

Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Master Thesis Bioinformatics

Comparison of canonical immune cells in tumour microenvironment

Anja Trumpp

28.02.2023

Reviewers

Prof. Dr. Manfred Claassen (Bioinformatik) Wilhelm-Schickard-Institut für Informatik Universität Tübingen	Prof. Dr. med. Christian Schürch (Medizin) Medizinische Fakultät Universität Tübingen
--	--

Trumpp, Anja:
*Comparison of canonical immune cells
in tumour microenvironment*
Master Thesis Bioinformatics
Eberhard Karls Universität Tübingen
Thesis period: 01.09.22-28.02.23

Abstract

The immune microenvironment of a tumour refers to the interaction between the tumour cells and the immune cells that are present in the surrounding tissue. The interactions between immune cells and tumour cells can have a significant impact on the growth and progression of the tumour. The specific composition and activity of the immune microenvironment can vary greatly depending on the type and stage of the tumour. Hepatocellular carcinoma and pancreatic ductal adenocarcinoma are both highly aggressive and lethal due to late diagnosis and lack of treatment options. To increase the survival time, new treatment strategies are needed, but a necessary prerequisite for that is a better underlying understanding of the tumour immune microenvironment. Therefore in this thesis, high-dimension proteomic imaging data of hepatocellular carcinoma and pancreatic ductal adenocarcinoma are analysed using unsupervised approaches to annotate cell types and using S³-CIMA a semi-supervised approach to detect disease-specific cell type compositions. This work shows that computational methods can not correct high-dimension proteomic imaging data with poor quality. Furthermore, it shows the unique character of different tumour types and the complexity they provide in understanding the tumour immune microenvironment. Furthermore, this work demonstrates that based on high-quality data, an in-depth characterisation of specific cell types is possible. S³-CIMA allows the identification of cellular interaction networks within the regulatory T cell and Macrophages cell niche. In conclusion, refined methods for individual cancer types, to generate high-dimension proteomic imaging data in the laboratory are needed. However, based on high-quality data it is possible to identify interaction networks between tumour and immune cells, which could improve the understanding of the tumour complexity and therefore, provide knowledge to develop new treatment strategies.

Zusammenfassung

Die immunologische Umgebung eines Tumors bezieht sich auf die Interaktion zwischen Tumorzellen und Immunzellen, die im Gewebe vorhanden sind. Die Interaktionen zwischen Immun- und Tumorzellen können sich signifikant auf den Wachstum und die Entwicklung des Tumors auswirken. Die spezifische Zusammensetzung und Aktivität der immunologischen Umgebung kann je nach Art und Stadium des Tumors stark variieren. Das hepatozelluläre Karzinom und das duktale Adenokarzinom des Pankreas sind beides sehr aggressiv und tödlich Krebsformen, da sie erst spät diagnostiziert werden und es an Behandlungsmöglichkeiten mangelt. Um die Überlebenszeit der Patienten zu verlängern, sind neue Behandlungsstrategien erforderlich, allerdings ist hiervor ein besseres Verständnis der immunologischen Umgebung des Tumors eine notwendige Voraussetzung. In dieser Arbeit werden daher hochdimensionale Proteom-Bildgebungsdaten des hepatozellulären Karzinoms und des duktale Adenokarzinom des Pankreas mithilfe von unsupervised Methoden zur Zuordnung von Zelltypen verwendet. Außerdem wird eine supervised Methode S³-CIMA verwendet, um krankheitsspezifische Zelltypzusammensetzungen zu erkennen. Diese Arbeit zeigt, dass computergestützte Methoden nicht in der Lage sind, geringe Qualität von hochdimensionale Proteom-Bilddaten zu korrigieren. Darüber hinaus zeigt die Arbeit die Individualität der verschiedenen Tumortypen und die dadurch einhergehende Komplexität, die das Tumorgewebe bereitet. Darüber hinaus wird demonstriert, dass auf der Grundlage qualitativ hochwertiger Daten eine Charakterisierung spezifischer Zelltypen möglich ist. S³-CIMA ermöglicht die Identifizierung von zellulären Interaktionsnetzwerken innerhalb der regulatorischen T-Zellen und Makrophagen Zellnische. Zusammenfassend lässt sich sagen, dass angepasste Methoden für einzelne Krebsarten erforderlich sind, um hochdimensionale Proteom-Bildgebungsdaten im Labor zu erzeugen. Auf der Grundlage hochwertiger Daten ist es jedoch möglich, Interaktionsnetzwerke zwischen Tumor- und Immunzellen zu identifizieren, die das Verständnis der Tumorkomplexität verbessern und somit Erkenntnisse für die Entwicklung neuer Behandlungsstrategien liefern könnten.

Acknowledgements

First a special thanks to my supervisor Dr. Sepideh Babaei, Universität Tübingen. Whenever I had a problem or a question she took the time and supported me. I am grateful that I was able to learn from her and at the same time had the freedom to find my path to work on the challenges that came up. Thank you for your understanding and especially for taking the time to support me in this process.

I would also like to extend my thanks to Prof. Dr. Manfred Claassen, Universität Tübingen, without his support and his ideas this thesis would not be the same. His door was always open for me and he supported my project whenever I ran into trouble. Thank you for the opportunity to learn from you.

In addition, many thanks to Dr. Christian Seitz, Universitätsklinikum Tübingen and Sophia Scheuermann, Universitätsklinikum Tübingen who did all the work in the lab and allowed me to further work with their results. Thank you for the chance to work with clinically relevant data produced by a new technology, which provided a lot of opportunities to learn. Thank you for your time and your support during this time.

I would also like to extend my thanks to Dr. Kilian Wistuba-Hamprecht and Nicola Herold for the interesting discussion on our Tuesday meetings. Thank you for the interesting conversations and the input regarding the biological background of my data. I was able to learn a lot from both of you and I am grateful for your input.

Finally, I would like to thank my family and friends, who supported me not only during this thesis but also during the last few years. Thank you for your support and encouraging words.

Contents

List of Figures	v
List of Tables	viii
List of Abbreviations	ix
1 Introduction	1
2 Methods and Material	8
3 Results	14
4 Discussion and Outlook	24
A Further Tables and Figures	29
Bibliography	41

List of Figures

2.1 Workflow of cell type annotation for pancreatic ductal adenocarcinoma (PDAC) samples. The raw cell segmented data was normalised with arcsinh and z-score additionally, the optimal co-factor was computed. The density of the raw and normalised data was computed and a bimodal distribution for the normalised data was expected. (A) The eight samples were merged, a neighborhood graph was computed and clustering with Leiden was visualised with a UMAP. In addition, a heatmap representing the median intensity of protein marker per Cluster. The resulting heatmap could not be annotated. (B) Instead of merging all samples, for every normalised ROI the neighborhood graph was computed and also clustering with Leiden was applied. The heatmap and UMAP were used to manually annotate the cell type for every cluster.	10
3.1 Merged pancreas carcinoma samples UMAP based on patients. A) UMAP of the samples displays a batch effect. The cells of one sample are clustering in the same areas. B) UMAP of samples after the batch correction with Harmony shows a distribution of the cells.	16
3.2 Median expression of every marker for every cluster in ROI12. Tumour cells are classified by a high expression of cytokeratin. All immune cells are classified by CD45 expression and further separated by their specific protein marker.	19
3.3 Frequency of the neighbour around the regulatory T cells for each patient and each tissue section. Comparison between the samples is computed by the Wilcoxon rank test ($p < 0.05$). The frequency of cells with a high filter response is significantly higher in the tumour core than in the normal liver ($p=0.0037$).	21

3.4	Waterfall plot displaying the enrichment score of the cell population of the regular T cell neighborhood. An enrichment score > 1 indicated an enrichment.	22
3.5	Enriched cells of M2 Macrophages PD-L1+ cell neighbors for the tumour core and the normal liver tissue. T cell CD57 and mixed immune CD45- cells are showing enrichment for both tissues.	23
A.1	Resulting heatmaps of the implemented workflow for all HCC samples after normalisation and clustering. Most of the cells displayed an interpretable expression pattern and were annotated based on Table A.1.	30
A.2	UMAP of merged pancreas carcinoma samples based on clusters. a) UMAP of the clusters before batch correction. b) UMAP of the clusters after the batch correction with Harmony shows a cloud with multiple clusters. Some clusters are clearly separated, but a lot of cluster overlap and no distinct cluster can be identified.	31
A.3	Merged pancreas carcinoma data displayed in a heatmap with the median intensity of every marker per cluster, after the batch correction using HARMONY. The heatmap, can not be used to annotate the cell types. Some protein marker shows no or weak expression, and others show high expression for all clusters of the same protein markers.	32
A.4	Overview of CD45 and Cytokeratin expression on all region of interest (ROI)s, who did show an overlapping expression for the same cells.	33
A.5	The density of CD45 and Cytokeratin for ROI6 for the raw data and the optimal computed co-factor. No bimodal distribution was achieved by the co-factor optimisation.	34
A.6	Median expression of all PDAC samples with an overlapping expression of CD45 and Cytokeratin. No differentiation between immune and tumour cells was possible for these samples and therefore resulting in an unreliable annotation.	35
A.7	UMAP of all PDAC samples with an overlapping expression of CD45 and Cytokeratin.	36
A.8	Expression of all PDAC samples with no overlapping expression of CD45 and Cytokeratin.	36

A.9 UMAP of all PDAC samples with no overlapping expression of CD45 and Cytokeratin.	37
A.10 Median expression of ROI10 PDAC sample with no overlapping expression of CD45 and Cytokeratin. The cells could be differentiated by these markers, however several other expression marker are overlapping resulting in an unreliable annotation.	37
A.11 Enriched cells of the regulatory T cell neighborhood. A) Waterfall plot displaying the enrichment score of the cell population of the regular T cell neighborhood for the entire tissue. An enrichment score > 1 indicated an enrichment. B) Bubble plot for regulatory T cells of the enriched cell types comparing tumour core and normal liver tissue.	38
A.12 The boxplots represent the frequency of high filter response for all neighbour cells in the three different tissue types. The cell frequency of M2 Macrophages PD-L1+ and M2 Macrophages PD-L1- anchor are shown. M2 Macrophages PD-L1+ have a significant difference between the background and the tissue. A significant difference between the tumour core and the normal liver is computed. M2 Macrophages PD-L1- does show no significance for the computed random background and the tissue areas. Using a different number of neighbours did not change the results.	39
A.13 Waterfall plot displaying the enrichment score of the cell population of the M2 Macrophages PD-L1+ cell neighborhood for the entire tissue. An enrichment score > 1 indicated an enrichment.	40
A.14 Bubble plot for M2 Macrophages PD-L1+ cells of the enriched cell types for the compared tissue area.	40

List of Tables

2.1	For all patients diagnosed with PDAC a frozen tissue sample was processed with MACSima TM . Multiple ROIs per sample were processed and biologically interesting ROIs were selected and further processed. Every patient was classified into an early and late relapse group based on survival time.	9
3.1	Overview of the tested parameter for k nearest neighbour and the number of used training data. The resulting best accuracy score for the training and the overall accuracy is documented. The highlighted values did result in the highest accuracy score and were further used for the downstream analysis.	20
A.1	Analysed protein marker and their corresponding cell type. . . .	29

List of Abbreviations

- CAF** Cancer Associated Fibroblasts
CAR-T Chimeric Antigen Receptor T
CNN Convolutional Neural Network
CODEX Co-Detection by Indexing
CT Computer Tomography
EMT Epithelial-to-Mesenchymal Transition
ES Enrichment Score
EUS Endoscopic Ultrasound
FISH Fluorescence In Situ Hybridization
HCC Hepatocellular Carcinoma
IHC Immunohistochemistry
IMC Imaging Mass Cytometry
MAIT Mucosal-Associated Invariant T
MAPK Mitogen-Activated Protein Kinase
MERFISH Multiplex Error Robust FISH
MICS MACSima Imaging Cyclic Staining
MRI Magnetic Resonance Imaging
NK Natural Killer
PDAC Pancreatic Ductal Adenocarcinoma
ROI Region Of Interest
S³-CIMA Supervised Spatial Single-cell Imaging Analysis
STAT3 Signal Transducer and Activator of Transcription 3
TIME Tumour Immune Microenvironment
WHO World Health Organization

Chapter 1

Introduction

The tissue microenvironment of cancer is complex and influenced by a wide variety of factors, which complicates in general the understanding of how the tumour develops and therefore, the development of new therapeutic strategies. At the same time, new treatment options are essential since cancer is one of the most prevalent diseases, with 10 million deaths and 23.6 million new cases per year (Kocarnik et al., 2022). Based on demographic changes, an increasing number of cancer cases is expected, which further rises the demand for new therapy strategies (Quante et al., 2016). New approaches, particularly patient-specific treatment such as immunotherapy already demonstrated success in several malignancies. However, this success can not be observed in all subtypes of cancer. To adapt these therapies, it is important to understand the underlying tumour immune microenvironment (TIME). The tumour is able to change the cell composition in general and influences the innate and adaptive immune cells to an immunosuppressive and poor immune infiltrated environment. These cell type compositions and immune responses differ in cancer subtypes, in other words for every cancer subtype a greater understanding of the TIME is needed (Binnewies et al., 2018; Morrison et al., 2018).

Pancreatic and Liver cancer

In this thesis we will analyse liver and pancreatic cancer in more detail. Liver and pancreatic cancers are among the most common and deadliest cancers globally (Allen et al., 2020). The TIME of pancreas carcinoma is poorly understood and highly differs between the subtypes, which increases the challenge of successful therapy. The pancreas is mainly composed of endocrine and exocrine cells. A small number of endocrine cells are located as islands in the exocrine area and are responsible for production of hormones, such as insulin. The major part of the pancreas are exocrine cells, which produce gastric enzymes for digestion (Schünke et al., 2022). Pancreas carcinomas can be distinguished

based on the following two histological types. A minority with less than 5% are diagnosed with endocrine pancreatic cancer (Mullan et al., 2001), with a significantly higher probability of survival than patients with exocrine pancreatic cancer (Fesinmeyer et al., 2005). The majority develop exocrine pancreatic cancer, where the most common subtype with 90% is PDAC (Cowgill and Muscarella, 2003). PDAC is resistant against classic chemo- and radiotherapy (Wang et al., 2016; Wild et al., 2020), therefore the most promising method is surgical resection with a 5-year survival time of 18% (Mayo et al., 2012). Nevertheless, surgical resection is only an option for 10 to 20% of pancreatic cancer patients (Gillen et al., 2010; Zhang et al., 2018). Lacking other treatment options, chemotherapy is still applied. FOLFIRINOX and gemcitabine are currently the most promising treatments with a median survival of 11.1 respectively 6.8 months. (Conroy et al., 2011). Late diagnosis due to lack of early symptoms, no general screening recommendations and poor treatment strategies (Canto et al., 2013), results in a highly mortal cancer with an overall 5-year survival of 9%. Additionally, the 12th most common cancer in the world is pancreas carcinoma with around 5000 new cases in 2020 (Allen et al., 2020; Wild et al., 2020). Novel treatment strategies are needed to increase life expectancy, therefore a fundamental understanding of the TIME is essential.

