UNIVERSITÀ DEL PIEMONTE ORIENTALE

School of Medicine

Department of Health Sciences

Master degree in Medical Biotechnologies

Laboratory of Immunomics

(Center for Translational Research on Autoimmune and Allergic Diseases - CAAD)

Development of a Bioinformatic Pipeline based on Imaging Mass Cytometry to evaluate Synovial Membrane in Patients affected by Osteoarthritis

Relator: Prof. Giuseppe Cappellano

Supervisors: Camilla Barbero Mazzucca, Riccardo Sinella

Candidate: Giada Carlucci

Matricula Number: 20033012

Firmato digitalmente da Giuseppe Cappellano
Data: 27.06.2024 14:43:01 CEST
Organizzazione: UNIVERSITA' DEGLI STUDI DEL PIEMONTE ORIENTALE/01943490027

Academic year 2023-2024

ABSTRACT

Osteoarthritis (OA) predominantly affects diarthrodial joints and is one of the leading causes of joint pain and disability. Approximately 240 million individuals currently suffer from this condition, a number expected to rise by 2030 due to the aging population. OA typically targets the synovial membrane, which is a highly vascularized and innervated connective tissue crucial in regulating synovial fluid volume, composition, and chondrocytes maintenance. In OA patients the synovium undergoes significant changes, becoming thicker and more vascularized, with an increased infiltration of inflammatory cells, including fibroblast-like-synoviocytes, playing a crucial role in OA pathogenesis, due to the release of pro-inflammatory mediators.

Identifying specific cell populations involved in OA pathogenesis is crucial for understanding the disease's cellular mechanisms and developing personalized therapies.

This thesis aims to develop a bioinformatics pipeline to characterize, by Imaging Mass Cytometry (IMC) data, cellular composition and spatial distribution in OA synovial samples. IMC is a high-throughput technology enabling the detection of more than 40 biomarkers simultaneously on a tissue slice, overcoming limitations raised by immunohistochemistry and immunofluorescence, such as background noise and signal overlap. To develop the bioinformatic pipeline, we analyzed synovial biopsies from 6 OA patients by employing a panel of 33 antibodies to stain the regions of interest (ROI) for each patient. Two ROIs were selected for bioinformatic analysis. Following the visualization of mean marker expression and the application of fastMNN correction to remove non-biological differences between samples, two approaches, unsupervised and semi-supervised, were employed for the analysis of immunophenotype. Both approaches identified similar specific cells populations, though the frequencies of clusters varied among patients due to the different stages of the disease, namely early and late onset. Our findings would suggest that the unsupervised approach is more suitable for exploratory analysis without the need of any input, due to its characteristic of randomly generating clusters; while the semi-supervised one is suggested when the target is known, albeit with the risk of missing some important information. Spatial analysis of cells also represents an important step that requires further investigation, being crucial for a detailed and comprehensive understanding of cellular interactions and behaviors. The development of a detailed bioinformatics pipeline capable of in-depth characterization of the synovial membrane would serve as a fundamental step for conducting further analysis on larger datasets, ultimately aiming to gain deeper insights into OA pathogenesis.

INDEX

1. INTRODUCTION

1.1 Synovial Joint Anatomy

The joint is a region of the skeleton where two or more bones interact and articulate. Its primary function is to allow movements between bones and facilitate growth in the early phase of life. Inflammatory processes in joint diseases often target synovial joint which is composed of:

1) The **articular capsule**, which surrounds the joint. It is constituted by the fibrous layer and the synovial membrane, also known as **synovium**. The fibrous layer connects the bones and supports the other layer, while the synovial membrane both produces and absorbs synovial fluid facilitating the nutrient exchange between the blood and the joint.
2) The **articular cartilage** plays a crucial role in protecting our joints from mechanical stress and impact. It is covered by a thin layer of hyaline cartilage.
3) The **synovial fluid**, located within the joint cavity of a synovial joint, is crucial for lubrification, nutrients distribution and shock absorption (https://www.ncbi.nlm.nih.gov/books/NBK507893/) (Figure 1).

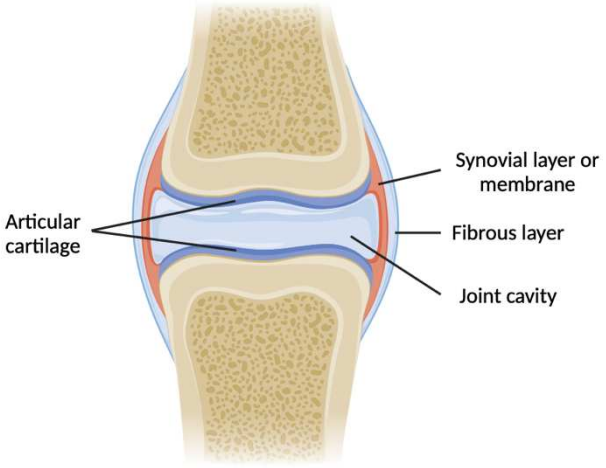


Figure 1: *Main structural features of synovial joint.*

1.1.2 Synovial Membrane

The synovial membrane or synovium is important for regulating the volume and composition of synovial fluid as well as maintaining chondrocytes. It is a connective tissue that covers the diarthrodial joints, surrounds tendons and forms the lining of bursae and fat pads, facilitating the movement of non-deformable tissues[1,2]. This membrane is crucial for the overall health and function

of joints due to its high vascularity, containing blood and lymphatic vessels, as well as being innervated[2,3].

The synovium consists of two main layers: the outer layer, also known as the subintima or sublining layer, and the inner one, referred to as the intima or lining layer.

The subintima layer is constituted by different types of connective tissue, including fibrous, adipose, and areolar tissues among the others, while the intima layer is located near to the joint cavity and it consists of 1-3 layers of synovial cells, namely type A and type B synoviocytes[1,4–6].

Type A synoviocytes, also known as synovial macrophages, play a crucial role in clearing excess materials and pathogens from the joint. They also produce and secrete enzymes, cytokines and chemokines that contribute to inflammation, cartilage degeneration and immune response. Although predominantly situated in the intima layer, type A synoviocytes can also be found in few numbers within the subintima layer.

Type B synoviocytes, or fibroblast-like synoviocytes (FLS) cells, produce hyaluronic acid (HA). They act as a barrier to maintain the synovial fluid in the capsule, and they regulate and produce extracellular matrix (ECM) components[2,3,7].

1.2 Osteoarthritis

Osteoarthritis (OA) is one of the most prevalent causes of joint pain and disability[1]. Currently, approximately 240 million people worldwide suffer from symptomatic OA, which occurs in individuals aged 60 years and older, with women being more susceptible compared to men[8–10].

The prevalence of this disorder has been constantly increasing over time and is expected to become one of the major causes of disability by 2030, correlating with the increase in the aging population[1,11].

In general, OA is attributed to a breakdown in the repair of joint damage, resulting from stresses or some tissue abnormalities[9]. This disorder can manifest in various regions of the body, including the knee, hip, hand, foot, and ankle[3,8]. Clinical manifestations include pain, stiffness, reduced joint motion and function, swelling, and muscle weakness. The long-term consequence involves reduced physical activity, deconditioning, impaired sleep, fatigue, depression, and disability[9,12].

Several risk factors can increase the possibility of developing OA, including older age, sex, overweight or obesity, as well as factors related to bone and joint shape, muscle strength, varus or valgus alignment, knee injury, and occupational factors involving repetitive activities. Additionally, genetic factors also may contribute to approximately 30-65% of cases[10,13,14].

1.3 Osteoarthritic Synovium

In healthy individuals, the synovium is a well vascularized fibro-collagenous tissue consisting of 1-3 layers of heterogenous cells fused together such as adipocytes, fibroblasts, mast cells, and macrophages. OA patients develop some abnormalities such as thickening of the lining layer, increased vascularity, and infiltration of inflammatory cells[1,15]. Hence, it is evident that the synovium assumes a pivotal role in the pathophysiological mechanisms underlying OA[3].

There are two stages of OA development: early and late stages.

The early stage is characterized by increased cell proliferation, macrophage infiltration and angiogenesis. Lymphoid aggregates are present in OA late stage, a feature absent both in healthy individuals and in the early stage[15]. However, synovial inflammation is present in both stages of OA[16]. When the synovium becomes inflamed, FLSs proliferate leading to hyperplasia in the lining layer, and macrophages and T cells accumulate in the sublining layer, creating a vascularized fibrotic environment. Conversely, B cells, mast cells and plasma cells are less abundant[1,17].

In general, macrophages aggregate and form multinucleated giant cells (MGCs) initiating the innate immune response and inducing the release of pro-inflammatory cytokines[18]. Also, FLSs and chondrocytes contribute to the development of inflammation. FLSs release the pro-inflammatory cytokines interleukin-1 beta (IL-1β) and tissue necrosis factor alpha (TNF-α), and the chondrolytic mediator such as matrix metalloproteinase (MMPs), leading to cartilage degradation. Additionally, FLSs are responsive to cytokines and toll-like receptors (TLRs) whose expression in chondrocytes may be upregulated, triggering a catabolic cascade[19].

1.3.1 Synovial Cells in OA

The most represented synovial cells are: FLSs, lymphocytes, macrophages, pericytes and nerve fibers.

FLSs

FLSs are specialized mesenchymal cells that facilitate the cartilage lubrication through the synthesis of synovial fluid, enriched in lubricin and HA[17].

FLSs express surface markers, including intercellular adhesion molecule (ICAM)-1, CD44, and β1 integrin. When FLSs are activated, they express CD90 (sublining) and Podoplanin (lining) markers[2,6];

in particular, CD90 is expressed also in mesenchymal stem cells (MSCs), and the $CD90^+$ FLSs population causes an increase of inflammation. Podoplanin is also a tumor biomarker, and it can be used to distinguish two types of FLSs populations based on the positive or negative expression of CD90. Specifically, the $CD90^+Podoplanin^+$ FLS population is associated with an inflammatory state, in contrast to those cells lacking CD90 expression ($CD90^-Podoplanin^+$)[20–22].

FLSs expressing specific markers have been already associated with different pathological conditions. Among them, CD34 is a surface protein, and it is used to distinguish three different cellular populations: (1) $CD34^-CD90^+$ cells are present in the intima layer of patients with rheumatoid arthritis (RA), (2) $CD34^-CD90^-$ cells surround blood vessels in the subintima layer of OA patients, and (3) $CD34^+$ cells are present in both RA and OA patients in the synovial membrane[2,22].

LYMPHOCYTES

Immune infiltration of T and B cells has been observed in inflamed OA tissues.

Specifically, T cells resulted to be increased in peripheral blood, synovial fluid, and synovial membrane of OA patients, suggesting a potential role in OA pathogenesis. Among T cells ($CD3^+$), $CD4^+$ and $CD8^+$ T subsets represent the most representative constituent of the synovial infiltrate (22%) in OA synovium[23].

B cells have been indicated as possible contributors to OA pathogenesis by producing autoantibodies, presenting autoantigens to autoreactive T cells, secreting pro-inflammatory cytokines and chemokines, and amplifying the inflammatory response[24,25].

MACROPHAGES

Macrophages represent 12-40% of synovial immune cells; specifically, those expressing $CD14^+CD16^+$ are the most abundant and active in the synovium[17]. Macrophages are classified into pro-inflammatory type 1 (M1) and anti-inflammatory M2 cells[16,26]. The typical markers of M1 cells are CD11c and CD16 ones, while the M2 cells express CD163 and CD14 markers[17,27].

In OA patients, the balance between M1 and M2 cells is altered, since M1 cells are the most prevalent and secrete MMP-1, aggrecans, cyclooxygenase, pro-inflammatory cytokines, such as IL-1β, IL-6 and TNF-α, promoting then cartilage degradation and OA progression[16,17].

PERICYTES

Pericytes are specialized cells crucial for vascular function since they regulate angiogenesis, provide mechanical support and stability to vessels, and control immune response, fibrosis, and inflammation[28,29]. Due to these functions, pericytes are found in specific areas, such as precapillary arterioles, capillaries, and postcapillary venules[28,30].

Pericytes are characterized by the expression of specific markers including CD146, PDGFR (platelet-derived growth factor receptors) or CD140a-b, NG2 and alpha-smooth muscle actin (α-SMA)[29,30]. In OA patients, the higher PDGFR expression is associated with aberrant subchondral angiogenesis, contributing to disease progression[29,31]. Pericytes also play an important role in the cartilage development, healing of osseocartilaginous tissues, and pathological processes related to cartilage formation.

These cells can undergo osteogenic, chondrogenic, fibrogenic, and adipogenic differentiations[30], and can originate from various cell types including smooth muscle cells, fibroblasts, endothelial cells, and bone marrow[28].

NERVES FIBERS

The nerve growth factor (NGF) is a neurotrophic factor produced by nervous system cells, lymphocytes and FLSs. It binds to the NGF receptor (NGFR), expressed in the cells of nervous system. NGF overexpression cause by the pro-inflammatory cytokines (as TNF-α and IL-1β) is associated with inflammation and amplified pain in arthritic joints, in RA patients and, to a lesser extent, in OA ones[32,33].

Synovial tissue contains two types of nerve fibers, namely sympathetic and sensory nerve fibers, that innervate the synovium.

Sensory nerves fibers express calcitonin gene-related protein (CGRP) which co-localizes with NGF in osteochondral channels[34,35]. Upregulation of NGF leads to increased CGRP expression, resulting in amplified sensory nerve fiber density and exacerbating pain and inflammation in OA patients[34,36].

Instead, sympathetic nerve fibers express tyrosine hydroxylase (TH) and are closely associated with blood vessels. Their presence is more abundant in OA patients compared with RA ones[35]. It has been shown that macrophages positive for TH and ADRB2 promote the shift from M1 toward M2 phenotype, with ADRB2 being an adrenergic receptor that activates the anti-inflammatory response in immune cells[37].

In conclusions, single-cell technologies facilitate the identification of the diverse immune and structural cell landscape and composition of the synovium[38], revealing heterogeneous and disease-specific cell patterns, which may be comprehensively appreciated by exploiting -omics techniques.

1.3.2 Soluble Inflammatory Mediators in OA

Soluble pro-inflammatory mediators and cytokines play an important role in OA pathogenesis[1]. Regarding cytokines, they can be classified in catabolic, anabolic, and regulatory types[3], having a central role in the inflammatory process[19].

Catabolic cytokines refer to signalling molecules that promote the breakdown of tissues, such as cartilage; this is a frequently observed condition in OA[39]. Indeed, these soluble mediators are increased in OA synovial fluid, synovial membrane, cartilage, and subchondral bone, having a synergistic effect on signaling pathways that increase inflammation and cartilage degradation[1,3]. Among them, IL-1β and TNF-α are pro-inflammatory mediators known to be involved in OA; other cytokines such as IL-6, IL-15, IL-17, and IL-18 play also a significant role[39].

IL-1β is produced by synovial macrophages and chondrocytes, leading to the suppression of ECM production and to the increase of MMP-1, which induces the degradation of cartilage[3,19].

Anabolic cytokines, such as transforming growth factor beta (TGF-β) and insulin-like growth factor-1 (IGF-1), are signalling molecules that promote tissue growth and repair, playing a role in maintaining the health and function of tissues like cartilage in joints[3,40]. In particular, TGF-β regulates the formation or degradation of ECM, maintaining its homeostasis. In patients affected by OA, a deficiency of anabolic cytokines is often reported, with an increase in cartilage degradation and damage[40].

Regulatory cytokines protect the joint from degradation; for example, IL-4, IL-10, and IL-13 exert anti-inflammatory properties on synovial macrophages and up-regulate the production of the natural inhibitors of catabolic cytokines[1,3]. IL-4 and IL-10 have also a chondroprotective effect by inhibiting the production of MMPs as well as downregulating the synthesis of TNF-α, IL-1β, and IL-6. Additionally, IL-10 stimulates the synthesis of the IL-1β antagonist (IL-1Ra) and the tissue inhibitor of metalloproteases-1 (TIMP-1)[39].

In OA patients an imbalance between pro-inflammatory and anti-inflammatory molecules is recurrent, with a bias towards the inflammatory side; thus, it is important to comprehensively analyze these soluble factors to get deeper insights in disease pathogenesis[1,3].

1.4 OA Treatment

OA treatments involve the use of non-pharmacologic or pharmacologic interventions[41].

Non-pharmacological therapies, including exercise and dietary adjustments, not only aimed at alleviating pain but also in reducing the risk of developing OA, while enhancing physical function[9,41].

On the other hand, current pharmacological options include acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDs) for treatment of moderate to severe symptoms, opioids for pain reduction, anti-inflammatory cytokines, methotrexate, and intraarticular (IA) therapies such as glucocorticoid or HA injections directly into the joint[1,9,41].

In those patients that do not respond adequately to the previous treatments, the only option is joint replacement surgery which is also the most invasive and less economically convenient[9,41].

Currently, FDA-approved disease-modifying drugs for OA capable of alleviating pain and impeding joint degradation are lacking[42].

Identifying specific subpopulations linked to the pathogenesis of OA is crucial for enhancing our comprehension of the cellular mechanisms underlying the disease and for developing personalized therapeutic interventions for OA.

Recent advancements in omics sciences and high-throughput omics technologies have the potential to significantly enhance our understanding of the distinct cell populations present in the joint tissues of each OA patient, paving the way for personalized therapy approaches[43].

1.5 Precision Medicine

In this context, the precision medicine approach offers the opportunity for patient-specific evaluations to select the best treatments or to study the pathological mechanisms for the development of tailored therapies. This approach aims to integrate high-throughput biological data generated through omics techniques with environmental and lifestyle factors to obtain a complete individual profiling able to provide patient-specific details that can be used as diagnostic, prognostic, and treatment response biomarkers or therapeutic targets[44,45].

Various authors describe precision medicine in different ways, characterizing it as a discipline that:

a) Customizes treatments for individual patients based on their unique characteristics, including genetics, biomarkers, phenotypic traits, and psychosocial factors, which differentiate them from others with similar clinical conditions.
b) Categorizes patients into novel subgroups to identify common disease susceptibility and manifestation patterns within these subcategories, thereby enabling more precise therapeutic interventions[46].

In summary, precision medicine aims to create personalized treatments for everyone[47], recognizing that every patient is unique, and what works for one person may not work for another[48].

For this purpose, omics approaches permit a patient-specific in-depth exploration of biological molecules and their interactions within biological systems. These approaches include genomics, transcriptomics, proteomics, metabolomics, and epigenomics.

Through omics investigation, substantial amounts of data are generated, far exceeding our capacity to process it. Many groups have developed tools aimed at connecting these gaps, addressing the need for instruments capable of analyzing and sorting the data. Bioinformatics techniques can help in the analysis of this vast amount of data[49]. The analysis of these data allows us to better understand disease pathogenesis, with the potential to develop more accurate predictive or prognostic models. Furthermore, these approaches enable us to anticipate patient responses to specific treatment protocols[48,50,51].

1.6 Imaging Mass Cytometry

For many years, immunohistochemistry (IHC) and immunofluorescence (IF) have been used for diagnosis of various diseases. Both IHC and IF utilize a specific primary antibody that binds to the antigen of interest. Subsequently, a secondary antibody is added, which binds to the primary one. The difference between these two approaches lies in their detection methods. In IHC the detection can be chromogenic or fluorogenic, whereas in IF involves the detection of a fluorescent label (fluorophore). Therefore, the secondary antibody can be linked to an enzyme that catalyzes a chromogenic reaction, or to a fluorophore that is observed under a fluorescent microscope.

Typically, we can use up to 5-7 different markers together, but this is limited by potential overlap between the antibodies, the quality and concentration of the antibodies, and the unstable staining

that may be occur. Additionally, the signal from fluorescent probes diminishes over time with exposure to light, a phenomenon known as bleaching[52,53].

Imaging Mass Cytometry (IMC) is an innovative multidimensional technique, representing a cutting-edge approach for studying complex tissue sections[54]. It configures as a combination between IHC and Mass Cytometry, allowing the simultaneous staining and analysis of markers in both formalin-fixed and paraffin-embedded (FFPE) tissues and frozen tissues, facilitating the study of various pathological processes and diseases[53–55].

As mentioned above, compared to the existing techniques, IMC adds a step forward permitting to detect and visualize simultaneously 35-40 different markers on a single tissue slide. Indeed, each antibody used for tissue staining utilizes stable metal isotopes linked with metal-chelating polymer chains rather than fluorophores or enzymes. Detection and quantification of metals are conducted using a time-of-flight (TOF) mass spectrometer[54], which, due to its high resolution, allows precise localization of proteins within cellular compartments such as nucleus, cytoplasm, or cellular membrane[56].

Since IMC can be performed on FFPE sections, it holds the potential for retrospective histological analysis of patients with known outcomes. Ultimately, leveraging IMC for retrospective analysis has the potential to refine our understanding of disease processes and improve patient stratification, ultimately leading to more targeted and personalized treatment approaches[55,57].

In particular, the advantages of IMC include high multiplexing, absence of autofluorescence, low interchannel crosstalk, stable staining, high quantitative, and high dynamic range. These attributes allow for a more in-depth study of specific diseases. In addition, with bioinformatic analysis, IMC enables the creation of different clusters (clustering approach), the study of the interactions between cells and cells populations (spatial analysis), and the identification of specific cell phenotypes (cell phenotyping).

