UNIVERSITAT DE BARCELONA

Final Degree Project

Biomedical Engineering Degree

"Study of fibroblasts activation kinetics and identification of fibroblast subpopulations in physiological and pathological situations"

Barcelona, 07/06/2023

Author: Lourdes Carreras Vidal

Director: Núria Gavara i Casas

Abstract

Fibroblasts undergo significant morphological and functional changes in response to specific environmental cues and functional changes. In wound healing processes and cancerous environments, fibroblasts undergo transformative activation adopting novel phenotype. The aim of this project was to understand the changes in cell morphology and cytoskeletal reorganization that occur for both normal-associated fibroblasts (NAFs) and cancer-associated fibroblasts (CAFs) activation. To achieve this, we used an innovative approach using biophysical biomarkers derived from epifluorescence imaging of the cell's cytoskeleton, employing CSKmorphometrics. Clustering algorithms have been applied to do a preliminarily identification of CAFs subpopulations. Our findings confirm that cytoskeletal reorganization occurs during both physiological and pathological activation. Non-tumoral fibroblasts experience larger morphological changes characterized by an increase in area and cell convexity, as well as changes in the total fluorescence of F-actin fibers during the 24 hours posterior to the administration of TGF-β. On the contrary CAFs exhibited sustained larger areas throughout the process regardless of TGF-β administration. They underwent most drastic changes in fiber length and showed a significant increase in nuclear volume. The application of logistic regression algorithms has allowed for a classification with 81% accuracy to differentiate between CAFs and NAFs, highlighting the differences in the cytoskeleton of these cell types in both study contexts. On the other hand, the intragroup analysis provided by clustering has enabled the identification of 5 clusters for non-activated CAFs, which converge at 72 hours into two larger clusters with significant differences. This study enhances understanding of the changes occurring in CAFs and NAFs during activation from a cytoskeletal point of view and remarks the need for a study of fibroblasts subpopulations as well as the need for novel biomarkers.

KEYWORDS: *cancer-associated fibroblasts, fibroblast subpopulations, biophysical biomarkers, cytoskeletal reorganization, machine learning.*

Acknowledgements

Firstly, I would like to sincerely thank Núria Gavara for welcoming me into her group lab and granting me the opportunity to carry out this project. Thanks to this opportunity, I have been able to learn plenty.

Secondly, I would like to express my most grateful thanks to África Martínez, whose support and trust in me have made this project possible. You have accompanied me every step of the way and for that I am extremely thankful.

I would also like to extend my thanks to all the members of the lab, who have consistently made me feel welcome throughout my time there.

Thanks as well to my friends and family, who have been a constant source of encouragement and support throughout this journey.

List of Figures

List of Tables

Glossary of abbreviations

AFM	Atomic Force Microscopy/Microscope
AU	Arbitrary Units
CAF	Cancer associated fibroblasts
CBUB	Comissió de Bioètica de la Universitat de Barcelona
CCiTUB	Centres Científics i Tecnològics de la Universitat de Barcelona
COPD	Chronic Obstructive Pulmonary Disease
CTGF	Connective Tissue Growth Factor
DDR2	Domain-containig Receptor 2
ECACC	European Collection of Authenticated Cell Cultures
ECM	Extracellular matrix
EMT	Epithelial-Mesenchymal Transition
FAF	Fibrosis associated fibroblasts
FAP	Fibroblast Activation Protein
FBS	Fetal Bovine Serum
FGF2	Fibroblast Growth Factor 2
FSP1	Fibroblasts-specific Protein 1
GEMM	Genetically Engineered Mouse Model
iCAF	Inflammatory Cancer Associated Fibroblast
IL	Interleukin
IPF	Idiopathic Pulmonary Fibrosis
ITS	Insulin-Transferrin-Selenium
myoCAF	Myofibroblastic Cancer Associated Fibroblast
NAF	Normal-associated or naïve fibroblasts
NCI	National Cancer Institute
NSCLC	Non-Small Cell Lung Cancer
NT	Non-Tumoral fibroblasts
P/S	Penicillin/Streptomycin
PDGF	Platelet Derived Growth Factor
scRNA-seq	Single cell RNA- sequencing
T	Tumoral fibroblasts
TGF-β	Transforming Growth Factor β
TME	Tumoral Microenvironment
TNF	Tumor Necrosis Factor
αSMA	Alpha smooth muscle actin

Content

1. Introduction

This project consists of a study of fibroblast activation kinetics and the identification of fibroblast subpopulations in physiological and pathological situations.

1.1. Objectives and scope of the project

The hypothesis of our project comes from the premise that fibroblasts are a highly versatile cell type that adapts morphologically to the environment and to the functionality they need to carry out, [1] which is why, fibroblasts found in a physiological context must differ phenotypically from fibroblasts found in a tumoral environment, such as is the case with lung cancer. In the same way, the activation of the fibroblasts due to the aforementioned circumstances must be accompanied of morphological changes.

To evaluate the morphological changes that the fibroblasts undergo, we aim to analyze the reorganization of their cytoskeleton during activation at different time points, to see whether cells go through different states during activation. This analysis will be carried out mimicking both a physiological situation (such as a wound) and a non-physiological situation (such as cancer). By doing so, we will be able to observe, not only the basal difference between normal fibroblasts (NAFs) and cancer-associated fibroblasts (CAFs), but also whether this reorganization follows the same pattern in the physiological fibroblasts and the CAFs.

The objective of this project also includes the identification of fibroblasts subpopulations in activated and non-activated CAFs, which may also vary depending on the physiological and pathological conditions. Therefore, we will observe which are the difference between the fibroblasts' subpopulations, but as well if these subpopulations are common between the different classes or if they are maintained throughout activation.

To accomplish said objectives, the scope of the project will include different aspects. Firstly, a cell culture experimentation phase, where the cells will be cultured and activated during a determined amount of time. Then, to visualize the cytoskeleton of the cells, an immunostaining protocol will be followed by the acquisition of pictures with the epifluorescence microscope. From this images, features of the cytoskeleton of the cells will be obtained. The programming of the code to extract features is not included in the scope of the project as our group has an already existing code to do so. Finally, the features will be analyzed statistically to obtain information of the activation process at different time points, as well as to assess difference between the initial state and the activated state and the NAFs and the CAFs. An analysis with machine learning will also be carried out to classify cells into different subpopulations, as well as to select relevant characteristics.

1.2. Motivation

The study of fibroblasts in cancer is crucial due to their association with the disease at all stages, including metastasis. CAFs are responsible for synthesizing various components in the extracellular matrix (ECM) and contributing to the structure, metabolic, and immune reprogramming of the tumor microenvironment.

The effect of CAFs on the environment and the tumor cells may vary due to their heterogeneous and highly plastic nature.[1]

Therefore, understanding the morphological changes and characteristics of fibroblasts in cancer is crucial in comprehending the mechanisms involved in the disease's progression and developing effective treatment strategies tailored to patients.

This need arises from a gap in the literature regarding the study of fibroblast in lung cancer, which is the type of tumoral environment we will be studying. While there have been some studies in other cancer types that have researched the activation of wound healing-associated activated fibroblasts and cancer-associated activated fibroblasts, none has found a unique biomarker that identifies all population of CAF. That is why, we propose a method based on the exploration of changes in the cytoskeleton morphology and reorganization via immunofluorescence imaging, instead of using biomarkers.

Over the past decade, significant advancements have been made in accurately and quantitatively characterizing the shape and morphology of cells and their cytoskeletons thanks to the development of image processing pipelines that can produce high-throughput, single-cell, and multiparametric outputs depicting the state of the cytoskeleton. As a result, image-based quantification of cytoskeletal state has become a valuable approach for describing the biophysical condition of cells.

Furthermore, there is no univocal classification and definition of fibroblastic subpopulations -it differs from publication to publication-, and their role in cancer is still unknown. Most of these studies on fibroblasts subpopulations are based on single-cell RNA sequencing, and flux cytometry biomarkers prove insufficient on subpopulations recognition based on functionality. That is why, we will do a study of this subpopulations based on cytoskeleton reorganization, as different functionalities will most likely require different cytoskeleton organization.

1.3. Spatial and temporal limitations of the project

This project has been carried out in in the group of "Mechanobiology of the cytoskeleton" at the Unit of Biophysics and Bioengineering in the Medicine Faculty of the University of Barcelona, whose aim is to apply the concepts and methodologies of physics and engineering to the comprehension of biological systems. Therefore, the project will make use of the resources available in the unit, including spaces, equipment, and material. These include a wet lab, a cell culture room with a cell culture hood and incubators, a dark room with a confocal microscope, amongst the other needed material and reagents to carry out the project.

Regarding the temporal limitation of the project, the project is started at mid-February and must be handed June 7th, providing us around four months to accomplish our objectives. There are several activities that might prove time-consuming. Firstly, the obtention of cells in a cell culture with good confluency and the optimal conditions to start experiments might take some time; specially, non-tumoral fibroblasts' proliferation in cell culture flask will be slower than tumoral cells proliferation. In addition, several hours have been spent in capturing images with the epifluorescence microscope suitable for

analysis, a task that proves arduous as the operation of the of the microscope must be learned and the eye must be trained to obtain in-focus images. On the other hand, image processing and image feature extraction requires high computing power and, therefore, the software will take some time to run.

2. Background

In this section, we will introduce the situation where the project is framed by describing what are fibroblasts and their main characteristics both in physiological and cancerous situations. We will also discuss the current state of the technology used to study cells, and specifically CAFs.

2.1. Fibroblasts and their role in wound healing and cancer

The extracellular matrix (ECM) is a highly dynamic complex that makes up the non-cellular aspect of tissues and varies in composition according to its tissue location and physiological circumstances. In the case of the lung, the ECM is typically divided into two basic compartments: basement membranes, which are thin specialized layers of ECM under epithelial and endothelial cell layers, and interstitial spaces, which form the parenchyma of the lung. It is in the lung interstitial where we find the resident fibroblasts, which are highly dynamic cells mainly responsible for ECM production, maintenance and remodelation.[2,3]

By creating an anatomically diverse array of ECM-rich connective tissues, fibroblasts support a broad range of essential organ functions, like resistance to stretching during breathing. Moreover, fibroblasts act as signaling niche cells for neighboring cells through the secretion of soluble mediators such as cytokines, growth factors, and metabolites, and via microarchitectural, biomechanical, and biochemical signaling in the ECM. Beyond producing connective tissue, fibroblasts serve as progenitors for specialized mesenchymal cell types during embryonic development or adult homeostasis, and injury, repair, and remodeling.[3]

It is in fact that fibroblasts play several critical roles to wound repair processes, both by synthesizing and depositing new ECM to repair the structural framework of the injury site and by signaling other key cell types. Any injury to the functional parenchyma induces the host response.[1] Following hemostasis, during which the intrinsic and extrinsic coagulation pathways are initiated to prevent blood loss, immune cells are recruited into the injury site, where they help to sterilize the wound and remove debris. The release of chemokines by platelets and early inflammatory cells in the wound attracts macrophages (derived from circulating monocytes), who engulf necrotic cellular debris and pathogenic material from the wound site. Then, they proliferate and move towards the wound, where they produce early extracellular matrix and regulate the inflammation.[4]

The early recruitment and activity of inflammatory cells and platelets activates and directs the migration of fibroblasts to the wound site around the fifth or seventh day. Subsets of fibroblasts recruited to the wound differentiate into myofibroblasts under the influence of mechanical tension and cytokines such as transforming growth factor β (TGF-β), which leads to the deposition of excess, poorly ordered matrix, resulting in fibrosis. The remodeling phase is the last stage of wound repair when fibroblasts continue to crosslink and turn over the initial disorderedly deposited granulation tissue. This lengthy process

strengthens and stiffens the provisional ECM until the mature scar is formed.[4] Once the wound is repaired, the number of activated fibroblasts decreases significantly owing to apoptosis, and the resting phenotype is probably restored. Such reversibility is the hallmark feature of tissue repair associated with wound healing.[1] This process is described in **Figure 1**.

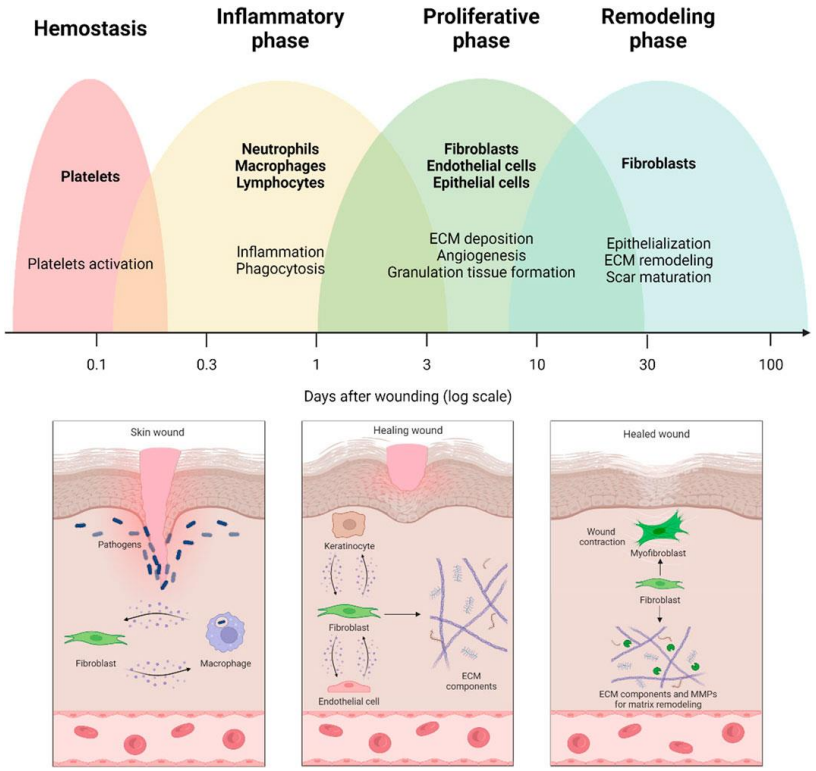


Figure 1. Fibroblast involvement in the various phases of the wound healing Journey. In the initial stages, fibroblasts contribute to the dynamic process of tissue repair through interactions with immune cells. Thereafter, they generate additional extracellular matrix (ECM) components and establish communication with endothelial cells and keratinocytes. Ultimately, fibroblasts actively participate in extracellular matrix remodeling by releasing matrix metalloproteinases (MMPs) and matrix constituents.[5]

If the injury is perpetual, the repair response does not cease, leading to a chronic wound healing condition that is known as tissue fibrosis. Activated fibroblasts are receptive to being further activated by growth factors, converting into fibrosis-associated fibroblasts (FAFs). The chronic tissue repair response also occurs in the setting of genetic damage encountered in cancer. Persistent emergence and accumulation of cancer cells in a given tissue represents an ongoing tissue injury, initiating a chronic wound healing response towards the cancer cells. This results in a chronic host repair response in tumors that is known as cancer fibrosis or stroma. In this case, CAFs are activated by growth factors released by the cancer cells and the infiltrating immune cells, such as transforming growth factor-β, platelet-derived growth factor (PDGF) and fibroblast growth factor2 (FGF2). CAF recruitment and local CAF proliferation and invasion is stimulated by TGF-β. [1]

The relationship between cancer progression and fibroblasts is still being unraveled. Their functional role in cancer progression and metastasis is emerging as being complex and bimodal, with both cancer-promoting and cancer-restraining actions. Although fibroblasts can function as positive or negative regulators of tumor growth, it is conceivable that fibroblasts may indirectly promote cancer progression.[1]

This is because, tumor progression has been found to depend on the components of the tumor microenvironment (TME) and the TME is mainly composed by CAFs. Specifically, CAFs are described to significantly increase tumor development by regulating several processes, including cancer cell proliferation, tumor cell invasion and migration, angiogenesis, and ECM remodeling due to the generation of growth factors and ECM. The abundance of CAFs in the TME is generally associated with poor prognosis, therapy resistance and disease relapse. In addition to their role in controlling cancer cell behavior, CAFs emerge as central players in shaping the TME toward an immunosuppressive and growth-promoting phenotype, via increased production of immunosuppressive cytokines and enhanced expression of immune checkpoint.[6] Dysregulated ECM remodeling by CAFs can lead to the desmoplastic reaction associated with poor outcome in breast, pancreatic and lung cancers. Loss of an organized and stable matrix is often considered a hallmark of these tumors, leading to extensive efforts to develop therapies targeting tumor ECM.[7]

2.2. Fibroblasts morphology

A key concept of cell biology lies in the intimate link between structure and function. Cell shape and the way cells interact with the surrounding environment influence the regulation of many cellular processes, including proliferation, differentiation, motility, and apoptosis. Thus, a change on the fibroblasts functioning due to the activation of the fibroblast activation is accompanied by a change in morphology and phenotype of the cell.[8]

Quiescent, or resting-state, fibroblasts present a classic spindle-shaped morphology: thin and elongated with frontal and back extensions. These fibroblasts are characterized by a lack of metabolic and transcriptomic activity. Their cell cycle is arrested on G0/G1, or they remain in a slow self-renewal cycle. They produce no ECM nor have a very active secretome, and they are fibroblast-specific protein 1 (FSP1) and α1β1 integrin positive. Quiescent fibroblasts can be activated by growth factors and become synthetic (myofibroblasts).[1]

Once they are activated, these cells become metabolically active, proliferative, and migratory, presenting the forementioned functions that include ECM synthesis, cytokines, and chemokines generation, immune cells recruitment, and the exertion of physical forces to modify tissue architecture. Morphologically, they are more planar with a cruciform or stellate shape. Activated fibroblasts can be identified predominantly by their expression of α-smooth muscle actin (αSMA), a cytoskeletal protein associated with smooth muscle cells, but they are also PDGFRβ and fibroblasts activation protein (FAP) positive. This type of fibroblasts is a major component of acute wound healing scars, but also of fibrotic tissue due to a chronic wound healing response in organ fibrosis and tumor growth.[1]

This way, an increasing number of CAFs which share a similar morphology with myofibroblasts observed in wound healing and present an activated phenotype is often found in the cancer regions [9]. Compared to NAFs and the myofibroblasts in wound healing, CAFs are perpetually activated, neither reverting to a normal phenotype nor undergoing apoptosis. However, we cannot talk about a unique CAF morphology,

as CAFs found in different cancers and TMEs are highly heterogeneous, and they are possibly derived from resident fibroblasts, epithelia cells, endothelia cells or mesenchymal cells.[10]

Figure 2 show the morphological as well as the molecular changes between quiescent fibroblasts, myofibroblasts and CAFs.