Liver cancer is the fourth leading cause of cancer deaths and the sixth most commonly occurring cancer worldwide. (Bray et al., 2018). It is estimated that liver cancer will affect 1.4 million people worldwide in 2040 (WHO, 2020). Hepatocellular carcinoma (HCC) is with 90% the most frequently observed subtype of liver cancer (Akinyemiju et al., 2017). Chronic infection with hepatitis B or C virus is the most prominent risk factor for developing HCC. Other risk factors especially in more developed countries are aflatoxin-contaminated food products, alcohol misuse, smoking, obesity and diabetes type 2 (Akinyemiju et al., 2017; Bray et al., 2018). Successful prevention of liver cancer initiated by hepatitis B is vaccination. Several other treatment options, depending on the pattern of the disease such as the size of the tumour, metastases and liver function are available. These treatment options could be liver transplantation, radiotherapy or immunotherapeutic approaches. Advanced HCC does poorly respond to classic chemotherapy. However, immunotherapy is a promising treatment strategy, by targeting the PD-1/ PD-L1 checkpoint (Akinyemiju et al., 2017; Gallage et al., 2021). Despite possible prevention and treatment options, an underlying understanding of the TIME is limited. The main reason for that is the general heterogeneity of genetic, histopathological and immunological development of HCC (Gallage et al., 2021). The liver contains a high number of immune cells such as T cells, which can promote or suppress the tumour progression. Tu et al. (2016) did observe a correlation between the enrichment of regulatory T cells in HCC and an immunosuppressive effect, which negatively impacts the survival prognosis. Another important cell type of the innate immune system are macrophages, which have

a highly diverse functionality. The majority of macrophages can be found in the liver, where based on the environmental stimuli, M1 and M2 macrophages are classified. M1 macrophages have a pro-inflammatory function, whereas M2 macrophages have an immunosuppressive effect (Huang et al., 2021). In general M2 macrophages are associated with poor survival in several cancer types, Yeung et al. (2015) did demonstrate *in vivo* and *in vitro* an enhanced invasion of HCC through M2 macrophages. Considering the heterogenetic TIME of HCC and the unclear function of the immune cells within the tumour, a greater understanding is needed to develop new immunotherapeutic treatment strategies.

In summary, the microenvironment of cancer is generally complex and not well understood. PDAC and HCC are both types of cancer with unclear interaction between the immune system and the developing tumour. However, both are potential targets for new immune-targeted therapies.

High Dimensional Imaging Technologies

Next-generation sequencing methods, such as bulk or single-cell RNA sequencing, highly expanded the understanding of the tumour-immune relation in the last few years (Li and Wang, 2021). With bulk RNA sequencing, cellular subpopulation composition can be identified, whereas single-cell RNA sequencing can identify single-cell phenotypes and rare cells in the tumour microenvironment. Single-cell RNA sequencing is also valuable to identify new biomarkers or drug resistance. On the other hand, with these techniques the cells dissociate from the tissue, resulting in a loss of spatial cell distribution and an incomplete picture of the highly heterogeneous TIME. Information on the cell distribution is valuable for an underlying understanding of the cell function and the presence of different cell types and their position within the tumour microenvironment (Li and Wang, 2021).

Image-based immunohistochemistry (IHC) applications have been established in the 1940s and were continuously further developed. In principle, an epitope of an antigen has been identified by a primary antibody. To visualise the interaction, a secondary antibody recognises the primary antibody, which is further linked with a fluorophore or enzymes. Applications of IHC methods are mainly used for cancer diagnosis. Also, a more detailed knowledge of the tumour microenvironment was obtained. However the cross-talk between immune- and tumour cells is highly complex, therefore more advanced methods with an improvement of resolution, multiplexing and modularity have been developed (Coons et al., 1941; Hofman and Taylor, 2013; Lewis et al., 2021).

The high dimensional imaging approaches highly differ in terms of multiplexing and resolution, but also in the underlying technology. The approaches can be, based on mass spectrometry, epitope detection, fluorescence or non-fluorescence technology. Imaging mass cytometry (IMC) is a mass spectrometry-

based approach, with enhanced resolution and multiplexing by using metal-labelled antibodies, which are read out by CyTOF a time-of-flight instrument to analyse proteins and metabolites up to a subcellular resolution. IMC is a fast method and able to detect 100 markers simultaneously IMC (Giesen et al., 2014). Fluorescence *in situ* hybridization (FISH) is an advanced transcriptomic imaging technique and is able to detect low abundance expression and rare mutations on a single-cell resolution. This technology has been further developed in order to resolve the limitation of multiplexing and resolution. Based on this technology, several methods such as multiplex error robust FISH (MERFISH), which overcomes a read-out error were developed (Xia et al., 2019). Iterative fluorescence-based techniques, such as co-detection by indexing (CODEX) and MACSima imaging cyclic staining (MICS) technology provide a multiplex imaging analysis for proteins and metabolites with a subcellular resolution. The cyclic staining process with CODEX allows the quantification of up to 60 markers. Despite an improvement in resolution and multiplexing, the high cost of purified antibodies can be a limiting factor (Black et al., 2021). Both methods are multiparametric and use antibodies conjugated to unique oligonucleotide sequence barcodes. High specificity is achieved by labelled target-specific barcodes (Black et al., 2021; Kinkhabwala et al., 2022). The problem of multiplexing with iterative fluorescence-based approaches on a single-cell or subcellular resolution is still challenging, as the cyclic staining and erasing process is error-prone. Additionally, the fluorescence-based methods tend to bleed through and therefore affect the signal of neighbour cells (Lewis et al., 2021). The MICS technology claims to avoid some of these problems. The problematic fluorescence erasing is handled by three different methods: photobleaching, REAdyelase and REAlease which can be used alone or in combination. In addition, the remaining signals are subtracted in the succeeding image to remove every signal left. In general, no degradation of the tissue has been reported, therefore a high number of markers could theoretically be used for one sample (Kinkhabwala et al., 2022). Both CODEX and MACSimaTM technologies are not tissue destructive allowing for further analyses. In addition, both technologies are highly sensitive which allows the detection of rare cells. They are also automated systems, which reduces the complexity of handling a high number of markers. CODEX, as well as MACSimaTM technology, were successfully used for the analysis of TIME. They also provide software for an in-depth analysis of the images, background correction and cell segmentation. In addition, several algorithms, such as clustering to further process the data are provided (Black et al., 2021; Kinkhabwala et al., 2022).

In conclusion, there are several approaches with different underlying techniques available for high-dimensional image-based approaches. Every method has advantages and disadvantages, based on the biological problem and the research question the appropriate tool has to be selected. These

methods are often used for diagnostics and also for the development of new treatment strategies. To analyse tumour tissue CODEX and MACSima™ technologies are well suited.

Unsupervised Computational Analysis

Multiplexed high dimensional imaging data needs to be further processed, where it is important to account for the unique technical noise of this data. There are several toolkits available for scRNA-seq data such as Scanpy¹ (Wolf et al., 2018) or Seurat (Stuart et al., 2019). However, high-dimensional imaging data cannot be processed in the same way. Based on Scanpy, Squidpy² (Palla et al., 2022) was developed for spatial single-cell analysis. These tools provide basic pre-processing and quality control functionalities, such as normalization, filtering and dimensionality reduction. Further several unsupervised clustering algorithms are provided (Dries et al., 2021). Squidpy is able to visualise spatial data and provides a standard workflow. In order to understand the relationship between different cell types, Squidpy provides additionally a neighborhood and cell-cell interaction analysis approach, which is based on enrichment methods. Based on these unsupervised analyses, meaning no external information is included, it is possible to manually annotate the cell types. A general description of specific cell types and their neighborhood can be realised (Kleino et al., 2022). However, to understand the complexity of a tumour and the microenvironment a comparative analysis of the samples is of interest. The differences across conditions are essential to understand the TIME of diseases. Thereby it is possible to identify specific cell types, which can affect the survival time of the patients. Unsupervised approaches alone can not explain differences associated with the disease status in the TIME. To overcome this limitation, Babaei et al. (2023) introduced a supervised approach supervised spatial single-cell imaging analysis (S³-CIMA). S³-CIMA is able to identify disease-associated cell type compositions (Babaei et al., 2023).

To summarise, high-dimensional spatial data provide challenges. The cell type composition can be identified with unsupervised approaches, provided by tools such as Squidpy. Based on the results the cell types are manually defined. Cell-type annotation is a challenging and time-consuming process, where expert knowledge is essential. The second main challenge is to identify the composition of cell types between the sample conditions, to gain a better understanding of higher-order properties of cellular organisations within the tumour. To overcome this challenge this semi-supervised approach S³-CIMA can be used.

¹<https://scanpy.readthedocs.io/en/stable/>

²<https://squidpy.readthedocs.io/en/stable/>

Supervised Computational Analysis

To identify cellular interaction networks between the cells, a semi-supervised approach S³-CIMA was developed, which already successfully demonstrated the identification of novel disease-associated interactions in colorectal cancer and diabetes type 1 (Babaei et al., 2023). S³-CIMA is based on a weakly supervised spatial learning method, implemented by a multiple instance learning approach (Amores, 2013) with a convolutional neuronal network (LeCun et al., 2015), using a protein expression profile image of single-cells as input. S³-CIMA comprises a global, anchor-based and functional spatial enrichment analysis, which either identifies cell subsets that are: enriched in the neighborhood across multiple conditions, condition specific enriched around the anchor cell or enriched nearby a specific functional activity in the tissue. The input for S³-CIMA is a multiplexed proteomic image. The data matrix is single-cell segmented, where the marker expression and the coordinates are represented in the columns. Rows describe the single cells. Using unsupervised approaches the cells are annotated with their cell type. Specific cell types of interest are selected as anchor cells. Based on the anchor a multi-cell input of the k nearest neighbor cells is generated and fed into cell convolutional neural network (CNN) (Arvaniti and Claassen, 2017). This model consists of a single convolutional layer, pooling layer and output layer. The convolution layer is composed of filter weights corresponding to the molecular profiles of the cell subset. The cell frequency is computed by the pooling layer and then used to either compute the regression or classification phenotype. The phenotypes (e.g., survival or patient outcome) are the ground truth and based on the molecular profiles, enrichment differences of cell subsets are determined. The associations are learned only based on the cell profiles and not on the annotated cell types (Babaei et al., 2023).

In summary, supervised approaches such as S³-CIMA are able to provide knowledge about higher-order properties of cellular organisations within the tumour. Based on this it is for example possible to identify tumour-suppressive or promoting cell types. Furthermore, these findings can be used to develop new treatment strategies.

General Structure

This thesis starts with the biological and technical background of the data, followed by the challenges of spatial image-based approaches and analytic tools used for high-dimensional image data. In the following chapter, the used datasets are introduced and all analysis steps are explained in more detail. The cell type annotation was implemented using an example HCC dataset. The developed workflow was applied to the PDAC dataset. In the next step S³-CIMA was applied and the results were analysed. Through the use of unsupervised and new supervised approaches, it is expected to extend the under-

lying TIME knowledge of HCC and PDAC. The immune-tumour interaction is highly complex and the role of specific cell types, such as T cells in HCC is uncertain. It is expected by using S³-CIMA to identify an impact of T cells and M2 macrophages on the tumour. All results are documented in chapter 3. A discussion of the resulting cell type annotation of the PDAC dataset and the enriched cells detected by S³-CIMA in HCC samples are documented in Chapter 4. To conclude this thesis a short outlook is provided.

Chapter 2

Methods and Material

The following chapter is an overview of the multiplexed high-dimensional imaging data and the computational steps. Three datasets were analysed in total, including two MACSima™ and one CODEX multiplex image dataset. A brief description of both imaging techniques is provided, followed by a more in-depth explanation of the cell type annotation. The CODEX data is further applied to S³-CIMA.

Multiplex Tissue Imaging

Datasets

One HCC sample and seven pancreatic cancer fresh frozen tissue samples were processed with MACSima™ imaging technology from Dr Seitz's group. An antibody panel of 159 protein markers for the HCC sample was developed, and three ROI were selected and further processed. All seven patients with pancreas cancer were diagnosed with PDAC. One ROI per patient was provided, except for patient six, where two ROIs were further analysed. An antibody panel of 134 protein markers was developed for the PDAC samples. Based on the time to progression, the patients were grouped into three patients with early and four patients with late relapse. Treatment strategies for all PDAC patients were an operative resection of the tumour and chemotherapy, either with mFOLFIRINOX, Gemcitabine or Capecitabin. The chemotherapy for all patients started after the resection. An overview of the patients and the further processed ROIs is listed in Table 2.1.

Tissue Staining with MACSima™ Imaging Technology

Dr. Seitz's group performed iterative staining with the MACSima™ imaging platform as described in the protocol by Kinkhabwala et al. (2022). The resulting stack of images for every single tissue was imported into the MACS iQ View Software version 1.1.0. Every image slide is segmented with the MACS iQ View Software to define the individual cells. Based on expert knowledge

Patient ID	ROIs	Provided ROI	Group
1	18, 19, 20, 21	19	Early relapse
2	13, 14	14	Late relapse
3	7, 8	8	Early relapse
4	22, 23	22	Late relapse
5	15, 16	15	Late relapse
6	10, 12	10,12	Late relapse
7	5, 6	6	Late relapse

Table 2.1: For all patients diagnosed with PDAC a frozen tissue sample was processed with MACSima™. Multiple ROIs per sample were processed and biologically interesting ROIs were selected and further processed. Every patient was classified into an early and late relapse group based on survival time.

the parameters with the best segmentation results were chosen. The provided segmentation options of MACS iQ are based on the watershed algorithm. The method 'Constrained Donut' was used for all ROIs. The detection sensitivity for cytoplasm and the nucleus was adjusted for every ROI. After the cell segmentation, the software computes based on the cells the mean intensity of the fluorescence and performs a background correction (Kinkhabwala et al., 2022). In the resulting images, artefacts at the corner of the images were observed. This error was corrected by adjusting the segmentation and excluding cells in the corner. The resulting CSV file containing the single-cell data was exported and used for further downstream analysis.

Cell Type Annotation for MACSima™ imaging datasets

The resulting CSV files contain coordinates for every cell and an expression value for every marker and cell. To obtain a comprehensive understanding of the cell distribution, each cell has to be annotated. A total of 15604 cells in 3 HCC ROIs and 78822 cells in seven PDAC samples were identified. Every single cell is normalised and then clustered. Based on the computed cluster and the expression pattern for the protein marker experts can define the cell type. In the following section, the detailed analysis steps are described. An overview of this workflow is visualised in Figure 2.1.