2. AIM

This internship was carried out at the Immunomics laboratory (Head Prof. Annalisa Chiocchetti), and it is part of the SINPAIN project (GA n.101057778 HORIZON-HLTH-2021-TOOL-06-02 "*A game changer for the treatment of osteoarthritis: a cost effective combined advanced therapy to treat knee osteoarthritis*"), which aims to develop a siRNA-based therapy pipeline for treating different stages of knee OA. This pipeline will be combined with current therapies and designed step-by-step to achieve successful management of inflammation and innervation therapy for the treatment of early and later stages of OA.

Within the SINPAIN project, our task is set on point a bioinformatic pipeline to analyze IMC data obtained from the staining of synovial samples from patients affected by OA.

We set up the antibody panel to reach optimal antibodies concentration, proceeding then to the samples acquisition and analysis, exploiting bioinformatic tools. Specifically, the comparison of different bioinformatic approaches represents the main focus of this thesis. The proposed pipeline will find application in future studies involving a bigger cohort and has the potential to improve the current knowledge about OA pathogenetic mechanisms and progression with an innovative approach.

3. MATERIALS AND METHODS

3.1 Human Specimens

Human synovia biopsies were taken by OA patients (n=6) from the Instituto de Investigacao e Inovacao Em Saude da Universidade do Porto (i3s), established in Rua Alfredo Allen, Porto (Portugal). Patients with OA were diagnosed based on clinical symptoms, examination, and radiographic findings of knee joints.

All patients participating in this study signed informed consent forms. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Eastern Piedmont (protocol OCEANIA 9/21).

3.2 IMC Workflow

The IMC workflow (Figure 2) can be divided in three parts as shown in Figure 2:

1) **Panel design**: markers and antibody selection, metal-antibody association, and antibody conjugation.
2) **Staining and tissue ablation**: samples processing, morphological and IMC staining and samples ablation to acquire high-dimensional spatial data.
3) **Data analysis**: biomarkers expression levels evaluation and single cell analysis.

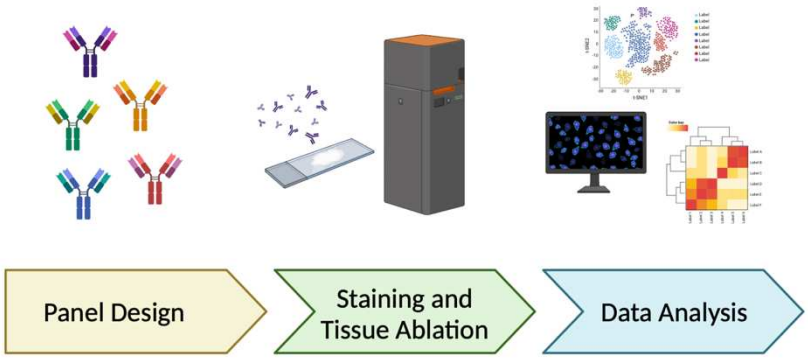


Figure 2: *Workflow of the study.*

3.3 Panel Design

Table 1 shows the markers included in the staining panel. Each antibody is linked to a specific metal based on its target expression in the tissue; antibodies have been selected following a literature

review to effectively characterize the synovial membrane and to determine the optimal dilution to maximize the signal and minimize the background.

MARKER	METAL	DILUTION
α-SMA	141 Pr	1:400
Collagen 1	144 Nd	1:400
CD14	148 Nd	1:400
CD146	160 Gd	1:400
CGRP	163 Dy	1:400
CD19	142 Nd	1:300
CD4	156 Gd	1:250
TH	143 Nd	1:200
HLA-DR	151 Eu	1:200
CD11c	154 Sm	1:200
IL-6	161 Dy	1:200
CD140a-b	164 Dy	1:200
Fibronectin	174 Yb	1:200
CD138	149 Sm	1:150
CD11b	146 Nd	1:100
CD163	147 Sm	1:100
CD66b	152 Sm	1:100
CD16	153 Eu	1:100
CD90	158 Gd	1:100
MMP-1	159 Tb	1:100
CD8	162 Dy	1:100
MMP-2	166 Er	1:100
NGFR	167 Er	1:100
NG2	171 Yb	1:100
IL-1β	173 Yb	1:100
Podoplanin	175 Lu	1:100
CD31	145 Nd	1:50
TIMP-1	150 Nd	1:50
TNF-α	155 Gd	1:50
CD34	168 Er	1:50
NGF	169 Tm	1:50
CD3	170 Er	1:50
NFH	172 Yb	1:50

Table 1*: Antibody panel used for IMC.*

3.3.1 Conjugation Process

Metals were dissolved in L-Buffer (StandardBioTools, USA), then loaded onto a polymer, specifically X8 (StandardBioTools, USA), and purified through multiple washes in C-Buffer (StandardBioTools, USA). The process involving the metal-loaded polymer utilized a filter device (Sartorius, UK) with a 3 kDa membrane to ensure proper handling. The X8 polymer is essential for retaining the metal, preventing its loss during washing, and facilitating effective antibody binding. Simultaneously, antibodies were prepared (in some instances, reconstituted in PBS), reduced using TCEP (tris-2-

carboxyethyl-phosphine), and purified through washes in C-Buffer. TCEP was necessary for antibody reduction to enable successful conjugation with the metal. Subsequently, the antibody was conjugated with the metal-loaded polymer, followed by additional washes in W-Buffer (StandardBioTools, USA), and the recovery rate of this process was quantified. All antibody-related procedures utilized a filter device (Sartorius, UK) equipped with a 50 kDa membrane.

Following the final wash with W-Buffer, 80 µl of W-Buffer were added to dilute the conjugate. Then, the conjugated antibodies were quantified using a NanoDrop spectrophotometer, measuring the absorbance of a 2 µl aliquot at 280 nm against a W-Buffer blank. To ensure stability, Antibody Stabilizer PBS supplemented with 0.05% sodium azide (Candor Bioscience, Germany) was added, and the conjugated antibodies were stored at +4°C in Protein LoBind tubes of 1.5 ml (Eppendorf, Germany) (Figure 3).

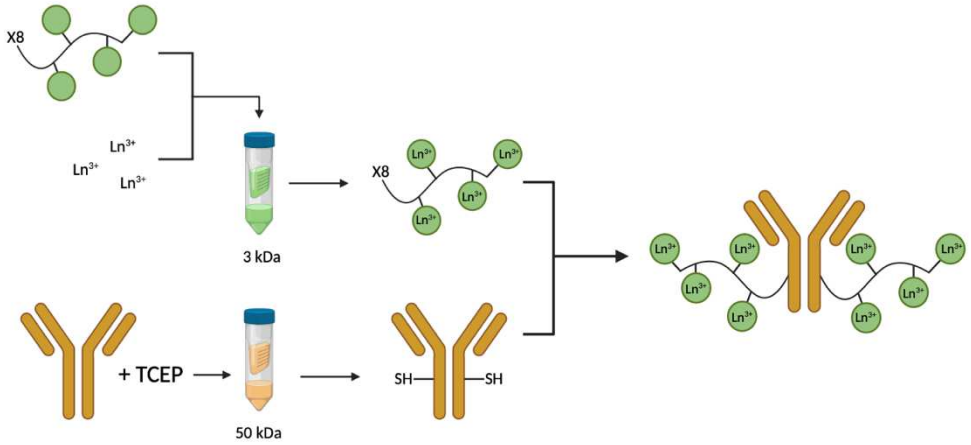


Figure 3: *Workflow of antibodies conjugation.*

3.4 Staining and Tissue Ablation

Synovial tissue biopsies were obtained from residual knee joint tissues of patients diagnosed with OA at various stages of disease.

Samples were processed within a class II containment hood, where the synovial tissue was carefully separated from the joint and placed into a 4% formalin solution for human tissue samples. The samples were fixed in 4% buffered formalin solution for 24 to 48 hours, before proceeding with tissue processing and embedding in paraffin.

3.4.1 Sample Processing

After sample fixation, synovial tissues were processed using the FFPE procedure.

Briefly, the samples were hydrated in ascending grades of alcohol (70%, 80%, 95%, 100%) for 90 minutes each, followed by immersion in xylene for 90 minutes to facilitate paraffin infiltration. Later, they were submerged into paraffin overnight to permit proper infiltration, and, the next day, the samples were embedded into paraffin blocks. This final step requires proper orientation of the sample to obtain observable sections suitable for IMC staining.

Overall, 6 blocks were obtained, one for each OA patient. Successively, serial sections for each sample of 5 µm thickness were cut using a microtome with S35 blade, optimized for soft tissue (Leica, Germany), and after a few minutes in the water bath at 37°C, they were positioned on charged glass slides (Bio Optica, Milan). The first section was used to perform Hematoxylin and Eosin staining (Bio Optica, Milan) for histomorphology, while the following sections were used to perform the IMC staining.

3.4.2 Hematoxylin and Eosin Staining

Carazzi's Hematoxylin was chosen for its less intense cytoplasm staining. It comprises a complex of hematein (hematoxylin oxidized by potassium iodate) and aluminum potassium sulfate, which carries a positive charge facilitating binding to anionic sites on chromatin histone proteins. The expected results are nuclei stained in purple and cytoplasm in red-pink.

The wax on sections was removed by soaking them in xylene for 20 minutes, followed by hydration in descending concentrations of alcohol (100%, 95%, 80%, and 70%) and rinsing in distilled water before staining with Carazzi's Hematoxylin for 12 minutes. A 5-minute tap water wash was performed to fix the Hematoxylin staining, due to the presence of salts. Subsequently, sections were immersed in aqueous 1% Eosin for 1 minute, followed by a final wash in distilled water to remove Eosin. Sections were then dehydrated in ascending concentrations of alcohol (70%, 80%, 100%) for 1 minute each, cleared in xylene for 10 minutes, and finally mounted using BioMount HM (Bio Optica, Milan).

3.4.3 IMC Staining

The slides underwent a series of steps as following: firstly, they were placed in an oven at 60°C for 2 hours to dissolve the paraffin, followed by dewaxing with xylene for 20 minutes and dehydrated in

descending concentrations of ethanol (100%, 95%, 80%, 70%) for 5 minutes each, and then washed in Maxpar Water for other 5 minutes (StandardBioTools, USA).

Subsequently, the slides were placed into a preheated antigen retrieval solution at 96°C for 30 minutes, for retrieving antigens masked by fixation and make them more accessible. This solution was prepared by diluting 4 ml of Dako target retrieval solution 50X (Dako, Denmark) in 36 ml of Maxpar Water.

Following incubation, the tubes containing the slides and the retrieval solution were removed from the oven and placed on a lab bench for 10 minutes to cool them to a temperature of 70°C. The slides were washed in Maxpar Water and then in Maxpar PBS (StandardBioTools, USA). The sections were encircled with an A-PAP pen (BioOptica, Italy) and blocked with 3% BSA in Maxpar PBS for 45 minutes at room temperature in a hydration chamber, to prevent nonspecific binding.

The antibody cocktail was prepared with PBS and BSA 0.5% and added onto the sections, which were left to incubate overnight at 4°C in a hydration chamber. The volume of the blocking solution and antibody cocktail was optimized for specific tissues and calculated depending on the size of each section and the number of total sections.

On the next day, the slides were washed two times in 0.2% Triton X-100 in Maxpar PBS for 8 minutes each to permeabilize the membrane, followed by two washes in Maxpar PBS for 8 minutes each. The tissues were stained with Intercalator-Ir (StandardBioTools, USA) in Maxpar PBS for 30 minutes at room temperature in a hydration chamber to label the nuclei. Finally, the slides were air-dried after a last wash in Maxpar Water.

3.4.4 IMC Ablation Test and Acquisition

To ensure adequate signal quality, it is important to adjust the ablation energy for the samples. The ablation energy refers to the energy of the laser ray delivered to the sample, and it can be adjusted using the energy attenuator. Optimal laser energy (measured in dB) varies depending on the tissue type and must be optimized before sample acquisition.

To determine the optimal signal-to-noise (SNR) ratio for the 193Ir (DNA-intercalator) channel in the MCD™ Viewer, a laser energy ramp was conducted across multiple regions of interest (ROIs). Five adjacent small ROIs of 50x50 µm were selected; increasing dB values were assigned to each ROI, starting from 0 dB, and incrementing to 1, 2, 3, 4, and 5 dB. The ablation energy was set to 2 dB for each ROI.

Sections were acquired using imaging IMC (Hyperion, StandardBioTools, USA), which uses a powerful laser to ablate the samples. Laser ablation involves the elimination of tissue plumes from a solid surface (such as a glass slide) by irradiating it with a pulsating laser beam. Each laser pulse ablates tissue from a spot measuring 1 μm^2, with the tissue plumes being carried by helium gas to an argon flow within the inductively coupled plasma argon system. Here, the tissue plumes are aerosolized, atomized, and ionized before entering into TOF chamber for detection (CyTOF instrument). The detector offers a resolution of 1 Da, and discrimination of lanthanides is based on their mass. Subsequently, the abundance of each isotope can be mapped back to the original coordinates for each tissue spot.

The tissue was examined systematically, spot-by-spot, while the slide was moved under the laser for scanning the whole ROI. The acquisition speed averaged around 100-120 minutes for each 1 mm^2 of tissue. Each ablation spot corresponds to an image pixel, with each pixel containing various metal ions. The resulting output consists of a reconstructed multi-channel multiparametric image.

Since it is not possible to ablate the whole sample, ROIs selection is a fundamental step for both acquisition and analysis. Thus, a standardized pipeline was developed, involving consecutive ROIs selection, ultimately resulting in the creation of a composite sample image. This approach is implemented for analyzing biopsies from each patient, facilitating a more in-depth evaluation, and eliminating bias associated with targeting specific structures of interest.
Each ROI covered an area of 717x661 mm and was acquired at a rate of 200 Hz, and the collected data were then exported in MCD format. The size of each ROI can be adjust based on the presence of the glass slide, that can cause interference during the ablation process.

3.5 Data Analysis

3.5.1 MCD™ Viewer

To evaluate the quality of the staining each marker was visualized using the MCD™ Viewer software (StandardBioTools). This step is crucial for the optimization of antibody dilutions and evaluation of staining specificity, background occurrence, and potential channel crosstalk allowing to visualize a maximum of 7 markers simultaneously. It also provides perceptions into the staining's cellular localization, whether it envelops the entire cell surface, is restricted to specific areas, or is

perinuclear. The colors observed in the resulting image are termed pseudo colors because they are selected by the operator.

High-quality staining is crucial for precise single-cell analysis clustering. To minimize background noise and enhance signal clarity, various antibody dilutions were validated, with adjustments of the blocking step duration and concentration.

3.5.2 IMC Segmentation Pipeline

To characterize our patients, we used the "IMC Segmentation Pipeline" for IMC data analysis provided by BodenmillerGroup (https://bodenmillergroup.github.io/IMCDataAnalysis/processing.html).

The IMC segmentation pipeline provides a relatively hands-on method for segmenting multi-channel images through a pixel classification-oriented approach. This pipeline is divided in three parts: (1) **pre-processing**, (2) **single cells analysis**, and (3) **spatial analysis**.

PRE-PROCESSING

In this step, images data were prepared for processing with Steinbock toolkit. It involves creating a panel file and extracting images starting from .txt files generated during Hyperion acquisition. Subsequently, the ilastik software was used for pixel classification, determining the probabilities of pixels belonging to a specific class (such as nucleus, cytoplasm, or background) for each image, based on a random forest approach. Then CellProfiler was employed for object segmentation. It is a pixel classification-based image segmentation, that requires probability images generated before. The final result was a greyscale object masks containing unique pixel values for each object.

Lastly, there was a step for object measurement, where different cell features like object intensities, object proprieties (such as area, centroid, major and minor axis length, and eccentricity) and object neighbors were measured. In particular, the neighbors were assessed based on distances between object centroids, borders, and by pixel expansion. The object centroid evaluation includes measuring the maximum distances between object centroids (dmax = 15) and identifying the k-nearest neighbour based on centroid distances (kmax = 5).

SINGLE-CELL ANALYSIS

In this section, the initial step involves importing single-cell data and images into R Studio after image processing and segmentation.

The segmentation quality control was used to observe the accuracy of our data using the *cytomapper* package. For the image-level quality control and the cell-level quality control it is important to calculate the SNR for individual channels and markers; it was calculated as SNR = I_s/I_n, where "I_s" is the mean intensity of pixels with true signal, and "I_n" is the mean intensity of pixels containing noise. The threshold to distinguish between signal and noise pixels was determined using the Otsu thresholding method[58], separating foreground (signal) and background (noise) pixels. SNR was then computed as the mean intensity of foreground pixels divided by the mean intensity of background pixels.

The next step involves performing batch effect correction that removes the non-biological differences between samples (such as different reagents, concentration of antibody), and facilitates the detection of cell phenotypes. There are three different types of correction:

- FastMNN correction tries to identify the mutual nearest neighbors (MNN) among cells from different samples and correct the differential expression between batches to align the data.
- Harmony correction aims to cluster and correct the positions of cells by removing the differences between patients, taking into consideration cell types rather than the specific conditions of the data.
- Seurat correction tries to identify the MNN in a low-dimensional space before correcting the expression values of cells. In summary, it uses some techniques of normalization, correction, and filtering to reduce biological variability between samples.

For this step, the fastMNN correction method was preferred because it takes into consideration variability between patients. This ultimately led to a better analysis and comparison of the data.

The final step of this section was to define cell phenotypes by clustering them using specific algorithms. We used two methods: semi-supervised, where we manually generated clusters based on our desired outcomes (gating cells), and unsupervised, where clusters were generated by a clustering approach without our influence (PhenoGraph).

The unsupervised approach uses the PhenoGraph clustering method, that considers a group of cells and tries to find the most similar cells based on the distance between them (using Euclidean distance). Then, it creates a graph based on these similarities to ultimately obtain clusters of cells with similar characteristics. The names of the clusters were assigned based on the expression of different markers.

For the semi-supervised approach, the *cytomapper* package was used for gating cells based on marker expression and visualizing selected cells directly on images. In this case, the clusters that were created, were influenced by our input because we gated the cells based on our interests and desired outcome. This approach allowed us to obtain information about the accuracy of cluster formation. We could visualize if each cell phenotype was correctly classified or misclassified and, in the end, visualize the probability of finding a specific cluster.

As a final outcome, it was possible to visualize single-cell data and markers abundance at both cell and sample level. For instance, to visualize the mean marker expression per cell type, we chose Z-score scaling.

SPATIAL ANALYSIS

In this section, single cells were analyzed in their spatial tissue context. It was possible to visualize the cells' centroids and cell-cell interactions as spatial graphs and perform a spatial community analysis or a cellular neighborhood analysis.

The spatial community analysis was based on cell interactions to identify the major population present in the samples. The cellular neighborhood created groups of cells based on the information contained in the direct neighborhood. It was conducted in two different ways: (1) determining the proportion of cells of a specific type among the neighbors of each cell, and (2) summing the expression counts across all neighboring cells for each individual cell.

Finally, the interaction analysis was utilized to perform a statistical test about the interaction between all cell types in the dataset. It took into consideration the interaction of a cell type with many images, or a specific number of cells of a certain type that were located around a target cell type.

4. RESULTS

4.1 Establishment of Staining and Acquisition Protocol

The analysis was performed on synovial biopsies taken from 6 OA patients. Tissues were embedded in paraffin, and they were processed and acquired with Hyperion.

Firstly, Hematoxylin and Eosin staining was performed, as shown in Figure 4, to orient the tissue and to identify within the biopsies the synovial regions (marked in red) to be subsequently selected during the ablation process (Figure 5).

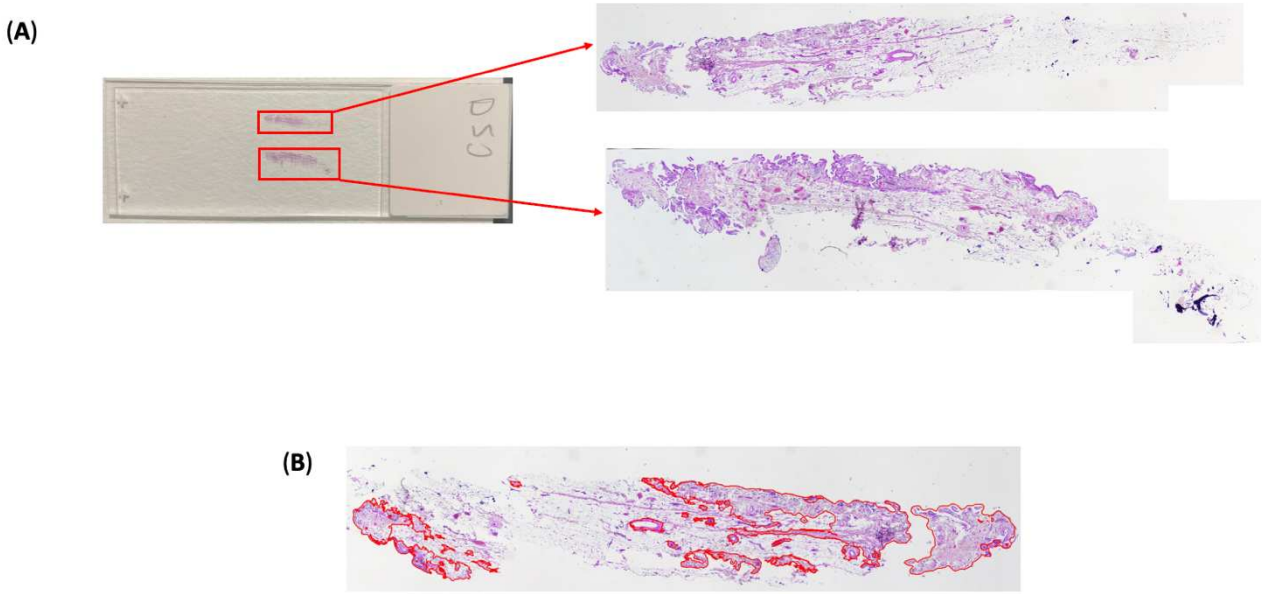


Figure 4: *(A) Hematoxylin and Eosin staining at higher magnification. (B) The red lines represent the part of tissue that is synovium.*

On that selected region, we performed IMC acquisition. We set the antibodies panel made of 33 antibodies to target the cells of interest. To optimize the staining and avoiding any issue in the signal/background ratio we performed serial dilution of each antibody, as shown in Table 2.