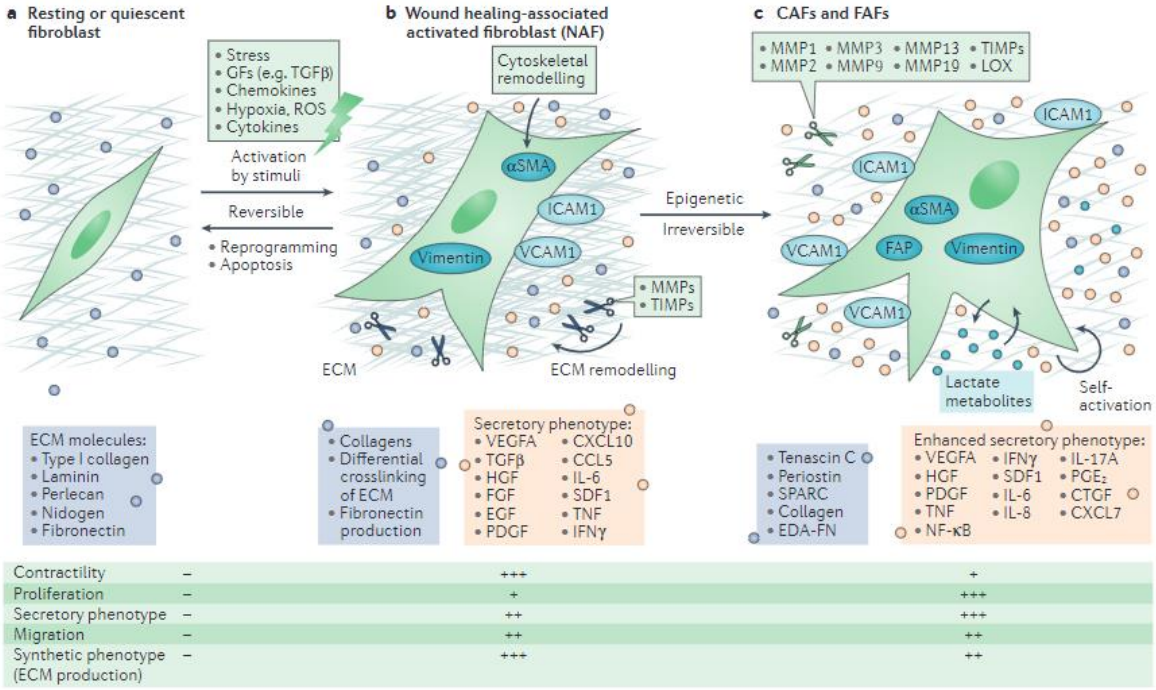


Figure 2. Multi-step activation of fibroblasts with their morphology and associated molecules. Characteristics are also shown. (a) shows quiescent or resting fibroblasts; (b) shows wound-healing associated activated fibroblasts; and (c) shows CAFs or FAFs.[1]

In this regard, the cytoskeleton plays a central role on determining cell morphology. Its dynamic nature allows for rapid rearrangement of its components to fulfill the cell's morphological requirements at particular time points in its life. Furthermore, the cytoskeleton and its associated proteins participate in a wide variety of signal transduction pathways, which can ultimately dictate cell fate. The cytoskeleton is composed of three main components; microtubules, intermediate filaments, and the actin network, all of which have the ability to resist deformation, reorganize in response to externally applied forces or stimuli, and maintain the spatial relationships between subcellular compartments.[8] Thus, the study of the cytoskeleton fibers elucidates morphological changes in cells.

2.3. CAF's heterogeneity

In recent years, it has become increasingly well understood that fibroblasts are an extremely heterogeneous cell type, exhibiting significant phenotypic and functional variability between- and even within- tissues. These cells are also highly plastic, transforming due to the biochemical and mechanical characteristics of the environment.[4]

Moreover, multiple findings show that show that CAFs are not a homogeneous cellular population.[6] The diversity of CAF functions and origins has led to the belief that CAFs are composed by multiple

subpopulation that only partially overlap.[7] Furthermore, it has been seen that fibroblasts exhibit considerable variation of cytoskeletal proteins (such as αSMA expression), surface markers and size, that suggests the existence of discrete cellular subsets, both within a localized site or in different locations.[11]

Such heterogeneity might result from numerous causes. It is conceivable that resting fibroblasts might be capable of differentiating into distinct subsets of functional fibroblasts, due to the reciprocal interactions between fibroblasts and neighboring cells or any other ECM component (or, in the case of CAFs, the interactions between CAFs, cancer cells and TME), in addition to the cytokines, chemokines and growth factors secreted into the ECM. The heterogeneity of fibroblasts could also depend on the origin of the precursor fibroblasts including the different cellular precursors[1,6] . It remains unclear whether different subsets of activated fibroblast perform unique functions.[1]

This heterogenicity impedes the universal identification of fibroblasts through a unique biomarker; some problems such as differences in marker expression, no unique population markers and heterogeneity of the population implying co-expression of different markers arise.[12] There are several biomarkers can identify activated fibroblasts, such as FSP1, vimentin, αSMA, FAP, PDGFRα, PDGFRβ, desmin and discoidin domain-containing receptor 2 (DDR2). However, none of these biomarkers is specific for fibroblasts or activated fibroblasts. Therefore, when using these markers, context, morphology, and spatial distribution should be taken into consideration to identify cells as resting or activated fibroblasts[1]

Furthermore, it is likely that many functionally activated fibroblasts may not express all these markers at the same time creating another degree of heterogeneity. Therefore, to identify such sub-sets, one may have to use multiple cell surface markers for detection. Once isolated, functional studies could be conducted to unravel specialized activities. CAF markers may also be associated with cells with diverse, and possibly opposing, functions in the context of specific TMEs.[13]

2.4. State of the art

The state of the art refers to the current level of advancement in the technological developments, methodologies, and key findings in a given area. In this section we will decipher the state of the art of the different methods used to study CAFs, as well as we will make a positive commentary on biophysical biomarkers, as a novel technology to study cell function.

2.4.1. Models to study CAFs

Different approaches have been used to study biological function and morphology of CAFs. Here we present the in vivo, in vitro and in silico approaches that have been taken to study fibroblasts.

2.4.1.1. In vivo models

The study of the properties of CAFs has heavily relied on observation of their phenotypes and transcriptomic landscape in various genetically engineered mouse models (GEMMs). Fibroblasts specific GEMMs and xenograft models were created. Among the most widely used GEMM models are those based on the CAF markers FSP1, Col1a1, Col1a2, FAP, ACTA2, PDGFRα, PDGFRβ, and Vimentin.[10]

In vivo models for fibroblast investigation still need to provide a specific markers trace fibroblasts lineage, and to identify CAFs independently from their cell origin (functional tracing), to capture the full heterogeneity of the CAF population.

2.4.1.2. In vitro models

In vitro studies provide a means to investigate the characteristics of CAFs and their interactions with other cells under controlled conditions. This has been made possible by the development of immortalized CAFs and NAF lines derived from human and mouse cells that can be cultured in a laboratory setting. In addition, advancements in tissue culture techniques have enabled the use of three-dimensional (3D) matrices with varying compositions, mimicking the stiffness of the organ from which the cells originated. By embedding cells within these matrices, they are exposed to physiological tension forces, creating an environment that profoundly influences the behavior of CAFs. These specialized fibroblasts exhibit precise mechanosensing capabilities, further emphasizing the impact of their surrounding environment.

2.4.1.3. In silico models

In addition to in vivo and in vitro modeling, the utilization of bioinformatic predictive tools has emerged as a valuable approach for studying CAF biology. These tools allow the identification of active pathways within specific cell types, enabling the formulation of hypotheses regarding their functions. Secondly, they enable the inference of relationships between different cell types, facilitating the mapping of dynamic crosstalk within the TME and its alterations over time or in response to perturbations. By leveraging bioinformatic predictive tools, researchers can gain insights into the complex interactions and regulatory mechanisms underlying CAF biology.[10]

2.4.2. Biophysical biomarkers for the cytoskeleton

Biomedical research has often focused on characterizing the progressive changes that take place at either the molecular scale, such as changes in genetic epigenetic or metabolic states, or at larger tissue-level scales, such as changes in organ physiology. In between, plenty of research has focused on studying biochemical changes at the cellular level, but few studies have focused on the cellular biophysical changes that accompany normal developmental processes or the onset of pathology.[14]

It has been proposed that prior to the observable functional changes at the tissue level, there may be initial alterations in the physical state of cells that play a role in driving these progressive changes. Likewise, the biophysical modifications occurring at the cellular level represent overall outcomes resulting from the concurrent up- and downregulation of numerous genes and signaling networks.[14]

The convenience of using comprehensive biophysical measurements of whole-cell readouts suggests that they could be used as optimal biomarkers of cellular state.

The concept of cellular biophysical and biomechanical properties (otherwise, referred as biophysical biomarkers) is broad and includes a large number of cellular readouts that can be obtained using nanotechnologies that measure physical changes of the cell, or computer vision approaches that allow morphological quantification of cellular structures from fluorescence images.

Over the past decade, many strategies and techniques have been proposed to support the characterization of biological samples down to the single cell level, either from the mechanical or morphological point of view. Among proposed approaches, Atomic Force Microscopy (AFM) is certainly the most adopted and effective to identify mechanical signatures of cellular systems.[15] Cellular biomechanical properties such as cellular stiffness, viscosity, migratory speed, contractile forces, and cell-cell adhesion strength have been reported.

Underlying those mechanical properties, the cellular cytoskeleton is the main structure regulating the mechanical state of a cell in a highly dynamic and adaptative manner. The study of the cytoskeleton allows and reflects numerous physiological properties and states of the cell, such as the cell biomechanical properties, proliferation and migration, and the cell morphology, which is why, the number of studies based on cellular cytoskeleton is growing.[16]

Over the past decade, significant advancements have been made in quantitatively and accurately characterizing the shape and morphology of cells and their cytoskeletons. These advancements have led to the development of image processing pipelines that enable high-throughput analysis and provide single-cell and multiparametric outputs of cytoskeletal states.[16,17]

A recent breakthrough has been the ability to integrate mechanical measurements, such as AFM, and cytoskeletal morphological data obtained from techniques like epifluorescence microscopy at the single-cell level.[18] This integration has confirmed the close relationship between cell mechanics and cytoskeletal organization. As a result, image-based quantification of cytoskeletal states has become a valuable approach for describing the biophysical condition of cells.

With the current automation and high-throughput capabilities of fluorescence microscopes, it can be argued that image-based quantification of cytoskeletal states is currently the most effective tool for generating biophysical biomarkers of cellular state in a single-cell, high-throughput, and multiplex manner.

3. Market analysis

Market analysis involves assessing the demand, supply, competition, and potential opportunities within a specific market. However, this is not a commercial project so the market this project is framed consists of the different publications that may appear on the topic. Hence, after identifying the sector the project is framed on, we will take a summarize some of the most novel findings, to assert the gap on the literature.

3.1. Targeted sectors

This project is basically fundamental research to investigate cellular function and morphology during the physiological process of fibroblast activation. By studying the activation process of both non-tumoral and tumoral fibroblasts, it aims to contribute to the understanding of key mechanisms underlying fibroblast behavior via biophysical markers obtained from epifluorescence images of fibroblasts' cytoskeleton. That is why, this research falls within the interdisciplinary field of biotechnology and biomedical research, where the intersection of biology, technology, and medical science plays a crucial role.

As tumoral fibroblasts are thought to play a role in tumor progression and metastasis[1], the project holds considerable relevance and potential applications within the field of cancer research. The findings of this research could lead to the elucidation of undiscovered tumoral mechanisms. This could bring the discovery of new therapeutic targets or drugs that could modulate fibroblast behavior, potentially leading to new treatment strategies for cancer or other diseases involving fibroblast activation. Biophysical biomarkers could also be used hence in the pharmacological field for the screening of drugs[19] by deriving their effect on the cytoskeleton of single cells.

3.2. Novel findings on fibroblasts activation and CAFs subpopulations

In this section, we will explore the most recent studies that have been published on this topic, regarding the study of CAFs' activation and of apparent fibroblasts' subpopulations in cancer.

3.2.2. Fibroblasts activation in pathological conditions

Although fibroblasts have been shown to possess the ability to revert to their quiescent state, in the case of chronic insult they enter a self-sustained activated state. To elucidate what happens during this activation process, current efforts are mostly being devoted to characterizing the epigenetic changes that accompany establishment of the activated fibroblast phenotype.[10]

Several review articles delve on the importance of epigenetic changes in fibroblasts to drive cancer progression. For the case of lung cancer, Alcaraz et al. (2021) summarizes the evidence of the epigenetic reprogramming of lung CAFs. They conclude that the activated/myofibroblast-like phenotype CAFs exhibit is driven and/or maintained by epigenetic modifications that may synergize with or modulate the potent fibroblast activator cytokine TGF-β1. [20]

Albrengues et al. (2015) define an epigenetic switch, whose initiation results in sustained pro-invasive activity of CAFs in carcinoma.[21]

Several studies attempt to study activation via the identification of activated CAFs through specific biological markers and transcriptional changes.[13] For example, Yang et al. (2022) analyzed FAP expression to understand sustained activation of lung fibroblasts and investigate whether FAP could be used as a potential readout.[22]

The heterogeneity of such markers in distinct tumor types and expression of some of these markers in healthy tissues, pose a significant challenge when studying the role of CAFs and their biological properties in cancer.[13]

Recent studies have demonstrated that fibroblast activation plays a role in the natural aging process and significantly impairs the tissue's ability to respond to injuries. In a comprehensive investigation conducted by Mahmoudi et al. (2019), fibroblasts from both young and old mice were extensively characterized, revealing age-associated transcriptional and metabolic changes induced by inflammation. Interestingly, the study found that older individuals, in both mice and humans, exhibited an increase in activated fibroblasts that secreted both pro- and anti-inflammatory cytokines, such as IL-6, tumor necrosis factor (TNF), and IL-4. [23]

3.2.2. CAF subpopulations

Characterization of CAF subpopulation studies have been carried out by the means of single-cell RNA sequencing technologies (scRNA-seq), immunostaining, in-situ hybridization, flow cytometry and fluorescence activated cell sorting. [7]

The advent of scRNA sequencing and other novel techniques has been a revolutionary technology in the stratification of a multitude of novel CAF subtypes through high resolution characterization of whole transcriptomes on a single cell level. [7] However, these techniques has resulted in a plethora of datasets and resulting classifications. No consensus has been reached in the scientific community. [10]

The first theories regarding CAFs subpopulation indicated the existence of two types of CAFs, myoCAF (myofibroblasts) and iCAFs (inflammatory CAFs). Various classifications converge on the two subtypes with different used terminologies. For example, Zeisberg and Zeisberg (2013) theorized that there are two types of fibroblast activation profile: 'reversible' and 'irreversible', determined partly by epigenetic regulation [24]. A few exemptions provide more diverse classifications.

In lung cancer, scRNA-seq of human non-small cell lung cancer (NSCLC) revealed five distinct CAFs types based on the expression of a unique repertoire of collagens and other ECM molecules. Further characterization showed high expression of genes associated with myogenesis (including ACTA2 which encodes for αSMA) in one subpopulation, whereas another showed a strong expression of genes linked to epithelial-mesenchymal transition (EMT) and the ECM[25]. Kim et al. (2021) scRNA-seq of tissues obtained from patients with lung adenocarcinoma at different stages of the disease identified seven subpopulations of fibroblasts, three of which were functionally related to ECM modulation: matrix fibroblasts and myofibroblasts. Matrix-related fibroblasts were the main subpopulations present in normal lungs and early-stage tumours. In contrast, myofibroblasts were dominant in advanced stage tumours (including in metastatic lymph nodes), reflecting a gradual change of fibroblast states associated with tumour progression.[26]

Other diseases have also been studied through scRNA-seq, such as the idiopathic pulmonary fibrosis (IPF) or chronic obstructive pulmonary disease (COPD). Adams et al. (2020) profiled cells from IPF and COPD patients and identified new aberrant cell subpopulations for IPF disease.[27]

Flow cytometry has been employed to annotate and sort different CAF subgroups from tumors, also enabling CAF subset isolation for further experimentation.[7] Kieffer et al. separated five cell clusters through flow cytometry and then validated its differences through scRNA-seq.[28]

In the aforementioned analysis by Mahmoudi et al. (2019), the study of the secretome revealed distinct populations of activated fibroblasts with different secretomes. The abundance of each population was predictive of wound healing efficiency in old mice.[23]

4. Concept engineering

In this section, we will be discussing the different methodological options we can consider to carry out the study of the cytoskeleton of fibroblasts cells in the context of both physiological and tumoral environment, from the obtention of the cells to the analysis of the results.

4.1. Cell obtention

To carry out our study, we firstly need to obtain the object of our study, the fibroblasts. We need to acquire both tumoral and non-tumoral fibroblastic cell lines associated from lung tissue to perform our research.

To obtain fibroblasts for research, there are different options available to scientist; the chosen one will depend on the availability, the nature of the research, and ethical considerations and regulatory requirements.

The most direct source is to obtain fibroblasts directly from the patients or donors, via a surgical resection or a lung biopsy, thanks to the collaboration with a hospital. The collected tissue is processed via enzymatic digestion to dissociate cells, which are subsequently cultured to allow fibroblasts growth and proliferation. Proper ethical approvals and informed consent from patients must be obtained previously.[29]

Being able to get cells from human patients is a privilege we have, being associated with Hospital Clínic de Barcelona. However, plenty of biomedical studies use animal models to obtain cells from their research. The advantage of working with animal models is that they can be genetically modified to mimic the desired conditions of the disease. Although human physiological responses can be mimicked, animal models may not always fully represent human biology, necessitating careful interpretation of the results. Ethical concerns also arise with the use of animals.

On the other hand, established cell lines derived from human lung fibroblasts can be obtained commercially through biotechnology companies, such as ATCC -to name one-, or through research repositories or cell banks, such as the National Cancer Institute (NCI) or the European Collection of Authenticated Cell Cultures (ECACC). These repositories often provide established, well-characterized and authenticated cell lines are derived from primary fibroblasts.

The last option is to obtain cells via collaborations, with other researchers that have already obtained or developed fibroblast cell lines associated with lung cancer. This is, in fact, how we have obtained our cells.

Duch et al.[30] study used immortalized cell lines from healthy lung tissue (which was used as a control) and tumoral lung tissue from the twenty patients. Informed consent was obtained, and ethical guidelines were followed to acquire the cells. **Table A1** in **12.1. Supplementary information regarding the obtained cells from Duch et al. (2022)** provides summarized information from the obtained cells from the patients. The cells that we ended up using in our experiment were the NAFs and tumoral cells from patient #13, a 59-year-old patient with an adenocarcinoma of type T2b and N2 and stage IIIA.

For the acquisition of these cells, one of our requirements was that the tumoral and non-tumoral cell lines should be in similar low passage number, as cell lines with higher passage numbers may exhibit alterations in cells morphology and functions.[31] As we aim to perform a comparative study, we wish to reduce any source of variability due to external factors, as a high-number pass would be. We were able to obtain two cryotubes of patient #13 tumoral and non-tumoral cells at passage #9 and #10 respectively.

4.2. Experiment design

After cell obtention, a well-designed experimental framework is crucial to address our research objectives effectively, encompassing careful consideration of variables, controls, and methodologies to ensure robust and reliable results.

4.2.1. Substrate or support

The substrate or support serves as a stable surface, on which cells can attach and grow during our experimentation. This substrate should allow for easy manipulation and handling of the cells, such as the ability to add or remove media or perform staining or other protocols easily. Furthermore, it must be compatible with the use on the epifluorescence microscope if we are to observe the attached cells. Optical transparency is hence a must to visualize the fibroblasts, both during the culturing in the cell culture room to monitor them and to see them through the epifluorescence microscope.

The material used should also be biocompatible and not interfere with the behavior or functionality of the cells. It should also be sterilizable as we are working in sterile conditions with the culturing of the fibroblasts. This ensures that the cells are not contaminated during the experimental process.

Some commonly used substrates include glass slides, coverslips, plastic dishes or plates, and specialized materials like polymer-coated surfaces or membranes.

Considering the previously mentioned requirements, we have chosen glass coverslips as our support material. Although they are quite tricky to handle because of their small size and fragility, they provide us with a very transparent vessel that generates no extra background in immunofluorescent images, as we experienced it happens with plastic supports. Furthermore, it allows for having replicas of the experiment.