The three HCC samples were used to test and develop the workflow for the cell type annotation. First, a set of protein markers relevant to the cell type annotation has been defined (Table A.1). For the following workflow only the selected protein markers were included. Next, each sample was normalised through the application of the arcsinh transformation and a z-score to the raw, cell-segmented single-cell data. Hickey et al. (2021) compared several normalisation strategies, where a combination of arcsinh transformation and z-score was most effective by reducing the noise of multiplexed imaging. In order to check the normalisation step the marker intensity distribution was computed and visualised before and after the normalisation. A bimodal distribution af-

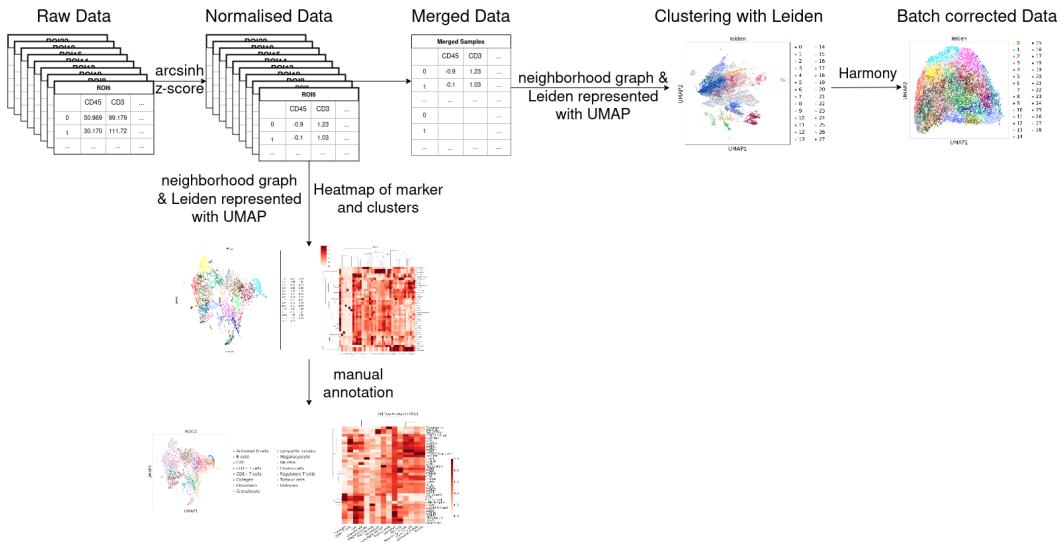


Figure 2.1: Workflow of cell type annotation for PDAC samples. The raw cell segmented data was normalised with arcsinh and z-score additionally, the optimal co-factor was computed. The density of the raw and normalised data was computed and a bimodal distribution for the normalised data was expected. (A) The eight samples were merged, a neighborhood graph was computed and clustering with Leiden was visualised with a UMAP. In addition, a heatmap representing the median intensity of protein marker per Cluster. The resulting heatmap could not be annotated. (B) Instead of merging all samples, for every normalised ROI the neighborhood graph was computed and also clustering with Leiden was applied. The heatmap and UMAP were used to manually annotate the cell type for every cluster.

ter the normalisation is expected. To scale the data in an optimised manner, multiple co-factors for every protein marker were computed. The resulting distributions were scored by the Dip test and the highest score, representing the highest probability of a bimodal distribution, was selected. In the next step, a neighborhood graph was computed with $k = 10$ neighbors. The neighborhood graph was embedded in two dimensions using UMAP projection. The expression for every protein marker is visualised with the computed UMAP. In order to annotate every single cell, similar expression patterns are clustered together using Leiden (Blondel et al., 2008; Traag et al., 2019), an unsupervised community detection clustering algorithm. A higher number of clusters than expected cell types was intended. This over-clustering allowed a clearer separation of the clusters. The clusters were manually annotated and the same cell types were mapped together. For all three HCC samples a resolution factor of $r = 2.2$ was used. Based on the computed clusters the median expression for every protein marker was computed and visualised in a heatmap. Based on the expression of the protein marker, the computed clusters and the heatmap the clusters were annotated.

The above-described workflow was implemented for the eighth PDAC ROIs. It was performed for all protein markers and also for the selected protein marker listed in Table A.1. Using the specified protein marker list results in a clearer result than using all markers. Therefore, only the selected protein markers are used for the final analysis. To summarise briefly the samples were normalised using arcsinh transformation and z-score. The marker intensity distribution before and after the normalisation and co-factor optimisation was computed and visualised in a histogram. In order to annotate the cells of all samples together, the ROIs were merged and batch-corrected on the patient label, using Harmony (Korsunsky et al., 2019) algorithm. A neighborhood graph was computed with $k = 15$ neighbors and an over-clustering was archived by applying the resolution factor $r = 0.6$. The resulting heatmap and UMAP did not show a clear protein marker pattern and therefore the clusters could not be annotated, a reason for this could be over-clustering. To avoid the potential over-clustering error, the workflow was applied to every ROI using $k = 15$ for the computed neighbors and a resolution $r = 2$ for the clustering. For every sample, a UMAP with the computed clusters and the expression for every marker, a heatmap and also a spatial plot using Squidpy was generated for every ROI. The resulting eight heatmaps are showing not a completely clear protein marker pattern. Also, the UMAPs did not separate the clusters. Clusters showing high expression for several protein markers were sub-clustered using the same workflow as described before. Every sample was annotated, but there are contradictory protein marker expressions for the majority of clusters and therefore the annotation is not reliable and can not be applied to S³-CIMA. In order to use S³-CIMA a new dataset was provided.

Spatial Cell Subset Enrichment Detection with S³-CIMA

Dataset

A dataset of 15 HCC human tissue samples processed by CODEX technology was provided by Dr. Ruf's team. In total, a panel of 37 oligonucleotide-barcoded antibodies were used to stain the tissue. The antibody panel was specifically designed to analyse the immune cells within the tumour. The surgical resection of specific areas allows the evaluation of tissue with different immune patterns. The large tissue section contains various histopathological distinct areas. The areas were divided into a non-tumour (adjacent), HCC invasive (rim) and tumour core area in the tissue (Ruf et al., 2023).

Tissue Staining with CODEX Imaging Technology

The detailed process of tissue staining is explained by Ruf et al. (2023). Briefly, a mix of commercially available and 21 newly developed antibodies were used for CODEX imaging. The focus of the newly developed antibodies was on (lymphoid) immune cells in the human liver. The staining process was performed by the recommended protocol of Akoya Biosciences. The images were transferred to

the CODEX processor software (version 1.7.0.6) and background corrected, a deconvolution and best focus selection were computed. In the next step, HALO was used to merge the data into a single file. The intensity of each marker was adjusted and bad-quality areas were excluded. Tissue regions (tumour, rim, adjacent liver) were manually identified. Further, a cell segmentation using the watershed algorithm was implemented in HALO. The cells were annotated by manual gating. Based on the cell segmentation and the spatial feature table, a CSV file containing the marker intensity and single-cell data was generated.

Unsupervised Clustering Analysis

The unsupervised clustering analysis was performed by Ruf et al. (2023). The cell-segmented data was normalised using arcsinh transformation. The cofactor was computed for every patient and marker. To account for the batch effect, Harmony (Korsunsky et al., 2019) was applied for the patient label. To scale the expression level to an interval of [0, 1], a MinMax-transformation was applied. The clustering algorithm Leiden was used based on the protein marker. The kNN-graph was built using $k = 30$ and the cluster size was determined based on the expected cell types. An over-clustering was intended, by using a resolution=2.2 to achieve a clear cluster separation of the various cell types. The cells were manually annotated by experts and the same annotated clusters are mapped together. To achieve a more specific identification of the immune cells, clustering was performed only based on CD45⁺ cells.

Spatial Cell Subset Enrichment Detection with S³-CIMA

The dataset provided by Ruf et al. (2023) are segmented, preprocessed and annotated. These data were further used for enrichment detection with S³-CIMA (Babaei et al., 2023). The provided data contains the cell intensity for each marker, the annotated cell type (phenotype) and the defined tissue (rim, adjacent liver and tumour) as labels. First, the input for S³-CIMA was generated by computing the nearest neighbors around an anchor with the Euclidean distance. The resulting multi-cellular input contains the intensity of the k nearest neighbour around the chosen anchor cell, the corresponding tissue and the phenotype. Regulatory T cells, M2 Macrophages PD-L1+ and M2 Macrophages PD-L1- were used as anchors. The generated input files were applied to a weakly supervised classifier S³-CIMA. The CNN trains filter weights. The cells with a high filter response were selected. These selected cells represent the specific anchor cell-niche enrichment and are mapped to the annotated cell types. The number of k nearest neighbors around an anchor and the number of used training and test data can be modified. In order to select the best model, several values for these parameters were tested. The training and validation sets were randomly split and 50 models were computed. Based on the validation accuracy the best model was selected. The test samples were used to assess the computed model. The best results for regulatory T cells are archived with $k = 10$ and nine training sets. For M2

Macrophages PD-L1+ and M2 Macrophages PD-L1- $n = 10$ training sets, $k = 40$ and respectively $k = 10$ nearest neighbour were applied.

The trained filters of the computed model are further used for the downstream analysis. The 5% highest scoring cells were selected and used. For every patient, the relative cell frequency in every tissue type is computed and compared by the Wilcoxon rank-sum test to a randomly selected background. The computation of the relative cell frequency is explained in more detail by Babaei et al. (2023). The computed results of the frequency and the comparison between the tissue sections calculated by the Wilcoxon rank-sum test are represented in a boxplot. To identify specific enriched cell subsets around the anchor an enrichment score (ES) was calculated across all patients for every cell type per tissue type. The ES is computed by the ratio of two factors. One factor represents the cell type frequency with a high filter response around the anchor. The second factor represents the cell type frequency with a high filter response outside of the anchor neighborhood. The detailed calculation of the ES is explained by Ruf et al. (2023). The enriched cell types are visualised in a waterfall and bubble blot.

Chapter 3

Results

The purpose of the used computational approaches was to identify the cell types in the PDAC and HCC microenvironment. Further, a more in-depth analysis to identify cell subsets in the cellular neighborhood which are enriched with respect to a phenotype was performed by S³-CIMA. Therefore, unsupervised approaches were applied to annotate the cell type for every PDAC sample. Due to the poor quality of the data, a more in-depth analysis to find specific cellular networks with S³-CIMA was not possible. The HCC dataset is a high-quality dataset and the annotated data was provided to identify the tissue type associated cellular interaction network within the regulatory T cell and Macrophages cell niche, using S³-CIMA.

Cell Type Annotation

Multiplex high dimensional MACSima™ imaging data provides its unique challenges. Several tools such as Scanpy are developed to analyse other high dimensional data like scRNA-seq data. However, these tools do not overcome the challenges of microscopy-derived data (Hickey et al., 2021). To solve these problems an adapted workflow was developed. First, the processed images were segmented. The CSV files containing numerical intensity values were normalised and unsupervised clustering analysis was performed. For every sample, the clusters were visualised with a UMAP representation. A heatmap displays the median intensity of the marker for each cluster. An over-clustering was achieved to manually annotate the cell types. The annotation is based on the expression level of specifically selected protein markers. The protein markers and the corresponding cell type is listed in Table A.1. For a broad classification of the cells, a tumour and immune cell class were defined by Cytokeratin and CD45.

Establishing the Workflow with HCC dataset

Three HCC datasets were used to test and develop the workflow described in Chapter 2. The resulting expression of the tumour and immune marker, the

UMAP and heatmap were used to classify the cell types. First immune and tumour cells were defined based on their expression. Tumour cells were identified by Cytokeratin, CD99, Ki-67 and CD99. Immune cells are overall defined by the expression of CD45. In all ROIs the expression of these two markers are on different cells and the respective cells could be separated into immune and tumour cells. T cells, plasma cells, endothelial cells, macrophages, fibroblasts and collagen were identified for every ROI. An overview of the resulting heatmaps and the corresponding annotation of all ROIs is displayed in Figure A.1. Overall 12 different cell types are defined for ROI14, where the majority of cells are classified as plasma cells. One-fifth of the cells were defined as T cell subtypes. Also, Natural killer (NK) cells could be identified by the expression of CD56. The majority of identified cells in ROI15 are tumour cells. In ROI16 half of the cells are classified as tumour cells. The next most present cell types are T cells with overall 789 cells and dendritic cells (414). Comparing ROI15 to ROI14 and ROI16, a high number of macrophages (481 of 4556) were identified, whereas in ROI14 overall only 76 and in ROI16 20 macrophages were found.

Annotation of Over-Corrected Merged Data

As described in Chapter 2 (Cell Type Annotation for MACSima imaging datasets) the eight PDAC samples were merged, normalised and clustered using Leiden (Blondel et al., 2008; Traag et al., 2019). UMAPs for the samples and the clusters were generated. The patient-based UMAP shows a grouping of the cells, indicating a batch effect based on the patients. By comparing the patient-based UMAP (Figure 3.1a) with the UMAP of the clusters (Figure A.2a), cell clusters appear in similar areas as the samples, showing a batch effect of the merged dataset. The batch effect was corrected using the Harmony algorithm as described in Chapter 2 (Cell Type Annotation for MACSima imaging datasets). The results of the samples before and after the correction are displayed in Figure 3.1. The UMAP based on the computed clusters after the batch correction is a cloud with no specific separated clusters and the majority of clusters overlap (Figure 3.1b).

To annotate the clusters based on the expression of the protein marker, a heatmap based on the median intensity per cluster was generated and is represented in Figure A.3. The heatmap of the batch-corrected data does not show an interpretable pattern of protein marker expression. Some clusters are overall not or weakly expressed, whereas a few protein markers show a high expression for all clusters. The data seems to be over-corrected by the batch correction algorithm. An analysis based on the over-corrected data is not possible, therefore every single sample was analysed.

In summary, the merged data shows a batch effect, which was corrected using Harmony. The resulting data appears to be over-corrected and could not further be used for the cell type annotation.

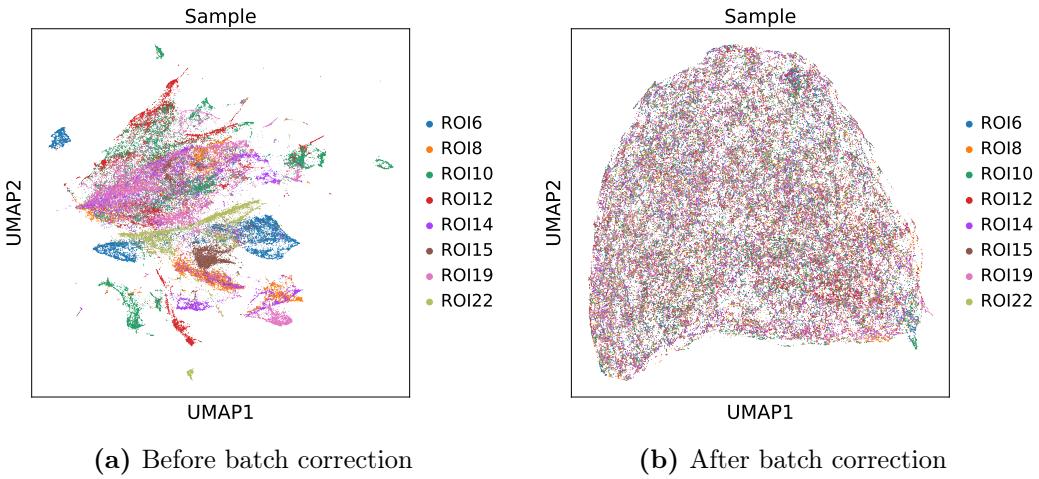


Figure 3.1: Merged pancreas carcinoma samples UMAP based on patients. A) UMAP of the samples displays a batch effect. The cells of one sample are clustering in the same areas. B) UMAP of samples after the batch correction with Harmony shows a distribution of the cells.

Cell Annotation with Overlapping Expressed Tumour and Immune Marker

Every single sample was processed as described in Chapter 2 (Cell Type Annotation for MACSima imaging datasets) and annotated based on the computed UMAP and heatmap. In ROI6 a total number of 9799 cells were identified, but for more than 10% of these cells, no cell type could be annotated. The annotation based on the heatmap and UMAP did not show a clear expression pattern of the protein marker. To differentiate the immune and tumour clusters, the expression of CD45 and Cytokeratin was visualised in a UMAP (Figure A.4a). As can be seen from the figure, both markers are expressed for the same cells, therefore no reliable annotation for the immune and tumour cells can be made. The density of CD45 and cytokeratin is normally distributed. In order to achieve a bimodal distribution of the marker an optimal co-factor for every marker was computed. However, the optimised density distribution does not show a bimodal distribution for CD45 or cytokeratin (Figure A.5). The overlapping expression of different protein markers was not only observed for CD45 and cytokeratin, but also for example in regulatory T cell marker CD152 CTLA4 and markers for dendritic cells (CD11c and CD123). The resulting annotation based on the marker expression in the heatmap is displayed in Figure A.6a. The annotation was also transferred to the generated UMAP shown in Figure A.7a. T cells are not only highly expressed for the typical protein marker (listed in Table A.1), but also for example in CD38, representing plasma cells or CD105 representing endothelial cells. Clusters with an expression for a single marker such as megakaryocyte, endothelial or granulocyte can be identified by this marker, but they represent only a small number with 68, 51, and 254 cells in the sample. The majority of cells are annotated as den-

dritic cells, nonetheless also non-distinct protein markers are highly expressed for this cluster. Subclustering the mixed expressed clusters did not result in an improved expression pattern.