MARKER	METAL	DILUTION	TARGET
HLA-DR	151 Eu	1:200	Antigen Presenting Cells
CD19	142 Nd	1:300	B cells
IL-1β	173 Yb	1:100	Cytokines
IL-6	161 Dy	1:200	Cytokines
TNF-α	155 Gd	1:50	Cytokines
CD31	145 Nd	1:50	Endothelium
CD34	168 Er	1:50	Endothelium
CD90	158 Gd	1:100	Fibroblast Like Synoviocytes
Podoplanin	175 Lu	1:100	Fibroblast Like Synoviocytes
CD11c	154 Sm	1:200	Macrophages M1
CD163	147 Sm	1:100	Macrophages M2
Collagen 1	144 Nd	1:400	Matrix
Fibronectin	174 Yb	1:200	Matrix
MMP-1	159 Tb	1:100	Matrix
MMP-2	166 Er	1:100	Matrix
TIMP-1	150 Nd	1:50	Matrix

MARKER	METAL	DILUTION	TARGET
CD14	148 Nd	1:400	Monocytes
NFH	172 Yb	1:50	Myelinated Nerve Fibers
CD11b	146 Nd	1:100	Myeloid Cells
NGF	169 Tm	1:50	Nerve Growth Factors
NGFR	167 Er	1:100	Nerve Growth Factors Receptors
CD16	153 Eu	1:100	Neutrophils
CD66b	152 Sm	1:100	Neutrophils
CD140a-b	164 Dy	1:200	Pericytes
CD146	160 Gd	1:400	Pericytes
NG2	171 Yb	1:100	Pericytes
CD138	149 Sm	1:150	Plasma Cells
CGRP	163 Dy	1:400	Sensory Nerves Fibers
TH	143 Nd	1:200	Sympathetic Nerves Fibers
CD3	170 Er	1:50	T Cells
CD8	162 Dy	1:100	T Cytotoxic Cells
CD4	156 Gd	1:250	T Helper cells
α-SMA	141 Pr	1:400	Vessels

Table 2: *Selected dilutions for each antibody and its respective target.*

IMC staining was performed simultaneously on all samples to facilitate comparison among samples by reducing the batch effect, which is a systematic variation caused by differences in reagent lots, sample preparation, staining conditions, or instrument use, rather than actual biological differences between samples.

Before each sample ablation, ROIs were chosen based on the Hematoxylin and Eosin staining shown in Figure 4. Each ROI was designed with a standardized dimension of 717x661 nm, which can be modified if the area of interest includes a large part of the glass slide (Figure 5).

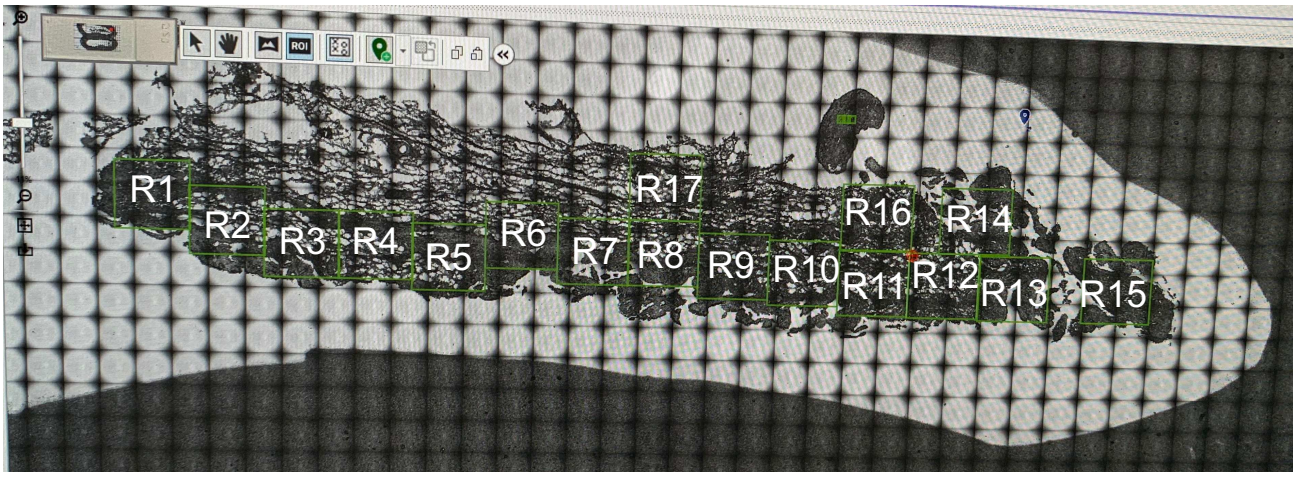
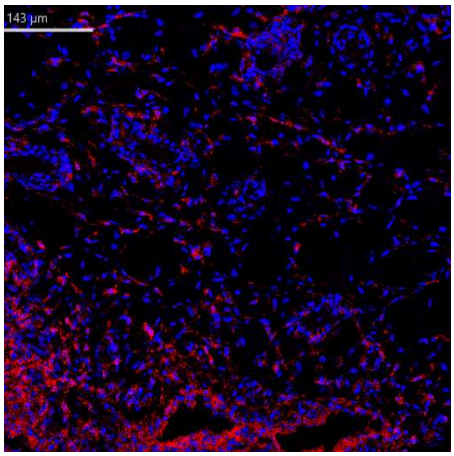


Figure 5: *ROIs design before acquisition.*

After the ablation, the acquired ROIs were firstly visualized using MCD™ Viewer. This software allows the visualization of up to 7 colors simultaneously for the evaluation of staining quality and efficacy.

As an example, Figure 6A shows the staining for CD14 (shown in red), which is a marker expressed by monocytes. Figure 6B shows a combination of 4 different markers allowing the discrimination of different type of cells simultaneously: CD14 (*red*, monocytes), CD140ab (*cyan,* pericytes), NGFR (*white*, nerve growth factor receptor), and IL-1β (*green*, a proinflammatory cytokine). DNA is marked in blue.

(**A**) 

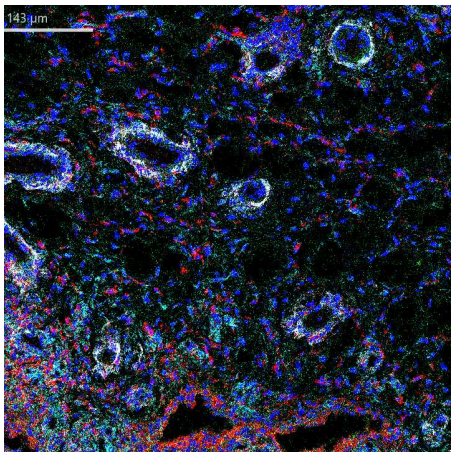
(**B**) 

Figure 6: *MCD™ Viewer images. (A) CD14 marker alone in red. (B) Multiple colors image in which CD14 is marked in red, CD140ab in cyan, NGFR in white, and IL-1β in green. In both images, DNA is colored in blue.*

4.2 Data Analysis

In order to perform, the bioinformatic analysis, 2 ROIs for each of the 6 patients have been chosen and the corresponding .txt files were generated by the IMC processing, used for the pre-processing step.

Initially, a Steinbock panel file was generated (Table 3); it contained information about the channels in an image, namely “channel” (related to the metal conjugated to the antibody) and “name” (related to the antibody used), and additionally the remaining columns allow us to select the channels used in different tasks of analysis (keep, ilastik, DeepCell, CellPose). In this step markers of interest to perform the analysis were selected.

channel,name,keep,ilastik,deepcell,cellpose
ArAr80,80ArAr,0,,,
I127,127I,0,,,
Xe131,131Xe,0,,,
Xe134,134Xe,0,,,
Ba138,138Ba,0,,,
Pr141,aSMA,1,,,
Nd142,CD19,1,,,
Nd143,TH,1,,,
Nd144,COL1,1,,,
Nd145,CD31,1,,,
Nd146,CD11b,1,,,
Sm147,CD163,1,,,
Nd148,CD14,1,,,
Sm149,CD138,1,,,
Nd150,TIMP-1,1,,,
Eu151,HLA-DR,1,,,
Sm152,CD66b,1,,,
Eu153,CD16,1,,,
Sm154,CD11c,1,,,
Gd155,TNFa,1,,,
Gd156,CD4,1,,,
Gd158,CD90,1,,,
Tb159,MMP-1,1,,,
Gd160,CD146,1,,,
Dy161,IL-6,1,,,
Dy162,CD8,1,,,
Dy163,CGRP,1,,,
Dy164,CD140a-b,1,,,
Ho165,ADRP2,0,,,
Er166,MMP-2,1,,,
Er167,NGFR,1,,,
Er168,CD34,1,,,
Tm169,NGF,1,,,
Er170,CD3,1,,,
Yb171,NG2,1,,,
Yb172,NFH,1,,,
Yb173,IL-1b,1,,,
Yb174,FIBRONECTIN,1,,,
Lu175,PODOPLANIN,1,,,
Ir191,DNA1,1,1,,
Ir193,DNA2,1,2,,

Table 3: *Steinbock panel file.*

After pre-processing the panel and the IMC images, the pixel classification step was performed using the ilastik software, which employs a random forest approach to execute this task. During this step we marked on each image in red the nucleus, in green the cytoplasm and in blue the background in order to train the machine for the differentiation of these distinctive regions in each ROI.

Finally, the whole batch of images was processed to obtain probability images, where each colour represents the probability of pixels belonging to the corresponding class.

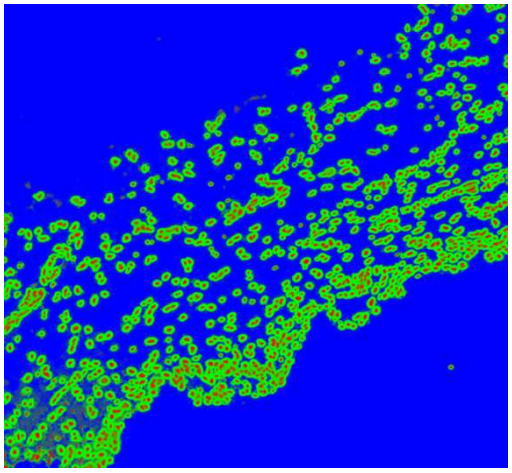


Figure 7: *Ilastik probability image after batch processing. In the image, cytoplasm (shown in green), nuclei (shown in red) and the background (shown in blue) are represented.*

Later, the object segmentation was performed with CellProfiler, which requires probability images generated by the preceding pixel classification step as input. In addition, after the segmentation step, grayscale object masks were created containing unique pixel values for each object.

Finally, the ROIs object intensities, region proprieties, and the object neighbours were evaluated. In particular, the neighbours can be assessed based on distances between object centroids, that measuring the maximum distances between object centroids (dmax = 15) and identifying the k-nearest neighbour based on centroid distances (kmax = 5).

Following image processing and segmentation, the data generated were read in R Studio.
The initial part involved the image and cell-level quality control. We evaluated all 12 ROIs to observe the accuracy of the segmentation. Using *cytomapper* function, as quality check, we overlapped the masks obtained from the segmentation process with the composite images allowing us to observe if the nuclei were centered within the segmentation masks and if all cell types were correctly segmented (Figure 8).

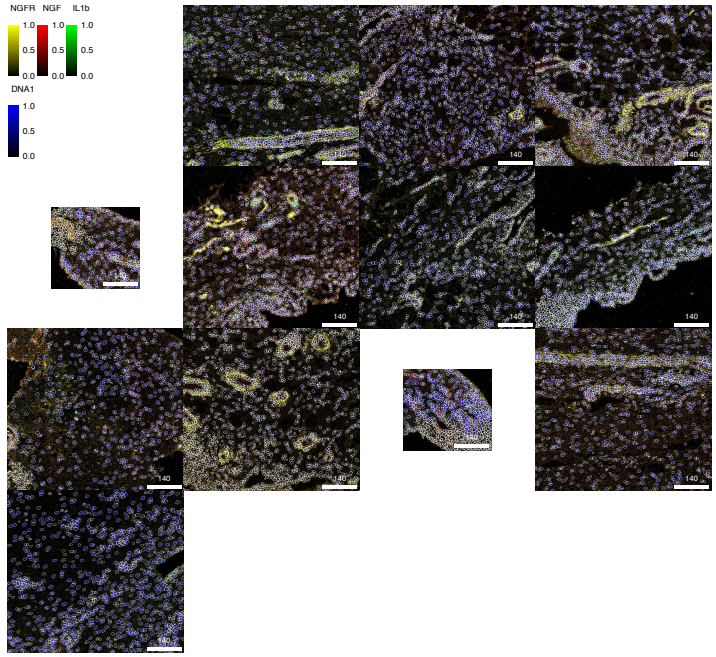


Figure 8: *Segmentation quality control. This image is used to observe whether the nuclei are centered within the segmentation masks and if all cell types are correctly segmented. In the image, DNA1 and DNA2 are visualized in blue, NGFR in yellow, NGF in red, and IL-1β in green.*

Additionally, single-cell expression of different markers can be visualized in the form of heatmap grouping the patients based on their indication. In this case, we choose to group all the patients under the same indication (OA) since the study is still blinded so we were not able to stratify the patients based on the grade of the disease (Figure 9).

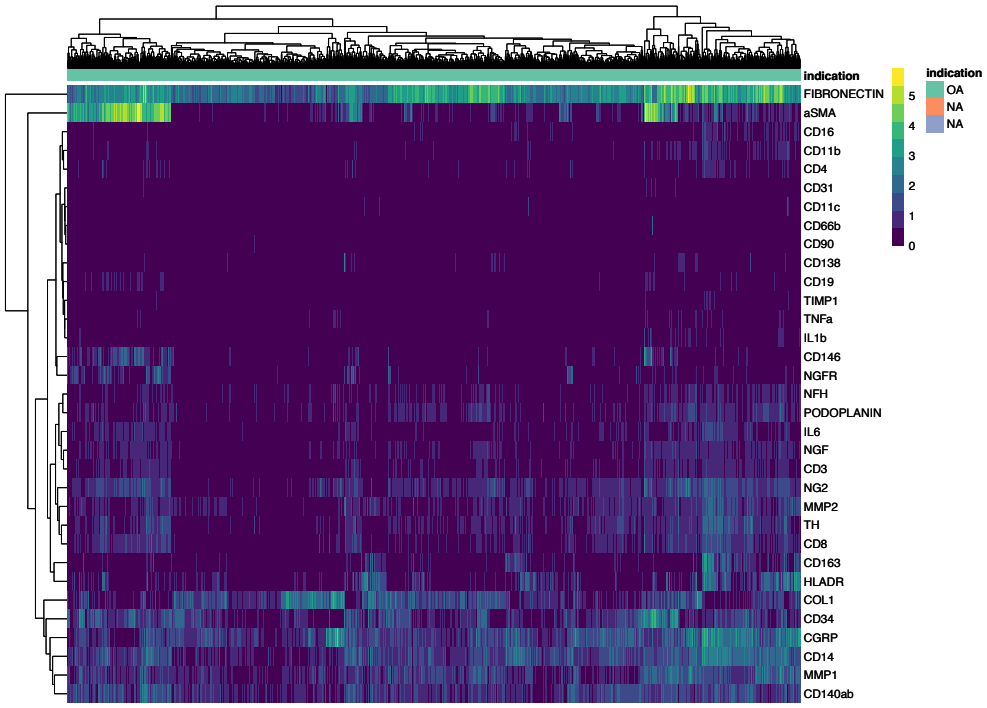


Figure 9: *Heatmap of single-cell expression based on the indication (OA) of patients. Blue denotes low expression, light blue intermediate expression, and yellow high expression. In the upper part of the heatmap are indicated the indication of the patients and the number of cells. The dendrograms on the left and on the top represent the similarity among markers and numbers of cells respectively.*

As shown in Figure 9, on a sample of 2000 cells among all patients, αSMA and Fibronectin markers were the most highly expressed ones. However, it is possible to observe that CD14, CGRP, and COL1 markers were expressed at low levels compared to the other ones.

To improve visualization, the mean marker expression per image can be displayed (Figure 10). This would help identify images with outlying marker expressions, highlighting the biological differences.

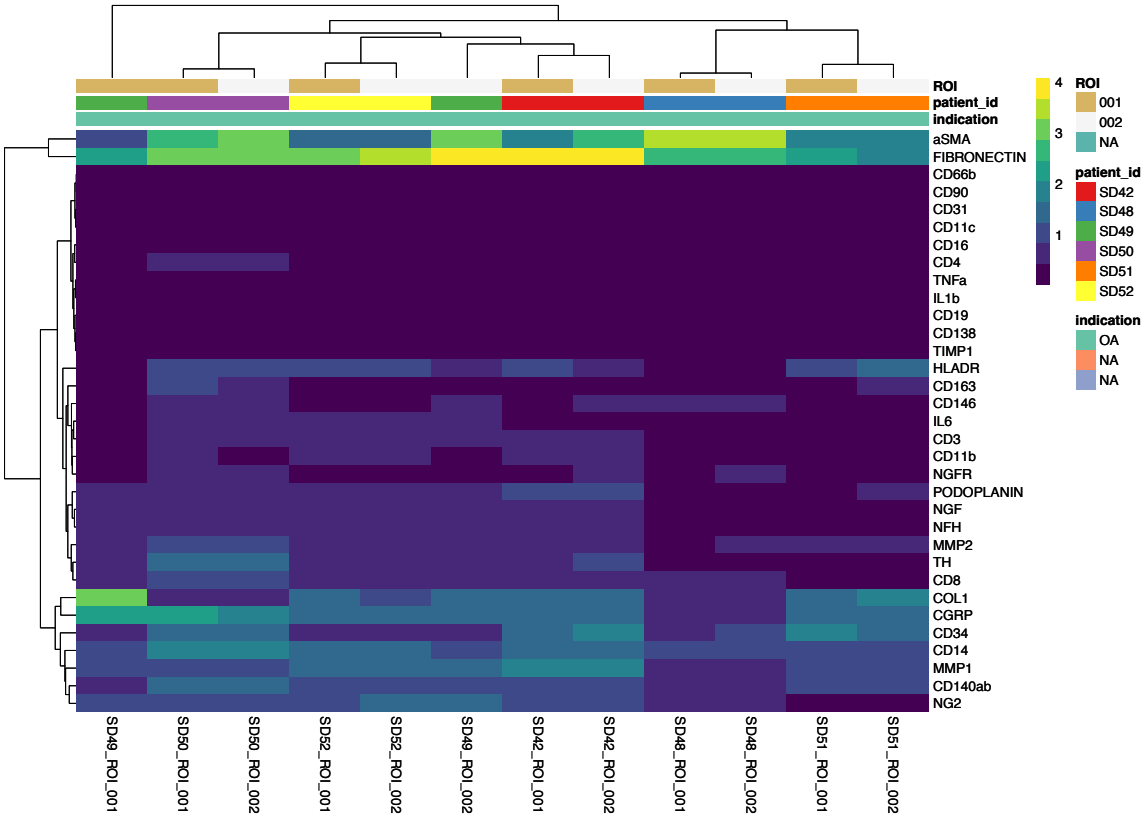


Figure 10: *Heatmap of mean marker expression per image. Blue denotes low expression, light blue intermediate expression, and yellow high expression. The upper part of the heatmap indicates ROIs, indication, and patient_id. The dendrograms on the left and on the top represent the similarity among markers and ROIs respectively.*

We found variable expression of each marker between patients and samples. For example, “SD49_ROI_001” and “SD49_ROI_002” show different marker expressions, even though they are samples from the same patient. In addition, the marker expression of patients SD49 and SD48 are completely different. This variability can be related to the different degrees of the disease in these patients.

To assess data quality at the single-cell level, we calculated the SNR as the ratio between the mean of the positive signal and the mean of the negative signal. As shown in Figure 11, we observe a relatively high signal intensity and a sufficient SNR ratio for all markers, except for DNA1 and DNA2

ones. These two markers likely have high intensity because some nuclei were considered as background.