The glass coverslips must be placed elsewhere so that they are safely handled. Two options, which can be seen in **Figure 3**, were discussed: to use a well plate and place a coverslip per well or use a petri-dish with several coverslips inside. For the first option we need as many wells as replicas we want to do per condition and control. In the second case, a petri dish per condition would be used and we would have as many replicas as many coverslips per petri dish fitted.

Although the first option seemed more comfortable because petri-dishes are more difficult to carry and keep safely, having to subtract the coverslip from the wells proved difficult and several coverslips broke. Using petri dishes also proved to be advantageous in optimizing both time and the volume of reagents used. By utilizing petri dishes, we were able to streamline the process since we could simply apply the reagent to the entire dish, allowing it to spread uniformly across all the coverslips within the petri dish. This eliminated the need for individually applying the reagent to each coverslip, saving us valuable time

and reducing the overall amount of reagents required. That is why, we ended up using 40-mm diameter petri dishes, which fitted a total of three 12-mm coverslips in them.

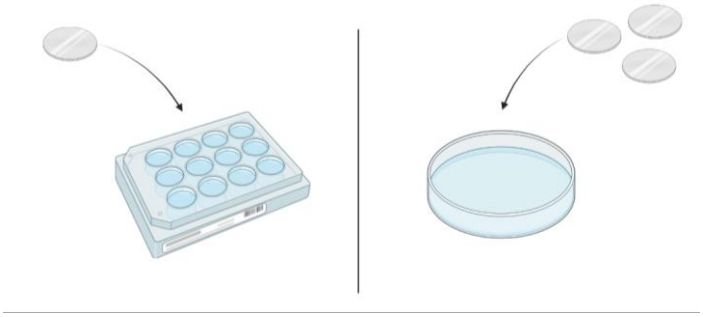


Figure 3. Support material options for our experiment [own figure]

4.2.2. Experimental design

Keeping in mind our goal is to study the activation profile of tumoral and non-tumoral fibroblasts at different time points during activation, as well as finding subpopulations with different activation profiles, we need to carry out an activation of fibroblasts experiment during a certain amount of time. Activation will be stopped at different time points and cells will be fixed, keeping the morphology of the time they were fixed at. There are several aspects we must consider when designing the experimentation stage.

4.2.2.1. Time frame of the experiment

Firstly, we need to decide what will be the time frame of the fibroblasts' activation. It is important to consider a time frame that will allow for observing and documenting morphological changes over time, so the expected kinetics of morphological changes associated with fibroblast activation should be taken into account. To determine so, we have followed the example of other studies. In this case, other studies such as Strutz *et al*. (2001), Midgley *et al*. (2013) or Ma et al. (2018)[32–34] chose 72 hours as the time of maximum activation, which seems to fit in our experimental.

With this time frames in mind, it is important to consider the appropriate time intervals between observations to capture meaningful changes in cell morphology.

Before starting the experiment, we have considered a time of 24 hours to let the cells set and attach to the bottom of the petri dishes, so that they could recuperate their normal morphology after being trypsinized and seeded. Then, we have opted for fixing the cells at the following time points: 0h (non-activated cells), 6h, 24h, 36h, 48h and 72h after activation.

4.2.2.2. Cell seeding

Considering that the timeline is of 72 hours (plus 24 hours we must wait for fibroblasts to set before starting activation), we must be very careful with the quantities of cells that are seeded in the coverslips.

As cells continue to proliferate during the experiment, we must seed a quantity of cells per cm^2 at the initial time point small enough, so that at the end of the experiment cells have not proliferated so much that we cannot image individual cells. This holds significant importance for us because the utilized code

for image analysis operates on individual cells, so we need to be able to clearly separate them. However, having a too small quantity of cells might affect the normal function and phenotype of the fibroblasts, as cells need cell-to-cell contact to develop normally.

After performing some cell counting experiments to check cell proliferation both in tumoral and non-tumoral populations, we have determined that a range of 1500-3000 cells was adequate.

4.2.2.3. Activation agent

The activation agent is the substance or factor that induces the activation or stimulation of fibroblasts cells.

In the body, activation of fibroblasts typically occurs through four distinct mechanisms: stimulation by cytokines and growth factors ("auto- and paracrine"), by direct cell-cell contacts, by extracellular matrix via integrins, and by environmental conditions.[35] This way, a number of growth factors have been associated with myofibroblast differentiation, for example, TGF-β, PDGF, IL-6, angiotensin II and connective tissue growth factor (CTGF). [1,36]

All these substances can be used experimentally to activate fibroblasts in the laboratory.[37]

TGF-β is multifunctional cytokine that regulates various cellular processes, including cell growth, differentiation, and extracellular matrix production. It is perhaps the most studied growth factor in fibrosis, as it has been shown to induce differentiation from quiescent fibroblasts to myofibroblasts and to stimulate ECM production directly. The main sources of TGF-β in wound healing conditions are activated immune cells, such as macrophages, and fibroblasts themselves, so activation by TGF-β is known as fibrotic stimulation.[37] Another profibrotic cytokine is CTGF, which is also indicative of fibroblasts activation.[38]

Immune cells secrete other cytokines such as interleukins (for example, IL-6) which can also stimulate fibroblasts. IL-6 is a pro-inflammatory cytokine that is implicated in skin fibrosis and has been seen to be able to activate fibroblasts and promote their synthesis of various extracellular matrix components.[37] IL-1 is also a pro-inflammatory cytokine that may also orchestrate fibroblasts responses in wound healing response.

PDGF, on the other hand, is a potent mitogen that can be applied to fibroblast cultures to induce their activation and promote cellular responses as, besides stimulating the migration and proliferation of fibroblasts, also drives fibroblast activation -at least to dermal fibroblasts [37,39]. PDFG is a growth factor induced due to vascular damage.

Angiotensin II stimulates cell proliferation and collagen type I synthesis in cardiac fibroblasts.[40]

From these mentioned options, it is no wonder that we chose TGF-β as our activation agent. This is because TGF-β is frequently associated with a myofibroblast αSMA phenotype in liver, lung, and kidney disease, which fits with the acquired lung adenocarcinoma fibroblasts. TGF-β been identified to be the preeminent growth factor responsible for fibroblast activation and matrix synthesis both in vitro and during vascular disease and fibrosis. Also, TGF-β plays a pivotal role in fibrogenesis, wherein several growth

factors, including PDGF and angiotensin II, exert their effects by directly stimulating TGF-β production. Additionally, TGF-β directly promotes myofibroblast development by inducing expression of αSMA phenotype. This is why, TGF-β is considered the foremost growth factor for induction of fibroblast activation in culture and in vivo, being its used very widespread in the scientific community.[36]

4.2.2.4. Cell morphology imaging techniques

The study of cell morphology is mainly carried out through the means of microscopy and imaging. Cytologists utilize microscopic techniques—light microscopy, phase contrast microscopy, interference microscopy, polarization microscopy, fluorescent microscopy, and electron microscopy—to investigate certain aspects of cell structure.[41]

As our aim is to study the morphology by analyzing the organization of the cytoskeleton of the cells, we need imaging techniques that allow the visualization of the cytoskeletal fibers. At present, there are three main techniques to study the cellular cytoskeleton: transmission electron microscopy, immunofluorescence microscopy and transmission X-ray microscopy. AFM has emerged as well as a newer technique used for the study of cytoskeletal fibers.[42]

Fluorescence imaging has proved to be the most useful technique for cytoskeletal imaging.[8] Using fluorophore-tagged antibodies to label the structure of interest, cellular location of the target protein can be detected by observing the fluorophore signal. Fluorescent labeling offers a high sensitivity and specificity, and most importantly capability for quantification.[8]

Table 1, show some of the imaging techniques used to visualize and analyze the cytoskeleton and focal adhesion in cells. Most of these techniques are more sophisticated ways to obtain fluorescence images, but electron microscopy is also mentioned.

Table 1. Examples of imaging techniques used to visualize and analyze the cytoskeleton and focal adhesions in cells.[8]

Imaging Technique	Basic Principle	Advantages	Disadvantages
Laser scanning confocal microscopy	Out-of-focus light is eliminated from sample via a pinhole	Relatively straightforward to use, optical sectioning allows imaging of thin optical slices in thick samples	Photobleaching, slow acquisition not well suited for living cells
Spinning disk confocal microscopy	Illumination via multiple pinholes	Rapid acquisition with minimal illumination of specimen, ideal for live-cell imaging	Pinhole crosstalk increases background signal
Interferometric photoactivated localization microscopy	Combines photoactivated localization microscopy with single-photon, simultaneous multiphase interferometry	Images a high density of specific fluorescent molecules with 3D nanoscale (10-20 nm) resolution	Limited by specimen thickness, complexity of microscope setup
Electron cryotomography	Electrons are used to produce projections of a sample at multiple angles which are back-projected to reconstruct the original object in 3D	Rapid freezing preserves sample in near-native state without need for fixation or staining	Limited number of images can be acquired due to the target area becoming electron irradiated

Scanning angle interference microscopy	Modified FLIC microscopy which actively scans the incidence angle of excitation	Provides nanometer precision and allows temporal sampling rates per second	Unknown refractive indices of sample structures cause minor errors in determining precise absolute height measurements
High content screenin	Automated fluorescence microscopy (confocal or non-confocal)	Quantitative cellular imaging producing large data sets in a relatively short time period	Lack of visual precision by researcher, potential issues with object segmentation during analysis
Epifluorescence microscopy	High intensity excitation light and emitted light pass through the same lens.	Easy imaging of intense signals, requiring a small amount of emitted light to be blocked	Precise location of fluorescence molecules is limited. No 3D interpretation.

In our case, we will use epi-fluorescence microscopy as it is a tool that we have available in the laboratory. In epifluorescence microscopy, a parallel beam of light from a high-intensity light source is passed directly upwards through the sample via an objective lens, maximizing the amount of illumination. The emitted light from the sample crosses the same lens, hence the term epi which means "same" in Greek. Epifluorescence microscopy is widely used in cell biology as the illumination beam penetrates the full depth of the sample, allowing easy imaging of intense signals and co-localization studies with multi-colored labeling on the same sample. In comparison to other forms of fluorescence microscopy, epifluorescence illumination has the advantage of only requiring a small amount of emitted light to be blocked. Epifluorescence imaging can, however, limit the precise localization of fluorescence molecules and does not allow the interpretation of three-dimensional spatial data, as any out-of-focus light will be collected.[43]

In the case of not being able to use the epifluorescence microscope in our unit, we could look into using the resources available in the "*Centres Científics i Tecnològics de la Universitat de Barcelona"* (CCiTUB).

By performing an immunostaining protocol to our samples with fluorescently labeled antibodies specific to cytoskeletal components, such as actin filaments, microtubules, or intermediate filaments, we will achieve the visualization of the cytoskeleton.

The second aim of our project is to study the level of activation of the different subpopulations of activated fibroblasts. By fluorescently labelling fibroblasts markers that indicate fibroblast activation, we will also be able to study the activation levels via epifluorescence microscopy.

It is of the utmost importance, hence, to correctly choose the targeted structures and the antibodies to accomplish the project's goals.

4.2.2.5. Molecular target for immunofluorescence protocol

As it has been previously mentioned, the first aim of this project is to define the morphological changes of fibroblasts through the characterization of cytoskeleton reorganization. That is why, we fist need to define which filaments will be stained during the immunofluorescence staining protocol: microtubules, intermediate filaments, or the actin network.

The major cytoskeletal protein of most cells is actin, which polymerizes to form actin filaments or microfilaments- thin, flexible fibers approximately 7 nm in diameter and up to several micrometers in length. Within the cell, actin filaments are organized into higher-order structures, forming bundles or three-dimensional networks that provide mechanical support, determines cell shape, and allows movement of the cell surface, thereby enabling cells to migrate, engulf particles, and divide. Actin can be labeled with a fluorophore by using phalloidin, which binds very specifically to F-actin, a certain type of microfilament. [44]

Intermediate filaments have a diameter of about 10 nm and appear to play basically a structural role by providing mechanical strength to cells and tissues. Intermediate filaments are composed of a diverse group of more than 50 different cytoskeletal proteins, including keratins, vimentin, neurofilaments, and others.[44] Antibodies specific to the intermediate filament protein of interest can be used for visualization.

The third principal component of the cytoskeleton are microtubules, which are rigid hollow rods approximately 25 nm in diameter. They function both to determine cell shape and in a variety of cell movements, including some forms of cell locomotion, the intracellular transport of organelles, and the separation of chromosomes during mitosis.[44] Microtubules are composed tubulin, a globular protein. Tubulin can be targeted using antibodies or fluorescently labeled tubulin-binding agents, such as antibodies against alpha- or beta-tubulin for fluorescent immunostaining.

On the other hand, we will need to use fibroblast markers to accomplish the second goal. These markers should be over-expressed when fibroblast get activated. However, as it has been mentioned, despite recent progress in the identification of multiple fibroblast markers, specificity remains a key challenge. **Table 2** provides a comprehensive list of fibroblast markers identified to date. Unfortunately, all these markers are also expressed by at least one other cell type. There remains a need to identify more selective fibroblast markers or identify the combination of markers to specifically define fibroblasts and activated fibroblasts.

Table 2. Some of the currently known fibroblasts' markers and the cells, other than fibroblasts, where they are expressed[34]

Markers	Location	Expressed by other cell type	Overexpressed in activated fibroblasts
α-smooth muscle actin	Intracellular	VSMCs, pericytes	Yes
Cadherin-9	Cell surface	Neurons, tumor vasculature	Yes
CD248 (endosialin)	Cell surface	Endothelial cells, pericytes	Unknown
Collagen I (α1 and α2)	Secreted	VSMCs, epicardial cells	Yes
Connective tissue growth factor (CCN2)	Secreted	Hepatic stellate cells, epithelial cells	Yes
Discoidin domain receptor 2	Cell surface	VSMCs, endothelial cells	Unknown

Fibronectin, ED-A splice variant	Secreted	Macrophages, endothelial cells	Yes
Embryonic smooth muscle myosin	Intracellular	Tumor cells, interstitial cells of Cajal	Yes
Fibroblast-activation protein-1 (FAP)	Cell surface	Activated melanocytes	Yes
Fibroblast surface antigen	Cell surface	Monocytes/macrophages	Yes
Fibroblast specific protein-1 (S100A4)	Intracellular	VSMCs, leukocytes, cancer cells	Yes
Periostin	Secreted	VSMCs	Yes
Prolyl-4-hydroxylase	Intracellular	Endothelial cells, epithelial cells	Unknown
Platelet derived growth factor receptor α	Cell surface	Platelets	Yes
Scleraxis	Intracellular	A wide range of cell types	Unknown
Tcf21 (epicardin, Pod1, capsulin)	Intracellular	A wide range of cell types	Unknown
Thymus cell antigen-1 (CD90)	Cell surface	Leukocytes, endothelial cells, various progenitor cells	Unknown
Vimentin	Intracellular	Endothelial cells, VSMCs	Yes[1,45]

This way biomarkers such as FSP1, vimentin, αSMA, FAP, PDGFRα and PDGFRβ, desmin and DDR2 would be of interest for us to study fibroblasts activation.

Taking all of this in account, we will perform two different immunostaining protocols to each of the different conditions.

Both protocols require of two mandatory staining's to visualize a set of cellular components which help us identify cell boundaries. The first one is the adding of DAPI, 4',6-diamidino-2-phenylindole, a DNA-specific probe which forms a fluorescent complex by attaching in the minor grove of A-T rich sequences of DNA.[46] Besides helping in recognizing cell location, DAPI is very important because it stains the nucleus of the cells in blue. From the nucleus of the cell, we can obtain several parameters that can also be analyzed.

The other important aspect is to determine overall shape of the cell. To do so, we will stain F-actin with phalloidin, which helps set context for other fluorescent labels within the shape of the cell. Phalloidin is a bicyclic peptide that belongs to a family of toxins isolated from the deadly *Amanita phalloides* that selectively binds to both large and small F-actin filaments of cells.[47] In the lab, we had phalloidin as an already a conjugated antibody with Texas Red dye (TXRED), a bright red-fluorescent dye with excitation ideally suited to the 561 or 594 nm laser lines. Phalloidin does not only provide a canvas to locate other fibers, but we also obtain very relevant information regarding morphology and cytoskeletal fibers.

As red and blue wavelengths are taken, additional green-labelled antibodies will be added to target the principal structures of interest. To study the cytoskeleton, vimentin will also be targeted to visualize intermediate filament. This way, staining both intermediate filaments with vimentin and actin filaments with phalloidin, we are able to cover a big part of cytoskeletal fibers components in our analysis.

To study fibroblasts activation, αSMA will be targeted as it is, overall, the most used marker. Furthermore, Kalluri et al (2020) states that myofibroblasts, induced by TGF-β-mediated signaling, proliferate and express vimentin and αSMA[1] , which supports our decision. Primary antibodies will be used, and secondary antibodies labelled with fluorescein (FITC), a bright green fluorophore, will bind specifically to those primary antibodies.

Used concentration for primary, secondary, and conjugated antibodies will be determined by manufacturer's indications. These ranges were adapted to our needs by trial-and-error of different immunostaining experiments at different concentrations to see which concentration provided the best balance between the amount of antibody used and the effectiveness of the tinction. Bibliographic references were used to guide this decision.

4.3. Data analysis

The obtained images must be analyzed: different algorithms and techniques must be applied to interpret and understand the visual content of the images. Feature extraction aims on identifying and extracting relevant features or patterns from images that can be used for further analysis.

To preprocess, to better their quality for posterior analysis, and to extract features from them, one can use different image analysis software. Some examples of this software tools can be ImageJ/FIJI, CellProfiler, Matlab (Image Processing Toolbox) or even Python has some image processing libraries, which are the ones we are most familiar with as we have employed their use before.

ImageJ[48] is a free and open-source software developed by the National Institutes of Health (NIH) for image analysis and processing. FIJI (a distribution of ImageJ) extends its capabilities with additional plugins and features specifically designed for biological image analysis. ImageJ/FIJI is highly versatile and supports a wide range of image formats and analysis tasks. It provides a user-friendly interface and a vast collection of plugins for tasks such as segmentation, quantification, colocalization analysis, and more.

CellProfiler[49] is an open-source software designed for high-throughput analysis of biological images. It is particularly useful for tasks such as cell segmentation, object identification, and measurement of various cellular features. CellProfiler provides a pipeline-based approach, allowing users to create customized workflows for their specific analysis needs. It has a user-friendly interface and is widely used in cell biology, drug discovery, and pathology research.

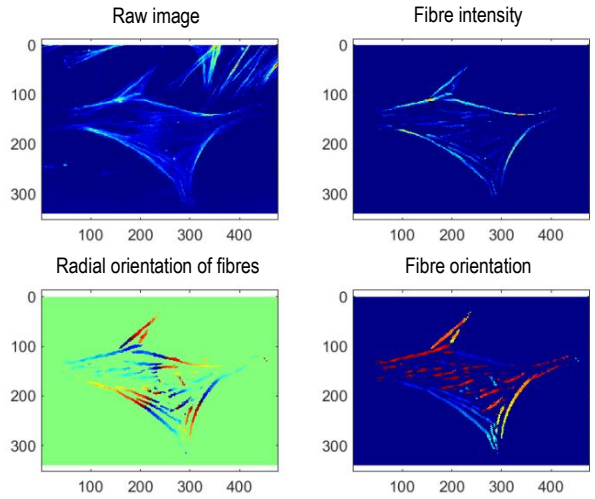
MATLAB[50] is a widely used programming language and environment for scientific computing. It offers the Image Processing Toolbox, which provides a comprehensive set of functions and tools for image analysis, including filtering, segmentation, feature extraction, and morphological operations. MATLAB's

extensive capabilities and flexibility make it a popular choice for researchers in the biological sciences who require custom analysis workflows or algorithm development.