Challenges, such as the overlapping expression of CD45 and Cytokeratin, as in ROI6 are also observed in ROI8, ROI14, ROI15, ROI19 and ROI22. The expression for these ROIs is shown in Figure A.4. For these ROIs no or only a small number of macrophages could be identified based on the expression pattern. In ROI8 the expression of immune and tumour markers is displayed in Figure A.4b. Also, other protein markers are overlapping such as CD33, CD117 and CD11c representing granulocytes, mast cells and dendritic cells. NK cells and also cancer associated fibroblasts (CAF) are more separated in the UMAP (Figure A.7b). Both are also expressed by their corresponding marker (Figure A.6b). However, a total of these two cell types represent only 365 of 10780 cells. Most cells are defined as tumour cells (4375) in the heatmap, although this cluster does show a high expression for cytokeratin but also a lower expression for several other protein markers. In general, all T cell clusters show expression of non-related markers. CD8 and CollagenIII are both highly expressed for the same clusters.

ROI14 also shows a high expression of tumour and immune cell markers on the same cells (Figure A.4c). The tumour cell cluster in the heatmap (Figure A.6c), is highly expressed for several protein markers. In addition, no separation of the tumour cells is displayed in the UMAP (Figure A.7c). Overall the UMAP has no separated distinct clusters. CollagenIII and CD8 protein markers are highly expressed for the same cells. Especially, Collagen and effector T cell clusters are highly expressed for these two protein markers. To classify T cells they need to be expressed for CD45 and also CD3, but they are never highly expressed in the same clusters. Tumour and collagen protein markers are highly expressed for the tumour cell cluster.

The overlapping expression pattern for CD45 and cytokeratin of ROI15 is displayed in Figure A.4d. Additionally, the UMAP (Figure A.7d) does not show a separation of a specific cluster. The heatmap displays also an expression of mixed protein markers (Figure A.6d). Plasma cells are highly expressed for CD38 but do show no expression for other general immune cell markers. The cluster for mast cells and dendritic cells are expressed for similar markers, both clusters are highly expressed for CD117 and CD11c and can therefore not be separated. Also in the UMAP, these two cell-type clusters are present in the same area. CD4+ T cells and also effector T cells show a mixed expression pattern and are not expressed for CD45. The granulocyte cluster is more separated in the UMAP and is also expressed by its corresponding marker.

In ROI19 187 cells are defined as unknown and overall 10770 cells are analysed. The majority area of CD45 expressed cells show also an overlapping expression of cytokeratin (Figure A.4e). 2202 tumour cells are identified, but

the cluster in the heatmap (Figure A.6e) is showing also an expression of other protein markers. Most of the defined CAF and tumour cells are separated in the UMAP as displayed in Figure A.7e. However, no other defined cluster shows a more distinctive separation in the UMAP. This is also reflected in the heatmap. The defined endothelial and CD4+ T cell clusters are expressed for the majority of the protein marker. Plasma cells are, as described above only expressed for their corresponding marker. Collagen and CD8 markers are both highly expressed for the same cells. Plasma and NK cells are expressed with their corresponding marker, but no general immune cell marker such as CD45 is expressed for these cells.

In ROI22 overall eight cell types were identified, where collagen represents nearly a third of the cells. Collagen is as described for other ROIs expressed at the same time as CD8 (Figure A.6f). However, for the defined collagen cluster, no other protein markers are expressed. Plasma cells and M1 macrophages are only expressed by their corresponding protein marker. CD8+ T cells and general T cells are clustered close by in the UMAP (Figure A.7f). The expression pattern differs for these clusters. Overall all T cell clusters show expression of several not distinctive protein markers. CD45 and cytokeratin show an overlapping expression for the same cells (Figure A.4f).

To summarise the mentioned ROIs of the pancreas carcinoma data can not be reliably annotated. In some samples, the protein marker expression was distinctive and did not overlap with another marker expression. However, most of the ROIs did show an overlapping expression. This results in a not interpretable UMAP and heatmap. The UMAP did not separate the cluster visually, therefore it could rarely be used for the annotation.

Cell Annotation with Different Expressed Tumour and Immune Marker

In ROI12, immune and tumour markers are expressed on different cells. The expression of these markers is displayed in Figure A.8b. Based on the separated expression of CD45 and Cytokeratin, a more detailed annotation for this sample was possible. Some clusters showed a high expression of different markers, these are further subclustered. Every subcluster was annotated based on the protein marker and then included in the annotated heatmap. The annotation for ROI12 is displayed in Figure 3.2. The heatmap shows the characteristic protein expressions of different clusters. Fourteen distinct cell types were identified, the majority with 2526 cells were distinguished as collagen. A total of 1333 tumour cells were identified by cytokeratin, CD99, HNF-4alpha and Ki-67. A cluster of CAF was distinguished by Fibronectin and Galectin9. However, also other protein markers such as collagen were highly expressed in the same cells. Twice as many cells of fibroblasts (SM Actin, Myosin SM) than CAF were found. Different lymphoid progenitor cells were identified, such as T cells which are first distinguished by the expression of CD45 and CD3. Highly-expressed cells of CD45 and CD3 were further divided into the T cell subtypes. In to-

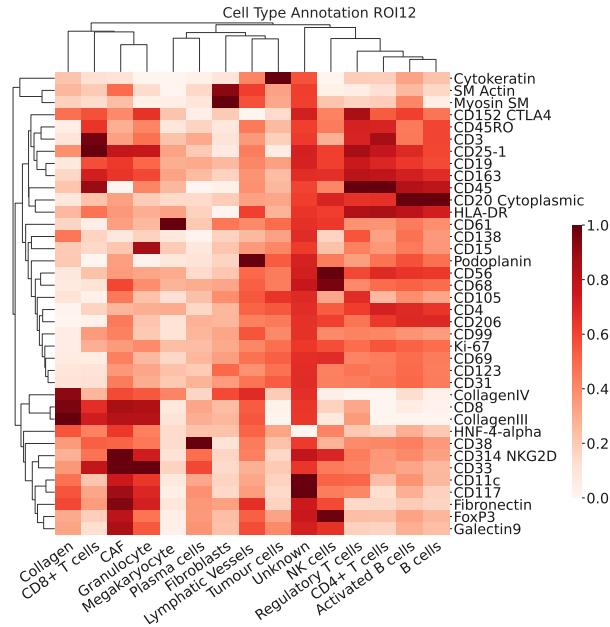


Figure 3.2: Median expression of every marker for every cluster in ROI12. Tumour cells are classified by a high expression of cytokeratin. All immune cells are classified by CD45 expression and further separated by their specific protein marker.

tal 1538 T cells are defined. B cells and activated B cells are clustered at the same area in the UMAP (Figure A.9b). In addition, their expression pattern is similar and therefore challenging to separate. Plasma cells are identified by a high expression of CD38. However, to identify plasma cells a co-expression with CD45 is expected. For every protein marker, the density distribution was plotted. The majority marker of ROI12 is normally distributed after the normalisation step. CD45, CD8, CD25-1, cytokeratin, CollagenIII and CollagenIV are as expected bimodal distributed. The UMAP in ROI12 does not distinctly separate the clusters.

ROI10 shows for the majority of the cells an expression of CD45 and cytokeratin on different cells (Figure A.8a). In the heatmap tumour cells are highly expressed for cytokeratin and CD99 (Figure A.10). A total of 2271 from 17249 cells are annotated as tumour cells. Collagen is with 6774 cells the most annotated cell type in this sample. The protein marker CollagenIII and CollagenIV is mainly expressed for the same cells and do not overlap with other protein markers. Activated T cells, CD8+ T cells and effector T cells do show a similar expression pattern in the heatmap. They also cluster in the same area in the UMAP (Figure A.9a), therefore it is challenging to separate these cell types from each other. 527 B cells are identified, in the heatmap and also in the UMAP these cells are distinctively separated by the other protein markers. CAF, endothelial and plasma cells show an expression in the same area on the

<i>Regulatory T cells</i>				<i>M2 positive</i>				<i>M2 negative</i>			
n_train	k	Ac.	Training	n_train	k	Ac.	Training	n_train	k	Ac.	Training
10	10	0.60	0.63	10	10	0.58	0.70	10	10	0.59	0.57
	20	0.58	0.58		20	0.59	0.72		20	0.58	0.55
	30	0.56	0.56		30	0.54	0.70		30	0.58	0.54
	40	0.54	0.54		40	0.67	0.73		40	0.58	0.54
	10	0.63	0.62		10	0.59	0.69		10	0.61	0.54
9	20	0.61	0.58	9	20	0.57	0.63	9	20	0.60	0.53
	30	0.59	0.55		30	0.58	0.65		30	0.59	0.53
	40	0.57	0.54		40	0.59	0.65		40	0.59	0.53
	10	0.59	0.59		10	0.65	0.58		10	0.56	0.54
8	20	0.57	0.55	8	20	0.66	0.53	8	20	0.56	0.53
	30	0.55	0.53		30	0.66	0.54		30	0.55	0.51
	40	0.55	0.53		40	0.70	0.55		40	0.54	0.52

Table 3.1: Overview of the tested parameter for k nearest neighbour and the number of used training data. The resulting best accuracy score for the training and the overall accuracy is documented. The highlighted values did result in the highest accuracy score and were further used for the downstream analysis.

UMAP. In addition, endothelial and CAF do not show a distinctive expression pattern. Plasma cells are highly expressed for CD38, but no other immune cell marker is expressed in this cluster. Fibroblasts are more separated in the UMAP and are highly expressed for SM Actin and Myosin SM.

To summarise, ROI12 and ROI10 are annotated based on different expressions of CD45 and cytokeratin. Even though the tumour cells are annotated, a more detailed annotation of immune and non-immune cells is more challenging. The overlapping expression of different protein markers and the crowded UMAP results in an unreliable cell annotation.

Detecting enriched T cells with S³-CIMA

PDAC and HCC are highly heterogeneous cancer types, with a poorly understood cellular interaction network. S³-CIMA can be used to identify enriched neighbor cells. Due to the poor quality of the PDAC data applying further analysing approaches would not be reasonable, since the results could not be confidential interpreted. However, the HCC data did show a high-quality, also a reliable annotation was provided. In order to identify cell subsets in the cellular neighborhood of 15 HCC patients S³-CIMA was applied. Several immune cell types in the TIME were identified, where especially T cells and macrophages are of interest as they are able to influence the microenvironment. The annotation of every single cell allows a description of their spatial position in the tumour. However, the spatial interaction between specific cells of interest and other cell types can not be captured. Babaei et al. (2023) developed a supervised learning method called S³-CIMA to analyse the higher-order properties of cell interactions within the tumour microenvironment. S³-CIMA was applied to HCC dataset. In order to find the best model, the several values of the k nearest neighbors and training set (n) were tested. The results of the tested parameters are listed in Table 3.1. The best accuracy for validation and test

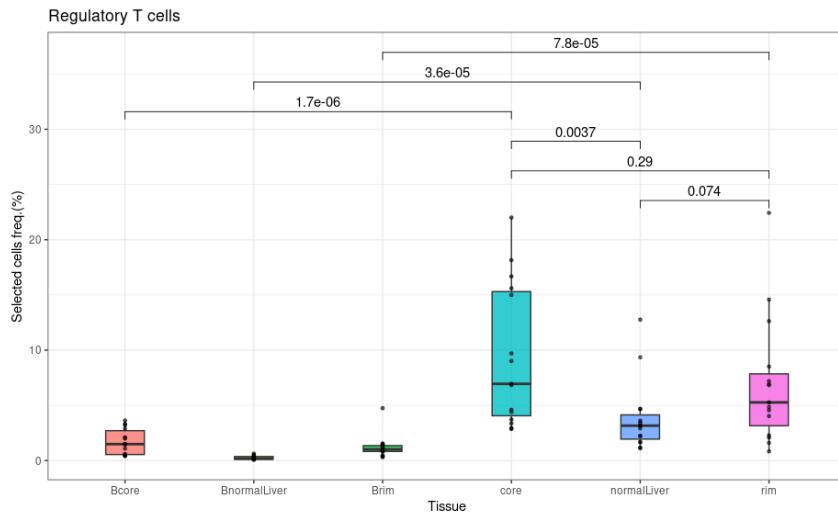


Figure 3.3: Frequency of the neighbour around the regulatory T cells for each patient and each tissue section. Comparison between the samples is computed by the Wilcoxon rank test ($p < 0.05$). The frequency of cells with a high filter response is significantly higher in the tumour core than in the normal liver ($p=0.0037$).

set is 0.63 respectively 0.62 for regulatory T cells using $k = 10$ and nine train sets. An accuracy of 0.73 for M2 Macrophages PD-L1+ ($n = 10$, $k = 40$) and 0.57 for M2 Macrophages PD-L1- ($n = 10$, $k = 10$) was achieved.

The application of S³-CIMA enables the detection of cell populations that are unique to the tissue in the anchor cell niche. In order to characterise the specific cell niche, the learned filters are used to select a specific cell population in the TIME. Analysing the cell frequency in the different areas of the tissue shows an overall significantly smaller frequency for the background of regulatory T cells. A significantly higher frequency is observed for regulatory T cell neighbors in the tumour core than in the normal liver ($p=0.0037$, Figure 3.3). In other words, more cells around the anchor with a high filter response were observed in the tumour core compared to the normal liver. In order to specify which cell types are more enriched an ES is computed, where cells with a score ≥ 1 are enriched. Enrichment of tumour, endothelial, and MHCII cells within the regulatory T cell niche for the tumour core was observed (Figure 3.4). Regular T cells show the lowest enrichment in the tumour core overall.

For the entire tissue, an enrichment of liver sinusoidal endothelial, immune, and endothelial cells was observed (Figure A.11a). The enrichment of the cells is also visualised in a bubble plot, where the tumour cells do show a positive enrichment for the tumour core but not for the normal tissue or the rim area (Figure A.11b).

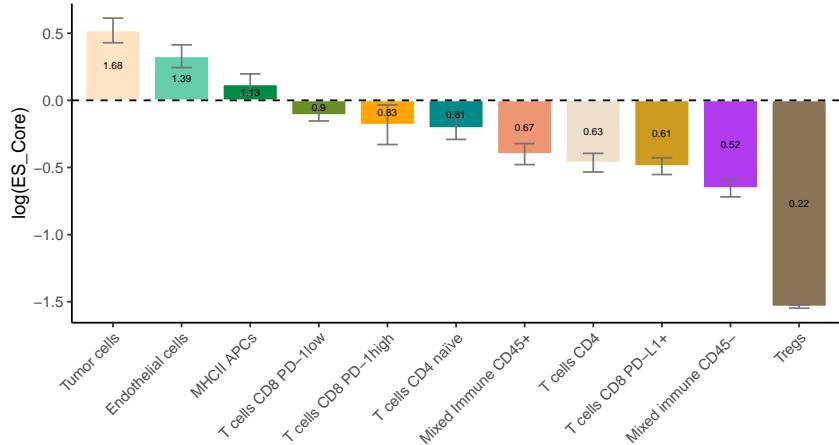


Figure 3.4: Waterfall plot displaying the enrichment score of the cell population of the regular T cell neighborhood. An enrichment score > 1 indicated an enrichment.