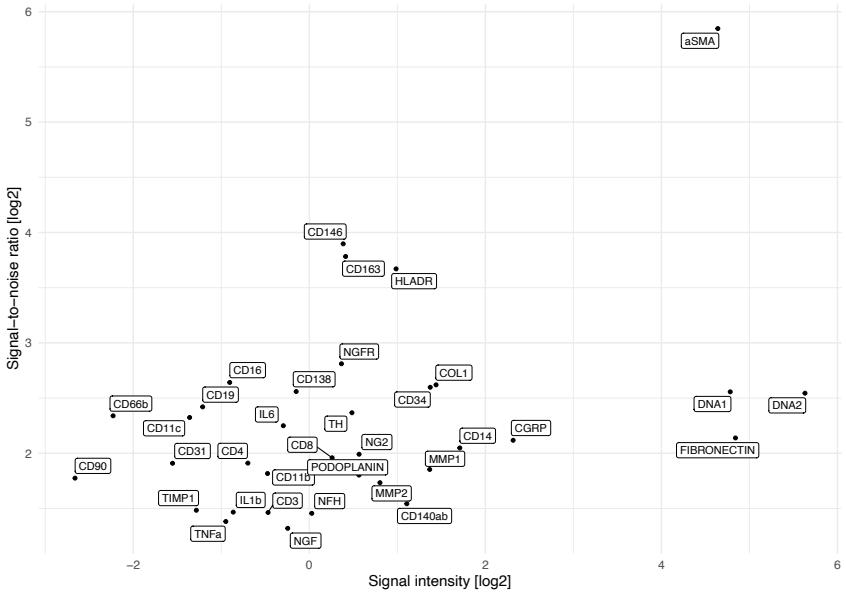
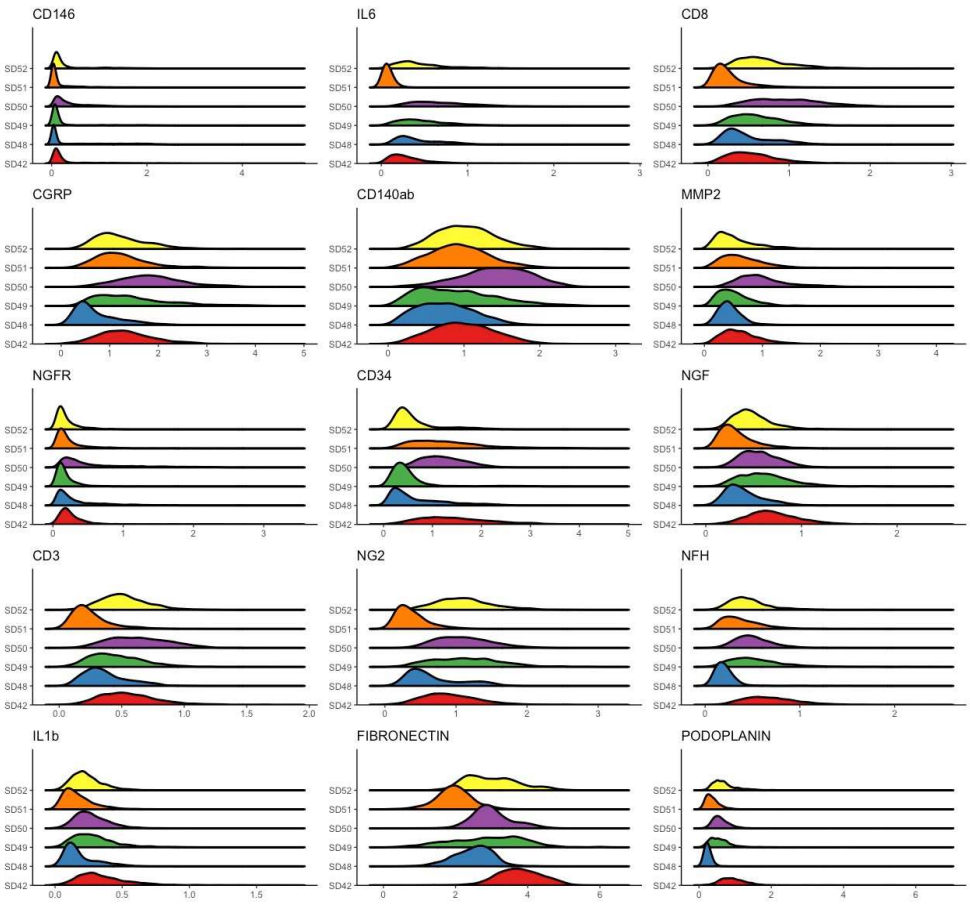


Figure 11: *Signal intensity and SNR ratio for various markers at single-cell level.*

As mentioned above, we found a variability among patients (Figure 10). To address it, we used ridgeline visualizations to compare the staining patterns of all markers across all samples (Figure 12).



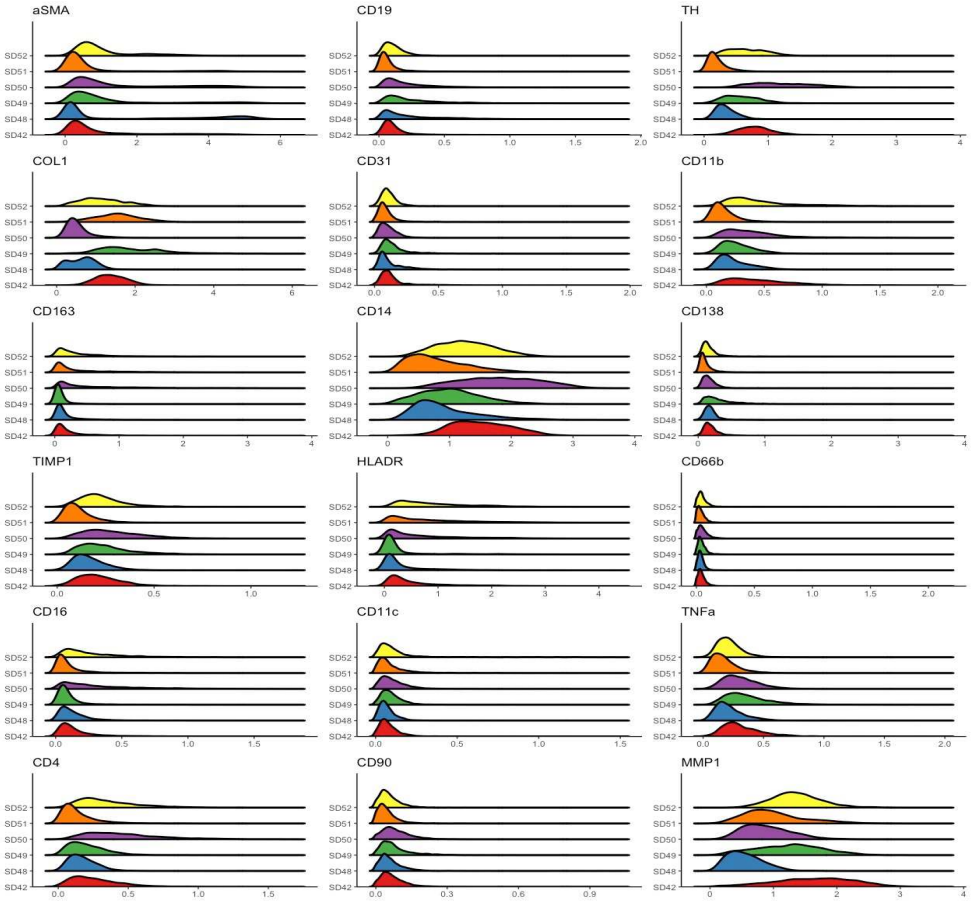


Figure 12: *Ridgeline visualization for marker expression among all patients.*

We found that differences in marker expression across patients can be ascribed to the different abundances of cells in each ROI.

Finally, a non-linear dimensionality reduction method was applied to project cells from high-dimensional down to low-dimensional space, allowing to visualize high dimensional data in a two bidimensional space. Within this step, we generated UMAP (Uniform Manifold Approximation and Projection) and TSNE (t-Distributed Stochastic Neighbor Embedding) graphs (Figure 13).

UMAP attempts to preserve the global structure of the data, better maintaining relative distances between data groups, which helps to interpret global relationships between clusters. On the contrary, TSNE focuses on preserving the local structure of the data, to maintain the similar point near to each other, thus making clusters clearly distinct.

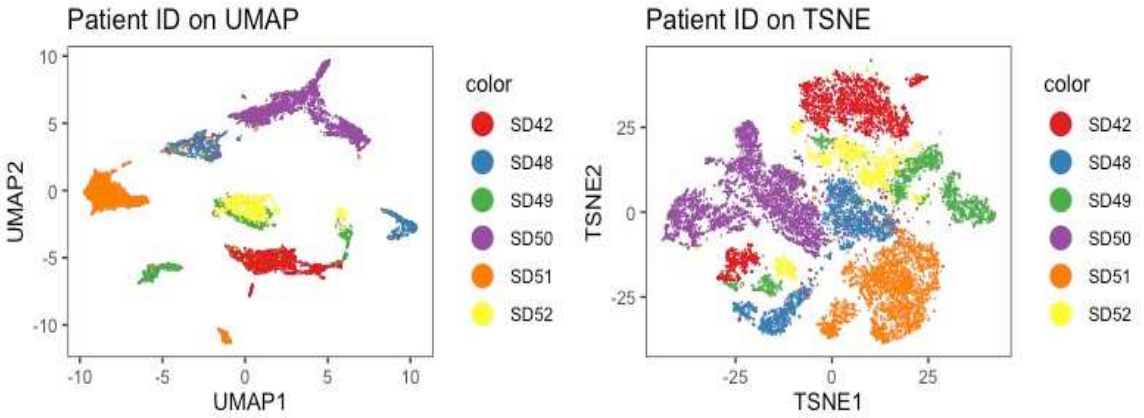


Figure 13: *Representation of UMAP, in the left part, and TSNE, in the right part, colored according to patient_id as indicated.*

Then, we performed a fastMNN correction, which identifies the MNN among cells from different samples and corrects differential expression between batches to align the data.

To assess the quality of the batch correction, the `batch.size` and `lost.var` entries are important. The `batch.size` entry reports the relative magnitude of the batch effect, while the `lost.var` entry represents the percentage of lost variance per merging step. A large `batch.size` and low `lost.var` indicate sufficient batch correction (Figure 14).

```
> merge_info[,c("left", "right", "batch.size")]
DataFrame with 5 rows and 3 columns
                left  right batch.size
              <List> <List>  <numeric>
1               SD50   SD48   0.558340
2          SD50,SD48   SD42   0.826151
3     SD50,SD48,SD42   SD52   0.861871
4 SD50,SD48,SD42,...   SD49   0.918032
5 SD50,SD48,SD42,...   SD51   0.839392
>
> merge_info$lost.var
            SD42        SD48       SD49        SD50       SD51        SD52
[1,] 0.000000000 0.039730142 0.00000000 0.025685791 0.00000000 0.000000000
[2,] 0.076755370 0.006260113 0.00000000 0.006591908 0.00000000 0.000000000
[3,] 0.003517314 0.002995257 0.00000000 0.004502786 0.00000000 0.112661570
[4,] 0.001664175 0.001323939 0.06820010 0.001483345 0.00000000 0.001791907
[5,] 0.013399824 0.008876651 0.02683737 0.011829576 0.08244903 0.010983250
```

Figure 14: *Batch.size and lost.var after fastMNN correction. The batch.size concerns to the initial part of the code, while the lost.var is related to the second part of the code. It is evident that the numbers for each patient are lower in the lost.var compared to the batch.size.*

As shown in the table generated with the code, the `lost.var` was lower compared to `batch.size`, indicating that the batch correction is effective. For a better and easier visualization of the correction, we generated a UMAP (Figure 15) which merged all patients.

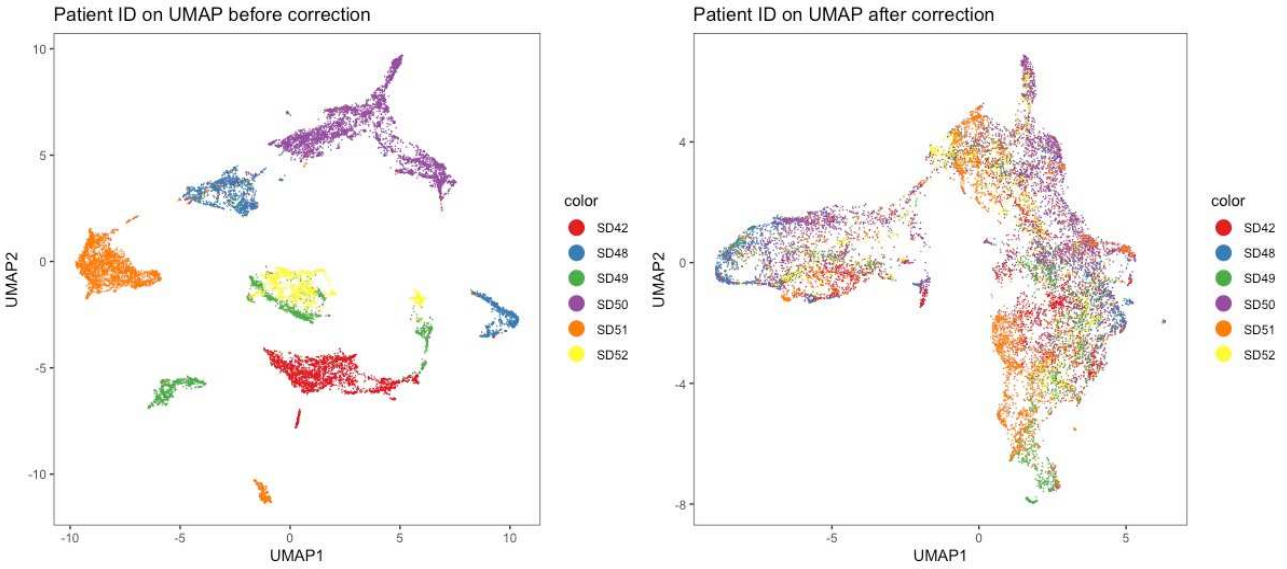
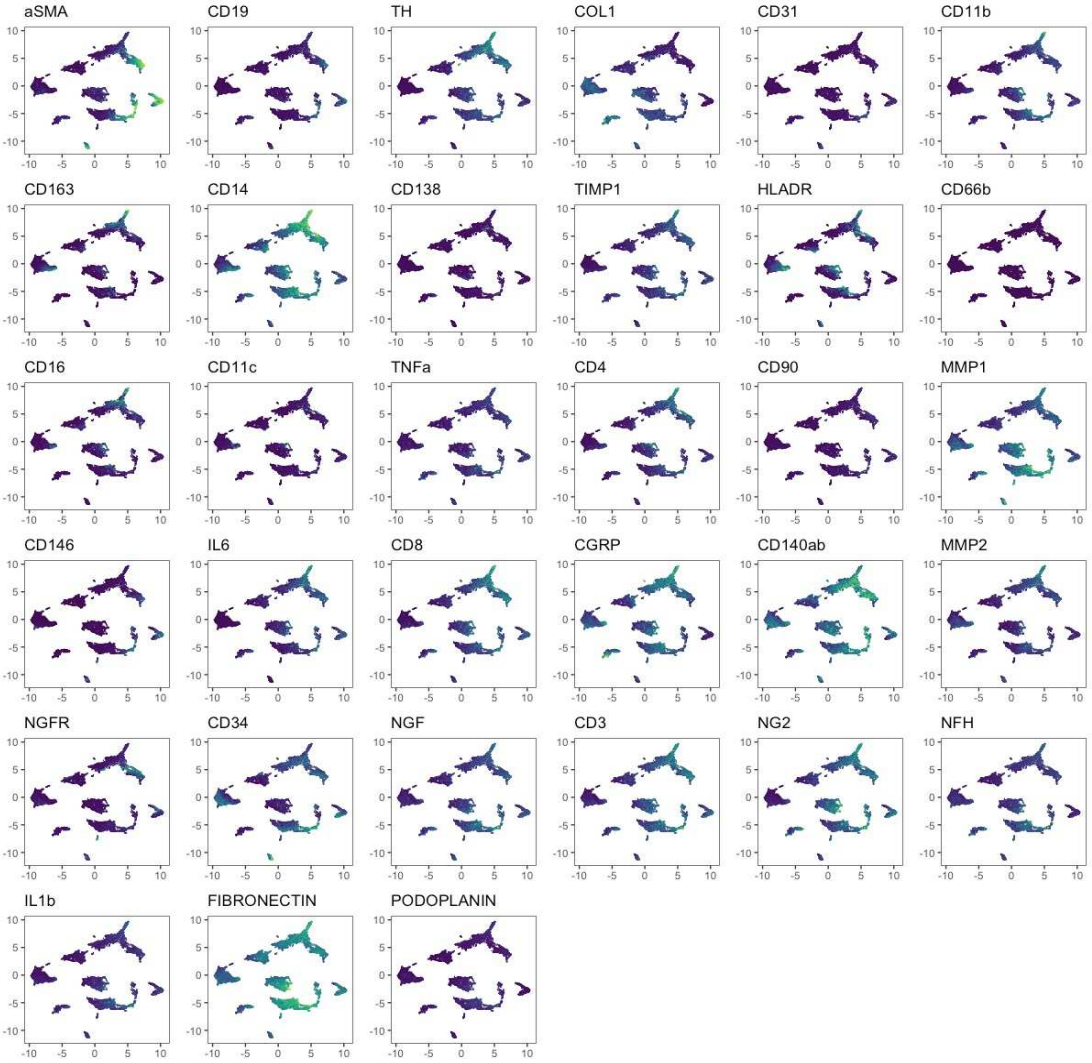


Figure 15: *UMAP before (left) and after (right) fastMNN correction, colored according to patient_id as indicated.*

The expression of markers in the UMAP before and after fastMNN correction can be visualized (Figure 16).

(**A**)



(**B**)

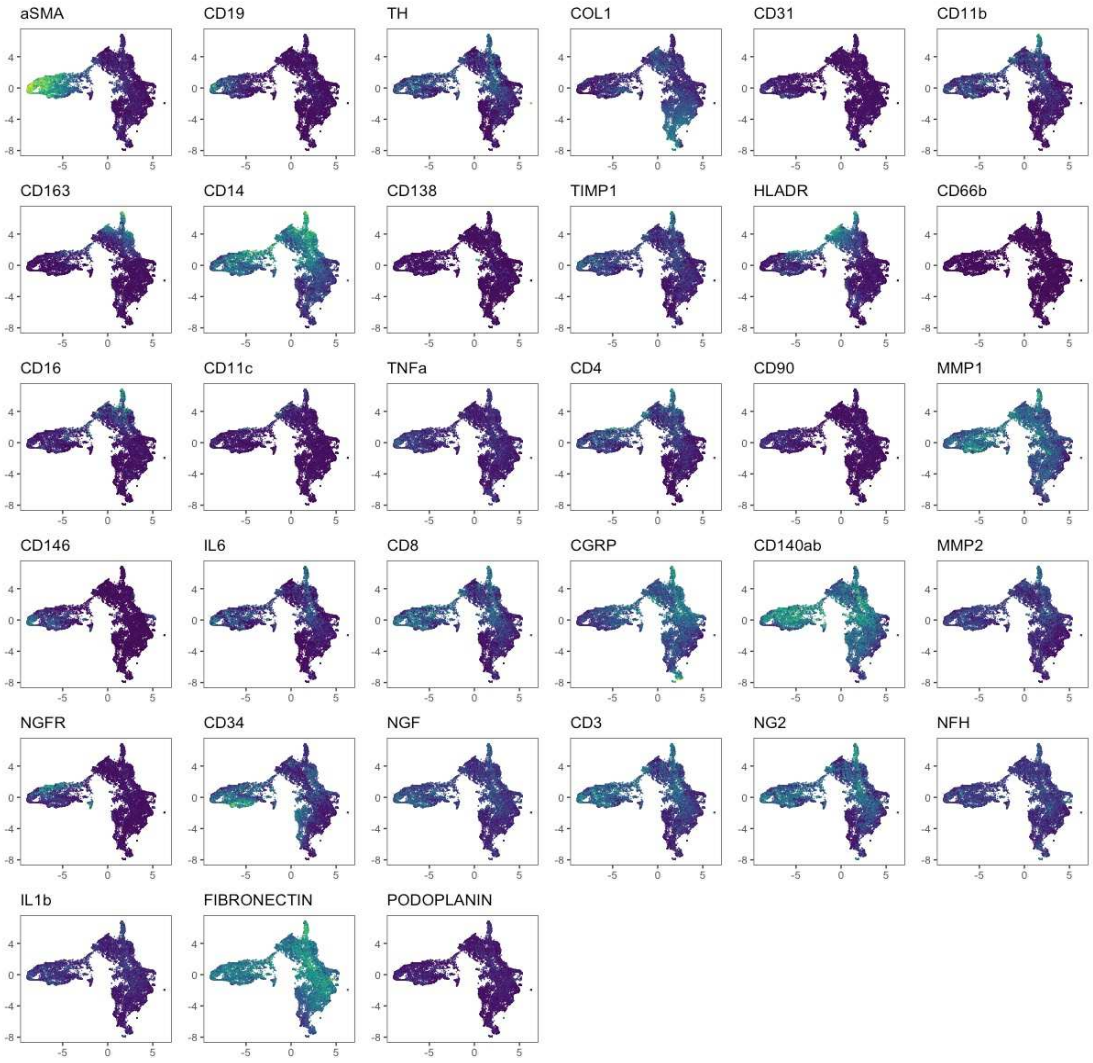


Figure 16: *UMAP of the expression markers (A) before and (B) after fastMNN correction. Blue represents lower expression level, while light green indicates marker expression. All markers are merged after batch correction.*

After fastMNN correction, all markers were merged without creating separate clusters, as well as all patients' data were well-integrated, as shown in Figure 15 where the UMAP after fastMNN correction showed a perfect overlap of patients.

The most important part involves cell phenotyping. It can be performed applying a semi-supervised or unsupervised approach. For this instance, we decided to use both approaches to compare the results and determine the best one based on the desired goal.