Python[51] is a popular programming language because of its versatility, extensive libraries, and ease of use. It provides several powerful libraries for image processing and analysis, such as NumPy, SciPy, OpenCV, scikit-image, and Python Imaging Library. These libraries offer a wide range of functions and algorithms for tasks like image loading, manipulation, filtering, segmentation, feature extraction, and visualization.

Thankfully, we must not worry about this aspect because our research group has already some MATLAB scripts for image analysis and feature extraction build specifically for the analysis of cell cytoskeleton: CSKmorphometrics.[52,53]

A



B

Gross Morphology	Individual Fibres
Area	Total protein in fibres
Aspect ratio	Thickness
Stellate factor	Curvature
	Length
Organization Fibres	**Nucleus**
Alignment	Chromatin condensation
Chirality	Volume
Radial location	Poisson's ratio
Spread	Aspect ratio

Figure 4. CSKmorphometrics outputs. (A) shows an example of processed fibroblasts using CSK morphometrics to extract position, brightness and directionality of its phalloidin-stained actin fibres. Top right shows the raw image; Top left shows the fibre intensity; Bottom right show radial orientation fiber; Bottom left shows fibre orientation. (B) Displays some examples of characteristics computed by CSKmorphometrics classified in the different categories as specified in the National Plan[19]

Only single cells that are clearly in focus, well attached and not damaged will be used for quantification. The pipeline for single-cell quantification of cytoskeleton and nuclear structures includes several steps. The first lasso of scripts includes image cropping to identify individual cells and homogenization of the obtained images to be run by the second lasso of scripts. This way, some image preprocessing is firstly performed, as well as selection of the cells of interest. Then, the cells are run through the second lasso of script which automatically segments the selected cells to create a cell mask. For each fluorescence channel, the pipeline follows three independent steps: initial fiber segmentation, fiber refinement and determination and subtraction of non-uniform background within the cells' boundaries. This has been explained in more detail in **Annex 12.2. Procedure for single cell analysis of immunofluorescence images**. The features that are extracted include gross morphology characteristics, characteristics of the individual fibers, information about the organization of the fibers, and characteristics of the nucleus, which indirectly

provide information about the cytoskeleton as well [17,19]. Some of the outputs of the CSKmorphology software can be observed in **Figure 4.**

The total number of extracted characteristics from the obtained .txt files can be found in **Table A2** of annex **12.2. Procedure for single cell analysis of immunofluorescence images**12. Annexes**.** An additional parameter related to the level of fibroblasts activation was also obtained, as specified in **Eq.2** of annex **12.2. Procedure for single cell analysis of immunofluorescence images**12. Annexes**.**

The obtained data will be analyzed statistically, to extract meaningful insights from our data. Some of the most know programming language for statistical computing is R[54], which offers a wide range of packages and functions specifically designed for biological data analysis, such as t-tests, ANOVA, regression analysis, and survival analysis.

However, we will use GraphPad Prism.[55]*,* a commercial software widely used in biological research that provides an intuitive and user-friendly interface for statistical analysis. It offers a wide range of statistical tests, visualizations, and curve fitting tools, making it easier to establish statistical tests and graphs. In GraphPad we will analyze our data using Ordinary One Way ANOVA, which is a statistical test that is used to compare the means of three or more groups to determine if there are any significant differences between them. The test works by calculating two types of variation: within-group variation and between-group variation. If the differences between the means of the groups are significant, the between group-variation will be larger than the within-group variation. Then, the ANOVA test generates an F-statistic, which is the ratio of the between-group variation to the within-group variation. This F-statistic is compared to a critical value, based on the significance level. If the calculated F-statistic is greater than the critical value, we can conclude that there is a significant difference between at least one pair of groups. To identify the specific group differences, a *post hoc* Turkey's multiple comparison test is performed.

To dig deeper into the extraction of the results, some machine learning algorithms will be applied to our data to extract conclusions on the differences amongst groups. For a first approach we have perform applied Logistic Regression [56] from the *scikit-learn* linear-model package in Python. LogisticRegression is implemented as a linear model for classification and prediction analysis, as it considers a categorical dependent variable. The algorithm to used to solve the optimization problem has been 'liblinear' as for smaller datasets it is a good choice. Z-score transformations have been applied to normalize data.

Regarding subpopulations analysis, R has been used to implement a preliminary identification of CAFs subpopulations. K-Means algorithm clusters data by trying to separate samples in n groups of equal variances, minimizing a criterion known as the inertia or within-cluster sum-of-squares. The issue with K-Means is that the number of clusters must be specified. To solve this problem, NbClust package.[57] has been used to determine the best number of clusters This package provides 30 indices for determining

the number of clusters and proposes to user the best clustering scheme from the different results obtained by varying all combinations of number of clusters, distance measures, and clustering methods.

4.4. Final solution

The final solution for the implementation of our project, derived from the different decisions we have been taking in the previous sections, is summarized **Figure 5**.

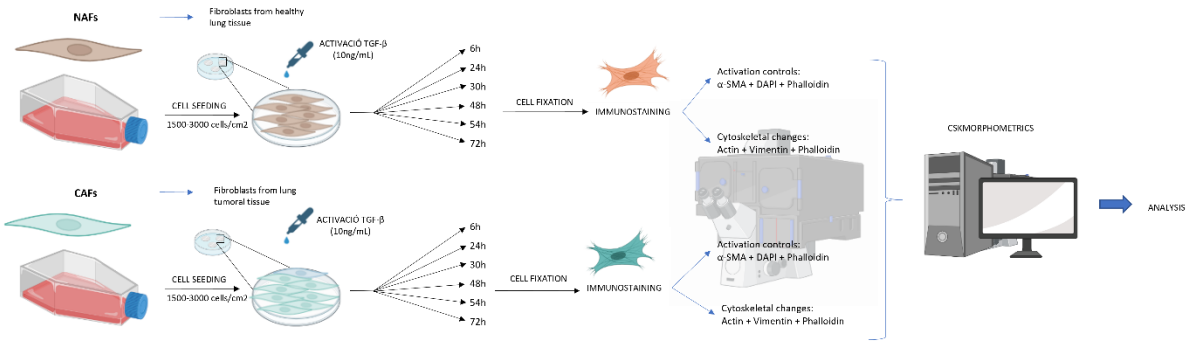


Figure 5. Final implemented solution. Cell acquisition, activation stage, immunostaining protocol, epifluorescence image acquisition, parameter extraction and data analysis are represented in the workflow. [own figure]

5. Detail engineering

In this section we show in detail the implementation of the chosen solution, as well as the obtained results and a discussion.

5.1. Material and methods

Cell Acquisition and Culture

Tumoral and non-tumoral cells from cryotubes of patient #13 were obtained at passage #9 and #10, respectively. The cells were cultured in T75 flasks until they reached a desirable confluency.

Experimental set up

Petri dishes, containing three coverslips each, were prepared. 22 petri dishes were set up for the tumoral conditions: 11 for staining with vimentin, phalloidin and DAPI tinction (to obtain information regarding cell morphology) and 11 for αSMA, phalloidin and DAPI staining (which will provide us with activation information and information about actin in fibroblasts). 22 more petri dishes were set up for the non-tumoral conditions for the same purposes: 11 for staining with vimentin, phalloidin and DAPI tinction and 11 for αSMA, phalloidin and DAPI staining.

These eleven petri dishes of each include the seven chosen timepoints: pre-activation, 6-hours activation, 24-hours activation, 30-hours activation, 48-hours activation, 54-hours activation and 72-hours activation. Besides this seven timepoints, four controls were established. Two controls were meant to check the effect of the vehicle of the activation of TGF-β, which is the addition of only the HCl 4mM that dissolves the TGF-β. Another control discarded the effect of the Insulin-Transferrin-Selenium (ITS) which is a component, we add when activating the cells, on the cell medium instead of Fetal Bovine Serum (FBS) to avoid the interference of FBS. The last control only aims to prove that cell changes are not due to the

pass of time, so the activation stimulus is not added. Due to time constraints neither 6-hour activation, 30-hour activation, 54-hour activation nor the additional controls were able to be stained nor imaged. This set up is represented in **Figure 6**.

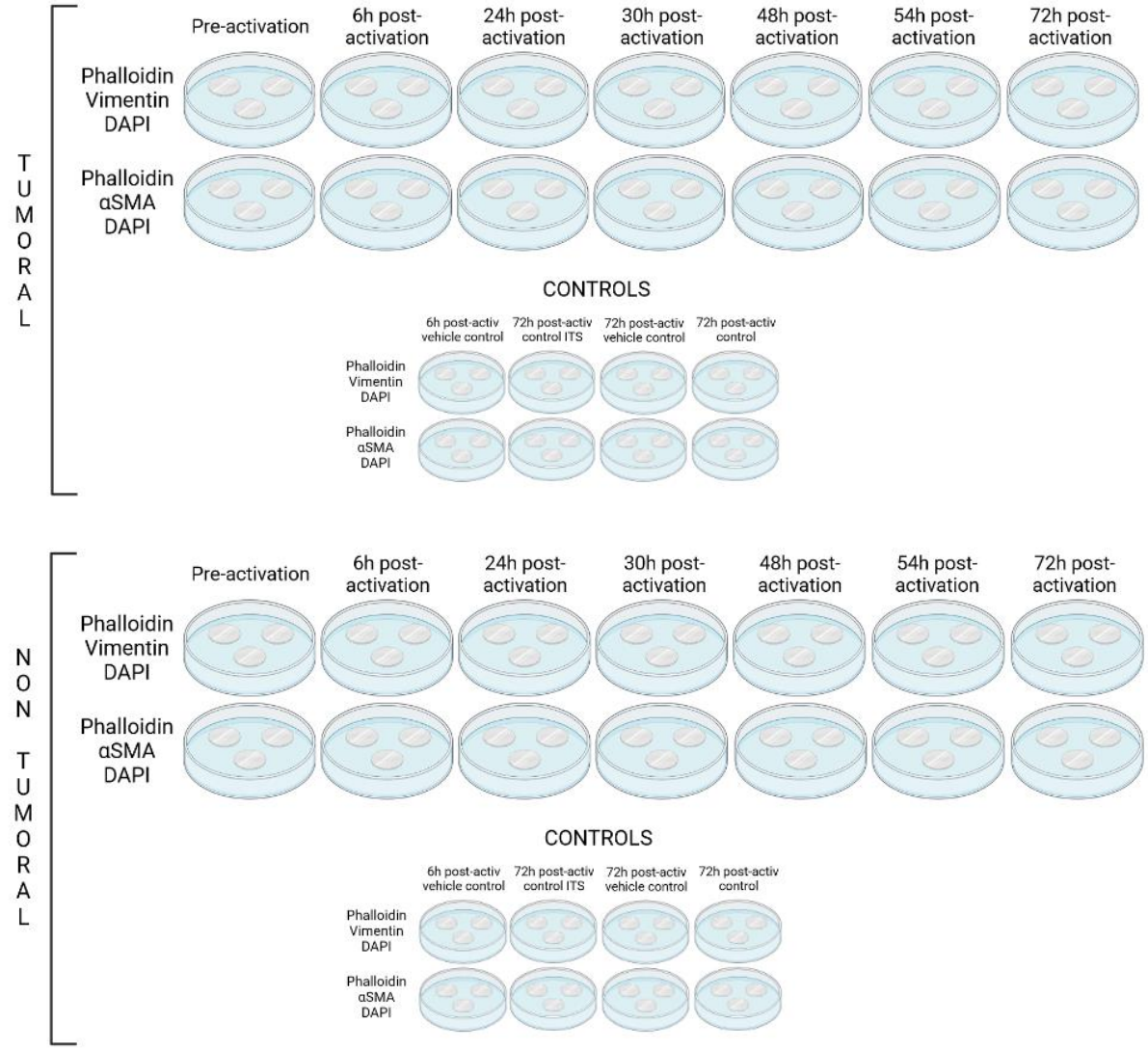


Figure 6. Set-up of the activation experiment for the tumoral and non-tumoral fibroblasts. 11 petri dishes are set up for the phalloidin, vimentin, DAPI tinction and 11 petri dishes are set up for the Phalloidin, αSMA, DAPI tinction. Post-activation refers to the hours after activation [own figure]

Sterilization and Conditioning of Petri Dishes

12 mm-glass coverslips were sterilized with 70% ethanol in 40-mm diameter petri dishes for approximately 20 minutes. Then, they were washed with PBS 1x and conditioned by adding cell media (DMEM, 10% FBS, 1% Penicillin/Streptomycin -P/S-).

Cell Seeding and Activation

Cells in T75 flasks were detached with 1mL of trypsin, after a wash of PBS 1x. Trypsin was neutralized with cell medium, and cells were centrifuged at 350G for 5'. The supernatant was discarded, and the

pellet dissolved in medium. Cells were mixed with Trypan Blue (1:2) to perform cell counting with the Neubauer chamber and the formula displayed in Eq. 1, which accounts for the dilution factors (represented as d) used.

$$Number\ of\ cells = \frac{\sum Number\ of\ cells\ on\ a\ grid}{Number\ of\ grids} \cdot 10^4 \cdot \mathrm{d} \qquad \text{Eq.1}$$

From the calculated number of cells per flask we have seeded the coverslips with a range of 1500-3000 cells per cm^2.

Cells were allowed to set and attach to the coverslips for approximately 24 hours. After 24 hours, the pre-activation coverslips were fixed with paraformaldehyde (PFA 4%), while the remaining petri dishes were conditioned with a special media without FBS (DMEM, 1% ITS, 1% P/S). Following a couple of hours, 2.5ng/ml of TGF-β was added per petri well to activate the required cells. At the predetermined time points, cell activation was stopped by fixing the cells with 2 mL of PFA 4% for 20 minutes. Three washes of PBS 1x were performed afterwards.

Immune staining protocol

Membrane Permeabilization and Blocking

A 0.2% Triton X-100 solution was used to permeabilize the cell membranes. The samples were incubated in an orbital rotator at room temperature for 20 minutes. Three PBS 1x washes were carried out afterwards. Subsequently, a blocking buffer (10% FBS in PBS 1x) was applied to block unwanted antibody binding sites and reduce background. The samples were left in the orbital rotator at room temperature for 45 minutes.

Primary Antibody Incubation

While the blocking buffer was acting, primary antibodies were prepared at concentrations of 1:1000 for anti-vimentin mouse monoclonal antibody and 1:500 for anti-αSMA rabbit monoclonal antibody. The primary antibody solution was added to the samples and left overnight at 4ºC in an orbital rotator.

Secondary Antibody Incubation

The appropriate secondary antibody was selected based on the species of the primary antibodies. Anti-rabbit secondary antibody was used for αSMA, and anti-mouse secondary antibody was used for vimentin. Both secondary antibodies were used at a concentration of 1:500. After three washes with blocking buffer, the secondary antibody solution was added to the samples, and they were kept in total darkness for 2 hours at 37ºC to promote union of the primary and secondary antibodies and increment the intensity of the fluorophore.

Phalloidin Staining and DAPI Staining

After the secondary antibody incubation, three washes with PBS 1x were carried out and the phalloidin staining solution was prepared by adding the conjugated antibody to a solution of PBS 1x and 1% BSA.

Phalloidin (1:1000) was added, and the samples were incubated in darkness at room temperature for 45 minutes. Following three more washes of PBS 1x under agitation, 1:1000 of NucBlue™ Live ReadyProbes™ Reagent were added and allowed to incubate with the samples for 10 minutes with agitation. The samples were washed with PBS 1x and stored in the refrigerator with parafilm until mounting.

Sample Mounting

Clean slides were prepared by adding two drops of Fluoromount™ onto each slide for condition. Two coverslips per condition were carefully lifted using tweezers. The excess of liquid was carefully dried off from the edges with paper slices. Then the coverslips were placed face-down over the Fluoromount™ drop and were sealed into place with clear nail-polish. The samples can be stored in a 4ºC refrigerator, ensuring they are in darkness.

Microscopic Analysis

Leica SP5 inverted microscope equipped with a CCD camera (C9100) with a Nikon CFI Plan Fluor 20XC MI objective with oil immersions was used for epifluorescence imaging. TXRED, FITC, and UV-2A channels were used to visualize the different fluorophores. The images were acquired via NIS-Elements software NIS-Elements (Nikon Corporation, Tokyo, Japan) for the three individual channels and were later exported with the same software in TIFF format as composite images for further manipulation.

Fifty pictures of single or multiple cells were obtained for each of the conditions. No cells were discriminated unless they were clearly damaged due to external factors or too agglutinated to be analysed by the software. Two coverslips for each of the conditions were sampled for each condition, randomly selecting different points to obtain cells from different parts. Due to time constraints, fifty cells were finally analysed from the acquired images for each of the conditions.

5.2. Results

In this section, we present the results of our study, focusing on different timepoints for both morphology and activation level analyses. Specifically, we have included data from the pre-activation stage and the timepoints at 24 hours, 48 hours, and 72 hours for morphology evaluation. For the assessment of activation level, we have only included data from the pre-activation and the 72-hour activation stage. From now on, in this section, CAFs are also referred as tumoral fibroblasts (or T) and NAFs as non-tumoral fibroblasts (NT).

5.2.1. Activation levels

Firstly, we aim to assess the effectiveness of the activation process through the application of a stimulus (TGF-β). To do so, we have measured the level of activation in both non-tumoral and tumoral fibroblasts before and after 72 hours of persistent activation stimulus.

To evaluate the success of induced activation, we calculated the activation level based on the analysis of different images. **Figure 7A** and **Figure 7B** depict a noticeable difference between the green αSMA

fibres between the pre-activation stage and the post-activation stage of the tumoral and non-tumoral fibroblasts. This visual observation suggests that the activation stimulus has indeed induced changes in the cells.

To further validate these visual differences, we quantified the levels of αSMA in both the pre-activated and activated tumoral and non-tumoral fibroblasts, as illustrated in **Figure 7C**. The analysis revealed a significant disparity in the calculated αSMA levels between the pre-activated and activated cells. This statistical evidence strengthens our conclusion that the activation stimulus has provoked cellular activation.

Interestingly, our finding also indicates that there is no significant difference in the activation levels between non-tumoral and tumoral fibroblasts. This implies that both cell types found themselves in similar activation levels before activation and that the application of the stimulus provided similarly effective across different cellular context.

