negative. This issue may arise from considering all cells with more than one-pixel intensity higher than the threshold as positives. To address this, a minimum area percentage will be later established.

Figure 5.22 provides a visual comparison between the Otsu and Triangle thresholding methods, highlighting how Otsu's model neglects a significant number of pixels. The corresponding histograms for these channels confirm a highly concentrated distribution of pixels around the zero value, with only a few having values near zero. The thresholds established by both methods are visually represented by vertical lines.
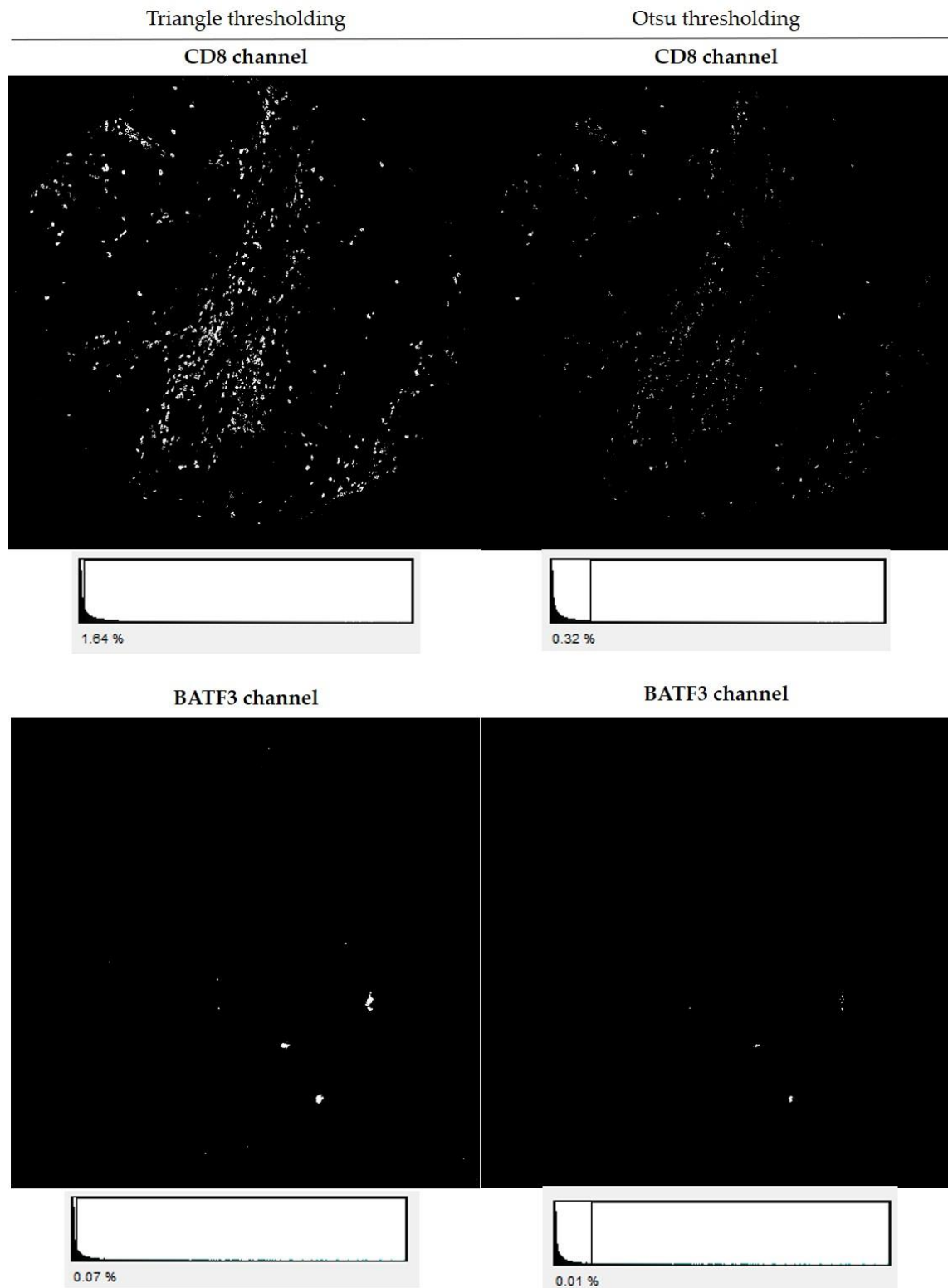
Figure 5.22: Triangle (left image) and Otsu (right image) thresholding methods applied in CD8 and BATF3 channels of TMA1_1_3_G image.

5.4.2. Optimal Minimum Area Percentage Selection for Cell Phenotype Classification

This analysis has been carried out using the threshold provided by the Triangle method, based on the results from the previous section.

Firstly, the minimum area percentage to classify CD8 T cells as positive has been determined. The results are illustrated in Figure 5.23, which shows that a minimum occupied percentage area of 10% provides the best CD8+ cell classification results, as evidenced by the F1 scores.
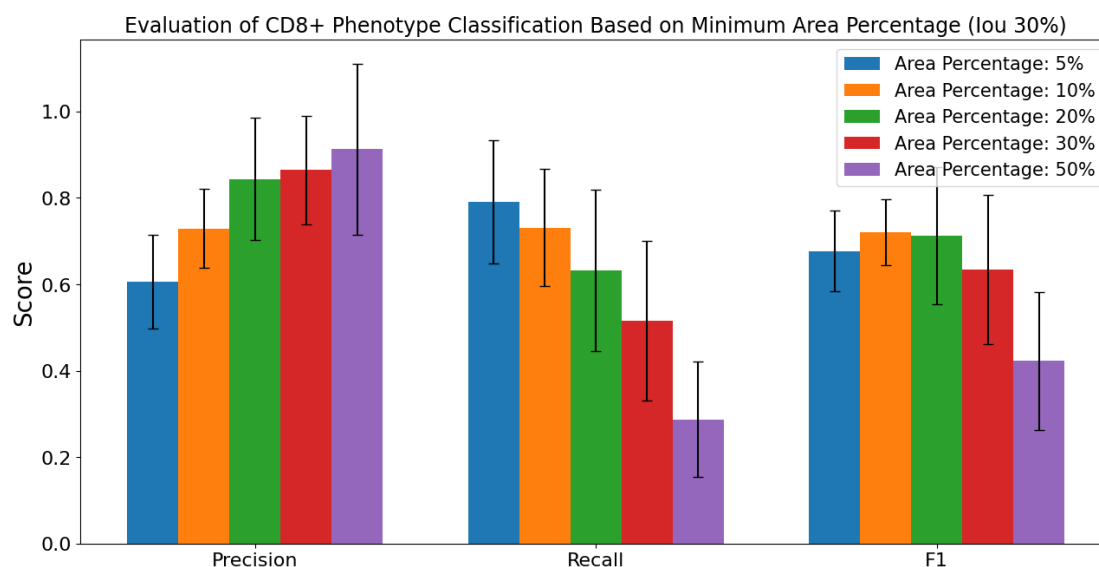
Figure 5.23: Minimum area percentage comparison for CD8+ phenotype classification.