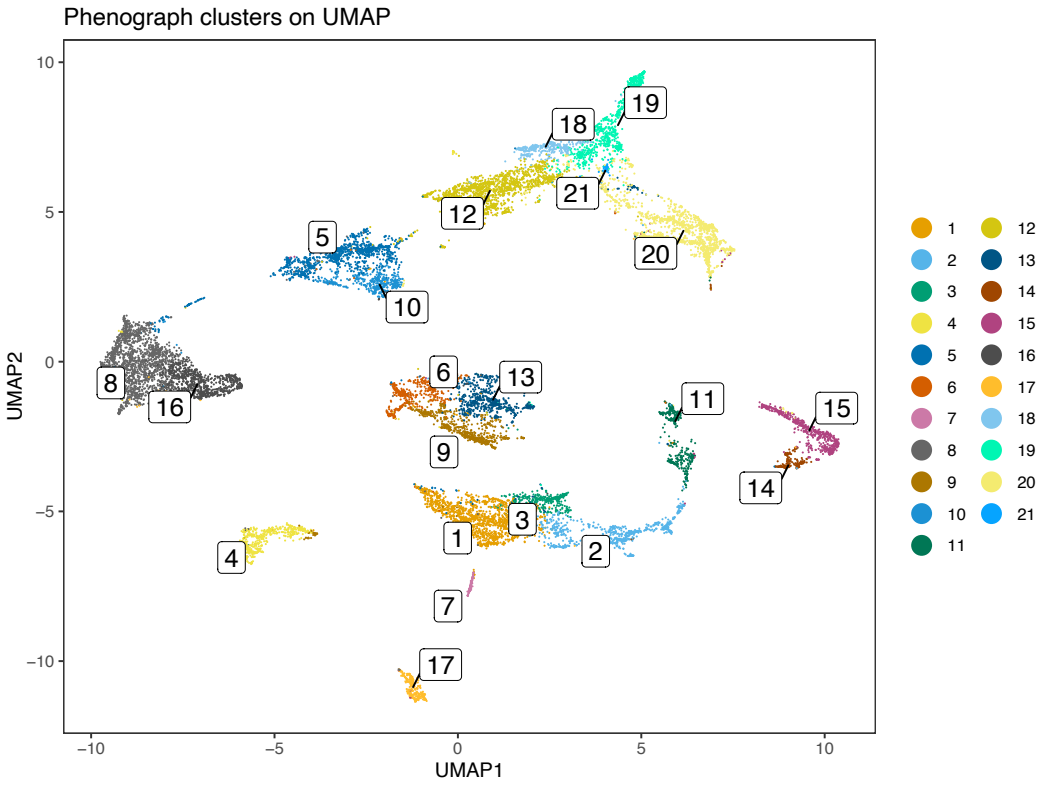
In a semi-supervised approach, we manually created clusters to align with our specific desired outcomes (known as gating cells). Conversely, in an unsupervised approach, clusters were formed autonomously by a clustering algorithm, such as PhenoGraph, without any manual intervention or predefined criteria.

4.2.1 Unsupervised Analysis

In an unsupervised analysis, the clustering approach groups cells based on their similarity in marker expression or by their proximity in low-dimensional space. For our purpose, we chose the PhenoGraph clustering approach which considers groups of cells and identifies the most similar cells based on their distances. This algorithm constructs a graph based on these similarities to ultimately form clusters of cells having similar characteristics.

For the PhenoGraph function, we choose 60 as “k” value (k = nearest neighbours) since this value reaches a plateau in the generation of clusters, resulting in the generation of 21 clusters. We generated the PhenoGraph clusters (Figure 17) on UMAP before and after the fastMNN correction.

(**A**)



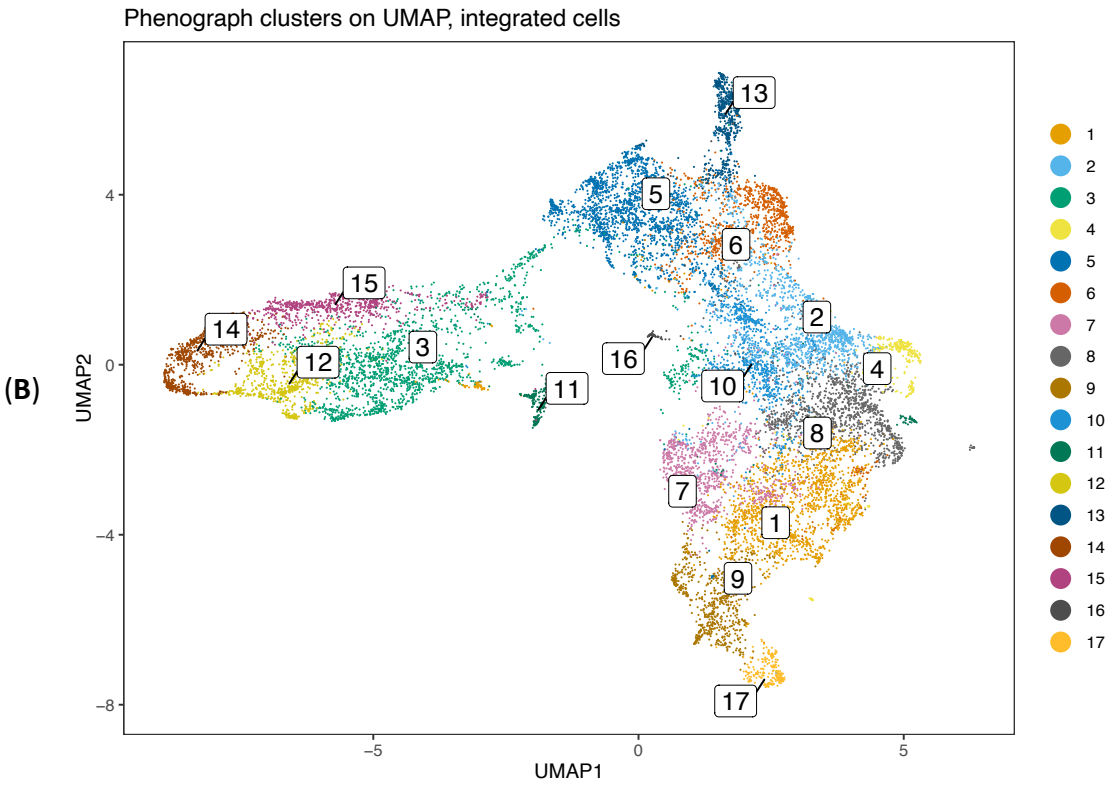


Figure 17: *Clusters obtained by PhenoGraph clustering approach. (A) UMAP visualization before applying PhenoGraph clustering approach and batch correction, showing 21 clusters. (B) UMAP visualization after applying PhenoGraphs approach and batch correction, showing 17 clusters.*

Cluster annotation has been performed by manually labelling each PhenoGraph generated clusters after fastMNN correction based on their marker expression visualized on heatmap (Figure 18).

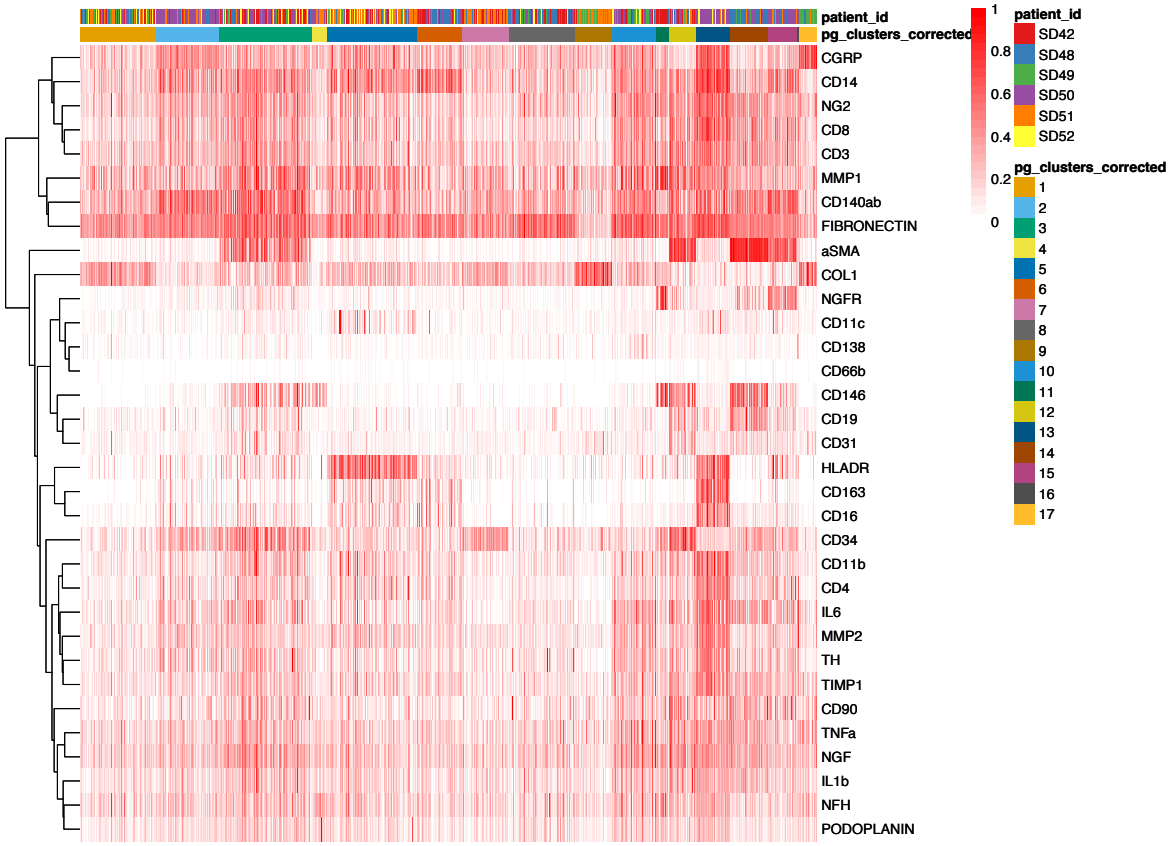


Figure 18: *Heatmap of marker expression after PhenoGraph clustering approach. The upper part of the heatmap indicates patient_id and pg_clusters_corrected, representing the 17 clusters generated after applying PhenoGraph and batch correction. Red color denotes marker expression, while white color indicates non-expression. The dendrograms on the left represent the similarity between marker expression.*

The clustering method allowed to group specific cells populations of interest such as FLSs, monocytes, macrophages, T cells, endothelial cells, pericytes, and stroma. Each cluster is named based on the expression of specific markers that represent a cell population:

- Stroma
- Pericytes
- Endothelium/Pericytes
- Classical Monocytes
- Monocytes
- Endothelium
- Pericytes/NGFR+/MMP1+
- Endothelium/Pericytes/FLS
- Endothelium/Pericytes/NGFR+
- T/Plasma/IL6/MMP2+
- Sensory nerve Fibers

Additionally, we labelled all cells not belonging to a defined cluster, as “undefined”.

Lastly, we generated a fastMNN-corrected UMAP (Figure 19) to visualize the disposition of the clusters.

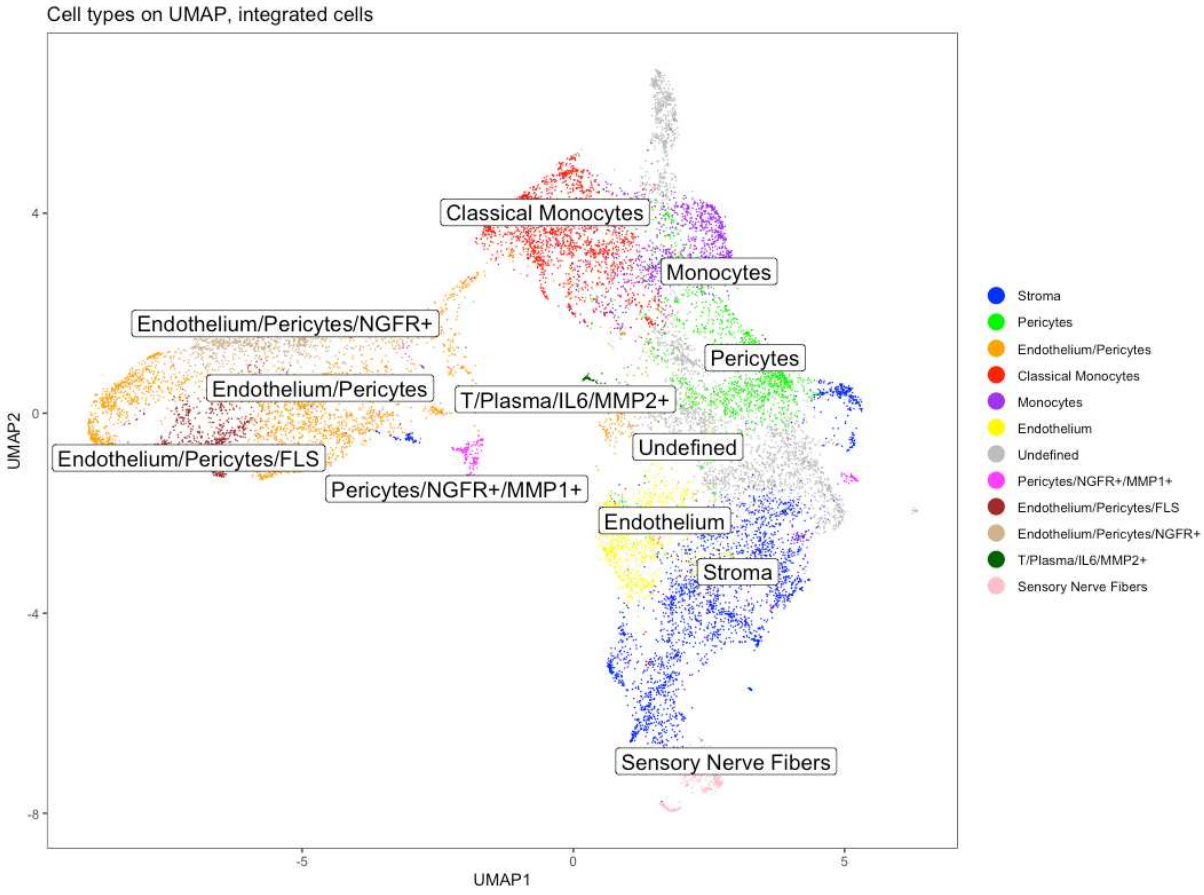


Figure 19: *fastMNN-corrected UMAP colored according to cell clusters as indicated.*

The disposition of clusters can be visualized in each ROI, with each point representing a specific cell that is part of a specific cluster (Figure 20).

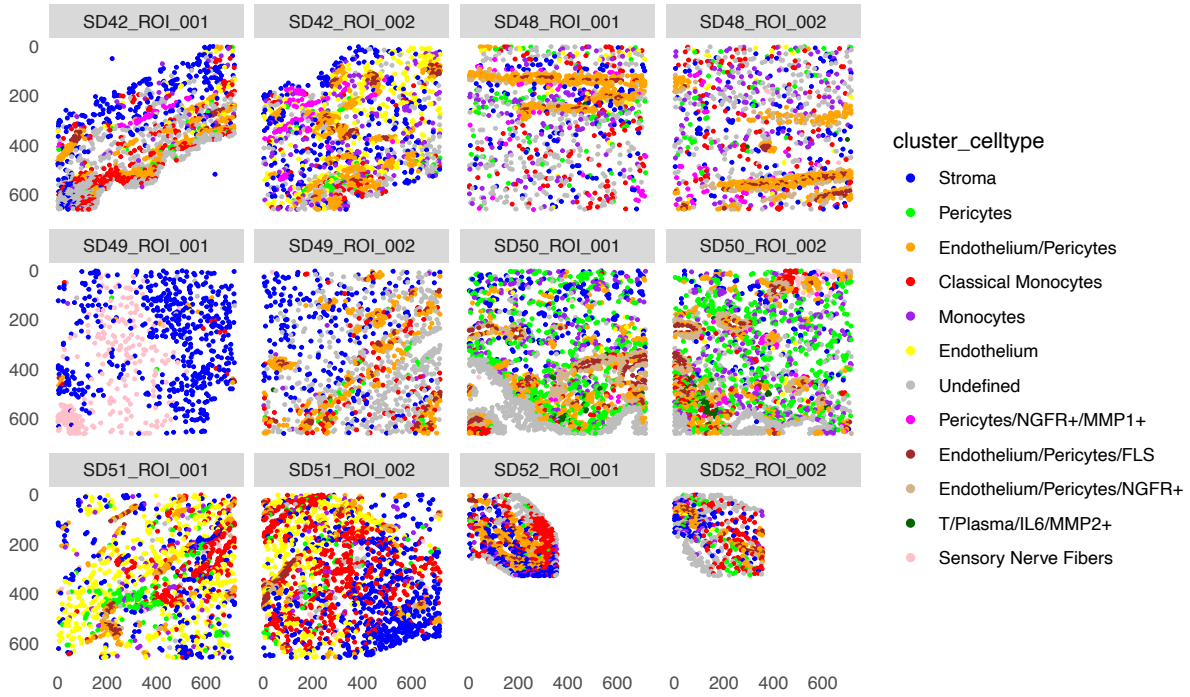
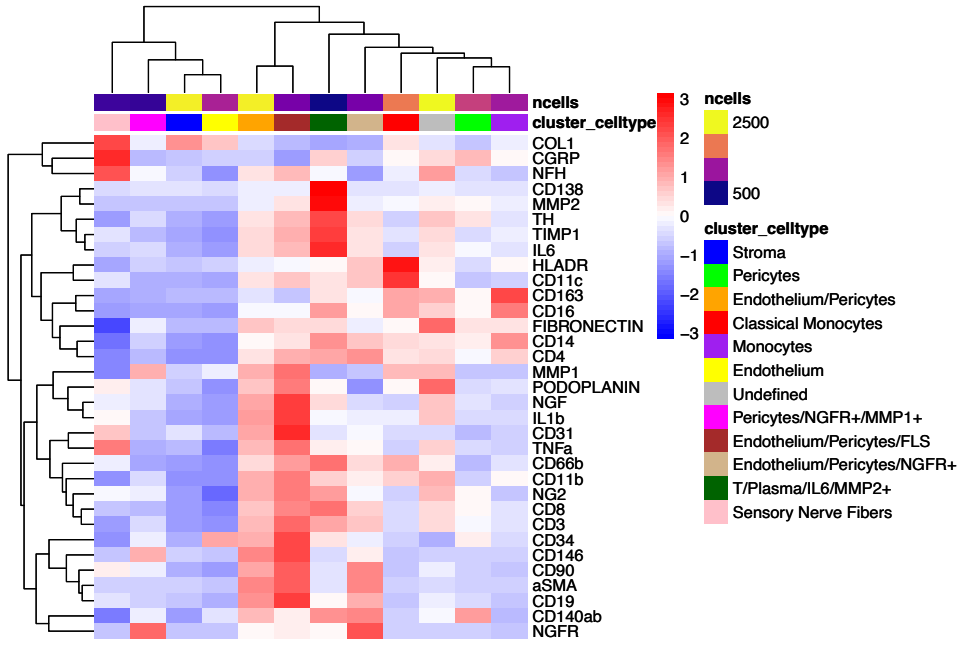


Figure 20: *Distribution of cells within specific cluster. Each image represents different ROIs of patients.*

For the visualization of each marker related to a specific cell type, we chose the Z-score scaling to better appreciate the differential expression of markers. In this scaling, those antibodies that were less expressed were marked in blue, while those markers that were higher expressed were coloured in red. We generated then two different graphs: one showing all the markers used in the panel (Figure 21A) and the other one showing our markers of interest (Figure 21B), which allowed us to visualize FLSs, monocytes, macrophages, T cells, endothelial cells, pericytes, and stroma populations.

(**A**)



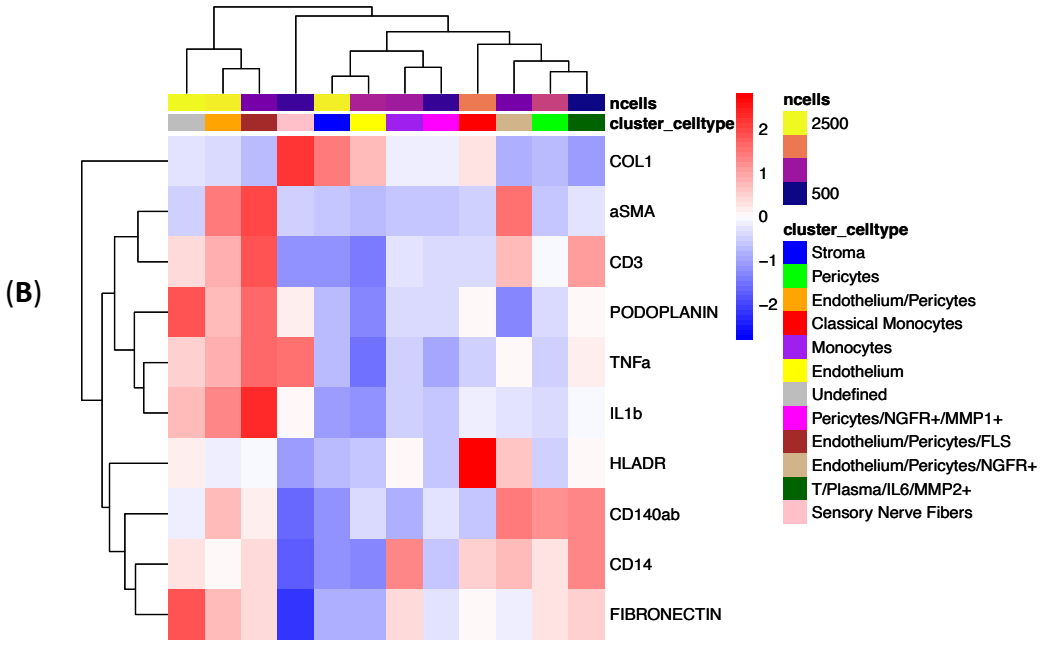


Figure 21: Heatmap with *Z-score scaling. (A) Z-score scaling with all markers. (B) Z-score scaling with markers of interest. The upper part of each heatmap indicates the number of cells (ncells), with blue representing fewer cells, purple representing a medium quantity, and yellow representing a high quantity of cells; it also indicates cluster_celltype, with each color corresponding to a different cluster. The dendrograms at the top and left represent the similarity between cell numbers and marker expression. In red are indicated markers with higher expression, while in blue markers with lower expression.*

As shown in Figure 20, we found differences related to the expression of some markers. For example, in Figure 20B the “Stroma” cluster (shown in blue) expressed only COL1 and not Fibronectin markers, while the “Endothelium/Pericytes” (shown in orange) and the “Endothelium/Pericytes/FLS” clusters (shown in dark red) expressed the markers specific to these populations, namely endothelial cells, pericytes and FLSs.