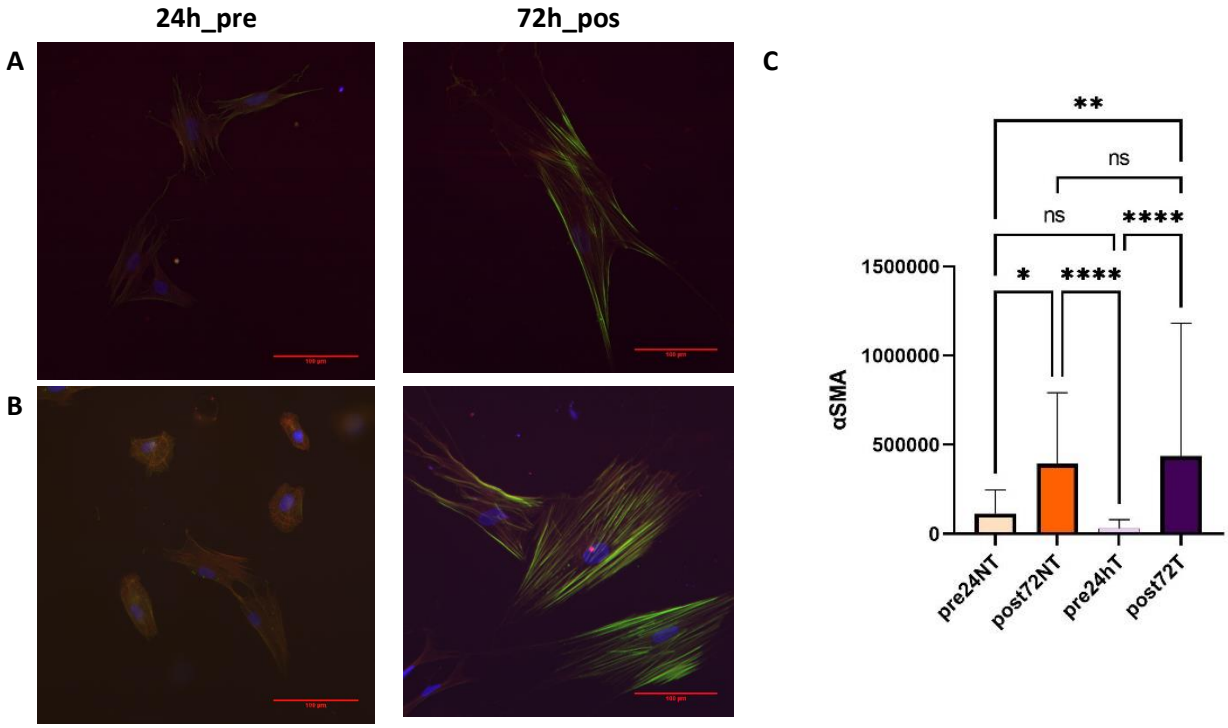


Figure 7. Activation levels of the T and NT fibroblasts at different timepoints. (A) shows NT fibroblasts prior to the activation (pre24h) and 72h after being under a persistent activation stimulus (pos72h). (B) shows T fibroblasts prior to the activation and 72h after being under a persistent activation stimulus. (C) displays the level of αSMA of the four different conditions, showing the statistical difference obtained from the One-Way ANOVA test.

In addition to the differences in activation levels, **Figure 7A** and **Figure 7B** start providing valuable insights into the morphological and cytoskeletal variances between tumoral and non-tumoral cells, as well as between pre-activated and post-activated cells.

5.2.2. Morphological changes during activation

To further investigate these differences, we analyzed the images in which the cytoskeleton was stained with actin and vimentin. **Figure 8A** and **Figure 8B** displays these images of the non-cancerous fibroblasts

and CAFs correspondingly at different activation stages: pre-activation and post-activation at 24h, 48h and 72h under persistent stimulus.

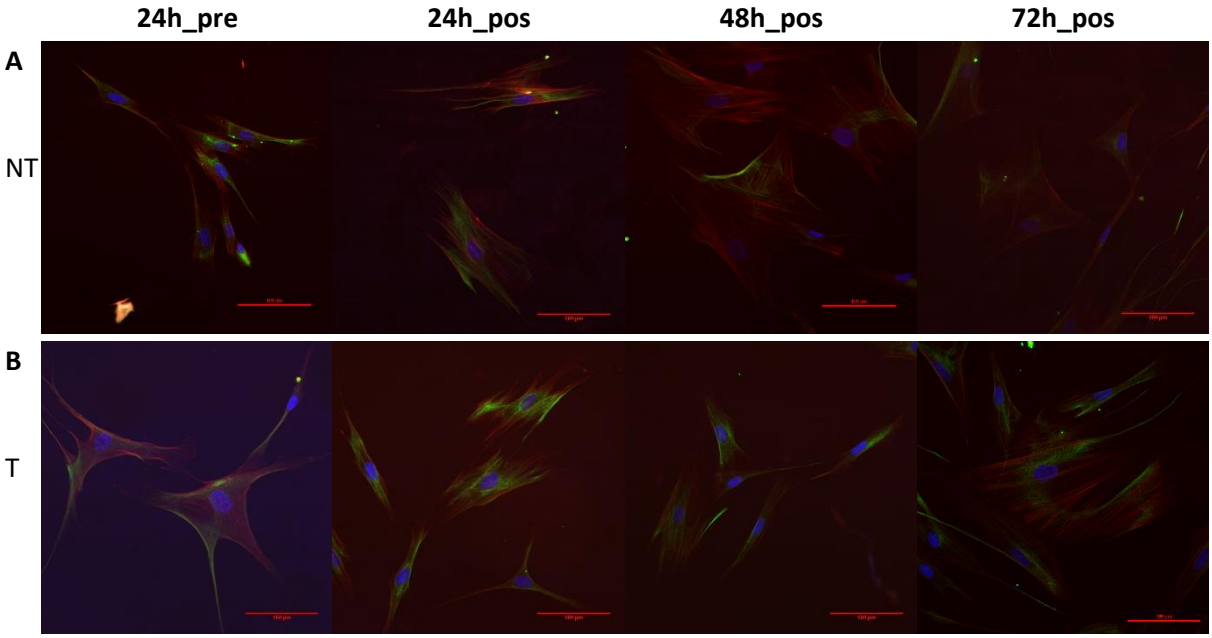


Figure 8. Non-tumoral (A) and tumoral (B) fibroblasts imaged at different activation stages: pre-activation (24h_pre), 24h under persistent stimulus (24h_pos), 48h under persistent activation stimulus (48h_pos) and 72h under activation stimulus (72h_pos).

From these images we can already derive some morphological changes the cells undergo when they are subjected to activation. When fibroblasts are activated, they lose their elongated fusiform aspect to become wider and more spread. They also become bigger, which is something we realized at the epifluorescence microscope, when fitting whole cells in the image frame became more difficult. While the change from fibroblasts to myofibroblasts, the change between the non-activated and activated state of CAFs is more difficult to appreciate.

To statistically confirm these visual changes, we have extracted various parameters pertaining to morphology, cytoskeletal fibers, fibers reorganization and the nucleus from the analyzed images.

In this discussion, we will firstly focus on the changes observed in gross morphology, specifically the fibroblasts area, cell aspect ratio and the convexity/concavity ratio of the entire cell outline. These changes are depicted in **Figure 9**, where one can observe the aforementioned parameters of the non-tumoral and tumoral cells at the different activation stages.

In the analysis of gross morphology parameters, an intriguing observation emerges when examining the cell area between different activation stages of non-tumoral cells. Notably, there is a significant change in cell area as the cells transition from a non-activated state to an activated state. Non-activated non-tumoral cells initially possess a considerably smaller cell area compared to their activated counterparts. These cells rapidly reach a maximum cell area at the 24-hour mark, and the cell area remains relatively stable thereafter at 48 hour and 72 hours post-activation. In contrast, the cell area of tumoral cells does not exhibit noticeable changes between the pre-activation and activation stages (**Figure 9A, 9D).**

Regarding aspect ratio, non-significant changes appear across the different activation stages of the non-tumoral cells. However, tumoral cells present a lower aspect ratio when activated compared to their non-activated counterparts. This means that non-activated ells are more elongated and activated cells are, on the other hand, more rounded (**Figure 9B, 9E).**

Furthermore, both tumoral and non-tumoral fibroblasts display an overall concave shape. Activated non-tumoral fibroblasts are slightly more concave than non-activated non-tumoral fibroblasts, and tumoral fibroblasts, whether in a non-activated or activate state, maintain a concave shape at a similar level to that of non-activated fibroblasts (**Figure 9C, 9F).**

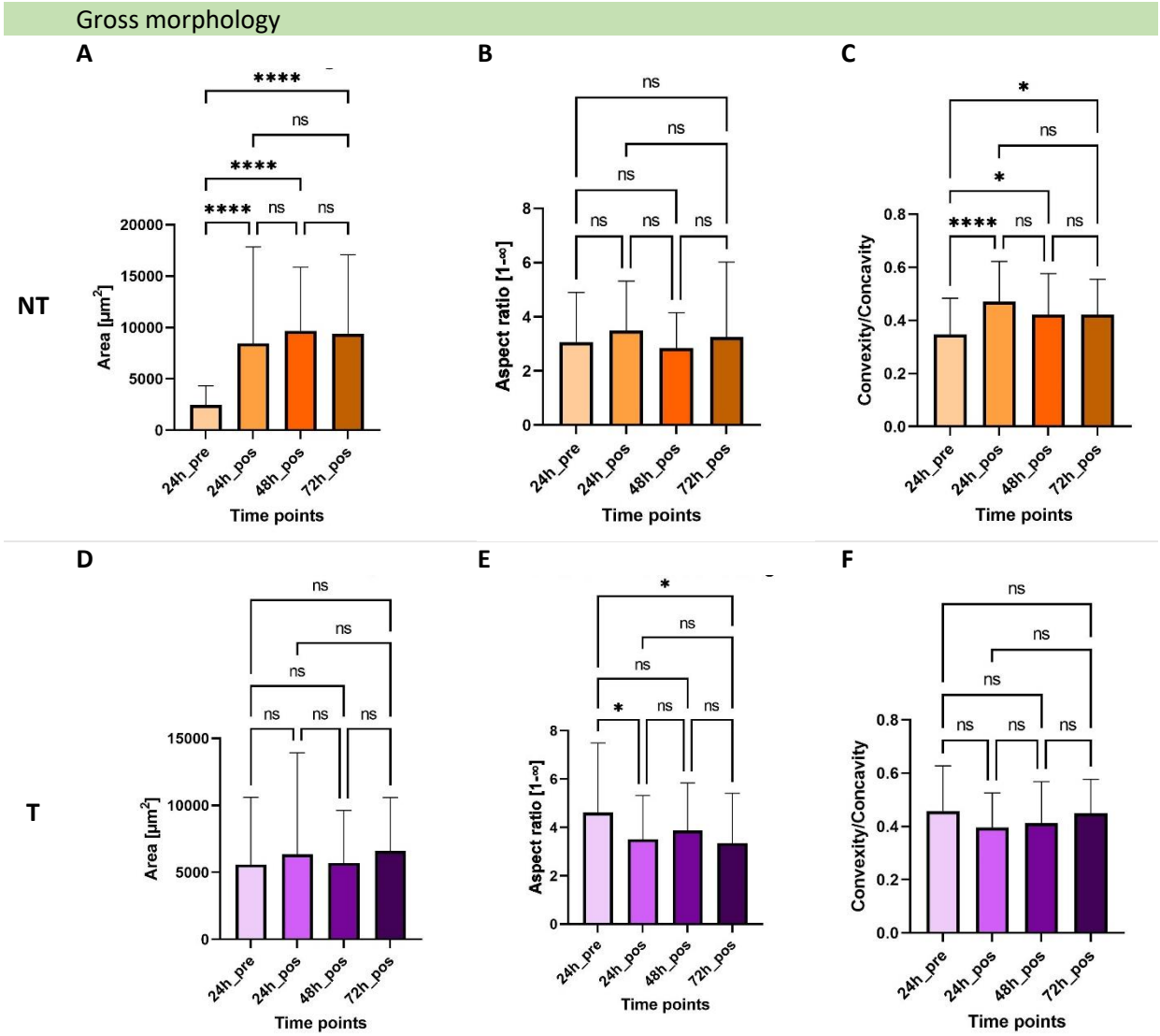


Figure 9. One-way ANOVA analysis of gross morphology parameters extracted from the cytoskeletal images of the cells at various activation stages. (A) and (D) represent the cell area of the non-tumoral and tumoral fibroblasts, respectively; (B) and (E) represent the aspect ratio of the non-tumoral and tumoral fibroblasts, respectively; and (C) and (F) represent the convexity/concavity ratio of the tumoral and non-tumoral cells respectively.

5.2.3. Cytoskeletal changes during activation

Changes in the gross morphology of the cells are ultimately provoked by changes observed in cytoskeletal fibers during the processes of activation. To gain insight into this transformation, we will assess several parameters. From parameters related to the individual fibers, we include in this section

the total fluorescence of the actin fibers and the fibers length. From studying the organization of the fibers within the cell, we focus on the fiber alignment and the fiber spread. Additional parameters have been analyzed and can be found in **12.3. All figures for the obtained compiled results of tumoral and non-tumoral fibroblasts.**

Figure 10 shows the analysis of the previously mentioned parameters of the tumoral and non-tumoral cells at different activation stages.

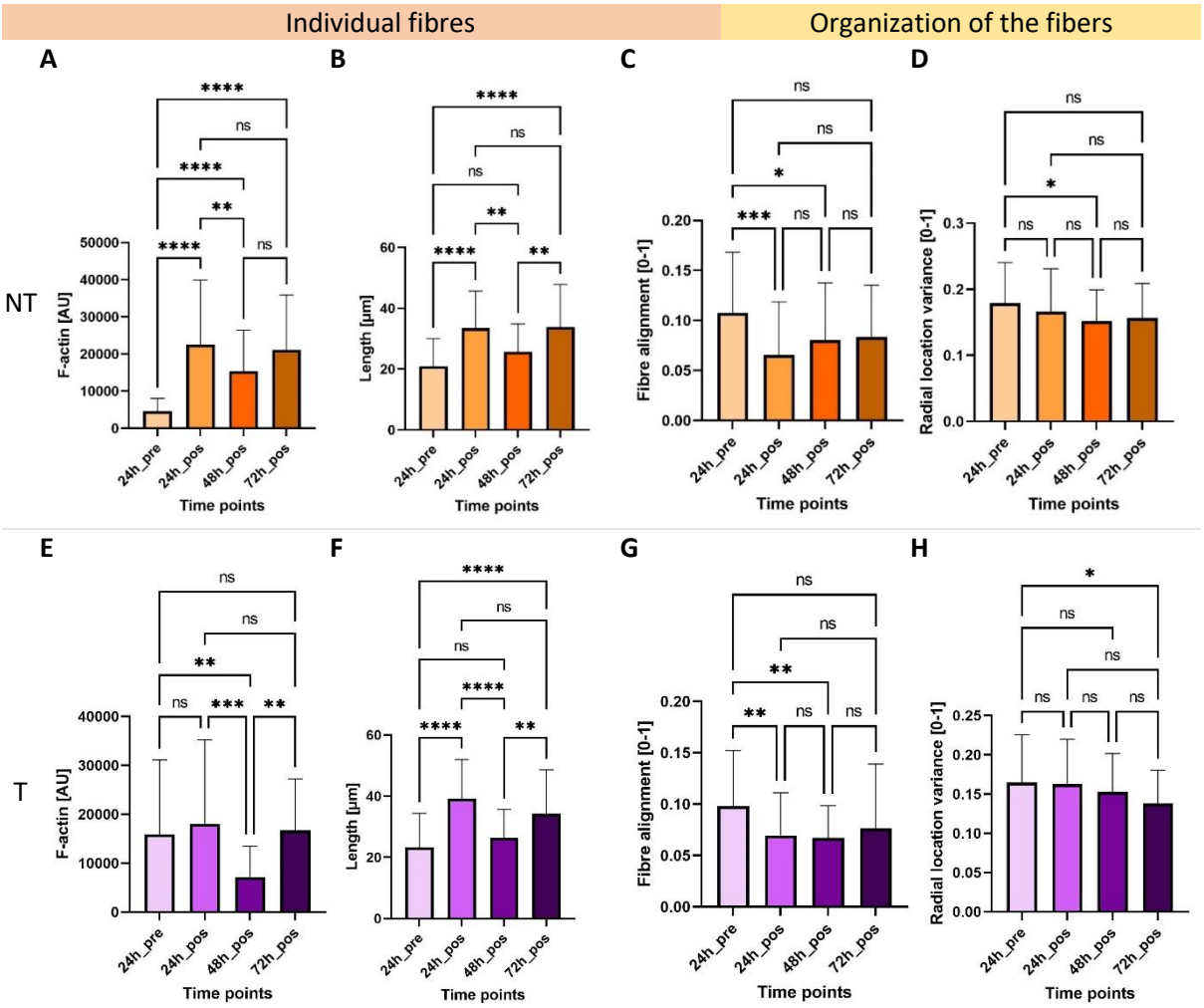


Figure 10. One-way ANOVA analysis of fiber-related parameters extracted from the cytoskeletal images of the cells at various activation stages. (A) and (E) represent the total fluorescence associated with the F-actin fibers of the non-tumoral and tumoral fibroblasts, respectively; (B) and (F) represent the length of the fibers of the non-tumoral and tumoral fibroblasts, respectively; (C) and (G) represent the convexity/concavity ratio of the tumoral and non-tumoral cells respectively; and (D) and (H) represent the fiber spread as radial location variance of the tumoral and non-tumoral cells respectively. AU means Arbitrary Units.

In analyzing the fibers individually, we observe a noteworthy increase in both parameters on non-tumoral cells. There is a significant augmentation in the total fluorescence of actin fibers and the length of individual fibers when non-tumoral fibroblasts go from a non-activated to an activated state. In similar lines, fibers in non-activated non-tumoral fibroblasts tend to be shorter than those of the activated cells **(Figure 10A, 10B).** An extremely similar behavior can be observed in the case of the fiber length of tumoral fibroblasts. On the other hand, although there are no changes in total actin fluorescence between

the non-activated and activated stage, we can observe significant changes in fluorescence at different activation time points, highlighting a drastic decrease at 48 hours. This phenomenon has also been observed in the analysis of non-tumorous samples **(Figure 10E, 10F).**

When examining the organization of fibers, although both study groups (tumoral and non-tumoral) maintain a consistent alignment of the fibers, with the fibers being mostly aligned, it is important to note that the most significant changes at this level occur during the first 24 hours of activation **(Figure 10C, 10G).** Fiber spread shows that fibers tend to be preferentially concentrated at a specific point of the cell, particularly closer to the cell periphery. Interestingly, in tumoral cells, as the activation progresses, the spread of the fibers becomes progressively more concentrated at a specific point of the cell. This suggests that the fiber distribution within tumoral cells becomes increasingly focused and localized during the activation process. **(Figure 10D, 10H).**

Additionally, we will explore the relationship between the nucleus and the cytoskeleton, examining aspects such as chromatin condensation, nuclear volume, nuclear stiffness, and nuclear centrality. By studying these parameters, we can gain a comprehensive understanding of the dynamic alterations occurring in cytoskeletal fibers during the activation processes.