For BATF3 population identification, similar comparisons have been conducted to establish the optimal minimum area percentage represented in Figure 5.24. The chosen minimum percentage of area is 20%, as it presents the best results in terms of F1 score.
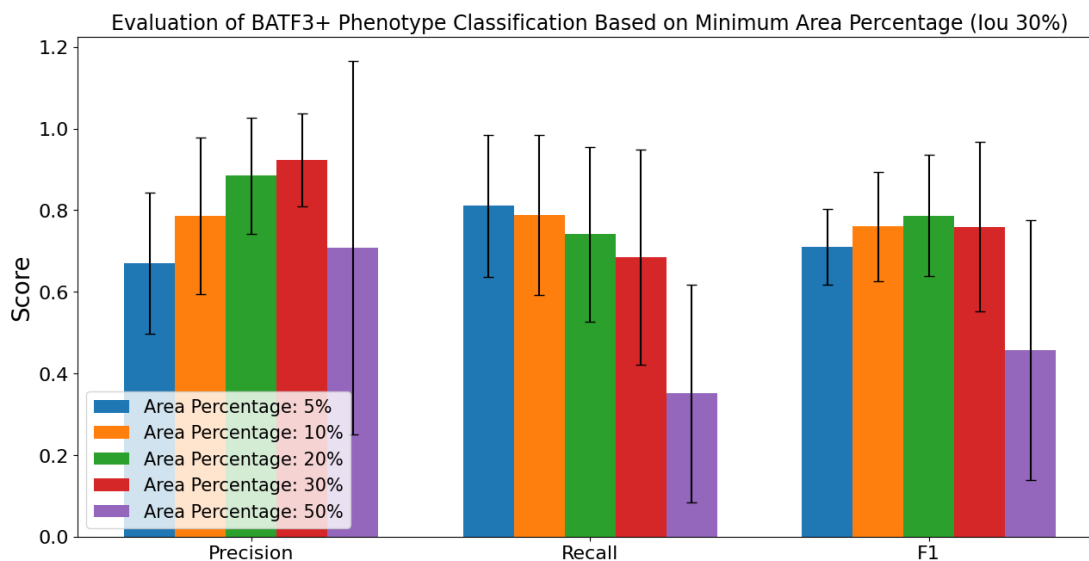
Figure 5.24: Minimum area percentage comparison for BATF3+ phenotype classification.

5.4.3. Optimal Correction Factor for Triangle Thresholding for Cell Phenotype Classification

Considering the optimal minimum area occupied results from the previous section and the Triangle thresholding method approach, a multiplier factor for this thresholding approach has been analyzed.

In the case of CD8+ phenotype classification, as observed in Figure 5.25, the best results are achieved when using the unaltered threshold obtained from the Triangle method observing the F1 evaluation metric.

Evaluation of CD8+ Phenotype Classification Based on Triangle Multiplier Factor (Iou 30%)
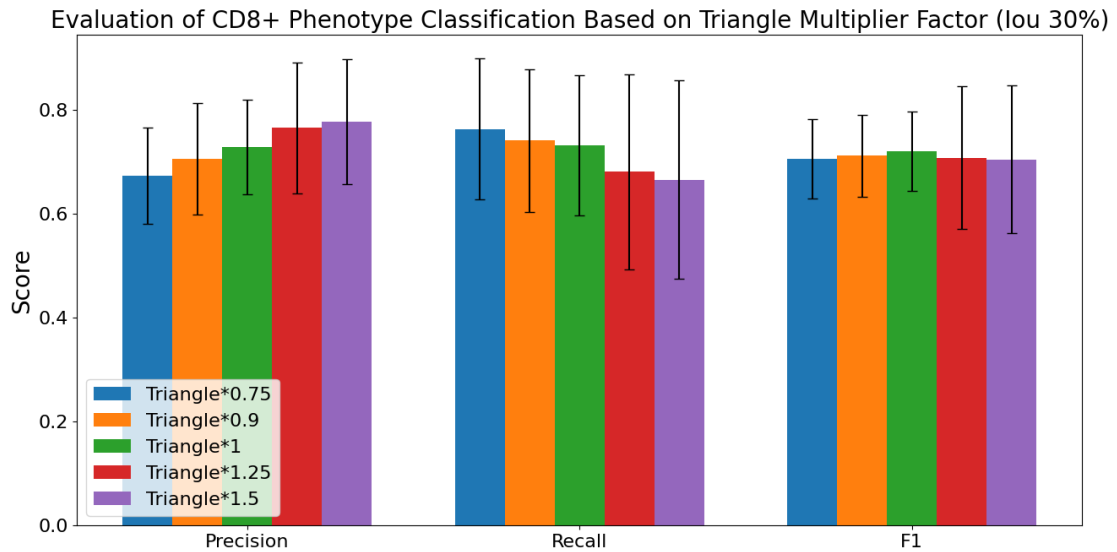
Figure 5.25: Automatic Triangle thresholding correction for CD8+ phenotype classification.