To display the frequencies of cell types per sample, we generated two different bar plots. The first plot (Figure 22A) had the indication of the disease on the X-axis, which in the same for all samples (OA). The second plot (Figure 22B) uses the patient_id on the X-axis, representing the name assigned to each ROI. The Y-axis in both cases represents the percentage of each cell type.

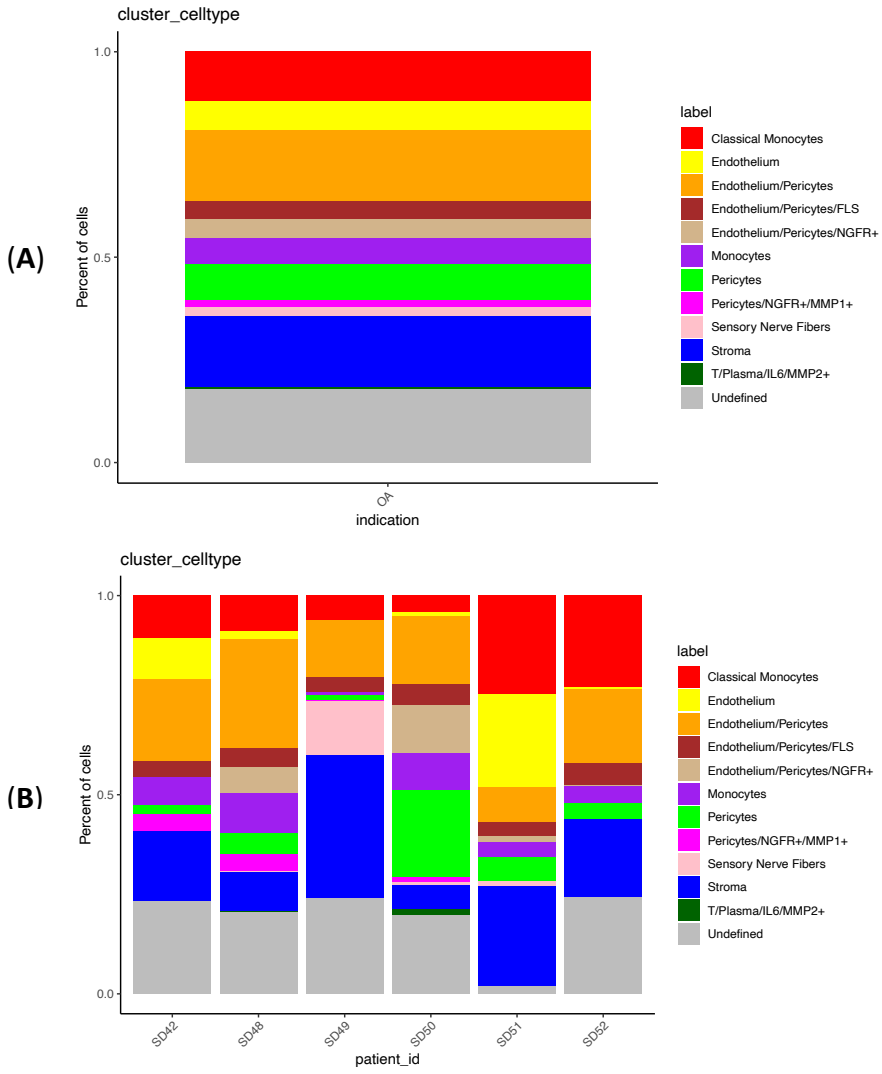


Figure 22: *Bar plot of cluster abundance. (A) Bar plot based on indication (OA). (B) Bar plot based on patient_id. In both bar plots, each color is associated with cluster names.*

Our findings confirmed that all patients expressed all clusters, except for the "T/Plasma/IL6/MMP2+ (shown in dark green) and the "Sensory Nerve Fibers" clusters (shown in light pink). This is probably due to the fact that we used different samples from different patients with a possible difference in the grade or severity of the disease. In addition, the consideration of only 2 ROIs for each patient could represent a factor of variability between each sample.

The analysis pipeline includes at the end a spatial analysis focusing on the interaction between all cell types in the dataset, particularly "from" cell type and "to" cell type (Figure 23), where red tiles indicate cell type pairs that significantly interacted in many images. Conversely, blue ones represent cell type pairs that tended to avoid each other in many images. These interactions can occur within the same cluster or between different clusters.

This function calculates how many times two cell types interact in different images and creates an average of these interactions. The cells were then grouped based on the previously generated cluster.

We can visualize the “from_label” (cell type from which the interaction originates) and the “to_label” (cell type which the interaction is destined to). The “sigval” indicates whether the cells significantly interact or avoid each other, or if there is no statistically significant interaction or avoidance.

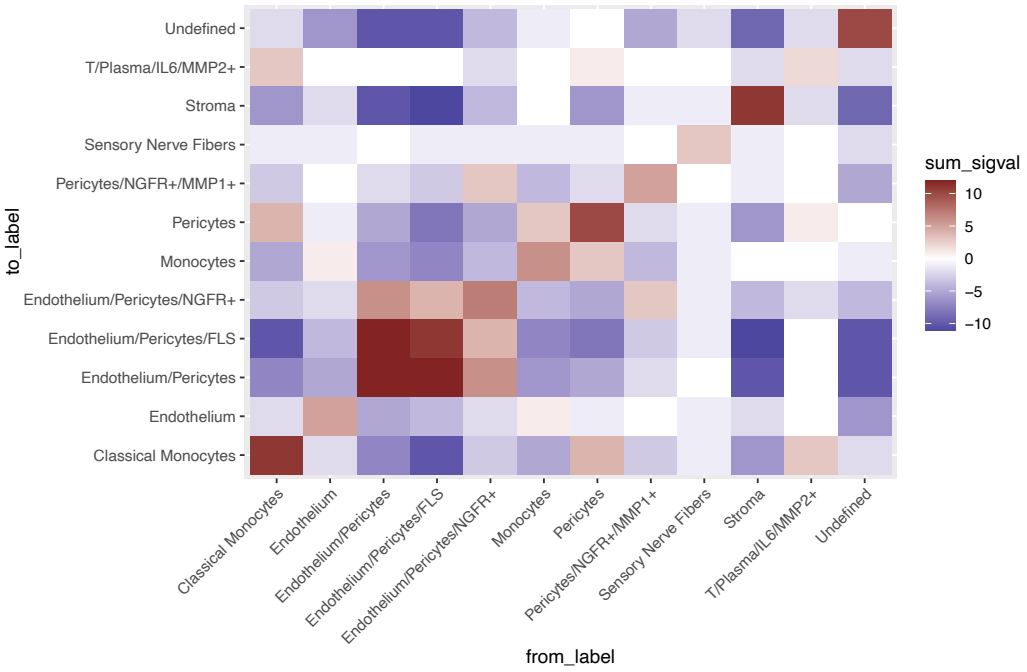


Figure 23: *Interaction graph. It shows the interactions between cell types, “from” cell type on X-axis and “to” cell type on Y-axis. Red tyles indicate cell type pairs that were detected to significantly interact on many images, while blue tyles show cell types that avoid each other in many images.*

In the interaction graphs, we observed that the “Endothelium/Pericytes” and “Endothelium/Pericytes/FLS” clusters interacted together and within themselves, while the “Stroma” cluster interacted only with itself.

Another possible way of interaction analysis was to use the patch method (Figure 24). The hypothesis to be tested in this approach is whether at least n cells of a certain type (to_label) are located around a target cell type (called "from_cell"). In summary, this approach checks whether a specific number of cells of a certain type are located around a target cell of another type.

The “patch_size” parameter in the function specifies the number of cells that must be present in the group (or "patch") for the hypothesis to be considered true.

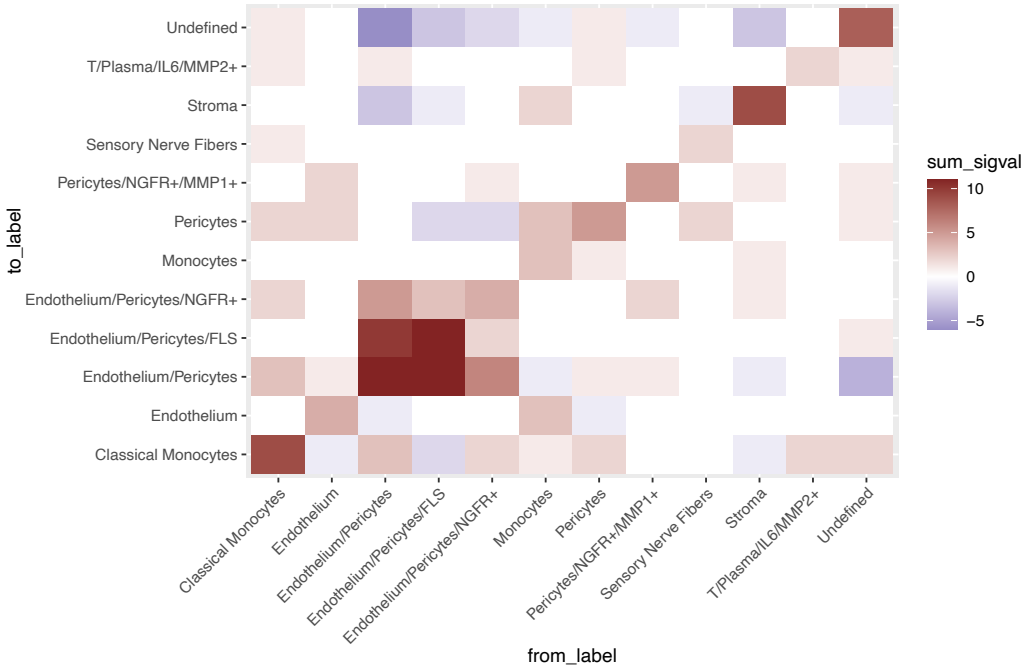


Figure 24: *Interaction graph based on patch method. It shows the interactions of n cell types (to_label) on Y-axis around a target cell type (from_label) on X-axis. Red tiles indicate cell type pairs that significantly interact in many images, while blue tiles show cell types that avoid each other in many images.*

The graph is comparable with the one generated before, shown in Figure 22. We observed that there is a strong interaction between “Endothelium/Pericytes/FLS” with the “Endothelium/Pericytes” and “Endothelium/Pericytes/NGFR+” clusters.

4.2.2 Semi-Supervised Analysis

The semi-supervised approach is based on ground truth labelling and random forest classification.

The *cytomapper* package provides a function that allows gating cells based on their marker expression and visualization of selected cells directly on images.

In this project, the choice of clusters was aimed at identifying cells expressing two cytokines: TNF-α and IL-1β.

The generated clusters were created taking into consideration these specific cells populations: FLS, monocytes, macrophages, T cells, endothelial cells, pericytes and stroma, detailed as following:

- FLS = IL-1β/TNFα negative + Podoplanin positive
- IL-1β_Peri_Endo = IL-1β positive + CD140ab/αSMA positive
- TNFα_Peri_Endo = TNF-α positive + αSMA/CD140ab positive

- Peri_Endo = IL-1β /TNF-α negative + CD140ab/αSMA positive
- IL1β_T = IL-1β positive + CD3 positive
- TNFα_T = TNF-α positive + CD3 positive
- T = IL-1β/TNF-α negative + CD3 positive
- Mono_IL1β = IL-1β positive + HLADR/CD14 positive
- TNFα_Mono = TNF-α positive + HLADR/CD14 positive
- Mono_Macro = IL-1β/TNF-α negative + HLADR/CD14 positive
- Stroma_Col1 = IL-1β/TNF-α negative + COL1 positive
- Stroma_Fibro = IL-1β/TNF-α negative + Fibronectin positive

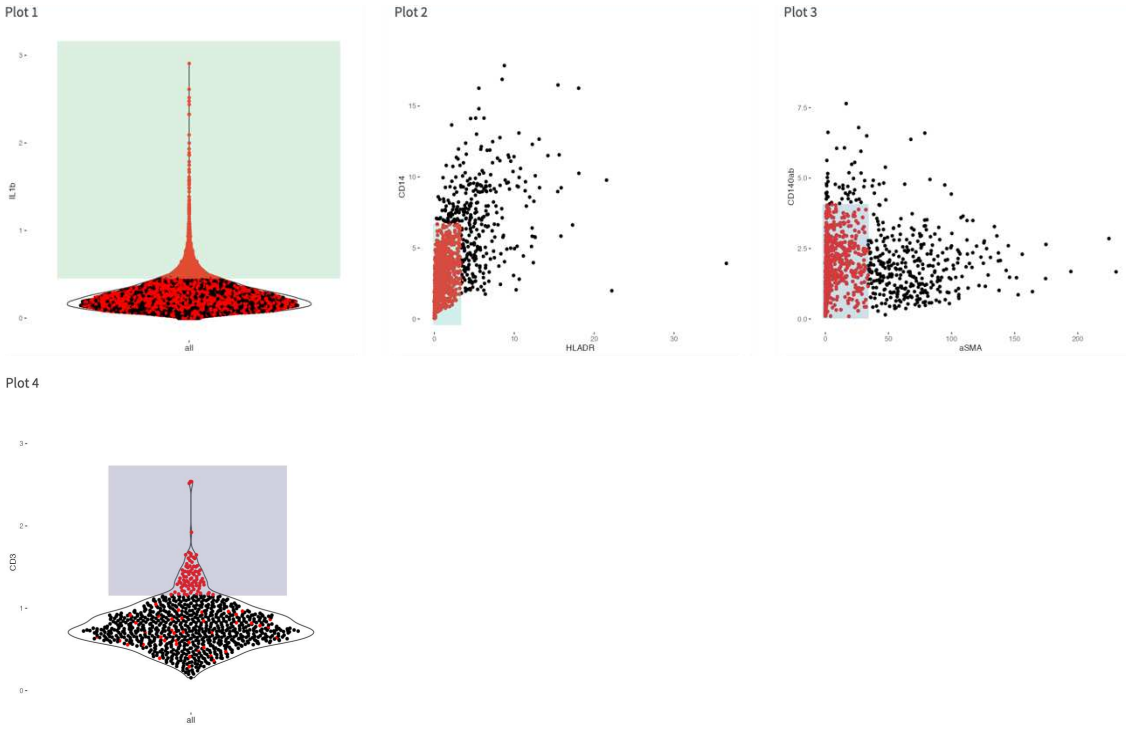


Figure 25: *Gating strategy. The image illustrates the gating strategy for the IL1β_T cluster, representing T cells positive for IL-1β expression. In the first plot, all cells expressing IL-1β are considered, followed by selection for double negativity for CD14 and HLA-DR, as well as for αSMA and CD140ab in subsequent plots. Plot 4 focuses on CD3-positive cells, resulting in the isolation of T cells expressing IL-1β.*

Additionally, we generated a cluster called “undefined” to include all cells that we are not specifically considering.

Firstly, we generated a table that identifies the labeled cells in each patient (Figure 26). Then, a random forest classifier was trained to classify all remaining cells as unlabelled cells.

	SD42	SD48	SD49	SD50	SD51	SD52
FLS	121	0	21	2	8	19
IL1b_Peri_Endo	77	6	5	1	2	0
IL1b_T	8	0	0	4	0	0
Macro_IL1b	14	0	4	17	4	1
Mono_Macro	93	31	47	212	376	124
Peri_endo	152	663	179	684	221	57
Stroma_Col1	6	0	249	1	92	6
Stroma_Fibro	798	155	289	123	33	308
T	10	0	3	122	1	8
TNFa_Mono	1	2	2	15	2	1
TNFa_Peri_Endo	3	1	18	7	1	0
TNFa_T	3	0	6	5	0	0
unlabeled	1472	1429	1097	2940	2315	758

Figure 26: *Table based on labeled cells.*

The data generated by the random forest were used to predict the cell phenotypes of the unlabelled data. The graph below illustrates the probability of finding each cell type in each cluster. To proceed with the analysis, it is important to choose a threshold that considers only cells after this parameter (Figure 27). Thus, we selected a maximum classification probability threshold of 0.4 (40%), because all clusters were considered without excluding any.

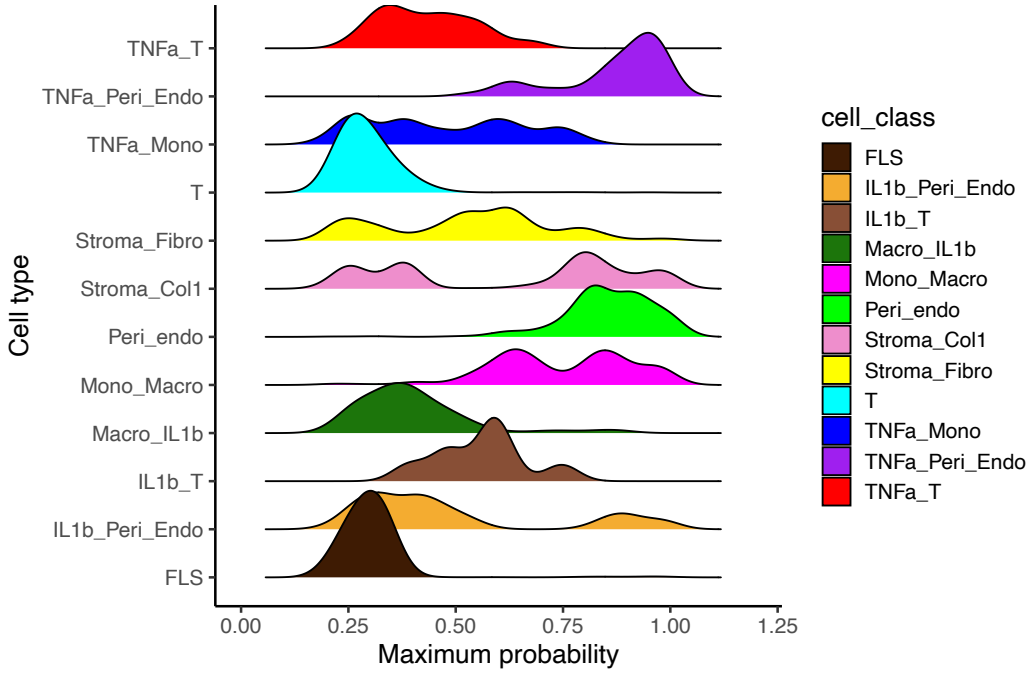


Figure 27: *Graphs of maximum probability. Each cluster is represented by a specific colour, with its probability spread along the X-axis. The selected maximum classification probability threshold is 0.4 (40%) because it takes in consideration the presence of all clusters.*

Thanks to the creation of these clusters through cell gating, it is now possible to generate the same graphs that were previously created using the unsupervised approach, but this time using the semi-supervised one.

The single-cell visualization was created through a fastMNN-corrected UMAP, based on semi-supervised generated clusters (Figure 28).