Computing the frequency of the filter response for M2 Macrophages PD-L1+ did show an overall significant difference between the randomly computed background and the frequency of the tissue areas. The filter response frequency of the neighborhood cells in the core is significantly higher than in the normal liver ($p=3.1e^{-5}$, Figure A.12a). Enrichment of T cells CD57 is observed as neighborhood cells of M2 Macrophages PD-L1+. Mixed immune CD45- and B cells CD38+ do also show an ES larger than one, but the error bar is reaching into the negative area (Figure A.13a). The entire tissue does show enrichment of endothelial, T cells, tumour and mixed immune cells, also MHCII, and NK cells (Figure A.13b). T cells CD57 are enriched in the tumour core and in the normal liver. The ES is higher in the tumour core than in the normal liver. Mixed immune CD45- cells are also enriched for tumour core and liver tissue. They have a higher enrichment in the normal liver than in the tumour core (Figure 3.5). The normal liver tissue and the rim are sharing multiple enriched cell types such as T cells, endothelial cells, and mixed immune and MHCII cells (Figure A.14a). The tumour core and the rim area are enriched for T cells, mixed immune and B cells (Figure A.14b).

M2 Macrophages PD-L1- do show an overall higher cell frequency in the background. Also, the frequency for M2 Macrophages PD-L1- with a model using $k = 20$ was computed and does also show no significant difference between the randomly selected background and the tissue areas (Figure A.12).

To summarise, enriched tumour cells were identified around regulatory T cells in the tumour core. In addition, T cells CD57 were identified as enriched around M2 Macrophages PD-L1+ cells in the tumour core. The enrichment of the M2 Macrophages PD-L1- neighbors could not be identified.

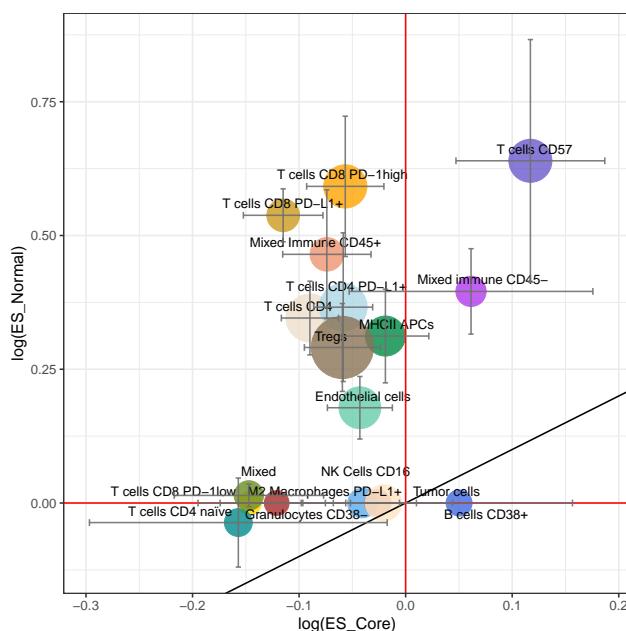


Figure 3.5: Enriched cells of M2 Macrophages PD-L1+ cell neighbors for the tumour core and the normal liver tissue. T cell CD57 and mixed immune CD45- cells are showing enrichment for both tissues.

Chapter 4

Discussion and Outlook

An in-depth characterisation of cell types based on high-dimensional proteomic imaging data is discussed in this thesis. The immune microenvironment of PDAC and HCC is not well known. The attempt in this thesis was to gain a better underlying biological understanding of these two cancer types, using unsupervised approaches to annotate the cell types and further using S³-CIMA to identify cell specific enrichment.

A workflow, based on approaches discussed by Hickey et al. (2021) was developed, to analyse a multicellular imaging dataset. To establish the workflow, three HCC samples were analysed. High-dimension proteomic imaging data have their specific noise, which needs to be addressed. To normalise these data arcsinh and z-score were applied, because they do seem to address this specific noise the best (Hickey et al., 2021). Leiden was used to cluster the data. Hickey et al. (2021) was using a combination of z or arcsinh normalisation with Leiden did result in the most accurate assigned cell types and also eliminated noise correctly.

The eight PDAC samples were merged and batch-corrected with Harmony (Korsunsky et al., 2019). Harmony was successfully applied to other multiplex imaging datasets such as Ruf et al. (2023). Comparing the batch-corrected results of the UMAP based on the samples, the data did correct the batch effect. However, the resulting UMAP based on the computed cluster displays a cloud of clusters close to each other. Cells are highly overlapping in some areas, which might indicate an over-correction of the data. Exploring the corresponding heatmap, no reasonable biological pattern could be identified. Based on the expression pattern of the computed heatmap no tumour cells could be identified, whereas collagen is highly expressed for all clusters. A reason for this uninterpretable result could be an over-correction of the data. However, applying the workflow for every sample did also result in conflicting results. Protein marker representing different cell types were expressed for the same cell. No reliable annotation is possible with these contradictory pieces of infor-

mation. To account for this problem the cofactor for the arcsinh normalisation was computed to achieve a bimodal instead of a monomodal distribution of the protein marker. However, the density distribution of the optimised cofactor did not result in the expected binomial distribution. No quality improvement could be achieved by optimising the cofactor.

An explanation of the failed attempts to computationally correct the common noise of the data could be the overall poor quality of the samples. The staining of the tissue is highly influenced by the resection of the tumour. In addition, the size of the tissue which was removed is patient-specific and highly variable. Especially the time frame in which the tissue was frozen impacts the quality of the sample. The assumption of poor-quality data is supported by comparing the HCC samples with the ROIs of PDAC. For both cancer samples, the same antibody panel and the same workflow in the laboratory was applied. The protein marker of HCC did mostly show a bimodal distribution and rarely overlapping protein marker expression. Additionally, comparing the ROIs within the PDAC samples ROI10 and ROI12 do show a separate expression of immune and tumour marker. Both ROIs are selected from the same patient, showing a slightly improved quality compared to the other PDAC samples. ROI6, ROI8, ROI14, ROI15, ROI19 and ROI22 could rarely be annotated and implying that the results are not trustworthy. Another factor influencing the work in the laboratory could be the different tissue structures of the tumour. The liver and pancreas are two different organs with completely different tumour development. In addition, the pancreas naturally contains multiple different enzymes, which are distributed when the tissue is cut. These enzymes will start to digest the tissue the moment they are released into it. An additional factor, which could result in a bad quality of the measured protein expression, is the antibody panel. In theory, it is possible to use a developed panel for different tumour types. But different antibody clones are differently well suited for the tissue staining. Therefore, the developed antibody panel may be not appropriate for PDAC samples.

The multiple expression of protein markers for one cell results in an untrustworthy cell type annotation, for every single cell. Overall tumour cells could be identified by cytokeratin for every sample. Also, CAFs are distinguished by Fibronectin and Galectin9. Ligorio et al. (2019) did demonstrate a relationship between CAF and the cancer heterogeneity. In addition, a connection of CAFs with epithelial-to-mesenchymal transition (EMT), mitogen-activated protein kinase (MAPK) and signal transducer and activator of transcription 3 (STAT3) were found. In other words CAFs are able to change the gene expression of cells and therefore change the cell activity. They also change the functionality of the cells and increase the metastasis activity. The present samples did overall show only a small number of macrophages. This could be due to the bad quality of the samples or the chosen antibodies. However, a high frequency of macrophages in PDAC patients was observed by Zhu et al. (2017).

Macrophages and monocytes appear to be the precursor of tumour-associated macrophages, promoting the PDAC progression (Zhu et al., 2017). Overall a high number of collagen was identified in all eight ROIs. PDAC is embedded in a collagen-rich environment, where little nutrients and oxygen is available. In this nutrient-poor environment, Olivares et al. (2017) observed tumour cells absorbing collagen fragments as a nutrition source. This additional energy and oxygen source could promote the survival of PDAC cells. In the present samples, multiple T cell subtypes were identified. T cells are overall important for the immune response. A high infiltration of CD4+, CD8+ and effector T cells can correlate with a significantly higher survival rate for PDAC patients (Carstens et al., 2017).

To summarise, normalisation, clustering and batch correction of the data were applied to the data. However, poor sample quality due to resection and further handling of the tissue could not be corrected by these approaches. Even though the cells could not be identified some relevant cell types such as T cells and CAFs for PDAC were expressed. The cell type composition is important because it influences tumour development and survival time. A more detailed analysis with samples of better quality is needed.

The annotation of all PDAC samples are not reliable and therefore they could not be analysed further. In order to apply S³-CIMA a new annotated HCC dataset was used. Regulatory T cells, M2 Macrophages PD-L1+ and M2 Macrophages PD-L1- were used as anchor cells and analysed in more detail. S³-CIMA computes high responsive filters. The trained filters reflect the molecular profile of the neighborhood surrounding the anchor and are used to find cell populations which identify a niche-specific cell type composition. Cell subsets that are spatially enriched show a specific association with the anchor cell in the corresponding spatial compartment.

The computed frequency regulatory T cells showed a significant difference between the core and the normal liver tissue. In addition, a high enrichment of tumour cells around the regulatory T cells in the core area was identified. Regulatory T cells are responsible to maintain peripheral tolerance. In other words, regulatory T cells do regulate the self-response and avoid an immune overreaction against their own body. However, this immunosuppressive regulation is also able to support tumour growth. The enrichment of tumour cells around the regulatory T cells could be an indication of an accumulation of immunosuppressive cells in the tumour core. Tu et al. (2016) did observe a worse prognosis for patients with high levels of regulatory T cells compared to those with lower levels of regulatory T cells in the tumour tissue. Besides the tumour cells also endothelial cells are enriched around the regulatory T cells in the tumour core. This enrichment can be expected as tumour tissue is mostly well supplied with blood. Matoba et al. (2005) observed a higher intrahepatic

recurrence for patients with a high gene expression encoding MHCII. Enrichment of specific MHCII-presenting cells are also observed for these samples. Regulatory T cells are negatively enriched in the tumour core, meaning a high number of regulatory T cells are identified in the tumour core, but they do not show a high filter response.

In the tumour core, CD57 T cells are highly enriched around the M2 Macrophages PD-L1+. Zhao et al. (2022) did show a prolonged survival time in patients with an increased number of CD4+, CXCR5+ and CD57 T cells in the serum. In addition, a positive expression of CD57 cells was observed by a IHC analysis. However, these two study designs highly differ, therefore more studies using high-dimensional imaging techniques are needed to compare the observed results. For M2 Macrophages PD-L1+ a positive enrichment of mixed immune cells CD45- and B cells CD38+ is computed. However, the variance of these two cell types is high. The ES is computed per patient so that the resulting score is not based on any outlier. The error bars of mixed immune and B cells are reaching into the negative area, therefore for some patients a negative ES was computed. These results can not reliably be interpreted.

M2 Macrophages PD-L1- did show no significant difference in the frequency between the random computed background and the tissue area. Meaning only a low number of cells in the neighborhood of M2 Macrophages PD-L1- with a high filter can be identified. The number of these cells is too low to show any significant difference computed by chance.

Overall, S³-CIMA was able to detect highly enriched neighbor cells for regulatory T cells and also M2 Macrophages PD-L1+ in the tumour core. Tumour cells are higher enriched around regulatory T cells. Regulatory T cells are immunosuppressive and could therefore have an influence on tumour development. A higher enrichment of CD57 T cells was observed around M2 Macrophages PD-L1+. CD57 T cells are associated with a prolonged survival time. However, more comparative studies are needed to compare the observed results.

All in all, high dimensional imaging technologies are promising to further understand the underlying biology of diseases. However, every tumour subtype differs and therefore it is challenging to apply the same laboratory methods to different tumour types. In addition, PDAC tissue is more difficult to work with, as the containing enzymes highly influence the analysing process. More studies are needed to develop a workflow resulting in good quality data for PDAC tissue sections. The process of cell annotation highly depends on the manual knowledge of experts. In addition, this process is highly time-consuming. Therefore new and more reliable tools need to be developed to computationally identify the cell types. In order to define enriched neighbor cells for specific anchors S³-CIMA was successfully applied. However, for further more profound

interpretation and new therapeutic designs, a larger patient cohort is needed. S³-CIMA seems to be a promising tool to find enriched cells around a specifically defined anchor, which allows for identifying cellular interaction networks between the cells. This allows a more profound knowledge about higher-order properties of cellular organisations within the tumour.

Appendix A

Further Tables and Figures

Cell Type	Marker	Cell Type	Marker
T cells	CD45 CD3	NK cells	CD56 CD314 NKG2D
Effector T cells	CD3 CD45RO	M1 Macrophages	CD68
CD4+ T cells	CD3 CD4	M2 Macrophages	CD206 CD163
CD8+ T cells	CD3 CD8	Dendritic cells	CD11c CD123 HLA-DR
Regulatory T cells	CD4 CD3 CD25-1 FoxP3 CD152 CTLA4	Granulocyte	CD15 CD33
Activated T cells	CD3 CD152 CTLA4 CD25-1 CD69	Mast cells	CD117
B cells	CD19 CD20 Cytoplasmic	Tumour cells	Ki-67 HNF-4-alpha Cytokeratin CD99
Plasma cells	CD38 CD138	Endothelial	CD31 CD105
Activated B cells	CD19 CD69	Lymphatic Vessels	Podoplanin
Megakaryocyte	CD61	Fibroblasts	SM Actin Myosin SM
		Collagen	CollagenIII CollagenIV
		Cancer-Associated Fibroblasts (CAF)	Fibronectin Galectin9

Table A.1: Analysed protein marker and their corresponding cell type.