For BATF3 population identification the optimal threshold multiplying factor has been determined following the same method. The results are represented in Figure 5.26. In this case, BATF positive cells identification achieves its best performance when using the threshold provided by the Triangle thresholding procedure multiplied by a 1.5 correction factor.
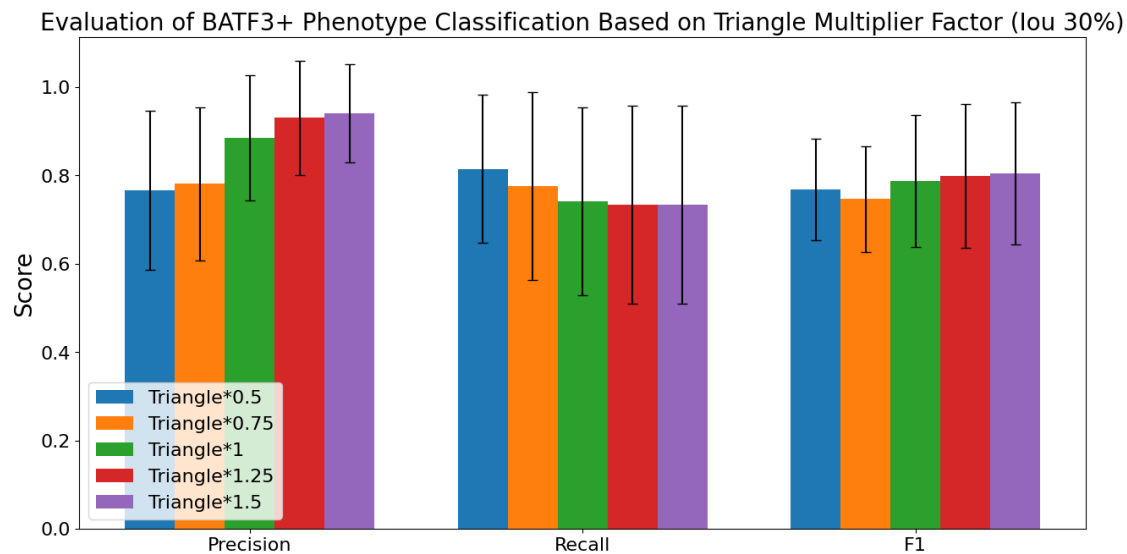
Figure 5.26: Automatic Triangle thresholding correction for BATF3+ phenotype classification.

5.4.4. Classification Results

The classification results are visually represented in each image by outlining the boundaries of CD8+ cells in red colour and BATF3+ cells in green, as shown in Figure 5.27.
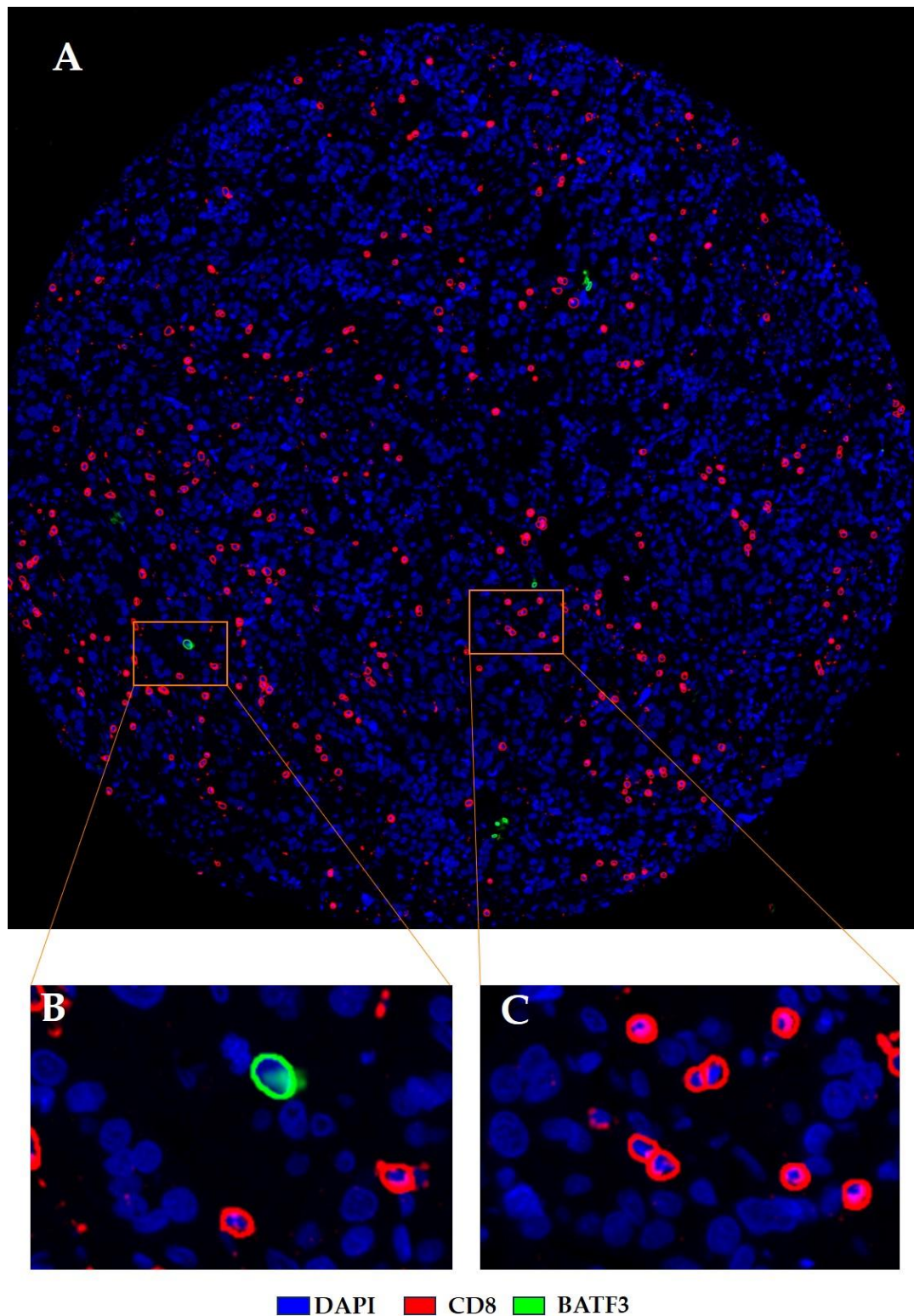
Figure 5.27: Classification results for the image TMA1_1_7_E. A. CD8+ cells outlined in red, while BATF3+ cells are outlined in green. B. Full view of the image. C. Zoom-in showing BATF3+ cell detected. D. Zoom-in showcasing CD8+ cells detected

The number of CD8 and BATF3 positive cells within the tissue area for each image have been exported to an Excel sheet. This enables a posterior analysis of the abundance of these cell types within the tumor microenvironment in this cohort of patients. As an example, a portion of the results table is displayed in Table 5.4, where each row corresponds to an image and each column represents the analysed tissue and the number of positive cells for CD8 and BATF3 found.

Image	Tissue area (um2)	#CD8+ cells	#BATF3+ cells
TMA1_1_1_A	3175574	606	11
TMA1_1_1_B	2092138	4	5
TMA1_1_1_C	2343634	93	5

Table 5.4: Sample of the Excel sheet indicating the presence of CD8+ and BATF3+ cells.

Similarly, for a later study of the spatial interactions between cells classified as CD8+ and BATF3+, an Excel sheet has been created for each image, registering the coordinates of all cells classified as positive for one of these biomarkers. The format of the sheets is represented in Table 5.5, where the columns indicate the horizontal and vertical positions of the centroids of positive cells for BATF3 or CD8. The number of rows varies depending on the quantity of positive cells encountered of each type.