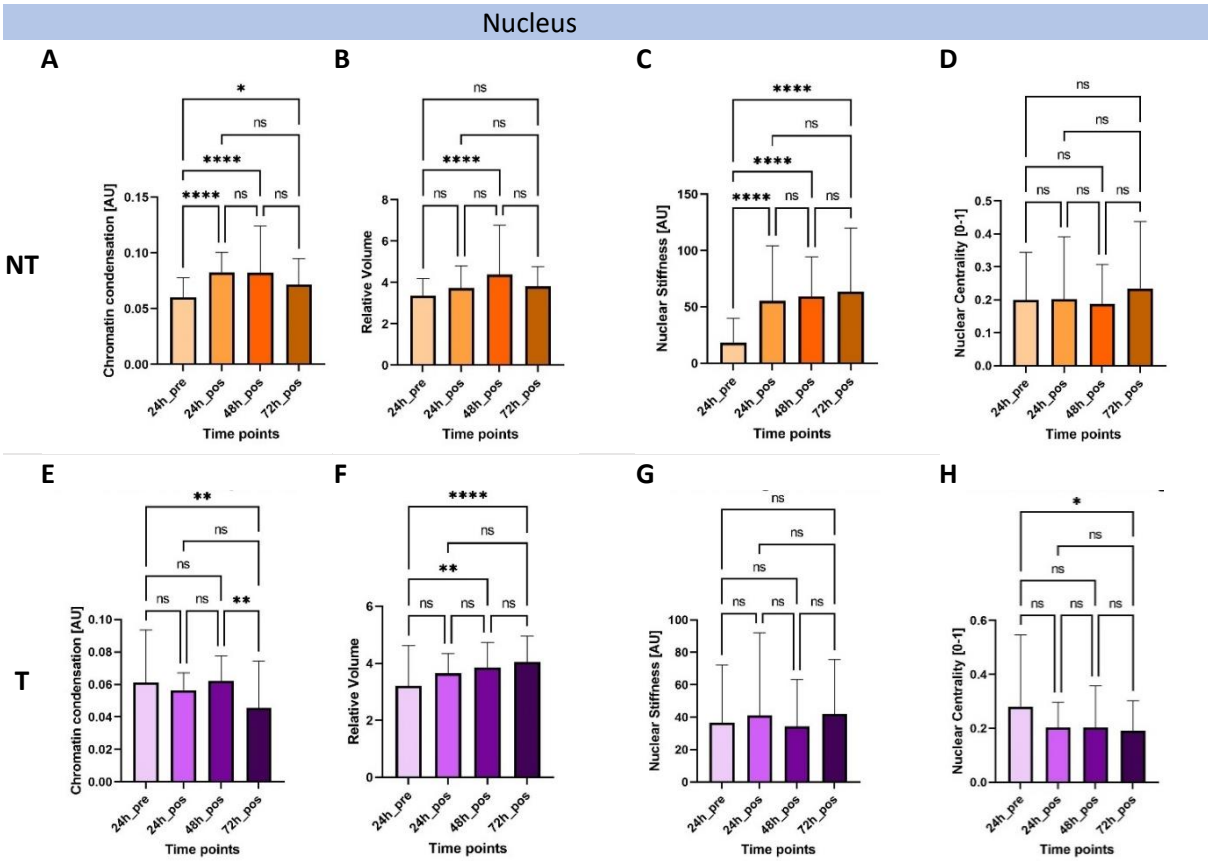


Figure 11. One-way ANOVA analysis of nucleus related parameters extracted from the cytoskeletal images of the cells at various activation stages. (A) and (E) represent the chromatin condensation of the nucleus of the non-tumoral and tumoral fibroblasts, respectively; (B) and (F) represent the nuclear volume of the non-tumoral and tumoral fibroblasts, respectively; (C) and (G) represent the nuclear stiffness of the tumoral and non-tumoral cells respectively; and (D) and (H) represent the nuclear centrality of the tumoral and non-tumoral cells respectively. AU means Arbitrary Units.

Regarding chromatin condensation, we observe contrasting trends between non-tumoral and tumoral cells during activation. In non-tumoral cells, chromatin condensation appears to increase as the activation progresses. However, in tumoral cells, there is a noticeable decrease in chromatin condensation throughout the activation process (**Figure 11A, 11**Figure 11**E).**

Furthermore, tumoral cells exhibit a significant increase in nuclear volume during activation, indicating a substantial expansion of the nucleus (**Figure 11F).** In contrast, while non-tumoral fibroblasts experience a significant increase in nuclear stiffness as they become activated, there are no significant changes in nuclear stiffness observed in tumoral fibroblasts (**Figure 11C, 11G).**

Concerning the alignment of the nucleus with the fibers, while we haven't seen changes in the non-tumoral fibroblasts stages, we see that in tumoral cells the cytoskeleton displaces the nucleus towards the center of the cell as the tumoral fibroblasts activate. (**Figure 11D, 11H).**

5.2.4. Refinement of the analysis through machine learning

To further refine the obtained results, we have applied machine learning techniques to our data. Logistic regression has been used to calculate an accuracy score, which measures how well the predicted labels of our data match the actual data labels. This analysis allowed us to make four different comparisons, the results of which can be seen in **Table 3**.

Table 3. Average accuracy score obtained from the different train-test partition of Logistic Regression algorithm

	Multiple-iteration average accuracy score
Prediction of NT activation stage	64%
Prediction of T activation stage	58%
Prediction of T or NT non-activated cells	81%
Prediction of T or NT 72-hour activated cells	68%

We have computed the ability of the algorithm to correctly predict the four different timepoints (24h pre-activation and 24h, 48h and 72h post-activation) in both the non-tumoral and tumoral groups. While the algorithm can discern with an average accuracy of 64% the different non-tumoral fibroblasts at the different activation timepoint, only an average accuracy score of 58% can be obtained when trying to classify tumoral cells.

These findings confirm that non-tumoral cells present more pronounced differences amongst different activation stages compared to tumoral cells. This difference becomes more evident when we simplify the classification into two different classes: pre-activation and post-activation. For non-tumoral fibroblasts, the classification score grows to 80%.

Additionally, we aimed to find differences between the tumoral and non-tumoral cells in both non-activated and activated fibroblasts (at 72 hours). As our observations have shown, we can conclude that the non-activated state of resting-state fibroblasts differs significantly from tumoral fibroblasts, with the algorithm achieving an accuracy of 81% in distinguishing between these categories. We do not find as

many differences between the activated stage of tumoral and non-tumoral fibroblasts, where the accuracy score drops to 68%.

5.2.5. Clustering to identify CAFs subpopulations

By means of K-Means clustering, we have studied CAFs subpopulation in their pre-activation stage and at 72h post-activation stage to gain insights on how CAFs subpopulations evolve during activation. **Figure 12A** illustrates that non-activated CAFs have been optimally clustered into five different subgroups. Upon investigating the area of the different clusters, we see significant differences between the different groups. It is noteworthy to mention that the majority of cells are classified into clusters with smaller area (45 cells) while only two clusters present a larger area (22 cells).

Interestingly, CAFs subjected to persistent activation stimulus for 72-hour exhibited a different clustering pattern. **Figure 12B** demonstrates that they have been optimally separated into two different groups, one with the majority of cells that present a larger area, and another with fewer cells that have a smaller area.

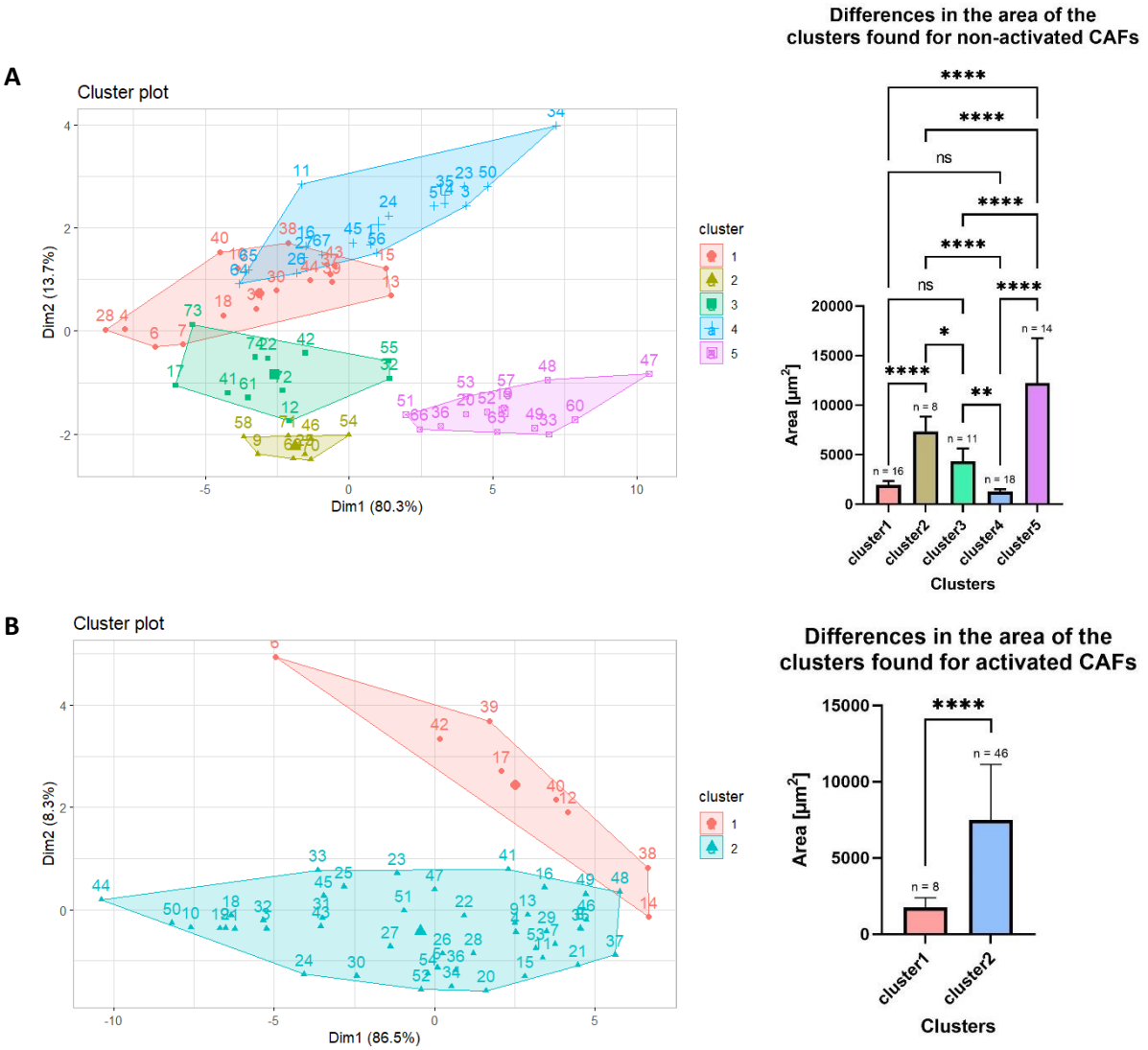


Figure 12. K-Means clustering for (A) 24h pre-activation tumoral fibroblasts and (B) 72h post-activation tumoral fibroblasts. The statistical differences on the area have also been displayed for (A) and (B).

5.3. Discussion

In this study, we have obtained various intriguing results, shedding light on the morphological and cytoskeletal changes observed in our experimental activation of fibroblasts. These findings provide valuable insights into the dynamic cellular processes and adaptations occurring during the activation stages. In the following discussion, we will delve into the implications of these results and explore their potential implications in the context of cellular behavior and function.

The changes in activation levels defined by αSMA validate the process of induced activation experienced by our cells. These activation processes give rise to distinct morphological and cytoskeletal alterations within both the non-tumoral and tumoral groups. These findings highlight the impact of activation on cellular characteristics and provide evidence of the specific changes induced by the activation process.

The transition of quiescent fibroblasts into a myofibroblastic state involves cytoskeleton reorganization and morphological changes.[1] Our findings support this knowledge as we have demonstrated that the area of fibroblasts does indeed increase upon activation. This increase in cell area can be attributed to the augmentation and elongation of actin fibers within the cells, which we have also characterized. This is consistent with the drastic changes fibroblasts undergo when transforming into myofibroblasts, who acquire a well-developed contractile apparatus with the presence of a robust network of actin stress fibers, which orchestrate ECM remodeling and enable the fibroblasts to gain new functionalities.[58]

What we have been able to note, is that the first 24 hours under persistent activation stimulus are crucial as major changes occur. Yang et al. also demonstrated that TGF-β induced activation reached its peak at 48 hours and maintained a stable level until 72 hours through the use of FAP, fibronectin, and αSMA gene expression.[22]

For tumoral cells, changes between pre-activated and activated stages are more subtle, specially morphologically wise. As CAFs already present a morphology like that of myofibroblasts[9], the presence of a persistent stimulus to mimic continuous activation produces faint effect to the morphology of the fibroblast. The cytoskeleton does see some changes, specifically on the length of the fibers and more curiously on the preferent location of the fibers at a certain point within the cell.

In addition to their impact on cytoskeleton and cellular morphology, these changes must also reflect significant alterations in the functional capabilities of the cells.[8] When fibroblasts transition into myofibroblasts, they acquire a multitude of functions, including the ability to produce extracellular matrix[1]. It is logical to assume that the observed morphological changes in CAFs are accompanied by corresponding functional changes. For instance, the concentration of fibers at the cell periphery may indicate an enhanced migratory capacity of CAFs [1].

The alterations in the nucleus may also contribute to the adaptation and adjustment of the nucleus to its new functionality, prompted by the acquisition of the activated status. Nuclear change may be attributed to changes in nuclear transcription, as the cells undergo modifications to produce the necessary proteins

required to induce the cytoskeletal changes or other proteins to support whichever new functionality has been gained.

In fact, multiple studies have described transcriptional differences in wound/damage and cancer contexts. In the context of injury, an enrichment of various pathways and regulators associated with inflammation has been observed, including IL-6, STAT3, and NFAT [59]. However, when discussing a cancerous context, fibroblasts not only express inflammatory genes but also exhibit significant and specific changes. These changes include an enrichment of genes related to cellular stress, remodeling, and extracellular matrix secretion [60].

The results of the logistic regression to classify the different study groups revealed the differences between subgroups in a combined manner. Thus, the best classification achieved with 81% accuracy is between tumoral and non-tumoral samples, indicating that both cell types have distinct cytoskeletal characteristics enabling clear differentiation. However, the predictions within different time points within the same group did not yield such good metrics, suggesting that intragroup changes are much smaller compared to intergroup differences. Additionally, it is important to consider that the dataset used for these predictions is limited by the number of cells, so we recommend validating these results with a larger dataset or with more complex algorithms, such as convolutional neural networks, some of examples of which include SigNet or FaceNet.[61,62]

As it can be appreciated in the obtained images, cells of various shapes and sizes appear on the different conditions. This denotes the high heterogeneity of CAFs. And the thing is, CAFs are actually defined as a heterogeneous group of mesenchymal cells due to their various potential cellular origins. Several markers have been individually tested in tumors. This quickly leads to the demonstration that CAFs are heterogeneous in cancer.[63]

To study this heterogenicity, a study on CAFs subpopulations has been conducted by the means of clustering. While our analysis is only preliminary, the diminishing of the optimal number of clusters in pre-activated and post-activated CAFs makes us think that CAFs converge morphologically under the application of a persistent activation stimulus. To confirm this convergence, in-between timepoints should be analyzed.

Clustering CAFs into functional subtypes and correlating each of them with other properties of the TME has demonstrated the complexity and heterogeneity of CAFs.[10] However results are very difficult to contrast with current published literature as the definition of CAFs subpopulation is very disparate between articles. It is worth mentioning that curiously we have achieved the two clusters classification diverse papers present under different names. [10] We should perform further studies to characterize these clusters.

In the case of pre-activation, we can observe that the majority of cells are in clusters with lower area indicating that they are indeed not activated. However, some population of non-activated CAFs already present an area that surpass that of activated fibroblasts. This might be because CAFs already are

considered to have a basal degree of activation and therefore were indeed already activated when subtracted from the patient. However, another feasible theory could be that the culturing of non-activated cells in stiff glass slides activate fibroblasts. There have been some articles that elucidate an activation response of the fibroblasts to the stiffening of the substrate. They confirm that fibroblast activation in response to substrate stiffening is associated with a myofibroblastic phenotype, furtherly characterized by prominent stress fibers rich is αSMA [64].

In the case of activated CAFs most cells are found in the cluster with a larger area, possibly indicating that the majority of cells have undergone the persistent activation. However, there are some cells that have not seen an increase on cell shape. These phenomena could be further investigated with functional techniques.

It is crucial to acknowledge that although these results offer substantial evidence supporting the changes in morphology and cytoskeletal reorganization resulting from the addition of an activation stimulus in tumoral and non-tumoral cells, several limitations need to be considered.

Time constraints have limited the number of cells and conditions we have included in this analysis, such as the intermediate time points at 6 hours, 30 hours, and 54 hours post-activation. Adding more cells to each condition would enhance the statistical power and reliability of our analysis, as a larger sample size reduces the impact of individual variations and sampling bias. Moreover, a larger cell population would enable a more detailed clustering analysis, as machine learning algorithms perform better with more data. On the other hand, exploring the 6-hour time point gains special importance to understand the transition between the inactive state of NAFs and the 24-hour activation stage of NAFs.

The findings could also potentially be influenced by the individual characteristics of the patient the cells were extracted from. Studying cells from other patients with the same or different tumor types and various stages would provide further insights into the role of fibroblasts in cancer.

6. Implementation schedule

In this section, we will detail how we will carry out to accomplish the project's objectives in terms of temporal and economical planification.

6.1. Work-breakdown structure (WBS)

A work-breakdown structure (WBS) in project management and systems engineering is a deliverable-oriented breakdown of a project into smaller components.

Our project is organized around the main activation experiment. The different work packages that lead up to it include the obtention of the required material and object of study, as well as the planification and preparation. A previous work packages regarding the experiments we performed to control the cell growth during a span of three days are also included.

The activation experiment has been divided into two main parts: the induction of activation in non-tumoral cells to evaluate what happens in physiological conditions and the induction of activation in tumoral cells

to study CAF's activation in lung cancer. A third set of work-packages related to the analysis of the data are needed to obtain results regarding the differences between tumoral and non-tumoral cells, analyze the evolution of activation in time and finally detect CAF's subpopulations.

Throughout this work-packages, we have established some transversal work-packages that must accompany the whole project, such as monitoring or redacting the project, or maintaining cells during cellular experiments.

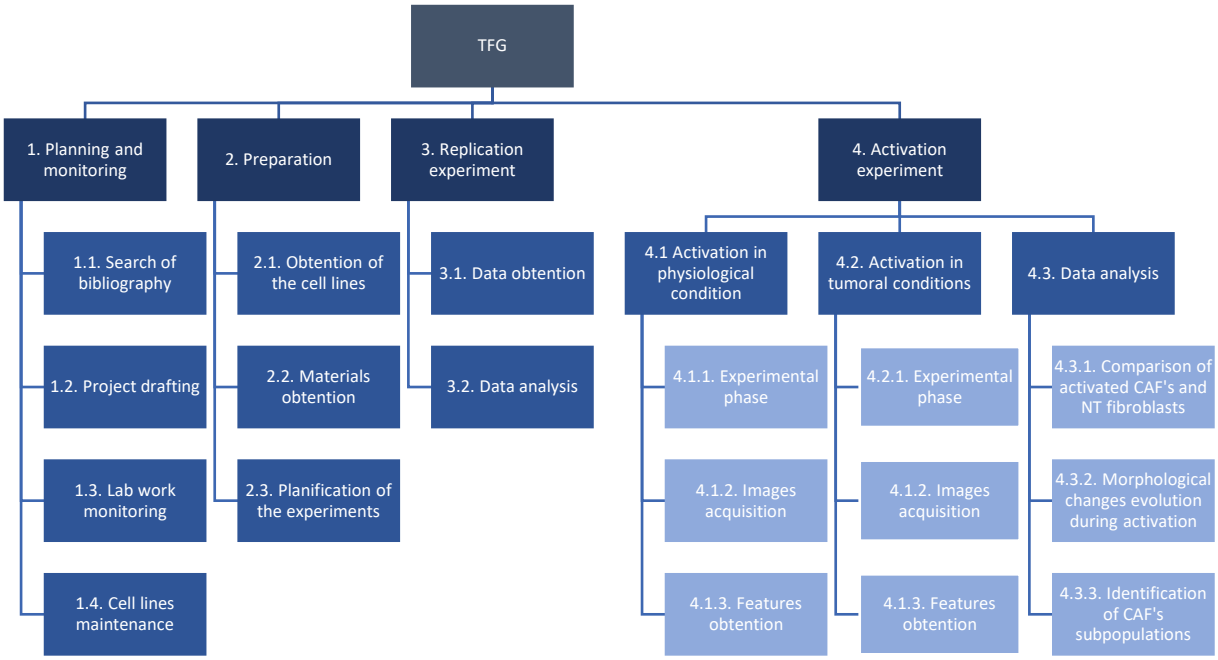


Figure 13. Work break-down structure of the project.