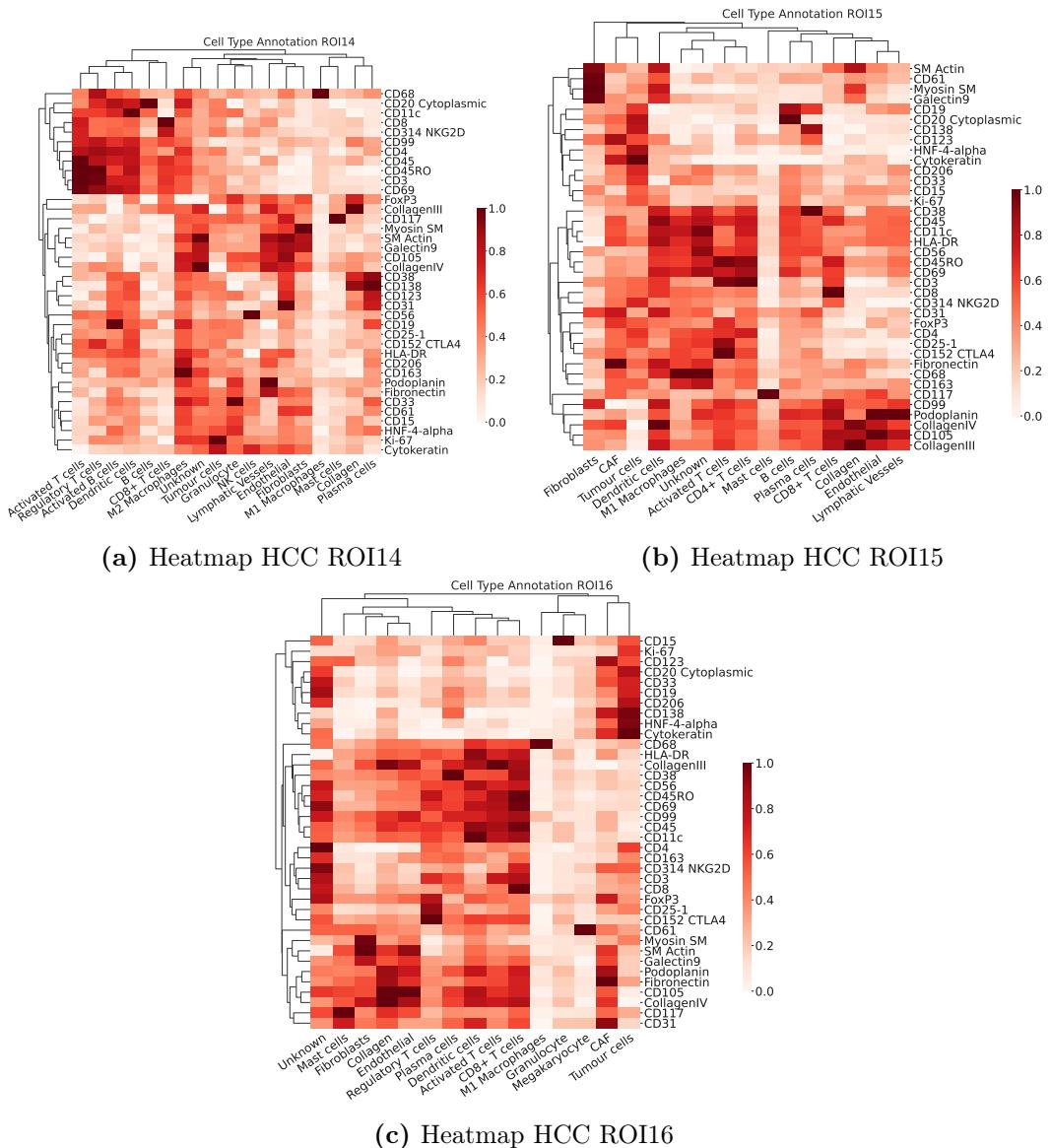


Figure A.1: Resulting heatmaps of the implemented workflow for all HCC samples after normalisation and clustering. Most of the cells displayed an interpretable expression pattern and were annotated based on Table A.1.

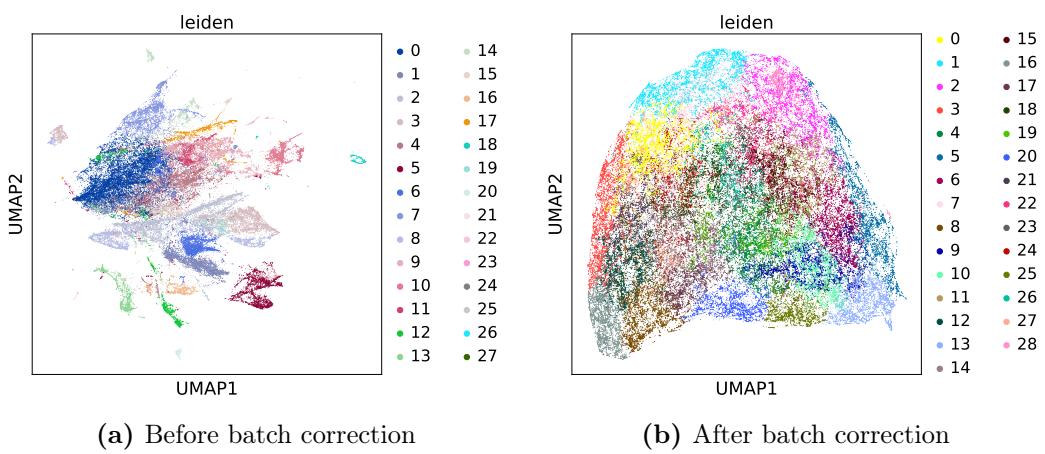


Figure A.2: UMAP of merged pancreas carcinoma samples based on clusters. **a)** UMAP of the clusters before batch correction. **b)** UMAP of the clusters after the batch correction with Harmony shows a cloud with multiple clusters. Some clusters are clearly separated, but a lot of cluster overlap and no distinct cluster can be identified.

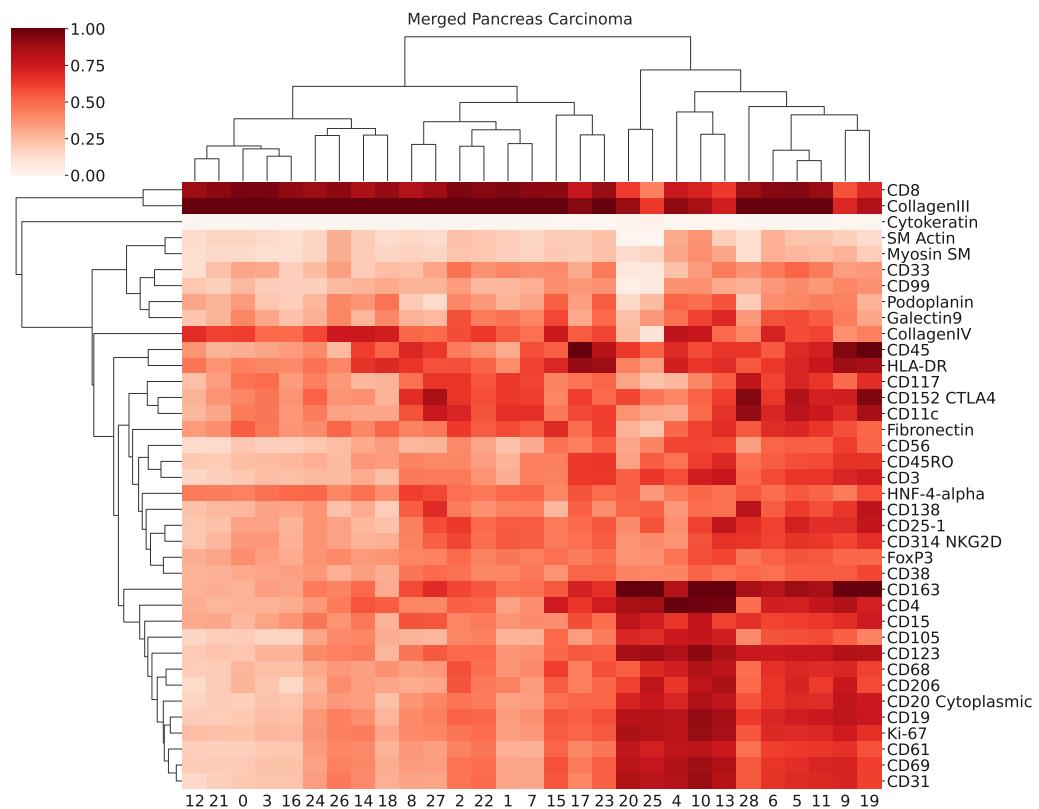


Figure A.3: Merged pancreas carcinoma data displayed in a heatmap with the median intensity of every marker per cluster, after the batch correction using HARMONY. The heatmap, can not be used to annotate the cell types. Some protein marker shows no or weak expression, and others show high expression for all clusters of the same protein markers.

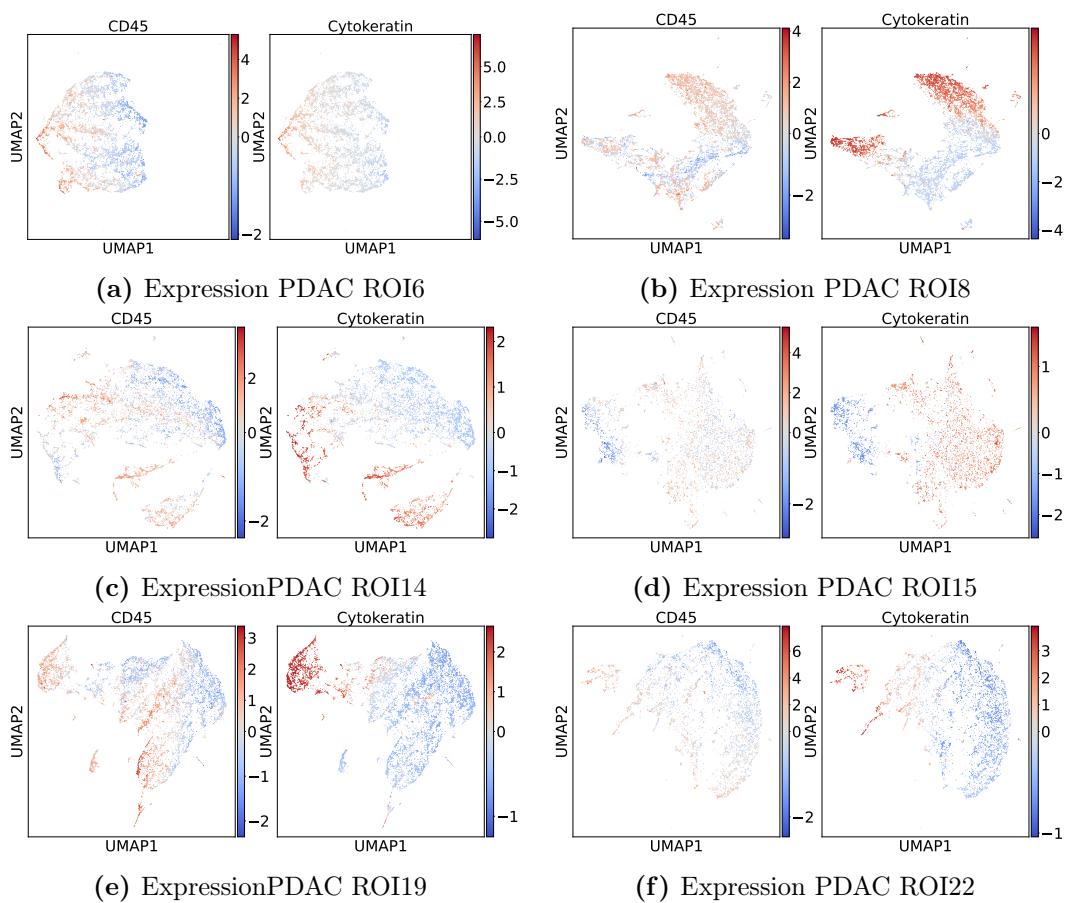


Figure A.4: Overview of CD45 and Cytokeratin expression on all ROIs, who did show an overlapping expression for the same cells.

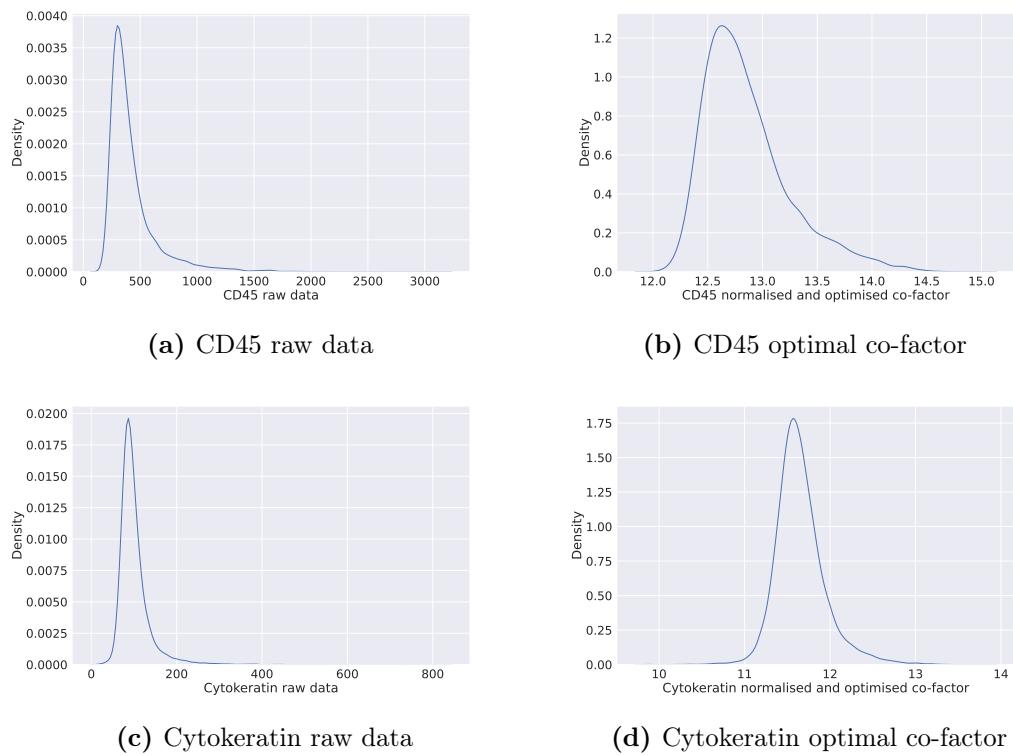


Figure A.5: The density of CD45 and Cytokeratin for ROI6 for the raw data and the optimal computed co-factor. No bimodal distribution was achieved by the co-factor optimisation.

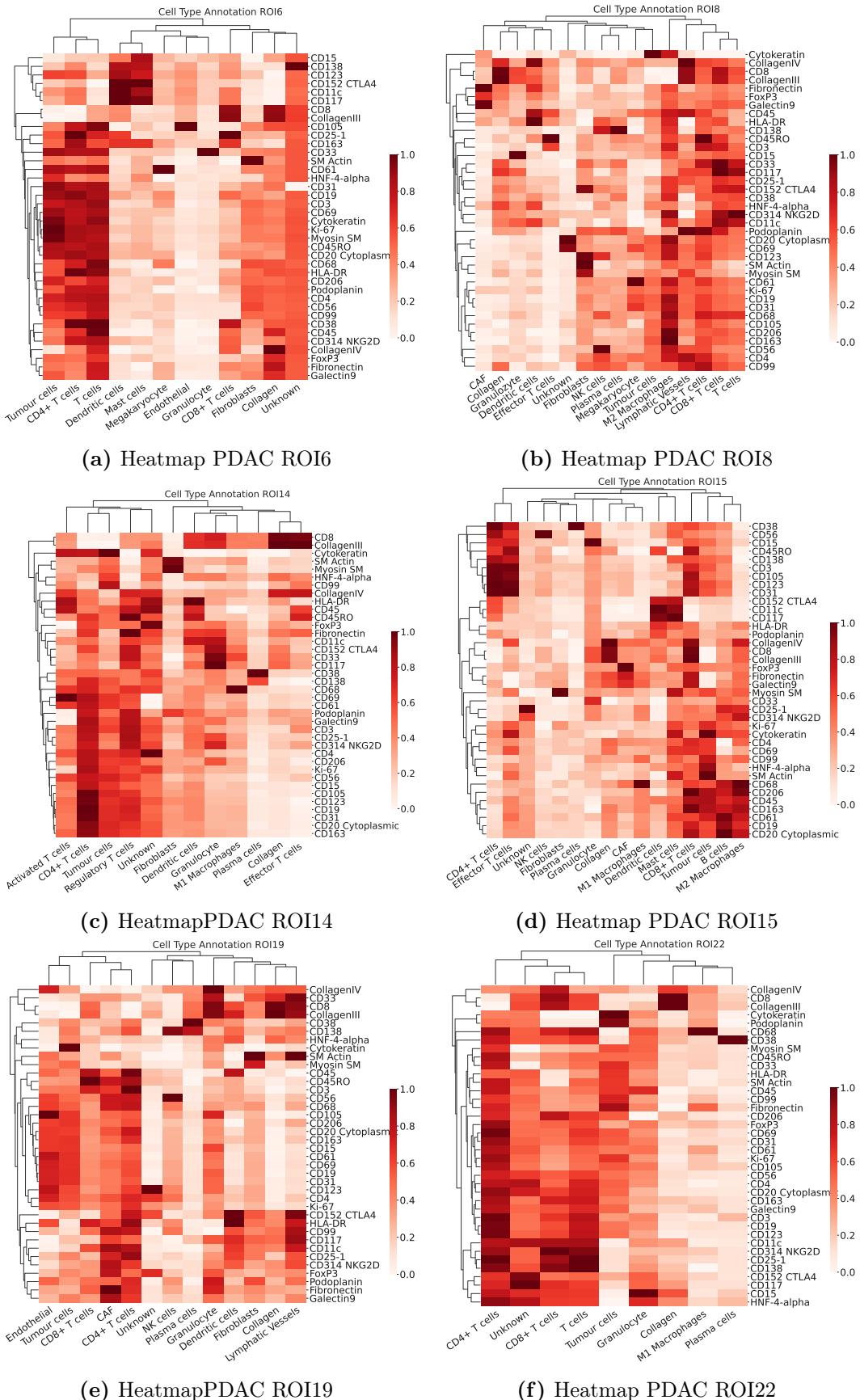


Figure A.6: Median expression of all PDAC samples with an overlapping expression of CD45 and Cytokeratin. No differentiation between immune and tumour cells was possible for these samples and therefore resulting in an unreliable annotation.