X_BATF3	Y_BATF3	X_CD8	Y_CD8
2375,83	1893,616	534,844	565,75
2398,69	1890,632	91,006	1218,823
1937,523	1857,863	1963,614	563,95

Table 5.5: Sample of the Excel table containing coordinates for the centroids of CD8+ and BATF3+ cells in the image TMA1_1_3_A.

5. 5. Study of Spatial Interactions Between Cell Phenotypes

The clinical response to the Atezolizumab immunotherapy is known for 61 patients in the studied cohort. However, four of them are excluded from the analysis due to presenting a not evaluable response.

Clinical responses have been divided into two groups: responders and non-responders. The responder group is composed of patients presenting partial (7) or complete (10) response to the therapy, while the non-responder group includes patients with stable (19) and progression (21) disease responses to treatment. In this study, stable disease cases, which are typically excluded from both responder and non-responder groups, have been included in the non-responder group. This decision was made due to the small sample size. Specifically, 19 patients present stable disease response to treatment, constituting a 33.33% of the clinical outcomes under investigation. Additionally, the predictive power of the biomarkers in this study increased when the patients presenting stable disease were categorized as non-responders.

Firstly, the correlation between clinical outcomes and the number of CD8+ cells per square micron has been assessed. There were 17 responder patients, while 40 belong to the non-responder group. The p-value probabilistic significance indicator has been calculated to measure the differences between the responders and non-responders groups throuth a statistical T-test [47]. Significant differences have been identified between patients responding to the treatment and those with progressing or stable disease in terms of the presence of CD8+ T cells in the tumor microenvironment ($p = 4.3342 \times 10^{-6}$), see Figure 5.28.
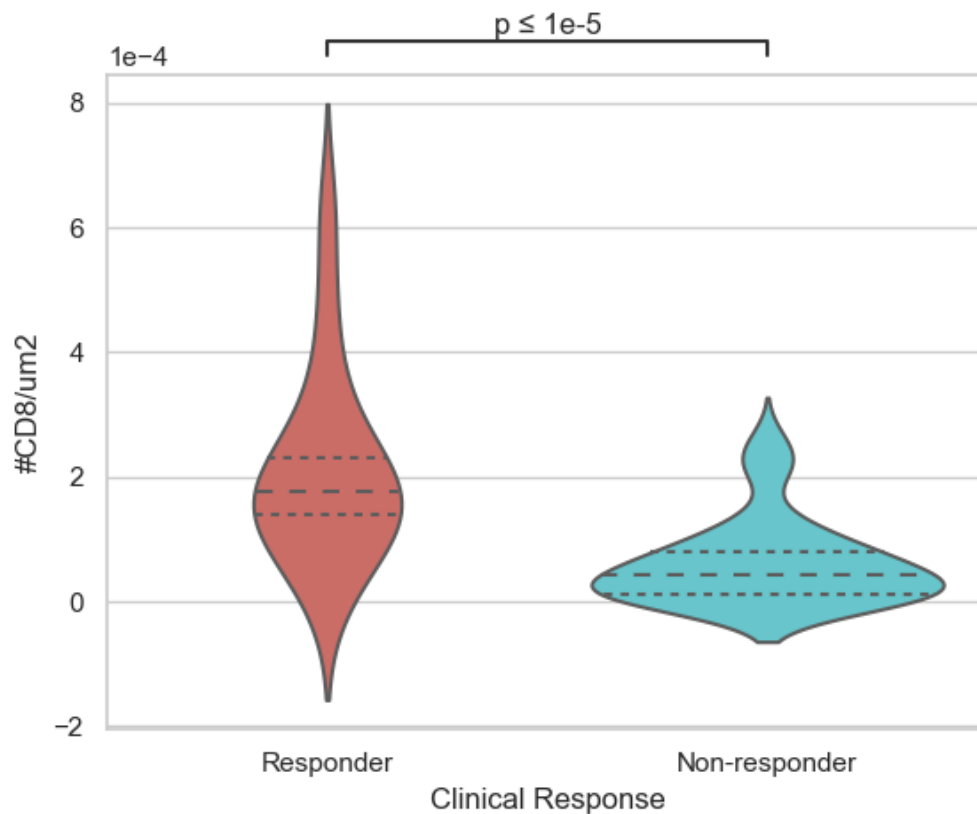
Figure 5.28: Correlation of the number of CD8+ T cells with the treatment response.

A visual representation of the correlation between treatment response and the presence of CD8+ T cells is depicted in Figure 5.29. It highlights the abundance of this cellular phenotype in responder patients and its absence in non-responder patients.
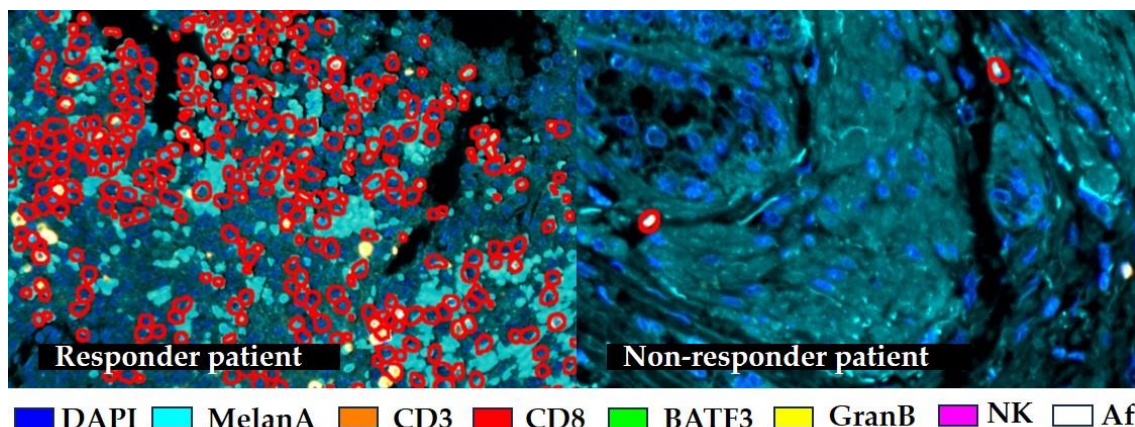
Figure 5.29: Visual comparison of CD8 presence in a responder versus non-responder patient. CD8+ T cells are outlined in red.