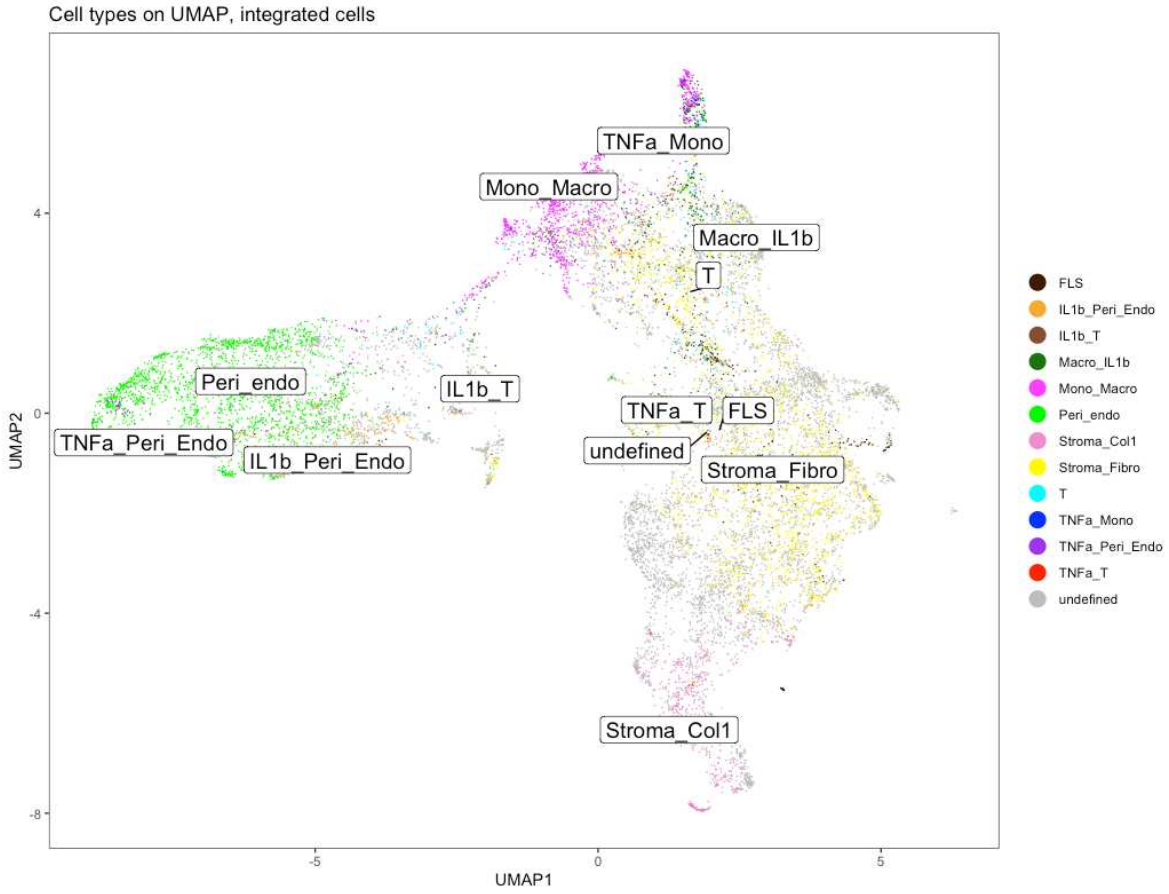


Figure 28: *fastMNN-corrected UMAP colored according to cell clusters as indicated.*

Clusters can be visualized in each ROI, where each point represents an individual cell belonging to a specific cluster (Figure 29).

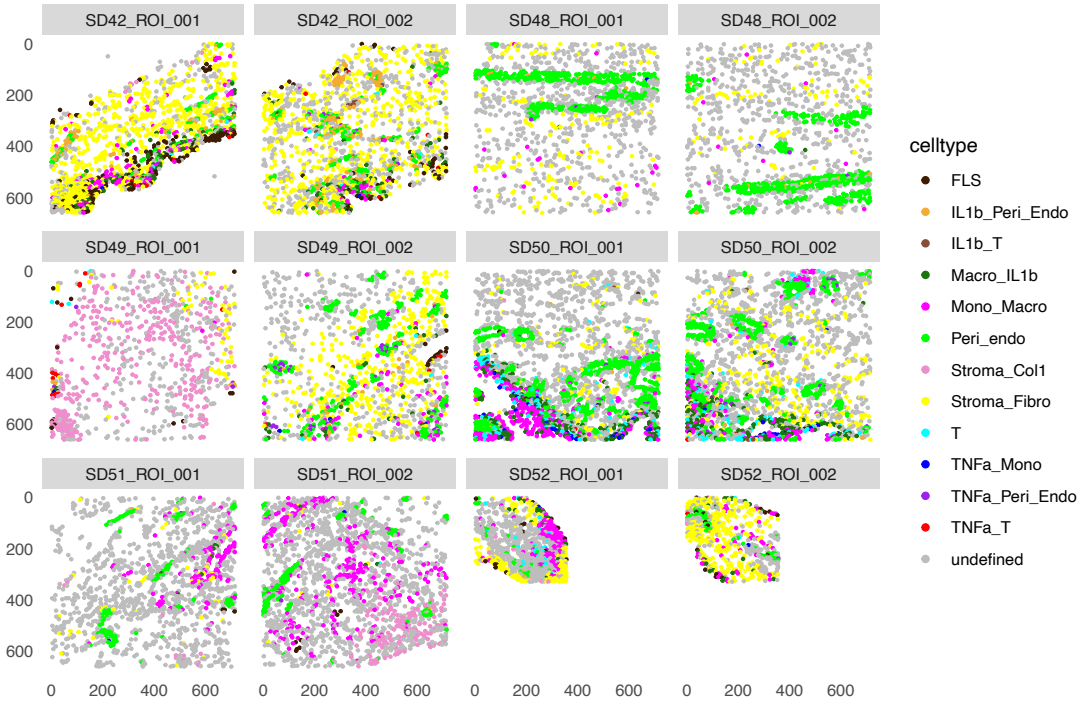


Figure 29: *Distribution of cells within specific cluster. Each image represents different ROIs of patients.*

To visualize the mean marker expression per cell type, we used Z-score scaling. As done previously, we generated two graphs: one with all markers used in the panel (Figure 30A) and the other showing the markers of interest (Figure 30B).

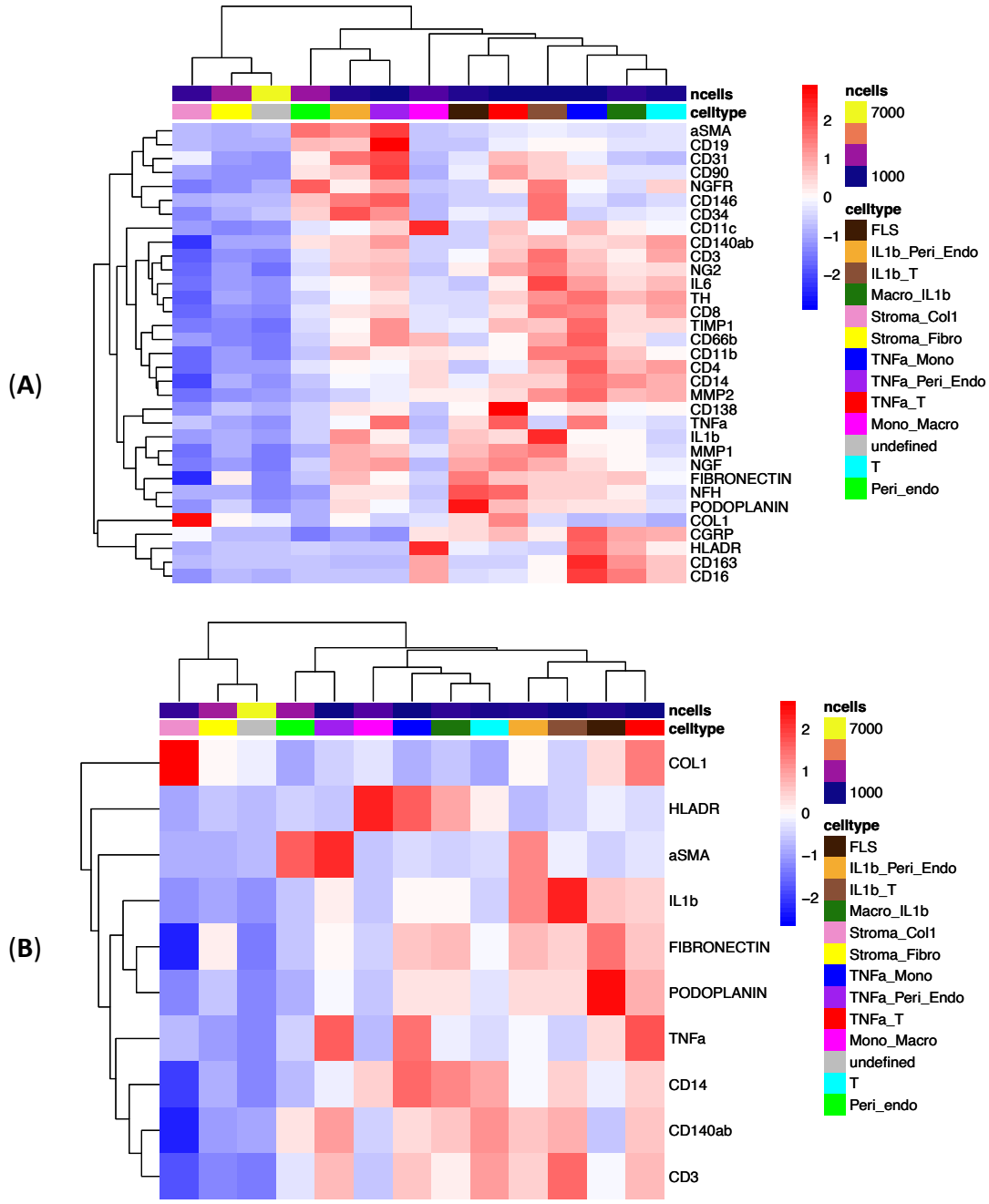


Figure 30: Heatmap with *Z-score scaling. (A) Z-score scaling with all markers. (B) Z-score scaling with the markers of interest. The upper part of each heatmap indicates the number of cells (ncells), with blue representing fewer cells, purple representing a medium quantity, and yellow representing a high quantity of cells; it also indicates celltype, with each color corresponding to a different cluster. The dendrograms at the top and left represent the similarity between cell numbers and marker expression. In red are indicated markers with higher expression, while in blue markers with lower expression.*

We found, as shown in Figure 27B, that the expression of CD3 marker was detectable in three clusters “TNFα_T” (red), “T” (cyan) and “IL1β_T” (light brown), as expected because CD3 is a marker of T cells and these clusters were created by considering only cells that were namely T cells, that might or not express TNF-α and IL-1β cytokines.

To display the frequencies of cell types per sample, we generated two different bar plots with the same characteristics as the graphs shown in Figure 22A and 22B. The first plot (Figure 31A) has the indication (OA) on X-axis, and the second plot (Figure 31B) has the patient_id on X-axis. In both cases, the Y-axis is represented the percentage of each cell type.

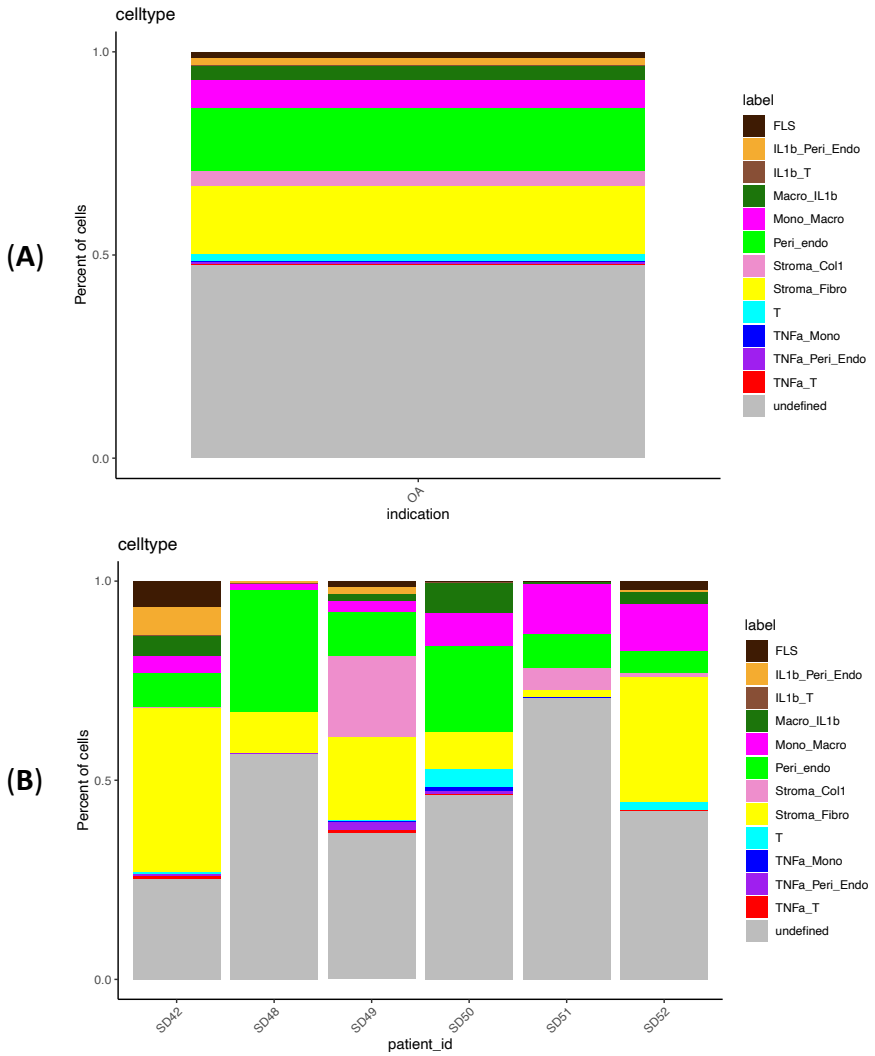


Figure 31: *Bar plots of cluster abundance. (A) Bar plot based on indication (OA). (B) Bar plot based on patient_id. In both bar plots, each color is associated with cluster names.*

In both graphs, we found several undefined cells (shown in grey). This was expected because we gated only specific cell populations (FLSs, monocytes, macrophages, T cells, endothelial cells, pericytes and stroma), so, all remaining cells were classified as “undefined”.

Additionally, we observed a similar distribution of cell types between patients, except for the “Strom_Col1” (pink), “IL1β_Peri_Endo” (ocher) and “FLS” (dark brown) clusters.

This analysis concludes with the spatial analysis, focusing on the interaction between all cell types in the dataset, with the same criteria previously described (Figure 32).

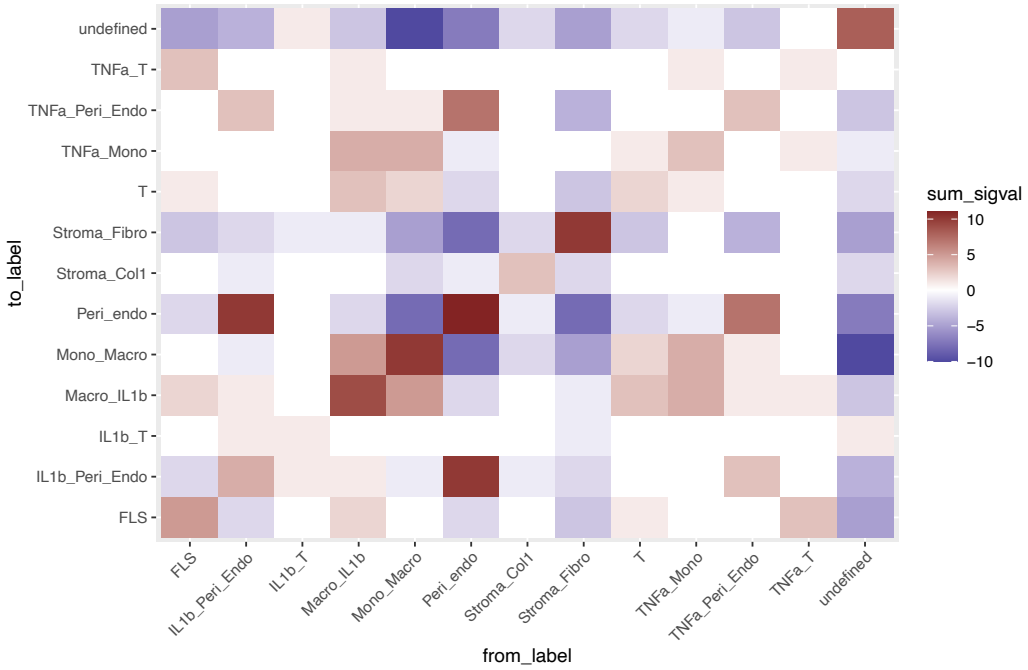


Figure 32: *Interaction graph. It shows the interactions between cell types, "from" cell type on X-axis and "to" cell type on Y-axis. Red tyles indicate cell type pairs that are detected to significantly interact on many images, while blue tyles show cell types that avoid each other in many images.*

In Figure 32 is possible to observe that "Macro_IL1b", "Mono_Macro", and "TNFα_Mono" clusters interact together, while "Stroma_Col1" and "Stroma_Fibro" clusters interact only with their respective labels.

In addition, the patch method (Figure 33) was used to check whether a specific number of cells of a certain type (to_label) were located around a target cell of another type (from_label).

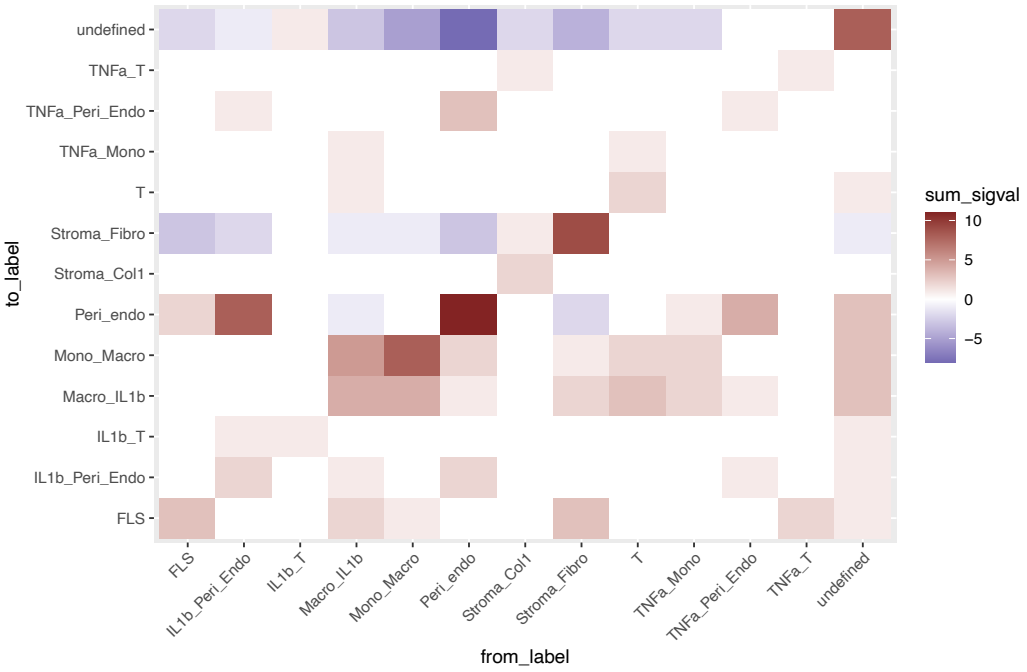


Figure 33: *Interaction graph based on patch method. It shows the interactions of n cell types (to_label) on Y-axis around a target cell type (from_label) on X-axis. Red tiles indicate cell type pairs that significantly interact in many images, while blue tiles show cell types that avoid each other in many images.*

Figure 33 is comparable with the one generated before (Figure 32). We found a strong interaction between the “Mono_Macro” cluster with the “Macro_IL1β” and “Peri_Endo”.

5. DISCUSSION

OA typically targets the synovium or synovial membrane, a highly vascularized and innervated connective tissue which regulates synovial fluid volume and composition, and chondrocytes maintenance. In OA patients, the synovium becomes thicker, more vascularized and there is an increase infiltration of inflammatory cells[1,15]. Among them, FLSs, lymphocytes, macrophages, pericytes, and nerve fibers are the predominant synovial cells, playing a crucial role in OA pathogenesis, thanks also to the involvement of soluble pro-inflammatory mediators and cytokines, which together contribute to the inflammatory process[19,3].

Identifying the aforementioned specific cell populations linked to OA pathogenesis is crucial for understanding the disease's cellular mechanisms and developing personalized therapies. Recent advancements in omics technologies, such as IMC, have shown great promise in understanding distinct cell populations within joint tissues of OA patients, paving the way for personalized therapy approaches[43].

Thus, this thesis aims to develop a bioinformatics pipeline for analysing IMC data, with the goal of elucidating cell composition and their spatial distribution in OA synovial samples.
IMC is an advanced technology enabling precise assessment of complex phenotypes and immune interactions in tissue microenvironments. It can overcome the limitations of traditional immunohistochemical analysis, such as the limited number of markers that can be visualized. IMC overcome this issue allowing the analysis of over 40 markers simultaneously and at single-cell level, thereby minimizing background noise and signal overlap[43].

We designed an antibody panel composed of 33 antibodies, able to characterize the infiltrated immune cells into synovial tissue as well as FLSs. The bioinformatic pipeline was build using IMC data from 6 OA samples. For each patient we selected 2 ROIs, and consequentially we analysed the data using the “IMC Segmentation Pipeline” provided by BodenmillerGroup.
Initially, a heatmap was created to visualize the mean expression of all markers. However, as shown in Figure 10, we found a high variability among patients, thus emphasizing the need for batch correction to eliminate non-biological differences between samples and enhance cell phenotype detection. The fastMNN correction method harmonized the data by identifying MNN among cells

from various samples and thus correcting differential expression between batches. Subsequently, in the UMAP (Figure 15), all patients were analyzed together; the expression of markers in Figure 16B shows their integration without distinct clusters being formed.

To build a bioinformatic pipeline, an important step involves cell phenotyping. For this purpose, we employed two different methods: unsupervised and semi-supervised. In both cases, we generated the same graphs to compare the outcomes of these two methods.
Using an unsupervised approach, we generated 17 clusters with the PhenoGraph function (Figure 17), in each cluster cells were grouped basing on their similarity in marker expression or their proximity in low-dimensional space, without any manual intervention or predefined criteria. These 17 clusters were then named based on marker expression shown in the heatmap (Figure 18), to identify populations of our interest such as FLSs, monocytes, macrophages, T cells, endothelial cells, pericytes and stroma.
On the other hand, in semi-supervised approach, cells were gated manually based on their marker expression and visualized directly on images. Using the *cytomapper* function, we generated 12 clusters, focusing on specific cells populations, such as FLSs, monocytes, macrophages, T cells, endothelial cells, pericytes and stroma, along with the expression of TNF-α and IL-1β in these populations.
In both cases, a cluster called “undefined” was created to group all cells that were not considered or could not be distinguished.