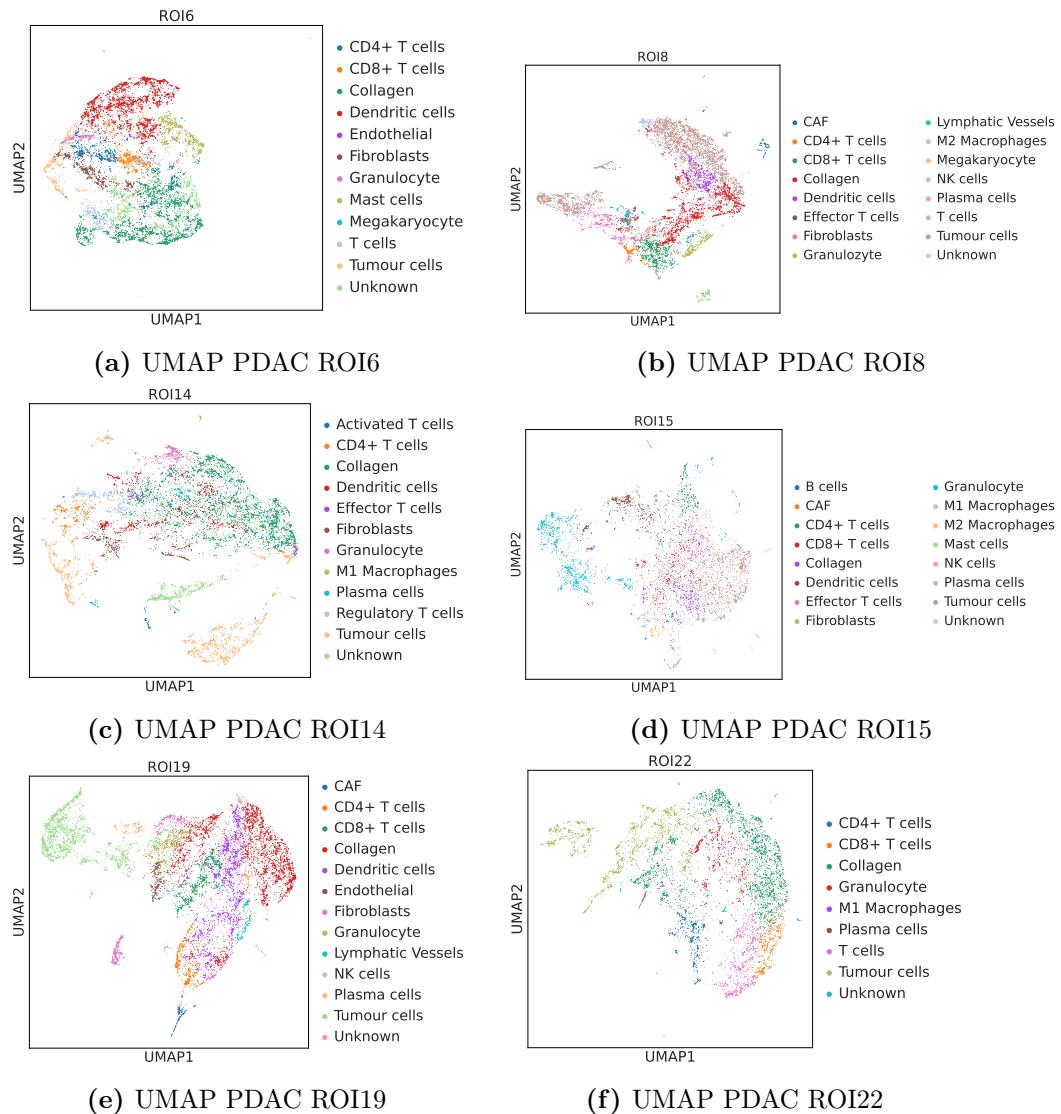


Figure A.7: UMAP of all PDAC samples with an overlapping expression of CD45 and Cytokeratin.

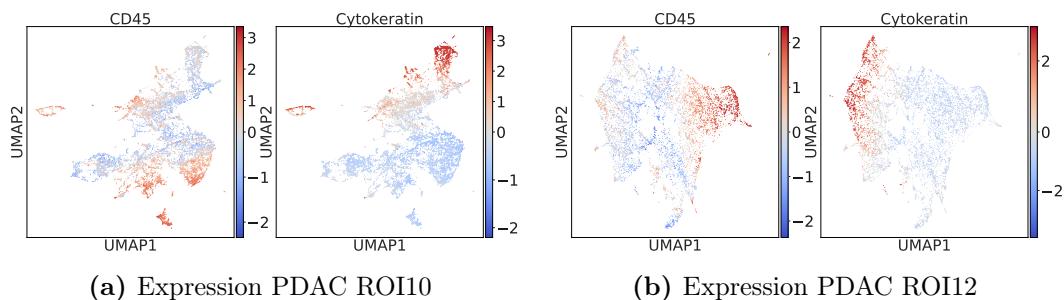


Figure A.8: Expression of all PDAC samples with no overlapping expression of CD45 and Cytokeratin.

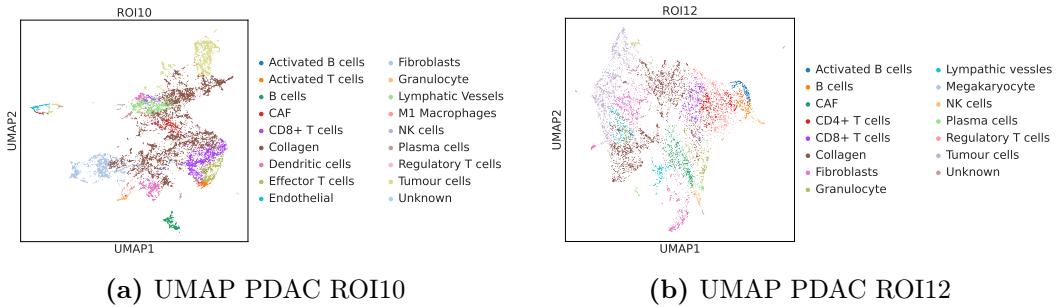
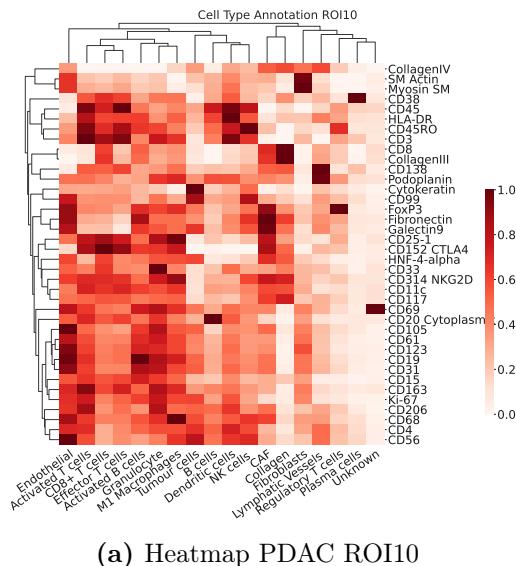


Figure A.9: UMAP of all PDAC samples with no overlapping expression of CD45 and Cytokeratin.



(a) Heatmap PDAC ROI10

Figure A.10: Median expression of ROI10 PDAC sample with no overlapping expression of CD45 and Cytokeratin. The cells could be differentiated by these markers, however several other expression marker are overlapping resulting in an unreliable annotation.

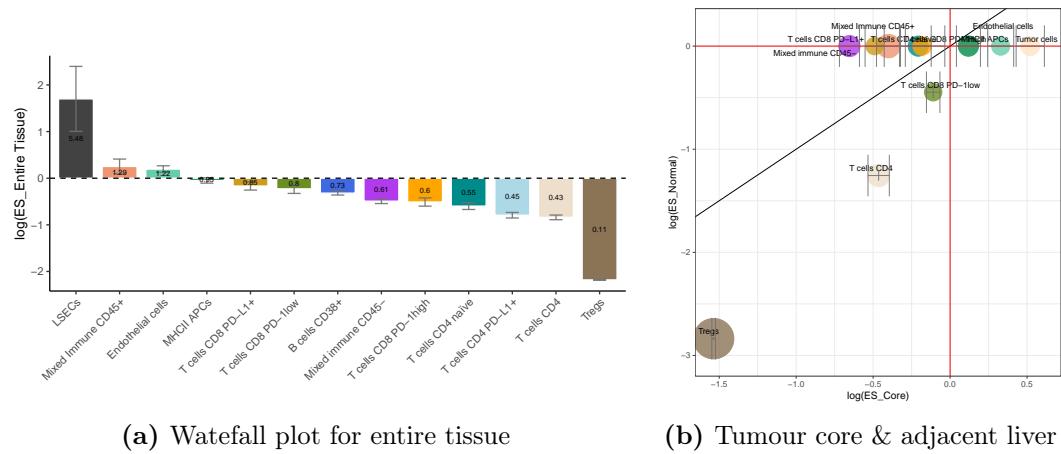


Figure A.11: Enriched cells of the regulatory T cell neighborhood. A) Waterfall plot displaying the enrichment score of the cell population of the regular T cell neighborhood for the entire tissue. An enrichment score > 1 indicated an enrichment. B) Bubble plot for regulatory T cells of the enriched cell types comparing tumour core and normal liver tissue.

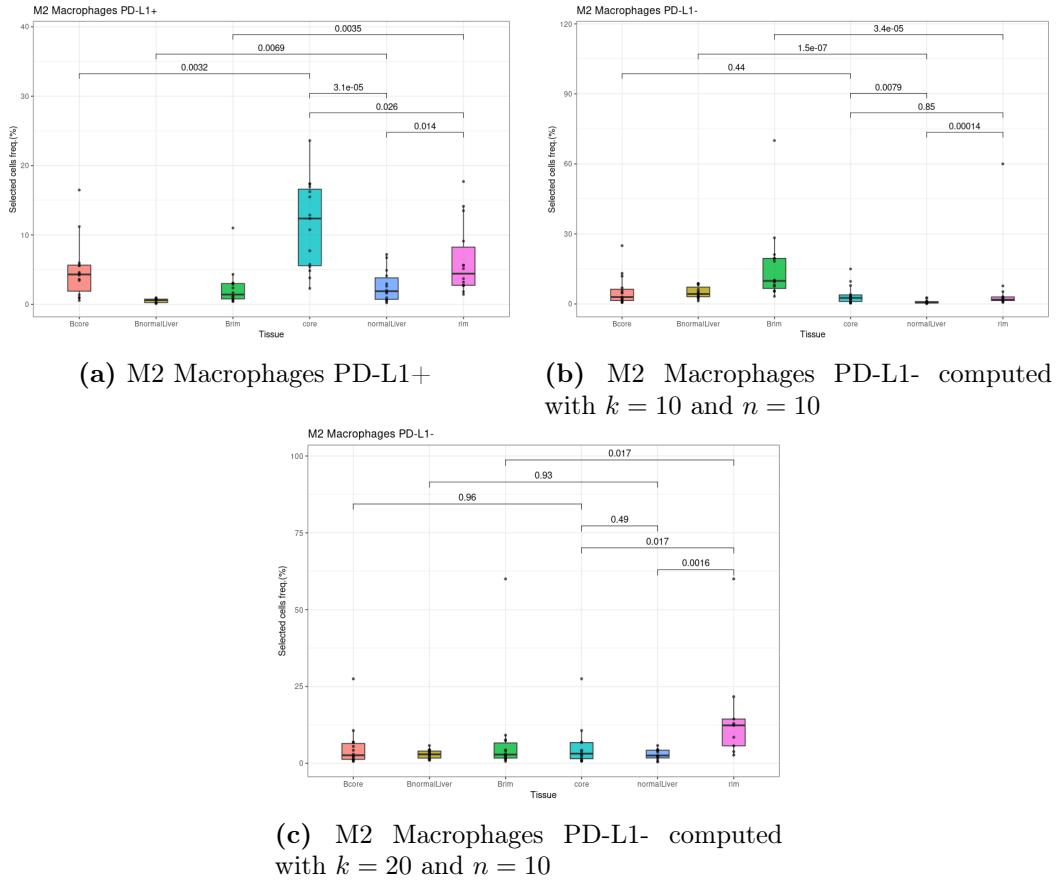


Figure A.12: The boxplots represent the frequency of high filter response for all neighbour cells in the three different tissue types. The cell frequency of M2 Macrophages PD-L1+ and M2 Macrophages PD-L1- anchor are shown. M2 Macrophages PD-L1+ have a significant difference between the background and the tissue. A significant difference between the tumour core and the normal liver is computed. M2 Macrophages PD-L1- does show no significance for the computed random background and the tissue areas. Using a different number of neighbours did not change the results.

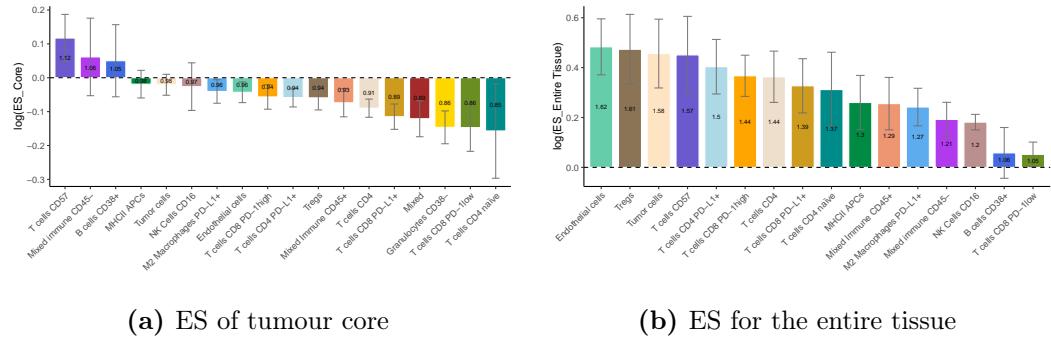


Figure A.13: Waterfall plot displaying the enrichment score of the cell population of the M2 Macrophages PD-L1+ cell neighborhood for the entire tissue. An enrichment score > 1 indicated an enrichment.