Secondly, differences in the number of BATF3+ cells per square micron of tissue in both groups of patients have been analysed. In this case, there were no significant differences between the positive or negative response to treatment, although a trend to significance can be appreciated (p=0.1401). The results are illustrated in Figure 5.30. This suggests that the number of BATF3+ cells in this patient cohort is not such a relevant factor in treatment response.
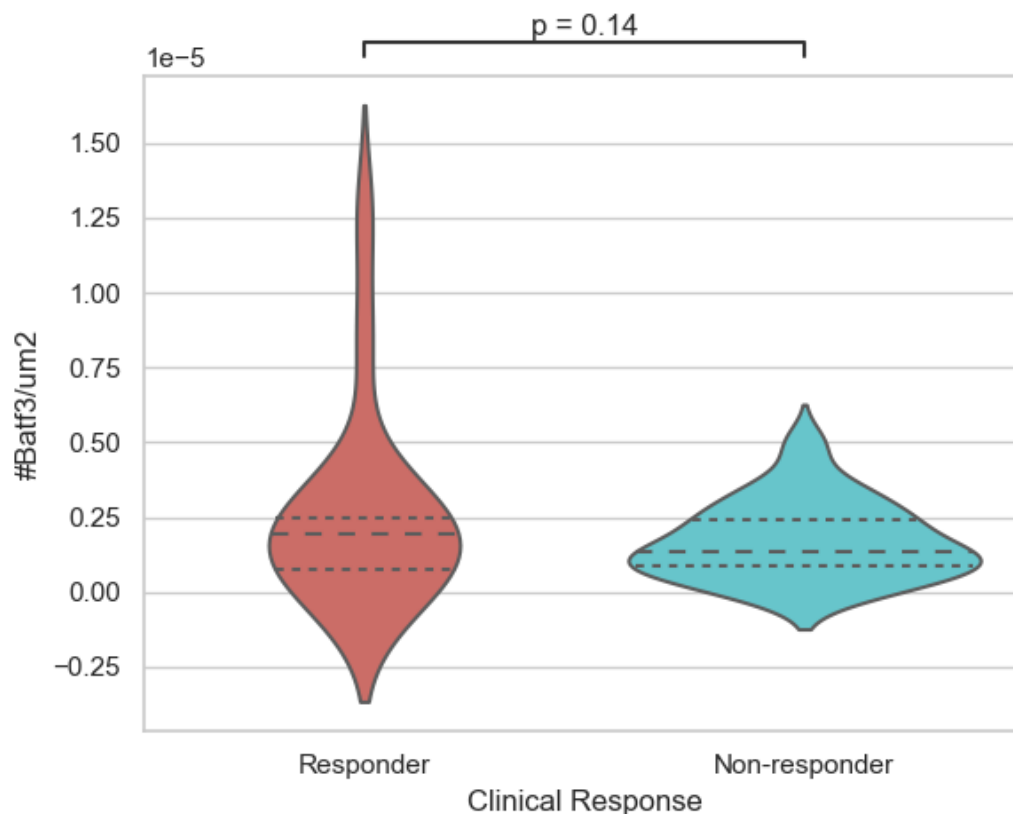
Figure 5.30: Correlation of the number of BATF3+ cells with the treatment response.

Finally, spatial interactions between the two cell populations have been studied. The analysis conducted was based on mean distances between CD8+ cells and their nearest BATF3+ cells, visually represented in Figure 5.31. It must be noted that the number of samples is lower than in the previous analysis due to the exclusion of those images that do not present either CD8+ or BATF3+ cells, resulting in 15 responder samples and 37 non-responder samples.
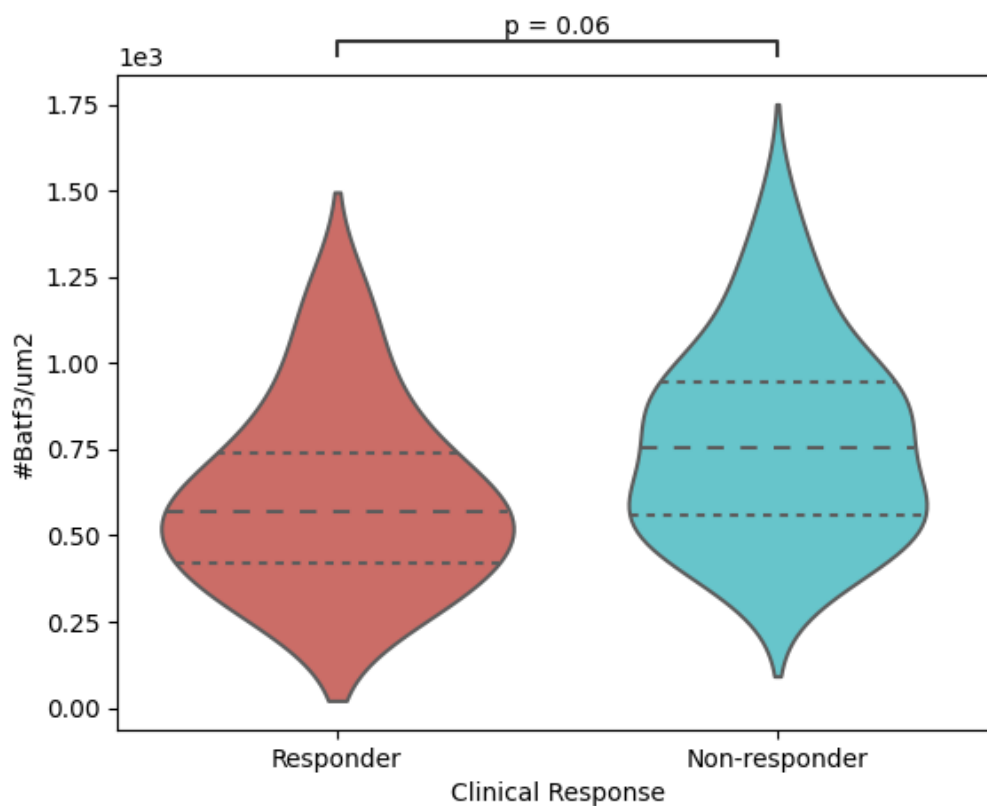
Figure 5.31: Correlation of the distances between CD8+ and BATF3+ cells with the treatment response.

In this case, it has been found differences that approach significance (p=0.0624). Furthermore, considering the reduced sample size due to the exclusion of the mentioned patients, there might be a potential relationship between the distances separating the CD8+ and BATF+ cells and the treatment response.

Chapter 6. Discussion and Conclusion

The implementation of Deep Learning and image analysis techniques in this Final Master project has permitted the quantification of CD8+ and BATF3+ cell populations and the examination of spatial interactions between them, thus, allowing to study their effect on the urothelial carcinoma patient outcomes to the Atezolizumab immunotherapy.