We used Z-score scaling to better appreciate the expression of markers in both unsupervised and semi-supervised approaches. In Figures 21B and 30B, we reported that the expression levels of the following markers such as COL-1, HLA-DR, αSMA, IL-1β, Fibronectin, Podoplanin, TNF-α, CD14, CD140ab, and CD3, did not change in both analyses. These markers identify our populations of interest (FLSs, monocytes, macrophages, T cells, endothelial cells, pericytes and stroma). However, we noted that these cells were better represented in the semi-supervised approach, which also shows the expression of the two cytokines of interest IL-1β and TNF-α. In contrast, when using the unsupervised approach, the expression of the two cytokines is not retained in the initial heatmap generated after applying the PhenoGraph function, as depicted in Figure 18. This initial heatmap shows that the cytokine expression data is lost, making it challenging to discern meaningful patterns. However, as shown in Figure 21B, the expression of these cytokines does reappear even within the

undefined cluster. This reappearance highlights a significant issue with the unsupervised approach: it is difficult to accurately distinguish and identify the specific populations of interest. The inability to discriminate between different populations correctly can lead to challenges in interpreting the data and identifying the correct population.

We also noted that FLSs were present in the heatmap generated by the semi-supervised approach; in contrast, in the unsupervised approach, the expression of these cells was found in the undefined clusters, indicating that with the unsupervised approach we may lose some important information.

However, these results suggested that both approaches can identify specific cells populations that are similar, taking into account, as shown in Figures 22 and 31, that the frequencies of the clusters in each patient were not homogeneous. This may be due to (a) the small number of ROIs chosen for each patient, and (b) the severity of OA that was unknow, which may be an important parameter to consider before applying a bioinformatic analysis. In line, Mimpen et al. (2023), reported different frequencies of immune and myeloid cells in 10 patients with advanced knee OA disease by immunofluorescence. They also observed a predominance of macrophages and T cells in each patient, and they were able to distinguish FLSs according to the positive or negative expression of CD34, CD90, Podoplanin and FAP markers[59,60].

As confirmed by Zou et al. (2023) and Mimpen et al. (2023), FLSs, macrophages and T cells are the most prevalent cells present in the synovium of OA patients. These cells play a crucial role in the inflammatory process affecting the synovial membrane. Specifically, FLSs and macrophages secrete large amounts of inflammatory factors, such as IL-1β and TNF-α, which are two of the most prevalent cytokines that promote synovial inflammation and possibly lead to cartilage lesions[60,61]. Our results showed that these two cytokines are expressed by pericytes, endothelial cells, monocytes and T cells populations[17,39]. In particular, as shown in Figure 31, by using the semi-supervised approach, the population most representative of IL-1β expression consists of macrophages. However, TNF-α shows a greater prevalence in the T cell populations, suggesting that while it is distributed among various cell types, its expression is more pronounced in T cells.

Therefore, by identifying more specifically which types of cells express these molecules, IL-1β and TNF-α can be targeted therapeutically, as the within the scope of SINPAIN project.

In addition, it could be interesting to better characterize the predominant cells population using specific markers and create specific clusters in the semi-supervised approach, as further prospective of this study.

Furthermore, based on our results, we can assess how the unsupervised approach is more suitable for exploratory analysis given its characteristic of random generation of clusters without the need of any input. Therefore, since no specific target is wanted, this method is perfectly suitable for the task. On the contrary, if the target is known, the semi-supervised approach may significantly outperform the unsupervised method, with the risk of losing some important information, regarding the co-expression of some markers on specific cells that we were not expecting. This method would provide information quicker and in a more specific way since inputs and outputs have already been given and the algorithm's only task is to fit the single-cell data into our indications.

Another important aspect in the bioinformatic pipeline to be considered, is the analysis of single cells in their spatial tissue. Ours allows for the analysis of interactions among all cell types using both semi-supervised and unsupervised approaches.

The interaction analysis shown in Figures 23, 24, 32 and 33 calculated the frequency of interactions between two cell types across different images by generating at the end an average of these interactions. The cells are then grouped based on the previously generated clusters. In these graphs, it is possible to observe that most clusters interact primarily with their respective cluster. However, a few clusters interacted with other ones, such as the “Macro_IL1β”, “Mono_Macro”, and “TNFα_Mono” clusters in the semi-supervised graph, and the “Endothelium/Pericytes” and “Endothelium/Pericytes/FLS” clusters in the unsupervised graph.

For the purposes of this thesis, the interaction analysis was sufficient to compare the two approaches. However, spatial analysis itself is far more intricate and precise, offering the potential to create more detailed and informative graphs aimed at achieving a deeper analysis.

Spatial analysis extends beyond examining interactions between cell types. There are several advanced approaches that can provide a more comprehensive understanding of cellular behavior and interactions within a sample. For instance, spatial community analysis can reveal how cells form distinct communities are spatially organized, providing insights into the structural organization of tissues. Cellular neighborhood analysis also allows for the examination of the immediate microenvironment of each cell, identifying how neighboring cells may influence each other’s behavior and function. Additionally, spatial context analysis can uncover patterns of cell distribution and localization within the broader tissue context, highlighting areas of cellular cooperation or conflict.

These sophisticated spatial analyses are crucial for a detailed and comprehensive view of cellular interactions and behaviours. They enable researchers to identify precise patterns and intricate networks that simpler methods might overlook. Emphasizing the importance of these sophisticated spatial analyses underscores their potential to provide significant insights and drive advancements in the field.

In conclusion, we have developed a detailed bioinformatics pipeline capable of deep characterization of the synovial membrane. This serves as a fundamental step for conducting further analysis on larger datasets to gain a deeper understanding in OA pathogenesis. Moreover, exploring interindividual variability among patients with the same disease could elucidate specific cell populations crucial to the pathogenetic process. This could lead to the development of new care strategies, prediction of treatment response, and guidance for precision medicine treatment plans.

6. BIBLIOGRAPHY

1. Mathiessen, A., and Conaghan, P.G. (2017). Synovitis in osteoarthritis: Current understanding with therapeutic implications. Preprint at BioMed Central Ltd., https://doi.org/10.1186/s13075-017-1229-9.
2. Maglaviceanu, A., Wu, B., and Kapoor, M. (2021). Fibroblast-like synoviocytes: Role in synovial fibrosis associated with osteoarthritis. Wound Repair and Regeneration *29*, 642–649. https://doi.org/10.1111/wrr.12939.
3. Monemdjou, R., Fahmi, H., and Kapoor, M. (2010). Synovium in the pathophysiology of osteoarthritis. Preprint, https://doi.org/10.2217/thy.10.72.
4. Scanzello, C.R., and Goldring, S.R. (2012). The role of synovitis in osteoarthritis pathogenesis. Preprint, https://doi.org/10.1016/j.bone.2012.02.012.
5. Smith, M.D. (2011). The Normal Synovium.
6. Wu, Z., Ma, D., Yang, H., Gao, J., Zhang, G., Xu, K., and Zhang, L. (2021). Fibroblast-like synoviocytes in rheumatoid arthritis: Surface markers and phenotypes. Preprint at Elsevier B.V., https://doi.org/10.1016/j.intimp.2021.107392.
7. Branco De Sousa, E., Ladeira Casado, P., Neto, V.M., Duarte, M.E.L., and Aguiar, D.P. (2014). Synovial fluid and synovial membrane mesenchymal stem cells: latest discoveries and therapeutic perspectives.
8. Allen, K.D., Thoma, L.M., and Golightly, Y.M. (2022). Epidemiology of osteoarthritis. Osteoarthritis Cartilage *30*, 184–195. https://doi.org/10.1016/j.joca.2021.04.020.
9. Sharma, L. (2021). Osteoarthritis of the Knee. New England Journal of Medicine *384*, 51–59. https://doi.org/10.1056/NEJMcp1903768.
10. Vina, E.R., and Kwoh, C.K. (2018). Epidemiology of osteoarthritis: Literature update. Preprint at Lippincott Williams and Wilkins, https://doi.org/10.1097/BOR.0000000000000479.
11. Gill, T.M. (2017). Do the tenets of late-life disability apply to middle age? Preprint at American College of Physicians, https://doi.org/10.7326/M17-2550.
12. Hawker (2019). Osteoarthritis is a serious disease. Clin Exp Rheumatol.
13. Neogi, T., and Zhang, Y. (2013). Epidemiology of Osteoarthritis. Preprint, https://doi.org/10.1016/j.rdc.2012.10.004.
14. Barbour, K.E., Hootman, J.M., Helmick, C.G., Murphy, L.B., Theis, K.A., Schwartz, T.A., Kalsbeek, W.D., Renner, J.B., and Jordan, J.M. (2014). Meeting physical activity guidelines and the risk of

incident knee osteoarthritis: A population-based prospective cohort study. Arthritis Care Res (Hoboken) *66*, 139–146. https://doi.org/10.1002/acr.22120.

15. Wenham, C.Y.J., and Conaghan, P.G. (2010). The role of synovitis in osteoarthritis. Preprint, https://doi.org/10.1177/1759720X10378373.
16. Yulin Chen, Wei Jiang, Huang Yong, Miao He, Yuntao Yang, Zhenhan Deng, and Yusheng Li (2020). Macrophages in osteoarthritis: pathophysiology and therapeutics.
17. Sanchez-Lopez, E., Coras, R., Torres, A., Lane, N.E., and Guma, M. (2022). Synovial inflammation in osteoarthritis progression. Preprint at Nature Research, https://doi.org/10.1038/s41584-022-00749-9.
18. Prieto-Potin, I., Largo, R., Roman-Blas, J.A., Herrero-Beaumont, G., and Walsh, D.A. (2015). Characterization of multinucleated giant cells in synovium and subchondral bone in knee osteoarthritis and rheumatoid arthritis. BMC Musculoskelet Disord *16*. https://doi.org/10.1186/s12891-015-0664-5.
19. Sokolove, J., and Lepus, C.M. (2013). Role of inflammation in the pathogenesis of osteoarthritis: Latest findings and interpretations. Ther Adv Musculoskelet Dis *5*, 77–94. https://doi.org/10.1177/1759720X12467868.
20. Krishnan, H., Rayes, J., Miyashita, T., Ishii, G., Retzbach, E.P., Sheehan, S.A., Takemoto, A., Chang, Y.W., Yoneda, K., Asai, J., et al. (2018). Podoplanin: An emerging cancer biomarker and therapeutic target. Preprint at Blackwell Publishing Ltd, https://doi.org/10.1111/cas.13580.
21. Croft, A.P., Campos, J., Jansen, K., Turner, J.D., Marshall, J., Attar, M., Savary, L., Wehmeyer, C., Naylor, A.J., Kemble, S., et al. (2019). Distinct fibroblast subsets drive inflammation and damage in arthritis. Nature *570*, 246–251. https://doi.org/10.1038/s41586-019-1263-7.
22. Mizoguchi, F., Slowikowski, K., Wei, K., Marshall, J.L., Rao, D.A., Chang, S.K., Nguyen, H.N., Noss, E.H., Turner, J.D., Earp, B.E., et al. (2018). Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. Nat Commun *9*. https://doi.org/10.1038/s41467-018-02892-y.
23. Li, Y. sheng, Luo, W., Zhu, S.A., and Lei, G.H. (2017). T cells in osteoarthritis: Alterations and beyond. Preprint at Frontiers Research Foundation, https://doi.org/10.3389/fimmu.2017.00356.
24. Wilkinson, M.G.L., and Rosser, E.C. (2019). B Cells as a Therapeutic Target in Paediatric Rheumatic Disease. Preprint at NLM (Medline), https://doi.org/10.3389/fimmu.2019.00214.

25. Zhu, W., Zhang, X., Jiang, Y., Liu, & X., Huang, L., Wei, Q., Huang, Y., Wu, W., and Gu, J. (2019). Alterations in peripheral T cell and B cell subsets in patients with osteoarthritis. https://doi.org/10.1007/s10067-019-04768-y/October.

26. Liu, B., Zhang, M., Zhao, J., Zheng, M., and Yang, H. (2018). Imbalance of M1/M2 macrophages is linked to severity level of knee osteoarthritis. Exp Ther Med *16*, 5009–5014. https://doi.org/10.3892/etm.2018.6852.

27. Mushenkova, N. V., Nikiforov, N.G., Shakhpazyan, N.K., Orekhova, V.A., Sadykhov, N.K., and Orekhov, A.N. (2022). Phenotype Diversity of Macrophages in Osteoarthritis: Implications for Development of Macrophage Modulating Therapies. Preprint at MDPI, https://doi.org/10.3390/ijms23158381.

28. Collett, G.D.M., and Canfield, A.E. (2005). Angiogenesis and pericytes in the initiation of ectopic calcification. Preprint, https://doi.org/10.1161/01.RES.0000163634.51301.0d.

29. Negri, S., Li, Z., Wang, Y., Cherief, M., Qin, Q., Xu, J., Salazar, M.A.G., Hsu, G.C.-Y., and James, A.W. (2022). The Role of Pericytes in the Articular Environment—a Review. Curr Tissue Microenviron Rep *3*, 77–82. https://doi.org/10.1007/s43152-022-00041-8.

30. James, A.W., Hindle, P., Murray, I.R., West, C.C., Tawonsawatruk, T., Shen, J., Asatrian, G., Zhang, X., Nguyen, V., Simpson, A.H., et al. (2017). Pericytes for the treatment of orthopedic conditions. Preprint at Elsevier Inc., https://doi.org/10.1016/j.pharmthera.2016.08.003.

31. Su, W., Liu, G., Liu, X., Zhou, Y., Sun, Q., Zhen, G., Wang, X., Hu, Y., Gao, P., Demehri, S., et al. (2020). Angiogenesis stimulated by elevated PDGF-BB in subchondral bone contributes to osteoarthritis development. JCI Insight *5*. https://doi.org/10.1172/jci.insight.135446.

32. Raychaudhuri, S.P., Raychaudhuri, S.K., Atkuri, K.R., Herzenberg, L.A., and Herzenberg, L.A. (2011). Nerve growth factor: A key local regulator in the pathogenesis of inflammatory arthritis. Arthritis Rheum *63*, 3243–3252. https://doi.org/10.1002/art.30564.

33. Raychaudhuri, S.P., and Raychaudhuri, S.K. (2009). The regulatory role of nerve growth factor and its receptor system in fibroblastlike synovial cells. Scand J Rheumatol *38*, 207–215. https://doi.org/10.1080/03009740802448866.

34. Aso, K., Walsh, D.A., Wada, H., Izumi, M., Tomitori, H., Fujii, K., and Ikeuchi, M. (2022). Time course and localization of nerve growth factor expression and sensory nerve growth during progression of knee osteoarthritis in rats. Osteoarthritis Cartilage *30*, 1344–1355. https://doi.org/10.1016/j.joca.2022.07.003.

35. Grässel, S., and Muschter, D. (2017). Peripheral nerve fibers and their neurotransmitters in osteoarthritis pathology. Preprint at MDPI AG, https://doi.org/10.3390/ijms18050931.

36. Walsh, D.A., Mapp, P.I., and Kelly, S. (2015). Calcitonin gene-related peptide in the joint: Contributions to pain and inflammation. Br J Clin Pharmacol *80*, 965–978. https://doi.org/10.1111/bcp.12669.

37. Ribeiro-da-Silva, M., Vasconcelos, D.M., Alencastre, I.S., Oliveira, M.J., Linhares, D., Neves, N., Costa, G., Henrique, R., Lamghari, M., and Alves, C.J. (2018). Interplay between sympathetic nervous system and inflammation in aseptic loosening of hip joint replacement. Sci Rep *8*. https://doi.org/10.1038/s41598-018-33360-8.

38. Schonfeldova, B., Zec, K., and Udalova, I.A. (2022). Synovial single-cell heterogeneity, zonation and interactions: A patchwork of effectors in arthritis. Preprint at Oxford University Press, https://doi.org/10.1093/rheumatology/keab721.

39. Wojdasiewicz, P., Poniatowski, Ł.A., and Szukiewicz, D. (2014). The role of inflammatory and anti-inflammatory cytokines in the pathogenesis of osteoarthritis. Preprint at Hindawi Publishing Corporation, https://doi.org/10.1155/2014/561459.

40. Blaney Davidson, E.N., van der Kraan, P.M., and van den Berg, W.B. (2007). TGF-β and osteoarthritis. Preprint, https://doi.org/10.1016/j.joca.2007.02.005.

41. Sinusas, K. (2012). Osteoarthritis: Diagnosis and Treatment.

42. Grandi, F.C., and Bhutani, N. (2020). Epigenetic Therapies for Osteoarthritis. Preprint at Elsevier Ltd, https://doi.org/10.1016/j.tips.2020.05.008.

43. Mobasheri, A., Kapoor, M., Ali, S.A., Lang, A., and Madry, H. (2021). The future of deep phenotyping in osteoarthritis: How can high throughput omics technologies advance our understanding of the cellular and molecular taxonomy of the disease? Osteoarthr Cartil Open *3*. https://doi.org/10.1016/j.ocarto.2021.100144.

44. Cohen, Z.D., and Derubeis, R.J. (2018). Annual Review of Clinical Psychology Treatment Selection in Depression. Annu. Rev. Clin. Psychol *14*, 209–245. https://doi.org/10.1146/annurev-clinpsy-050817.

45. Louvel, S., Schwarz, J., Delpierre, C., and Lefèvre, T. Precision and personalized medicine: What their current definition says and silences about the model of health they promote. Implication for the development of personalized health.

46. König, I.R., Fuchs, O., Hansen, G., von Mutius, E., and Kopp, M. V. (2017). What is precision medicine? Preprint at European Respiratory Society, https://doi.org/10.1183/13993003.00391-2017.

47. Ashley, E.A. (2016). Towards precision medicine. Preprint at Nature Publishing Group, https://doi.org/10.1038/nrg.2016.86.

48. Ota, M., and Fujio, K. (2021). Multi-omics approach to precision medicine for immune-mediated diseases. Preprint at BioMed Central Ltd, https://doi.org/10.1186/s41232-021-00173-8.

49. Wassermann, A.M., Lounkine, E., and Glick, M. Supporting Information Bioturbo Similarity Searching: Combining Chemical and Biological Similarity to Discover Structurally Diverse Bioactive Molecules.

50. Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. Preprint at BioMed Central Ltd., https://doi.org/10.1186/s13059-017-1215-1.

51. Sachs, M.C. (2015). Statistical principles for omics-based clinical trials. Preprint at AME Publishing Company, https://doi.org/10.3978/j.issn.2304-3865.2015.01.02.

52. Odell, I.D., and Cook, D. (2013). Immunofluorescence techniques. Journal of Investigative Dermatology *133*. https://doi.org/10.1038/jid.2012.455.

53. Schlecht, A., Boneva, S., Salie, H., Killmer, S., Wolf, J., Hajdu, R.I., Auw-Haedrich, C., Agostini, H., Reinhard, T., Schlunck, G., et al. (2021). Imaging mass cytometry for high-dimensional tissue profiling in the eye. BMC Ophthalmol *21*. https://doi.org/10.1186/s12886-021-02099-8.

54. Milosevic, V. (2023). Different approaches to Imaging Mass Cytometry data analysis. Preprint at Oxford University Press, https://doi.org/10.1093/bioadv/vbad046.

55. Chang, Q., Ornatsky, O.I., Siddiqui, I., Loboda, A., Baranov, V.I., and Hedley, D.W. (2017). Imaging Mass Cytometry. Preprint at Wiley-Liss Inc., https://doi.org/10.1002/cyto.a.23053.

56. Le Rochais, M., Hemon, P., Pers, J.O., and Uguen, A. (2022). Application of High-Throughput Imaging Mass Cytometry Hyperion in Cancer Research. Preprint at Frontiers Media S.A., https://doi.org/10.3389/fimmu.2022.859414.

57. Baharlou, H., Canete, N.P., Cunningham, A.L., Harman, A.N., and Patrick, E. (2019). Mass Cytometry Imaging for the Study of Human Diseases—Applications and Data Analysis Strategies. Preprint at Frontiers Media S.A., https://doi.org/10.3389/fimmu.2019.02657.

58. Jun, Z., and Jinglu, H. (2008). Image segmentation based on 2D Otsu method with histogram analysis. In Proceedings - International Conference on Computer Science and Software Engineering, CSSE 2008, pp. 105–108. https://doi.org/10.1109/CSSE.2008.206.

59. Labinsky, H., Panipinto, P.M., Ly, K.A., Khuat, D.K., Madarampalli, B., Mahajan, V., Clabeaux, J., MacDonald, K., Verdin, P.J., Buckner, J.H., et al. (2020). Multiparameter Analysis Identifies Heterogeneity in Knee Osteoarthritis Synovial Responses. Arthritis and Rheumatology *72*, 598–608. https://doi.org/10.1002/art.41161.

60. Mimpen, J.Y., Hedley, R., Ridley, A., Baldwin, M.J., Windell, D., Bhalla, A., Ramos-Mucci, L., Buckley, C.D., Coles, M.C., Alvand, A., et al. (2023). Cellular characterisation of advanced osteoarthritis knee synovium. Arthritis Res Ther *25*. https://doi.org/10.1186/s13075-023-03110-x.

61. Zou, Z., Li, H., Yu, K., Ma, K., Wang, Q., Tang, J., Liu, G., Lim, K., Hooper, G., Woodfield, T., et al. (2023). The potential role of synovial cells in the progression and treatment of osteoarthritis. Preprint at John Wiley & Sons Inc, https://doi.org/10.1002/EXP.20220132.