Figure 13 shows the WBS of our project. From it, a WBS dictionary can be elaborated to detail each of the work-packages, with their activities and deliverables. The required resources and risks of the different are also detailed in the dictionary. **Table 4** shows the WBS dictionary.

Table 4. Work break-down structure dictionary of the project

1. Planning and Monitoring	
1.1. Search for bibliography	
Description	Before defining the idea, it is key to know what has been done in this area, so that we can define in which ways we can widen the knowledge in the field. Furthermore, this bibliography will be very useful when writing the project.
Deliverable	Bibliography of the articles read and used for the writing of the project
Activities	- Search for articles related to the topic. - Reading of said articles. - Extraction of ideas from these articles.
Duration	This work packet will accompany the writing of the project
Required resource	Internet access and journal access to the articles (provided by Universitat de Barcelona)
Risk	-
1.2. Project drafting	

Description	Defining the main ideas of the project and writing down the project memory as it is being carried out.
Deliverable	TFG document
Activities	- Definition and presentation of the main ideas of the project. - Redaction of the TFG document with the main ideas, the carrying out of the project and the results.
Duration	1 week (at the beginning of the project) and throughout the carrying out of the project.
Required resource	-
Risk	-
1.3. Lab work monitoring	
Description	Writing out everything that is being done daily in the lab, as well as keeping record of the performed experiment to make sure that the results are believable.
Deliverable	Lab notebook
Activities	- Writing down what has been done after each day. - Keeping track of the procedure and the different variables that may affect our experiments.
Duration	Throughout the wet lab phase of our project
Required resource	-
Risk	Due to the combination of this project with mandatory assistance to academic classes, the experiments must sometimes go on without my presence there. Although we were on the lab for the maximum time as possible, certain fixation timepoints were carried out by the PhD student. To keep track of what has been done, we ask the PhD student for her notes, as she writes down in her own notebook what is being done at every step of the process.
1.4. Cell line maintenance	
Description	During the wet lab experimentation, we must preserve our cell lines to have enough cells to perform our experiments. To do so, several steps must be carried out: change of media, cell passing, cell freezing and good laboratory practices.
Deliverable	Lab notebook with a daily diary

Activities	- Cell media preparation
	- Non-tumoral fibroblasts line
	  o Change of media every two days.
	  o Cell passaging once the confluence is above 80%.
	  o Cell freezing to have cells at a determined passage.
	- Tumoral fibroblasts line
	  o Change of media every two days.
	  o Cell passaging once the confluence is above 80%.
	  o Cell freezing to have cells at a determined passage

Duration	Throughout the wet lab phase of our project
Required resource	For the cell media: DMEM high-glucose, Fetal Bovine Serum (FBS), P/S (antibiotics) Trypsin PBS DMSO (Dimethyl Sulfoxide) for cell freezing. Cell culture hood and lab equipment (centrifuge, pipettes, T75 flasks…)
Risk	- Delay on the lab orders: not being able to obtain the required components on time. The provider responded quickly to the provision of material. Meanwhile, we can borrow material from adjacent labs. - Cell culture contamination and cell death due to external causes. By freezing cells at different time point we will be able to recuperate them.
2. Preparation	
2.1. Obtention of the cell lines	

Description	To do our experiments, we need to obtain cell lines. We need two different cell lines: tumoral fibroblasts and non-tumoral fibroblasts from the same patient. If possible, we need them to be in a similar cell passage number. Once they are obtained, the cells must reach good enough confluency and conditions before starting the experiments. Cryotubes were obtained from another lab group.
Deliverable	Cell flasks
Activities	- Obtention of tumoral cell line cryotube - Thawing of the tumoral cell line - Culturing of the tumoral cell lines - Obtention of non-tumoral cell line cryotube - Thawing of the non-tumoral cell line - Culturing of the non-tumoral cell line
Duration	2 weeks
Required resource	Cell flasks For the cell media: DMEM high-glucose, FBS, P/S (antibiotics) Cell culture hood and lab equipment (centrifuge, pipettes, T75 flasks…)
Risk	- Inability to obtain fibroblasts cell lines. This risk is unlikely as other groups are also working with fibroblasts. - The primary vials had to undergo an immortalization process carried out by collaborator lab group. This process can be harsh on cells, which could entail the risk of losing the cells. - Freezing and thawing of cells may also entail the risk of losing cells, as DMSO is toxic for them.
2.2. Materials obtention	
Description	To do our experiments (and as well maintaining our cell lines), we need reactants and other materials that we must ask for in a regulated process through the university or the different lab grants.
Deliverable	-
Activities	- Inscription to the orders Excel, budget revision, buying from the providers, delivery of the asked materials - Obtention of the cell culturing material (DMEM, FBS, trypsin, P/S…) - Obtention of the activation experiment reactants (HCl, TGF-β…) - Obtention of the immunostaining protocol reagents (primary and secondary antibodies, DAPI…)
Duration	1 to 2 weeks
Required resource	-
Risk	- Non arrival of the orders in time due to the slowness of the process to ask for material as a regulated process in the lab. To avoid this, we asked for material to adjacent labs while material was on its way. - Non arrival of the orders due to the capacity of the budget - Non arrival of the orders due to the provider
2.3. Planification of the experiments	
Description	Concretization on how we are going to perform the different experiments: from writing down the protocols that we are going to follow, to deciding in which dates we are going to do them, to organizing the methodology and different materials we are going to use.
Deliverable	Experiment protocols
Activities	- Writing down the protocol we are going to follow to perform the experiment. - Programming in the calendar the carrying out of the experiments and booking the cell culture hood for the specific hours we will need it. - Gathering and organizing the needed material
Duration	Throughout the wet lab experimental phase of the project
Required resource	Access to the cell culture hood calendar

Risk	- High booking of the cell culture hood. The solution to this risk is to book the culture hood shared between the different groups in the third floor of the faculty.
3. Replication experiment	
3.1. Data obtention	
Description	Replication of the conditions of the activation experiment in different wells where cells will be plated to count how many cells do we have at different timepoints. This way, we will control that, when we do the activation experiment, we do not have too many or too little cells per well.
Deliverable	Document with the number of cells on each well
Activities	- Cell plating a determined number of cells on each well of some well plates. - Cell activation and setting the controls of the designated wells. - Stopping cell replication at the required time. - Cell counting with the Neubauer chamber
Duration	5 days
Required resource	Cell flasks (tumoral and non-tumoral) Well-plates Cell medium, trypsin, PBS HCl, TGF-β Tripan Blue, Neubauer chamber, tally counter Cell culture hood, microscope, pipettes and other cell culture material
Risk	- Death or contamination of the cells in the wells. The solution of this would be the repetition of the experiment.
3.2. Data analysis	
Description	Obtention of quantity, time-evolution graphs, and conclusions from the data obtained during the experimental phase of the experiment.
Deliverable	Graph of the evolution of the quantity of cells in time depending on the different conditions.
Activities	- Calculation of total number of cells per well - Plot the evolution of the quantity of cells in time depending on the different conditions. - Derivation of conclusions of how many cells can we plate for the activation experiment.
Duration	1 day
Required resource	Excel sheet Data obtained from cell counting with the Neubauer chamber.
Risk	-
4. Activation experiment	
4.1. Activation under physiological conditions	
4.1.1. Experimental phase	
Description	Cell plating the desired quantity of non-tumoral cells to perform the activation experiment to simulate a wound. This will consist of activating certain wells with TGF-β and stopping the activation with PFA 4% at different time points. Controls must also be set.
Deliverable	Coverslips/wells with non-tumoral cells in different conditions.
Activities	- Cell plating of cells (non-tumoral) separating for the different conditions and control setting - Cell activation with TGF-β and control setting - Stop cell activation at certain time points by fixing them with PFA 4% and performance of three washes with PBS. - Storage of cells in the 4ºC refrigerator in PBS
Duration	5 days
Required resource	Cell flasks (non-tumoral) Well-plates/petri dish, coverslips Cell medium, trypsin, PBS, DMEM, ITS, P/S HCl, TGF-β
Risk	- Cell contamination or cell death. The solution to this is to repeat the experiment.

	- Not enough cells to culture all conditions. The solution to this is to wait until we have enough cells.
4.1.2. Images acquisition	
Description	To obtain data for our experiment, we must obtain image with an epifluorescence microscope. To do so, we first need to perform an immunostaining protocol on our fixed cells and then we will proceed by taking images with the epifluorescence microscope.
Deliverable	Epifluorescence images of the cells in different conditions.
Activities	- Performance of the immunostaining protocol of different conditions - Taking images with the epifluorescence microscope
Duration	2 weeks.
Required resource	Cell plates with the fixed cells. Immunostaining reagents Lab material Confocal/Epifluorescence microscope
Risk	- Bad state of the reagents may lead to the non-staining of the cellular components. We would need to start the experiments again. - No access to the epifluorescence microscope if it is too concurred or if it breaks for any reason. The solution would be to try to find another epifluorescence microscope as it is essential for the continuation of the project. CCiTUB has a microscopy service.
4.1.3. Features obtention	
Description	Running the images on a software to obtain the characteristics of the cytoskeleton from the images of the cells. This step requires the manual selection of the cell, and the program runs by itself to obtain the features.
Deliverable	Data set with the different cell features per cell.

Activities	- Learning how to use the software
	- Running the software on the images
	  o Image segmentation
	  o Features extraction

Duration	21 days
Required resource	Epifluorescence images taken previously of the cells Software: CSKmorphometrics Remote access to a computer with higher CPU will be granted
Risk	- This step is a bottle neck in our process as it takes a while to run the software on all images. - Too many cells per image would impede correct cell segmentation. We would have to repeat experiments with less cells. - Unfocused, saturated, bad quality images.
4.2. Activation under tumoral conditions	
4.2.1. Experimental phase	
Description	Cell plating the desired quantity of tumoral cells to perform the activation experiment to simulate a tumor. This will consist of activating certain wells with TGF-β and stopping the activation with PFA 4% at different time points. Controls must also be set.
Deliverable	Coverslips/wells with tumoral cells in different conditions.
Activities	- Cell plating of cells (tumoral) separating for the different conditions and control setting - Cell activation with TGF-β and control setting - Stop cell activation at certain time points by fixing them with PFA 4% and performance of three washes with PBS. - Storage of cells in the 4ºC refrigerator in PBS
Duration	5 days
Required resource	Cell flasks (tumoral) Well-plates/petri dish, coverslips Cell medium, trypsin, PBS, DMEM, ITS, P/S

	HCl, TGF-β
Risk	- Cell contamination or cell death. The solution to this is to repeat the experiment. - Not enough cells to culture all conditions. The solution to this is to wait until we have enough cells.
4.2.2. Images acquisition	
Description	To obtain data for our experiment, we must obtain image with an epifluorescence microscope. To do so, we first need to perform an immunostaining protocol on our fixed cells and then we will proceed by taking images with the epifluorescence microscope.
Deliverable	Epifluorescence images of the cells in different conditions.
Activities	- Performance of the immunostaining protocol of different conditions - Taking images with the epifluorescence microscope
Duration	2 weeks.
Required resource	Cell plates with the fixed cells. Immunostaining reagents Lab material Confocal/Epifluorescence microscope
Risk	- Bad state of the reagents may lead to the non-staining of the cellular components. We would need to start the experiments again. - No access to the epifluorescence microscope if it is too concurred or if it breaks for any reason. The solution would be to try to find another epifluorescence microscope as it is essential for the continuation of the project.
4.2.3. Features obtention	
Description	Running the images on a software to obtain the characteristics of the cytoskeleton from the images of the cells. This step requires the manual selection of the cell, and the program runs by itself to obtain the features.
Deliverable	Data set with the different cell features per cell.

Activities	- Learning how to use the software
	- Running the software on the images
	  ○ Image segmentation
	  ○ Features extraction

Duration	21 days
Required resource	Epifluorescence images taken previously of the cells. Software: CSKmorphometrics Remote access to a computer with higher CPU will be granted
Risk	- This step is a bottle neck in our process as it takes a while to run the software on all images. - Too many cells per image would impede correct cell segmentation. We would have to repeat experiments with less cells. - Unfocused, saturated, bad quality images.
4.3. Data analysis	
4.3.1. Comparison between activated tumoral fibroblasts and non-tumoral fibroblasts	
Description	Statistical analysis of the data from the basal and activated stages of tumoral and non-tumoral fibroblasts to evaluate the differences. Conventional data visualization in graphs and tables.
Deliverable	Graphs and visual representation of the data analysis
Activities	- Data exploration and manipulation - Data visualization - Statistical analysis of the features
Duration	2 weeks
Required resource	Obtained data set of cytoskeleton characteristics and activation level characteristics.
Risk	-
4.3.2. Morphological and cytoskeletal changes evolution during induced activation	

Description	Statistical analysis of the data from the different activation stages to see the evolution of the morphological characteristics and cytoskeletal characteristics during the activation process. Conventional data visualization in graphs and tables.
Deliverable	Graphs and visual representation of the data analysis
Activities	- Data exploration and manipulation - Data visualization - Statistical analysis of the features
Duration	2 weeks
Required resource	Obtained data set of cytoskeleton characteristics and activation level characteristics.
Risk	-
4.3.3.	Identification of CAF's subpopulations
Description	Application of Machine Learning techniques to the data set obtained to detect and try to identify CAF's subpopulations.
Deliverable	Machine learning model
Activities	- Feature reduction - Application of different Machine Learning algorithms
Duration	2 weeks
Required resource	Obtained data set of cytoskeleton characteristics
Risk	-

6.2. Phases and milestones

This project has been divided in five phases, each of which comprise different milestones, events that are crucial for the successful development of the project.

The first phase focuses on acquiring all the necessary materials, particularly obtaining, and preparing the cell lines for initiating work on the project. Within this phase, there are several milestones associated with acquiring the cells and material. The second phase involves conducting experiments on tumoral and non-tumoral fibroblast activation. The objective is to obtain various coverslips of cells where the activation process has been halted at different timepoints, along with corresponding control samples. In the third phase, the goal is to obtain epifluorescence images of the cells' cytoskeleton. This phase encompasses two milestone: performing the immunostaining protocol to obtained immunostained cells and acquiring the actual images. Proceeding to the fourth phase, the focus shifts towards extracting features from the obtained images. This process also consists of two milestones. Initially, the images undergo preprocessing, including segmentation of different cells within the images. Subsequently, a second stage generates a dataset containing the extracted features. Lastly, the fifth stage is the data analysis of the obtained dataset with features. In this phase, the goals we aimed to accomplish in this project will be reached. These include comparing the differences between activated tumoral and non-tumoral cells, the analysis of the changes in the organization of the cytoskeleton and in the morphology during the different stages of activation., and finally, the detection of CAF's subpopulations.

Figure 14 shows the explained phases with their associated milestones, which must be accomplished between the two main dates that define the project: the beginning of the project, which is the 13/03/2023; and the end of the project, which is the 07/07/2023.

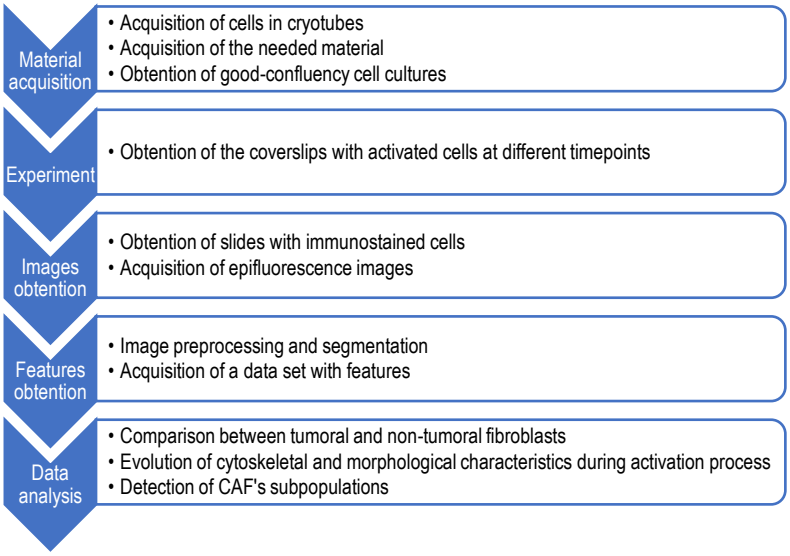


Figure 14. Phases and milestones of the project

6.3. Precedence Analysis and Critical Path Diagram (PERT-CPM)

From the WBS dictionary we can define and outline the specific tasks required to achieve our goals. Additionally, it provides valuable information regarding the estimated durations associated with each task. Furthermore, by analyzing the precedence relationships among these tasks, we can establish their natural sequence.

To visually represent this sequence and gain a better understanding of the project's flow, we can construct a PERT-CPM diagram. This diagram serves as a powerful tool that showcases the interdependencies among tasks. It allows us to determine early start and finish times as well as late start and finish times for each task. By identifying these time intervals, we can identify critical tasks, which are those that must be completed within a specific timeframe to prevent project delays. Conversely, tasks that have flexibility in their completion times are considered flexible tasks, from which we can calculate a clearance period. This period represents the amount of time that can be allocated to these tasks without impacting the overall project schedule.

Table 5. Definition of tasks and times, precedence analysis and extracted parameters from the PERT-CPM diagram.

	Tasks	Precedence	Time (days)	Early start	Early finish	Late start	Late finish	Margin
A	Idea presentation	-	1	0	1	0	1	0
B	Project redaction	A	30	1	31	55	85	54
C	Materials obtention	A	14	1	15	1	15	0
D	Tumoral cell line obtention	C	7	15	22	22	29	7
E	Non-tumoral cell line obtention	C	14	15	29	15	29	0
F	Tumoral replication experiment	D	7	22	29	29	36	7
G	Non-tumoral replication experiment	E	7	29	36	29	36	0
H	Non Tumoral activation experiment	H	7	36	43	36	43	0
I	Non Tumoral immunostaining	I	4	43	47	43	47	0
J	Non Tumoral image acquisition	J	3	47	50	47	50	0
K	Non Tumoral features obtention	K	21	50	71	50	71	0
L	Tumoral activation	G	7	29	36	36	43	7
M	Tumoral immunostaining	M	4	36	40	43	47	7

N	Tumoral image acquisition	N	3	40	43	47	50	7
O	Tumoral features obtention	O	21	43	64	50	71	7
P	Data analysis	L, P	14	71	85	71	85	0
Q	Machine Learning	L, P	14	71	85	71	85	0

Table 5 displays the precedence analysis as well as the extracted information from the PERT-CPM diagram, which can be observed in Figure 15. Critical Path Diagram of the project.**Figure 15**.