The nuclei segmentation has been carried out by training StarDist on our image dataset. Training the model significantly improved its performance with respect to the pre-trained version, especially evident when evaluating with a restrictive IoU of 75%, resulting in a 92%, 149%, 77%, and 84% improvement in accuracy, precision, F1 and recall metrics, respectively. This indicates that the trained model not only detects nuclei more effectively, but also exhibits an improvement in the accuracy of the nuclear segmentation. Furthermore, the evaluation highlighted that the trained StarDist model outperformed other state-of-the-art Deep Learning based nuclear segmentation models. This superiority suggests promising implications in biological tasks where accurate nucleus detection is crucial for understanding cellular and pathology behaviour.

Whole cell has been conducted using the DeepCell Mesmer Deep Learning model, which utilized nuclear and cytoplasmic input images to outline entire cell boundaries. The nuclear image corresponded to the DAPI channel, reflecting nuclei shape expression. Conversely, due to the absence of a biomarker matching the entire cell shape, a combination of the GranB membrane marker and CD8 cytoplasmic marker was used as the cytoplasm input image. However, generating ground truth images for whole cell evaluation is challenging due to the difficulties in delineating cell boundaries, for this reason it could not be possible to evaluate the whole-cell segmentation model performance.

Once cells and their nuclei have been segmented, the next step involved classifying cells into CD8+ or BATF+ populations based on optimal intensity and area percentage parameters. The evaluation of the classification results, measured through F1 scores, highlighted the superiority of categorizing CD8+ cells based on positive pixels occupying over 10% of the entire cell, and BATF3+ cells when positive pixels exceeded 20%. Accurately determining the minimum area percentage occupied by positive pixels of each marker is crucial for ensuring precise cell classification, as it directly influences the reliability of the classification process.

Attempts to refine the obtained thresholds through Triangle thresholding confirmed that unaltered thresholds provided the best performance for CD8 cell classification. However, better results for BATF3 cell identification were achieved when the threshold provided by the Triangle method was increased by a correctional factor of 1.5. By increasing the threshold, the classifier becomes more selective, resulting in a reduction of false positives and a higher precision in identifying BATF3+ cells. Moreover, it's worth noting the superior performance of the Triangle method compared to Otsu, possibly due to its effectiveness in handling unimodal histograms.

In summary, the results of the optimal classifier for our patient cohort are achieved using an automatic Triangle threshold and a minimum area of 10% for CD8+ cell classification, along with a Triangle threshold factorized by 1.5 and a 20% area for BATF3+ cells. This configuration results in F1 scores of 0.72 and 0.80, respectively, for these parameters.

The results, including cell counts within tissue areas and coordinates for spatial analysis, were exported to Excel sheets for further investigation and visually represented. The visual representation of classification offers useful information regarding the spatial distributions of CD8+ and BATF3+ cells within the tumor microenvironment. In conclusion, this classification process provided valuable insights for studying treatment response and spatial interactions among cell populations in the patient cohort.

Correlations between the presence of CD8+ and BATF+ cell populations, their interactions, and clinical outcomes in response to Atezolizumab immunotherapy for this cohort of patients have been

measured based on classification results and available clinical data. The correlation between the response to the immunotherapy treatment and CD8+ cell density per square micron revealed significant differences between responders and non-responders, as indicated by a p-value of 4.3342×10^{-6}, demonstrating an important impact of CD8+ T cells on treatment response. Consistent with the current state of the art literature, which recognizes CD8+ T cells as potent pathogen killers, our findings perfectly align with prior research, highlighting the crucial role of CD8+ T cells in treatment response.

However, the analysis of BATF3+ cell density presented a p-value of 0.1401, which, despite being low, suggests that no significant differences were observed between positive and negative treatment responses. This finding contradicts the existing state-of-the-art literature, which suggests a correlation between the presence of BATF3+ cells and patient outcomes. Further analysis is required to determine whether this discrepancy is due to a classification issue with BATF3+ cells or if it truly reflects the unexpected behaviour in this specific patient cohort.

While BATF3+ cells alone may not directly impact on Atezolizumab treatment response in this cohort of urothelial carcinoma patients, their interaction with CD8+ cells appear to positively influence clinical response. Regarding the spatial interactions between both populations, nearly-significant differences (p-value of 0.0624) were observed between the non-responder and responder group. Despite not being below the significance threshold, the spatial interactions between CD8+ and BATF3+ cells are expected to have a direct impact on the patient's response to treatment, with those patients whose CD8+ cells are closer to BATF3+ cells presenting a better response to treatment. However, the limited number of samples may be the reason why the results found in this study are at the edge of significance. All in all, these findings provide valuable information to predict the disease behaviour after Atezolizumab treatment, contributing to a deeper understanding of how the immune system combats the tumor.

In conclusion, a comprehensive image analysis pipeline has been constructed, incorporating and retraining deep learning-based methods, enabling a thorough analysis of the tumor microenvironment in multiplex fluorescence images. This analysis has revealed associations between

image parameters and clinical variables that align with previous findings reported in the literature, thereby validating its utility for the scientific community.

Chapter 7. Future Work

Due to the limited time available for the development of this Final Master project, it was not possible to address all the ambitious ideas that could have enhanced the results and facilitated the procedure. These ideas will be detailed in this chapter.

Improving the normalization process represents an interesting avenue for future research. In the context of this Final Master project, the normalization method employed for all images was received from prior analyses conducted on the same patient cohort. Consequently, enhancing the normalization technique to adapt it specifically to our objectives, i.e, BATF3 and CD8 detection, would be of considerable interest.

Regarding nuclei segmentation, it would be interesting to train the CellPose model with our image dataset, due to the good performance demonstrated by its pretrained model for this cohort of patients. The laborious process of annotating nuclei has already been completed, making the training of the model a feasible task. It would only require the preparation of the data in the input format demanded by the model and an understanding of the training process.

State of the art methods utilize neural networks to classify cell phenotypes in multiplexed images as it enhances phenotyping performance [48]. Training a machine learning-based classifier for this purpose would avoid using parameters such as intensity thresholds or minimum percentage of positive pixels for cell classification. These parameters have been well adjusted to this dataset, but may need to be tuned if this image analysis pipeline is to be applied to another patient cohort. A machine learning-based cell phenotype classification approach would be expected to enhance the results obtained, but it requires having annotated data. As mentioned earlier, the annotation process is time-consuming and would require the assistance of experts in image labelling. Furthermore, it is highly recommended that more than one expert annotated the images.