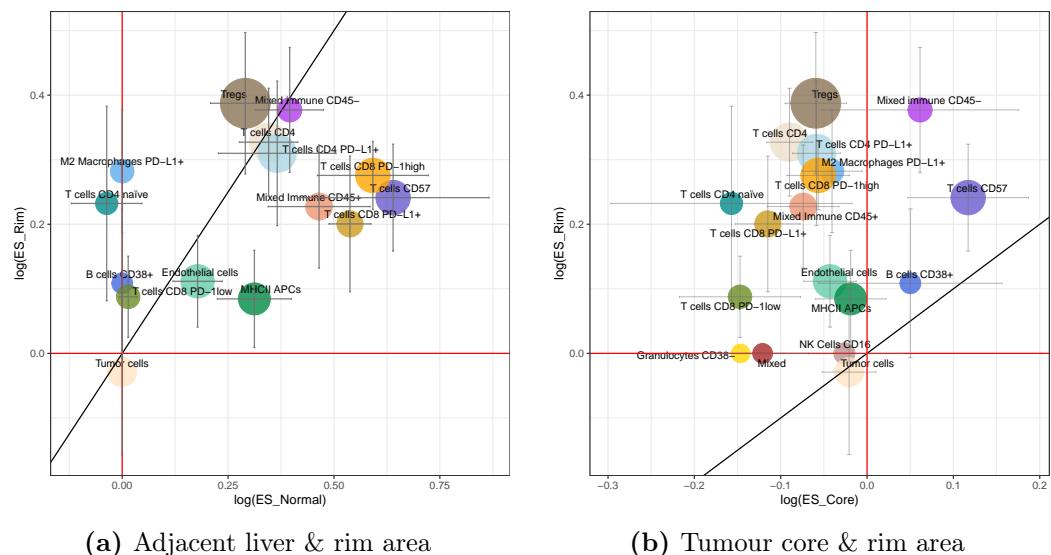


Figure A.14: Bubble plot for M2 Macrophages PD-L1+ cells of the enriched cell types for the compared tissue area.

Bibliography

- Akinyemiju, T., Abera, S., Ahmed, M., Alam, N., Alemayohu, M. A., Allen, C., Al-Raddadi, R., Alvis-Guzman, N., Amoako, Y., Artaman, A. et al. (2017), ‘The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015’, *JAMA oncology* **3**(12), 1683–1691.
- Allen, K., Mitrou, P. and Croker, H. (2020), ‘Worldwide cancer data; global cancer observatory by the world health organization, international agency for research on cancer’, <https://www.wcrf.org/cancer-trends/worldwide-cancer-data/>. Accessed: 29-09-2022.
- Amores, J. (2013), ‘Multiple instance classification: Review, taxonomy and comparative study’, *Artificial intelligence* **201**, 81–105.
- Arvaniti, E. and Claassen, M. (2017), ‘Sensitive detection of rare disease-associated cell subsets via representation learning’, *Nature communications* **8**(1), 1–10.
- Babaei, S., Christ, J., Makky, A., Zidane, M., Schürch, C. and Claassen, M. (2023), ‘S3-cima: Supervised spatial single-cell image analysis for the identification of disease- associated cell type compositions in tissue’, *Submitted to Cell Patterns* .
- Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C. et al. (2018), ‘Understanding the tumor immune microenvironment (time) for effective therapy’, *Nature medicine* **24**(5), 541–550.
- Black, S., Phillips, D., Hickey, J. W., Kennedy-Darling, J., Venkataraaman, V. G., Samusik, N., Goltsev, Y., Schürch, C. M. and Nolan, G. P. (2021), ‘Codex multiplexed tissue imaging with dna-conjugated antibodies’, *Nature protocols* **16**(8), 3802–3835.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008), ‘Fast unfolding of communities in large networks’, *Journal of statistical mechanics: theory and experiment* **2008**(10), P10008.

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018), 'Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA: a cancer journal for clinicians* **68**(6), 394–424.
- Canto, M. I., Harinck, F., Hruban, R. H., Offerhaus, G. J., Poley, J.-W., Kamel, I., Nio, Y., Schulick, R. S., Bassi, C., Kluijft, I. et al. (2013), 'International cancer of the pancreas screening (caps) consortium summit on the management of patients with increased risk for familial pancreatic cancer', *Gut* **62**(3), 339–347.
- Carstens, J. L., Correa de Sampaio, P., Yang, D., Barua, S., Wang, H., Rao, A., Allison, J. P., LeBleu, V. S. and Kalluri, R. (2017), 'Spatial computation of intratumoral t cells correlates with survival of patients with pancreatic cancer', *Nature communications* **8**(1), 15095.
- Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., de la Fouchardière, C. et al. (2011), 'Folfirinox versus gemcitabine for metastatic pancreatic cancer', *New England journal of medicine* **364**(19), 1817–1825.
- Coons, A. H., Creech, H. J. and Jones, R. N. (1941), 'Immunological properties of an antibody containing a fluorescent group.', *Proceedings of the society for experimental biology and medicine* **47**(2), 200–202.
- Cowgill, S. M. and Muscarella, P. (2003), 'The genetics of pancreatic cancer', *The American Journal of Surgery* **186**(3), 279–286.
URL: <https://www.sciencedirect.com/science/article/pii/S0002961003002265>
- Dries, R., Chen, J., Del Rossi, N., Khan, M. M., Sistig, A. and Yuan, G.-C. (2021), 'Advances in spatial transcriptomic data analysis', *Genome research* **31**(10), 1706–1718.
- Fesinmeyer, M. D., Austin, M. A., Li, C. I., De Roos, A. J. and Bowen, D. J. (2005), 'Differences in survival by histologic type of pancreatic cancer', *Cancer Epidemiology Biomarkers & Prevention* **14**(7), 1766–1773.
- Gallage, S., García-Beccaria, M., Szydlowska, M., Rahbari, M., Mohr, R., Tacke, F. and Heikenwalder, M. (2021), 'The therapeutic landscape of hepatocellular carcinoma', *Med* **2**(5), 505–552.
- Giesen, C., Wang, H. A., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolimund, D., Buhmann, J. M., Brandt, S. et al. (2014), 'Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry', *Nature methods* **11**(4), 417–422.

- Gillen, S., Schuster, T., Meyer zum Büschenfelde, C., Friess, H. and Kleeff, J. (2010), 'Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages', *PLoS medicine* **7**(4), e1000267.
- Hickey, J. W., Tan, Y., Nolan, G. P. and Goltsev, Y. (2021), 'Strategies for accurate cell type identification in codex multiplexed imaging data', *Frontiers in Immunology* p. 3317.
- Hofman, F. M. and Taylor, C. R. (2013), 'Immunohistochemistry', *Current Protocols in Immunology* **103**(1), 21–4.
- Huang, Y., Ge, W., Zhou, J., Gao, B., Qian, X. and Wang, W. (2021), 'The role of tumor associated macrophages in hepatocellular carcinoma', *Journal of Cancer* **12**(5), 1284.
- Kinkhabwala, A., Herbel, C., Pankratz, J., Yushchenko, D. A., Rüberg, S., Praveen, P., Reiß, S., Rodriguez, F. C., Schäfer, D., Kollet, J. et al. (2022), 'MacSima imaging cyclic staining (mics) technology reveals combinatorial target pairs for CAR T cell treatment of solid tumors', *Scientific reports* **12**(1), 1–16.
- Kleino, I., Frolovaitė, P., Suomi, T. and Elo, L. L. (2022), 'Computational solutions for spatial transcriptomics', *Computational and Structural Biotechnology Journal* **20**, 4870–4884.
URL: <https://www.sciencedirect.com/science/article/pii/S2001037022003786>
- Kocarnik, J. M., Compton, K., Dean, F. E., Fu, W., Gaw, B. L., Harvey, J. D., Henrikson, H. J., Lu, D., Pennini, A., Xu, R. et al. (2022), 'Cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019: a systematic analysis for the global burden of disease study 2019', *JAMA oncology* **8**(3), 420–444.
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-r. and Raychaudhuri, S. (2019), 'Fast, sensitive and accurate integration of single-cell data with Harmony', *Nature methods* **16**(12), 1289–1296.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015), 'Deep learning', *nature* **521**(7553), 436–444.
- Lewis, S. M., Asselin-Labat, M.-L., Nguyen, Q., Berthelet, J., Tan, X., Wimmer, V. C., Merino, D., Rogers, K. L. and Naik, S. H. (2021), 'Spatial omics and multiplexed imaging to explore cancer biology', *Nature methods* **18**(9), 997–1012.

- Li, X. and Wang, C.-Y. (2021), 'From bulk, single-cell to spatial rna sequencing', *International Journal of Oral Science* **13**(1), 1–6.
- Ligorio, M., Sil, S., Malagon-Lopez, J., Nieman, L. T., Misale, S., Di Pilato, M., Ebright, R. Y., Karabacak, M. N., Kulkarni, A. S., Liu, A. et al. (2019), 'Stromal microenvironment shapes the intratumoral architecture of pancreatic cancer', *Cell* **178**(1), 160–175.
- Matoba, K., Iizuka, N., Gondo, T., Ishihara, T., Yamada-Okabe, H., Tamesa, T., Takemoto, N., Hashimoto, K., Sakamoto, K., Miyamoto, T. et al. (2005), 'Tumor hla-dr expression linked to early intrahepatic recurrence of hepatocellular carcinoma', *International journal of cancer* **115**(2), 231–240.
- Mayo, S. C., Nathan, H., Cameron, J. L., Olino, K., Edil, B. H., Herman, J. M., Hirose, K., Schulick, R. D., Choti, M. A., Wolfgang, C. L. et al. (2012), 'Conditional survival in patients with pancreatic ductal adenocarcinoma resected with curative intent', *Cancer* **118**(10), 2674–2681.
- Morrison, A. H., Byrne, K. T. and Vonderheide, R. H. (2018), 'Immunotherapy and prevention of pancreatic cancer', *Trends in cancer* **4**(6), 418–428.
- Mullan, M. H., Gauger, P. G. and Thompson, N. W. (2001), 'Endocrine tumours of the pancreas: review and recent advances', *ANZ journal of surgery* **71**(8), 475–482.
- Olivares, O., Mayers, J. R., Gouirand, V., Torrence, M. E., Gicquel, T., Borge, L., Lac, S., Roques, J., Lavaut, M.-N., Berthezène, P. et al. (2017), 'Collagen-derived proline promotes pancreatic ductal adenocarcinoma cell survival under nutrient limited conditions', *Nature communications* **8**(1), 16031.
- Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., Rybakov, S., Ibarra, I. L., Holmberg, O., Virshup, I. et al. (2022), 'Squidpy: a scalable framework for spatial omics analysis', *Nature methods* **19**(2), 171–178.
- Quante, A. S., Ming, C., Rottmann, M., Engel, J., Boeck, S., Heinemann, V., Westphalen, C. B. and Strauch, K. (2016), 'Projections of cancer incidence and cancer-related deaths in germany by 2020 and 2030', *Cancer medicine* **5**(9), 2649–2656.
- Ruf, B., Bruhns, M., Babaei, S., Kedei, N., Ma, L., Revsine, M., Ma, C., Heinrich, B., Subramanyam, V., Qi, J., Wabitsch, S., Green, B., Bauer, K., Myojin, Y., Benmebarek, M.-R., Greten, L. T., McCallen, J., Pouzolles, M. C., Kleiner, D. E., Telford, W., Dadkhah, K., Rushinkas, A., Stovroff, M., Kang, J., Oza, K., Kroemer, A. H., Wang, X. W., Claassen, M., Korangy, F. and Greten, T. F. (2023), 'Csf1r+pd-l1+ tumor-associated macrophages trigger mait cell dysfunction at the hcc invasive margin', *Submitted to Cell*.

- Schünke, M., Schulte, E., Schumacher, U., Voll, M. and Wesker, K. (2022), *1.13 Bauchspeicheldrüse (Pancreas)*. In: Schünke M, Schulte E, Schumacher U, Voll M, Wesker K, ed. *Prometheus LernAtlas - Innere Organe*. 6., vollständig überarbeitete und erweiterte Auflage, Thieme, Stuttgart.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalex, E., Mauck III, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019), ‘Comprehensive integration of single-cell data’, *Cell* **177**(7), 1888–1902.
- Traag, V. A., Waltman, L. and Van Eck, N. J. (2019), ‘From louvain to leiden: guaranteeing well-connected communities’, *Scientific reports* **9**(1), 1–12.
- Tu, J.-F., Ding, Y.-H., Ying, X.-H., Wu, F.-Z., Zhou, X.-M., Zhang, D.-K., Zou, H. and Ji, J.-S. (2016), ‘Regulatory t cells, especially icos+ foxp3+ regulatory t cells, are increased in the hepatocellular carcinoma microenvironment and predict reduced survival’, *Scientific reports* **6**(1), 1–8.
- Wang, C., Liu, B., Xu, X., Zhuang, B., Li, H., Yin, J., Cong, M., Xu, W. and Lu, A. (2016), ‘Toward targeted therapy in chemotherapy-resistant pancreatic cancer with a smart triptolide nanomedicine’, *Oncotarget* **7**(7), 8360.
- WHO (2020), ‘International agency for research on cancer. globocan 2018. iarc’, https://gco.iarc.fr/tomorrow/en/dataviz/isotype?types=0&sexes=0&mode=population&group_populations=1&multiple_populations=1&multiple_cancers=0&cancers=11&populations=903_904_905_908_909_935&single_unit=50000. Accessed: 11-01-2023.
- Wild, C. P., Weiderpass, E. and Stewart, B. W. (2020), ‘World cancer report: Cancer research for cancer prevention’. Accessed: 18-09-2022.
- Wolf, A., Angerer, P. and Theis, F. (2018), ‘Scanpy: large-scale single-cell gene expression data analysis’, *Genome biology* **19**(1), 1–5.
- Xia, C., Fan, J., Emanuel, G., Hao, J. and Zhuang, X. (2019), ‘Spatial transcriptome profiling by merfish reveals subcellular rna compartmentalization and cell cycle-dependent gene expression’, *Proceedings of the National Academy of Sciences* **116**(39), 19490–19499.
- Yeung, O., Lo, C.-M., Ling, C., Qi, X., Geng, W., Li, C., Ng, K., Forbes, S., Guan, X.-Y., Poon, R. T. et al. (2015), ‘Alternatively activated (m2) macrophages promote tumour growth and invasiveness in hepatocellular carcinoma’, *Journal of hepatology* **62**(3), 607–616.
- Zhang, L., Sanagapalli, S. and Stoita, A. (2018), ‘Challenges in diagnosis of pancreatic cancer’, *World journal of gastroenterology* **24**(19), 2047.

- Zhao, C., Wu, X., Chen, J. and Qian, G. (2022), 'The therapeutic effect of il-21 combined with ifn- γ inducing cd4+ cxcr5+ cd57+ t cells differentiation on hepatocellular carcinoma', *Journal of Advanced Research* **36**, 89–99.
- Zhu, Y., Herndon, J. M., Sojka, D. K., Kim, K.-W., Knolhoff, B. L., Zuo, C., Cullinan, D. R., Luo, J., Bearden, A. R., Lavine, K. J. et al. (2017), 'Tissue-resident macrophages in pancreatic ductal adenocarcinoma originate from embryonic hematopoiesis and promote tumor progression', *Immunity* **47**(2), 323–338.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Masterarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Ort, Datum

Unterschrift