In addition, examining the expression of GranB and its correlation with the response to immunotherapy treatment in urothelial carcinoma patients would offer new insights into this biomarker's relevance within the tumor microenvironment. Literature suggests, despite the limited number of studies, that GranB correlates with tumor progression, potentially promoting cancer cell invasion [49].

Finally, the pipeline implemented during the Final Master project encloses four principal sections: nuclei segmentation, whole-cell segmentation, phenotype classification and the study of the relevance of CD8 and BATF3 biomarkers within the tumor microenvironment. Therefore, the development of a Fiji plugin, written in Java programming language to accelerate the processing, and encapsulating all the methods, would significantly enhance the efficiency of the entire procedure. Furthermore, it would offer the scientific community a valuable tool for the complete analysis of a multiplexed image dataset.

Bibliography

[1] 'Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017', *Lancet Lond. Engl.*, vol. 392, no. 10159, pp. 1789–1858, Nov. 2018, doi: 10.1016/S0140-6736(18)32279-7.

[2] H. Nagai and Y. H. Kim, 'Cancer prevention from the perspective of global cancer burden patterns', *J. Thorac. Dis.*, vol. 9, no. 3, pp. 448–451, Mar. 2017, doi: 10.21037/jtd.2017.02.75.

[3] T. I. A. for R. on Cancer (IARC), 'Global Cancer Observatory'. Accessed: Jan. 25, 2024. [Online]. Available: https://gco.iarc.fr/

[4] J. Miyazaki and H. Nishiyama, 'Epidemiology of urothelial carcinoma', *Int. J. Urol.*, vol. 24, no. 10, pp. 730–734, 2017, doi: 10.1111/iju.13376.

[5] 'What Is Bladder Cancer? - NCI'. Accessed: Dec. 10, 2023. [Online]. Available: https://www.cancer.gov/types/bladder

[6] H. Kaseb and N. R. Aeddula, 'Bladder Cancer', in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2023. Accessed: Dec. 15, 2023. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK536923/

[7] H. Sung *et al.*, 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[8] 'Cancer today'. Accessed: Jan. 25, 2024. [Online]. Available: http://gco.iarc.fr/today/home

[9] 'Cancer of the Urinary Bladder - Cancer Stat Facts', SEER. Accessed: Jan. 25, 2024. [Online]. Available: https://seer.cancer.gov/statfacts/html/urinb.html

[10] A. Cassell, B. Manobah, and S. Willie, 'Diagnostic and Therapeutic Challenges of Rare Urogenital Cancers: Urothelial Carcinoma of the Renal Pelvis, Ureters and Urethra', *World J. Oncol.*, vol. 12, no. 1, pp. 20–27, Feb. 2021, doi: 10.14740/wjon1360.

[11] Y. Zhang, H. Rumgay, M. Li, H. Yu, H. Pan, and J. Ni, 'The global landscape of bladder cancer incidence and mortality in 2020 and projections to 2040', *J. Glob. Health*, vol. 13, 2023, doi: 10.7189/jogh.13.04109.

[12] E. B. Cone *et al.*, 'Assessment of Time-to-Treatment Initiation and Survival in a Cohort of Patients With Common Cancers', *JAMA Netw.*

Open, vol. 3, no. 12, p. e2030072, Dec. 2020, doi: 10.1001/jamanetworkopen.2020.30072.

[13] B. R. da Cunha *et al.*, 'Cellular Interactions in the Tumor Microenvironment: The Role of Secretome', *J. Cancer*, vol. 10, no. 19, pp. 4574–4587, Aug. 2019, doi: 10.7150/jca.21780.

[14] H. Ungefroren, S. Sebens, D. Seidl, H. Lehnert, and R. Hass, 'Interaction of tumor cells with the microenvironment', *Cell Commun. Signal. CCS*, vol. 9, p. 18, Sep. 2011, doi: 10.1186/1478-811X-9-18.

[15] D. S. Chen and I. Mellman, 'Elements of cancer immunity and the cancer–immune set point', *Nature*, vol. 541, no. 7637, Art. no. 7637, Jan. 2017, doi: 10.1038/nature21349.

[16] K. J. Hiam-Galvez, B. M. Allen, and M. H. Spitzer, 'Systemic immunity in cancer', *Nat. Rev. Cancer*, vol. 21, no. 6, Art. no. 6, Jun. 2021, doi: 10.1038/s41568-021-00347-z.

[17] A. P. Cogdill, M. C. Andrews, and J. A. Wargo, 'Hallmarks of response to immune checkpoint blockade', *Br. J. Cancer*, vol. 117, no. 1, pp. 1–7, Jun. 2017, doi: 10.1038/bjc.2017.136.

[18] M. L. Disis, 'Immune Regulation of Cancer', *J. Clin. Oncol.*, vol. 28, no. 29, pp. 4531–4538, Oct. 2010, doi: 10.1200/JCO.2009.27.2146.

[19] D. Jiménez-Sánchez, M. Ariz, H. Chang, X. Matias-Guiu, C. E. De Andrea, and C. Ortiz-de-Solórzano, 'NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images', *Med. Image Anal.*, vol. 78, p. 102384, May 2022, doi: 10.1016/j.media.2022.102384.

[20] L. Jia, Q. Zhang, and R. Zhang, 'PD-1/PD-L1 pathway blockade works as an effective and practical therapy for cancer immunotherapy', *Cancer Biol. Med.*, vol. 15, no. 2, pp. 116–123, May 2018, doi: 10.20892/j.issn.2095-3941.2017.0086.

[21] J.-R. Lin, M. Fallahi-Sichani, and P. K. Sorger, 'Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method', *Nat. Commun.*, vol. 6, p. 8390, Sep. 2015, doi: 10.1038/ncomms9390.

[22] W. H. Fridman, F. Pagès, C. Sautès-Fridman, and J. Galon, 'The immune contexture in human tumours: impact on clinical outcome', *Nat. Rev. Cancer*, vol. 12, no. 4, Art. no. 4, Apr. 2012, doi: 10.1038/nrc3245.

[23] S. Jarosch *et al.*, 'Multiplexed imaging and automated signal quantification in formalin-fixed paraffin-embedded tissues by

ChipCytometry', *Cell Rep. Methods*, vol. 1, no. 7, p. 100104, Oct. 2021, doi: 10.1016/j.crmeth.2021.100104.

[24] B. Bodenmiller, 'Multiplexed Epitope-Based Tissue Imaging for Discovery and Healthcare Applications', *Cell Syst.*, vol. 2, no. 4, pp. 225–238, Apr. 2016, doi: 10.1016/j.cels.2016.03.008.

[25] A. L. Ji *et al.*, 'Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma', *Cell*, vol. 182, no. 2, pp. 497-514.e22, Jul. 2020, doi: 10.1016/j.cell.2020.05.039.

[26] K. Iwahori, 'Cytotoxic CD8+ Lymphocytes in the Tumor Microenvironment', in *Tumor Microenvironment: Hematopoietic Cells – Part A*, A. Birbrair, Ed., in Advances in Experimental Medicine and Biology. , Cham: Springer International Publishing, 2020, pp. 53–62. doi: 10.1007/978-3-030-35723-8_4.

[27] H. Raskov, A. Orhan, J. P. Christensen, and I. Gögenur, 'Cytotoxic CD8+ T cells in cancer and cancer immunotherapy', *Br. J. Cancer*, vol. 124, no. 2, Art. no. 2, Jan. 2021, doi: 10.1038/s41416-020-01048-4.

[28] E. N. Scott, A. M. Gocher, C. J. Workman, and D. A. A. Vignali, 'Regulatory T Cells: Barriers of Immune Infiltration Into the Tumor Microenvironment', *Front. Immunol.*, vol. 12, p. 702726, 2021, doi: 10.3389/fimmu.2021.702726.

[29] H. Cheng, K. Ma, L. Zhang, and G. Li, 'The tumor microenvironment shapes the molecular characteristics of exhausted CD8+ T cells', *Cancer Lett.*, vol. 506, pp. 55–66, May 2021, doi: 10.1016/j.canlet.2021.02.013.

[30] A. Teijeira *et al.*, 'Depletion of Conventional Type-1 Dendritic Cells in Established Tumors Suppresses Immunotherapy Efficacy', *Cancer Res.*, vol. 82, no. 23, pp. 4373–4385, Dec. 2022, doi: 10.1158/0008-5472.CAN-22-1046.

[31] J. P. Böttcher and C. Reis e Sousa, 'The Role of Type 1 Conventional Dendritic Cells in Cancer Immunity', *Trends Cancer*, vol. 4, no. 11, pp. 784–792, Nov. 2018, doi: 10.1016/j.trecan.2018.09.001.

[32] M. Zidane *et al.*, 'A review on deep learning applications in highly multiplexed tissue imaging data analysis', *Front. Bioinforma.*, vol. 3, p. 1159381, Jul. 2023, doi: 10.3389/fbinf.2023.1159381.

[33] E. Meijering, 'Cell Segmentation: 50 Years Down the Road [Life Sciences]', *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 140–145, Sep. 2012, doi: 10.1109/MSP.2012.2204190.

[34] M. S. van der Heijden *et al.*, 'Atezolizumab Versus Chemotherapy in Patients with Platinum-treated Locally Advanced or Metastatic Urothelial Carcinoma: A Long-term Overall Survival and Safety

Update from the Phase 3 IMvigor211 Clinical Trial', *Eur. Urol.*, vol. 80, no. 1, pp. 7–11, Jul. 2021, doi: 10.1016/j.eururo.2021.03.024.

[35]   N. M. T. Jawhar, 'Tissue Microarray: A rapidly evolving diagnostic and research tool', *Ann. Saudi Med.*, vol. 29, no. 2, pp. 123–127, Mar. 2009, doi: 10.4103/0256-4947.51806.

[36]   U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, 'Cell Detection with Star-Convex Polygons', in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, vol. 11071, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds., in Lecture Notes in Computer Science, vol. 11071. , Cham: Springer International Publishing, 2018, pp. 265–273. doi: 10.1007/978-3-030-00934-2_30.

[37]   P. Bankhead *et al.*, 'QuPath: Open source software for digital pathology image analysis', *Sci. Rep.*, vol. 7, no. 1, p. 16878, Dec. 2017, doi: 10.1038/s41598-017-17204-5.

[38]   J. Schindelin *et al.*, 'Fiji - an Open Source platform for biological image analysis', *Nat. Methods*, vol. 9, no. 7, p. 10.1038/nmeth.2019, Jun. 2012, doi: 10.1038/nmeth.2019.

[39]   N. F. Greenwald *et al.*, 'Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning', *Nat. Biotechnol.*, vol. 40, no. 4, pp. 555–565, Apr. 2022, doi: 10.1038/s41587-021-01094-0.

[40]   C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, 'Cellpose: a generalist algorithm for cellular segmentation', *Nat. Methods*, vol. 18, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41592-020-01018-x.

[41]   A. E. Carpenter *et al.*, 'CellProfiler: image analysis software for identifying and quantifying cell phenotypes', *Genome Biol.*, vol. 7, no. 10, p. R100, Oct. 2006, doi: 10.1186/gb-2006-7-10-r100.

[42]   F. Quan *et al.*, 'Annotation of cell types (ACT): a convenient web server for cell type annotation', *Genome Med.*, vol. 15, no. 1, p. 91, Nov. 2023, doi: 10.1186/s13073-023-01249-5.

[43]   M. T. Dayao, M. Brusko, C. Wasserfall, and Z. Bar-Joseph, 'Membrane marker selection for segmenting single cell spatial proteomics data', *Nat. Commun.*, vol. 13, p. 1999, Apr. 2022, doi: 10.1038/s41467-022-29667-w.

[44]   N. Otsu, 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[45]   G. W. Zack, W. E. Rogers, and S. A. Latt, 'Automatic measurement of sister chromatid exchange frequency.', *J. Histochem. Cytochem.*, vol. 25, no. 7, pp. 741–753, Jul. 1977, doi: 10.1177/25.7.70454.

[46]   X. Ying, 'An Overview of Overfitting and its Solutions', *J. Phys. Conf. Ser.*, vol. 1168, no. 2, p. 022022, Feb. 2019, doi: 10.1088/1742-6596/1168/2/022022.

[47]   M. Krzywinski and N. Altman, 'Significance, P values and t-tests', *Nat. Methods*, vol. 10, no. 11, Art. no. 11, Nov. 2013, doi: 10.1038/nmeth.2698.

[48]   Y. Amitay, Y. Bussi, B. Feinstein, S. Bagon, I. Milo, and L. Keren, 'CellSighter – A neural network to classify cells in highly multiplexed images'. bioRxiv, p. 2022.11.07.515441, Nov. 08, 2022. doi: 10.1101/2022.11.07.515441.

[49]   D. D'Eliseo *et al.*, 'Granzyme B is expressed in urothelial carcinoma and promotes cancer cell invasion', *Int. J. Cancer*, vol. 127, no. 6, pp. 1283–1294, 2010, doi: 10.1002/ijc.25135.