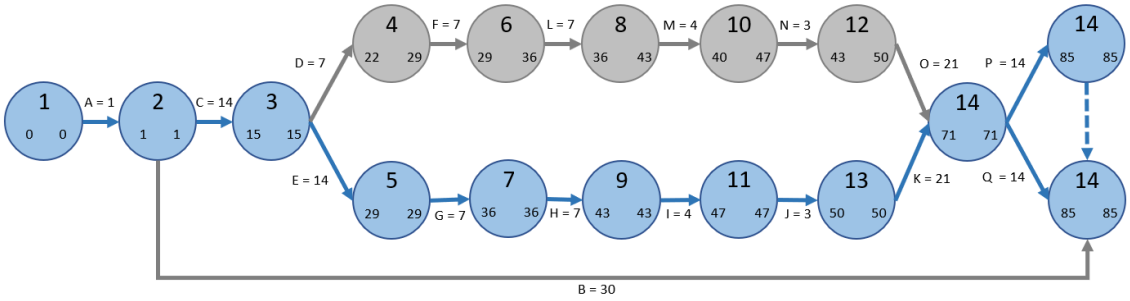


Figure 15. Critical Path Diagram of the project.

6.4. GANTT diagram

The different tasks that were established in the precedence analysis, which have also been represented in the PERT diagram, can also be effectively visualized using a GANTT diagram. The GANTT diagram provides a clear and intuitive representation of the temporal progression of the tasks, allowing us to easily identify the start and end dates of each task. This visual representation also highlights the dependencies and overlaps among different tasks. Additionally, the Gantt diagram enables us to visualize critical tasks, which are essential for meeting project deadlines, as well as flexible tasks that offer some leeway in terms of scheduling. **Figure 16** shows the GANTT diagram of the project.

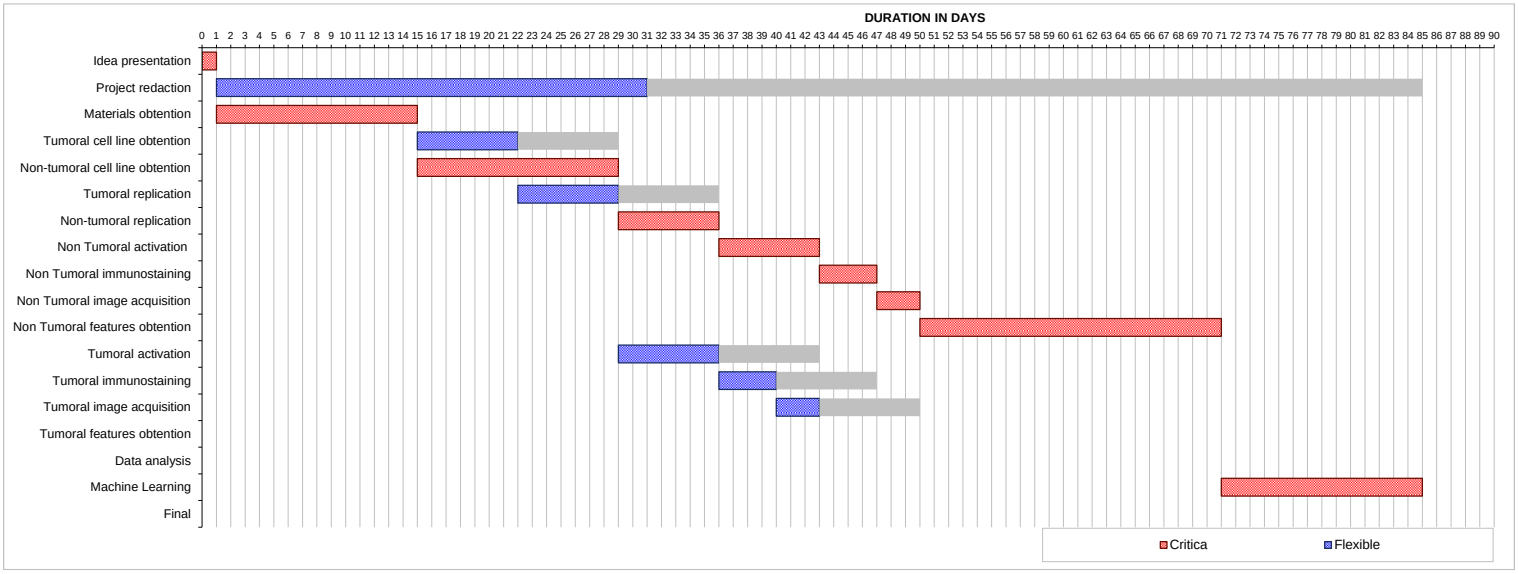


Figure 16. GANTT Diagram of the project.

7. Technical feasibility

Technical feasibility is an assessment of whether the proposed project, product or service can be successfully implemented using current or available technologies. Our project faces specific technical challenges, primarily related to the image acquisition using epifluorescence microscopy and the

extraction of features from said images. Cell culturing might also be considered a technical requirement, as sterile systems must be used to avoid cell contamination.

To address these challenges, this project has been carefully tailored to make use of the available technologies in the lab. For instance, the epifluorescence microscopy or the cell culturing hoods are vital equipment for our research that are already at our disposal in the lab. If any of this equipment became unavailable, we could always explore other options in the CCiTUB.

Regarding the extraction of parameters from images, we already have at our disposal a dedicated pipeline that has proven successful in working with our specific image data. This pipeline effectively solves the technical challenge associated with extracting relevant features.

Additionally, we have undergone rigorous training in both cell culturing techniques to ensure sterile conditions and using the confocal microscope to obtain correctly focused good-quality images. This ensures that we have the necessary expertise to handle these technical aspects of the project effectively. A background on data analysis and the application of machine learning techniques makes the obtention of results also achievable.

To assess the internal and external factors that can impact the project's success, a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis has been performed.

The strength of this project lies in its unique approach to studying fibroblast activation through the analysis of cytoskeleton reorganization and biophysical biomarkers. This approach offers valuable insights that cannot be captured solely through traditional biomarkers or RNA-sequencing. Additionally, the project adopts a multidisciplinary approach, integrating techniques from various fields and scientific disciplines, including cell biology or computer vision to allow morphological quantification of cellular structures from fluorescence microscopy images.

The fact that we can make use of existent equipment on the lab allows for a much smoother implementation of the project. As well, having already available image analysis tools supplies a degree of sophistication on data interpretation and extraction of meaning results. Having already worked with fibroblasts and being familiar with the protocols, as well as leveraging the expertise of researchers who are well-versed in research and cell-morphology analysis really helped overcome several challenges.

However, we have had limitations of sample size due to time constraints and the consideration of only one patient. Time constraints also determined the number of time-points we ended up analyzing. Furthermore, there are several steps on this project that require of random sampling of the cells. Although we tried being as indiscriminatory as possible, unconscious bias may appear when selecting which of the cells on the images to analyze.

As it has been mentioned, there is an existing gap in the literature regarding the study of activation stages of fibroblasts and cancerous wound repair situations, and even less by using morphological descriptors or cytoskeletal reorganization. Morphological descriptions of fibroblasts have often been limited to two

states, quiescent fibroblasts and myofibroblasts. Moreover, study of subpopulations is very heterogenous and non-specific, differing from study to study. Our project would provide a new classification, based instead on biophysical biomarkers.

The recent advances on current fluorescence microscopes and image preprocessing pipelines that yield single-cell multiparametric outputs of the cytoskeletal state in a high throughput manner make image-based quantification of cytoskeletal state the best positioned tool to yield biophysical biomarkers of cellular state.

However, it is omics data, who has experienced a growth in the most recent years. Today's studies put more interest in single cell RNA sequencing to discover fibroblasts population, and not so much on morphological and biophysical biomarkers data.

Table 6 gathers the Strengths, Weaknesses, Opportunities and Threats that have been previously mentioned in a summarized manner.

Table 6. SWOT analysis of the project

SWOT ANALYSIS	
Strengths	**Weaknesses**
- We present a novel approach to study fibroblasts activation. - We make use of multidisciplinary techniques. - We have available the equipment and required software in our unit - We have had aid from experienced researchers, as well as personal experience	- Time constraints reduces the sample size of our experiment. - Unconscious bias may happen when randomly sampling the cells to analyze
Opportunities	**Threats**
- There is a gap in the study of fibroblasts' activation via morphological descriptors and cytoskeletal reorganization. - Study of fibroblasts subpopulations is highly heterogeneous. - Image-based quantification of cytoskeletal state is the best tool to yield biophysical markers.	- There is a high interest in scRNA sequencing as the way to analyse subpopulations

8. Economic viability

Carrying out this project incurs certain economic costs for the Unit of Biophysics and Bioengineering. In this section, we will discuss the financial aspects of the project.

The costs of the materials used in the experiments have been divided into three categories. Firstly, various reagents are required, ranging from cell culture media to antibodies for the immunostaining protocol. Secondly, laboratory supplies and consumables are utilized. It is important to note that only the price of single-use materials has been considered, while the cost of reusable glassware and other common-use devices in the lab has been excluded.

Additionally, the expenses associated with the use of specific equipment have been taken into account. This includes the cost of utilizing culture hoods, incubators, and the epifluorescence microscope. An estimated cost for the use of this equipment has been calculated by referencing the hourly rate of utilizing similar equipment at CCiTUB.[65] This allows us to factor in the expenses associated with the wear and tear of the culture hoods, incubators, and the epifluorescence microscope, even though they are already available in the Unit. An approximate cost for buying the CFI Plan Fluor 20XC MI objective that we have used in the epifluorescence microscope.

The cost of working at the laboratory bench and of using the available devices in the laboratory, such as the centrifuge, has not been considered.

By categorizing the costs in this manner in **Table 7**, we can better understand the financial implications of the project and assess the overall expenditure involved.

Table 7. Study of the costs of the used materials

Item	Cost
Laboratory reagents	
DMEM, high glucose, pyruvate (500 mL)	25,34€
Fetal Bovine Serum (FBS) (500 mL)	408,00€
Penicillin-Streptomycin (P/S) (100mL)	18,65€
Trypsin-EDTA (0.05%), phenol red (100mL)	15,04€
Phosphate-Buffered Saline (PBS) (500mL)	43,99€
Trypan Blue (20mL)	15,00€
Human TGF-beta 1 Recombinant Protein	794,87€
Hydrochloric acid 2M (1L)	62,00€
Cell Freezing Medium-DMSO Serum free 1x (50mL)	207,00€
Paraformaldehyde (500G)	36,90€
Insulin-Transferrin-Selenium (ITS -G) (100X) (10 mL)	25,72€
Triton™ X-100 (250mL)	148,00€
Bovine Serum Albumin (BSA) (10G)	92,20€
Recombinant Anti-alpha smooth muscle Actin (acetyl E3) + ACTG2 (acetyl E3) antibody [E184] (ab32575)	569,48€
Anti-Vimentin antibody [RV202] – Cytoskeleton Marker (ab8978)	581,05€
Phalloidin-iFluor 555 Reagent (ab176756)	253,10€
Donkey Anti-Rabbit IgG H&L (Alexa Fluor® 488) preadsorbed (ab150061)	253,10€
Goat Anti-Mouse IgG H&L (Alexa Fluor® 488) (ab150113)	212,84€
NucBlue™ Live ReadyProbes™ Reagent (6 vials)	162,00€
Clear nail polish	0,98€
Leica™ Immersion Oil	111,00€
Fluoromount-G™ Mounting Medium (25mL)	81,75€
SUBTOTAL COST	**4118,01€**
Consumables and single-use laboratory supplies	
Serological pipettes 10 mL (200 units)	27,50€
Serological pipettes 50 mL (100 units)	172,00,
Serological pipettes 5 mL (200 units)	77,20€
Serological pipettes 2 mL (100 units)	241,00€

100-1000 µL pipette tips (1000 units)	12,25€
5-200 µL pipette tips (1000 units)	7,95€
0,1-10 µL pipette tips (1000 units)	16,95€
40mm Petri Dishes (740 units)	155,00€
12 mm Microscope Glass Coverslip, circular (1000 units)	117,58€
Glass Microscope Slide, frosted, beveled edges, clipped corners, 75 x 25mm (144 units)	28,00€
T75 cell culture flasks with filter (100 units)	165,14€
Fisherbrand™ Comfort Nitrile Gloves (200 units)	29,67€
Kimberly-Clark Professional™ Kimtech Science™ Precision Wipes™ Tissue Wipers	28,85€
14-well well plate (75 units)	191,00€
Centrifuge tubs 15mL (50 units)	28,40€
Centrifuge tubes 50 mL (100 units)	65,50€
Eppendorf 1.5 mL (1000 units)	14,35€
Eppendorf 2 mL (500 units)	9,50€
SUBTOTAL COST	**1188,34**
Equipment and instruments	
Use of culture hoods and incubators	2,60€/hour x 29h = 75,40€
Cell maintenance	0,57€/hour x 672 h = 383,04€
Use and capture of images in the epifluorescence microscope, autoservice from 9:00 to 18:00	2,47€/hour x 40h = 98,80€
CFI Plan Fluor 20XC MI objective	850€
SUBTOTAL COST	**1407,24€**
FINAL COST	**6713,59€**

Based on the information provided, a final cost of €6713,59 has been calculated. This cost appears to be high due to the consideration of acquisition of materials in larger quantities than might be necessary. It is important to remark once again that most of this material and equipment is already available for common use in the laboratory and it is not a specific cost of this project, although we have considered it as so.

In addition to the final cost, it is important to consider expenses related to infrastructure and utilities, including electricity, lighting, and water.

The budget for material is of about 10.000€ per research per year for the average, so taking into account the different consideration that have been mentioned, the project would be within budget.

Personnel costs have not been taken into account. If we estimate the job equivalent to an entry-level research assistant position, which typically has an average salary of approximately €18 per hour, the personnel costs would amount to €5,400 for 300 hours.

Therefore, considering these factors, the total cost, including both materials and personnel expenses, would be €12,113.59.

9. Regulations and legal aspects

Research projects have ethical implications that need to be assessed by specific committees. These committees ensure that the experimental design complies with current standards, collaborate with researchers in analyzing the ethical implications of projects, and issue reports for required calls. The University of Barcelona considers it highly important that all members of the university community are aware of the ethical implications underlying research and hence it has several ethics committees, such as the "Comissió de Bioètica" or the "Comitè Ètic d'Investigació Clínica de l'Hospital Clínic de Barcelona", amongst others.

Specifically, the University of Barcelona's Bioethics Commission (CBUB) has the aim to evaluate research projects, engage in reflection and debate on ethical issues raised by scientific research, and promote bioethics training and research integrity among doctoral researchers and trainee investigators. The CBUB advises and evaluates projects to ensure compliance with ethical standards outlined in documents such as the Nuremberg Code, the Declaration of Geneva, the Helsinki Declaration, the Belmont Report, and the Universal Declaration on Bioethics and Human Rights by UNESCO, among others.

Obtaining a favorable opinion from the CBUB is a legally established requirement (Law 14/2007) to initiate any research involving human subjects, human biological samples, or personal data. These requirements include the principles of beneficence, respect for autonomy in obtaining consent to participate in research, justice, the primacy of the individual's well-being over other interests, and the protection of non-autonomous individuals and vulnerable populations.

The CBUB has available in its web page a series of documents regarding bioethics in research that must be followed by all biomedical research performed in the University of Barcelona. In this section we will detail some of it.

The University of Barcelona's Code of Conduct for Research Integrity[66] aims to provide a positive definition of scientific integrity, which is defined as "honesty in the commitment to truth; independence in the preservation of freedom of action in relation to pressures outside the profession; and impartiality in the neutrality of professional practice in relation to private interests outside the research".[67] The code is based on the six principles of research integrity: honesty, responsibility and accountability, reliability, rigor, respect, and independence; to improve the quality, impact and results of research in all fields.

Furthermore, research with people, human biological material and personal data must respect internationally recognized human rights and the principles of autonomy, beneficence, non-maleficence, and justice. At the same time, the relevant information and informed consent processes required from an ethical perspective must be applied: research staff must request and obtain the explicit and informed consent of those people who have provided biological samples for our project, as well as maintain anonymity of the participants.

Taking into account that we are dealing with patients who have voluntarily given part of their biopsy for research (and not with voluntaries), our data was obtained with the informed consent of the patients, following the protocols of the ethical committees of *Hospital Clínic* and *Universitat de Barcelona*.

Regarding the ethics involving data protection, a document was drafted by a panel of experts at the requires of the European Commission (DG Research and Innovation) to raise awareness about data protection.[68] Research projects must consider personal data and sensitive data when recording, storing, and analyzing patient's data. This way data of the patient's was pseudonymized and anonymized to protect the data subject's privacy and minimize the risk of unauthorized access.

Data processing must be lawful, fair, and transparent. It should involve only data that are necessary and proportionate to achieve the specific task or purpose for which they were collected. Only data that is needed to achieve our research objectives should be collected, otherwise it is unethical and unlawful.

10. Conclusions and future lines

In this project, we have successfully activated tumoral and non-tumoral fibroblasts and observed differences in the activation process between NAFs and CAFs. Activation led to significant changes in the cell cytoskeleton and morphology of NAFs, while CAFs exhibited less pronounced changes in their cytoskeletal fibers across different activation stages. Additionally, we have concluded both activated NAFs and activated CAFs displayed similar characteristics.

Clustering analysis has provided preliminary insights into CAFs subpopulations in both the activated and non-activated stages, that seems to indicate that the number of subpopulations upon persistent activation converge. Furthermore, some non-activated CAFs already exhibited characteristics surpassing those of activated fibroblasts. Substrate stiffness of glass coverslips may play an important role in provoking fibroblast activation of non-activated fibroblasts. That is why we propose conducting a similar analysis but seeding cells in substrates with tunable stiffness instead of glass, such as hydrogels, to mimic the actual extracellular matrix and, most importantly, tumoral microenvironment.

To create a more realistic cellular environment, 3D cell culturing using a scaffold would be an even better resource. This approach allows for more accurate cell-cell and cell-matrix interactions, facilitates the presence of nutrient and oxygen gradients, and improves cell differentiation. Bioreactors and microfluidic systems could also mimic the physical conditions of the cellular environment. Incorporating AFM could also provide additional biophysical biomarkers on a single-cell basis, which, when combined with the obtained parameters, could elucidate changes in cell plasticity or stiffness.

Moving beyond biophysical biomarkers, it would be interesting to study the epigenetic characteristics that define the different groups of cells. Understanding how these cytoskeletal changes translate into functional differences would provide valuable insights.

Furthermore, exploring the deactivation process is crucial, as the reversibility of activation distinguishes myofibroblasts from CAFs. By removing the stimulus that persistently activates tumoral and non-tumoral fibroblasts, we can study the deactivation process using the same parameters analyzed in this project.

In conclusion, this project has shed light on the activation process of fibroblasts and revealed distinctions between NAFs and CAFs. By addressing the mentioned limitations and pursuing future research directions, we can gain a deeper understanding of the role of fibroblasts in cancer and explore potential therapeutic targets and strategies.

11. Bibliography

1. Kalluri R. The biology and function of fibroblasts in cancer. *Nature Reviews Cancer 2016 16:9*. 2016;16(9):582-598. doi:10.1038/nrc.2016.73

2. White ES. Lung extracellular matrix and fibroblast function. *Ann Am Thorac Soc*. 2015;12 Suppl 1(Suppl 1):S30-S33. doi:10.1513/ANNALSATS.201406-240MG

3. Plikus M V., Wang X, Sinha S, et al. Fibroblasts: Origins, definitions, and functions in health and disease. *Cell*. 2021;184(15):3852-3872. doi:10.1016/j.cell.2021.06.024

4. Talbott HE, Mascharak S, Griffin M, Wan DC, Longaker MT. Wound healing, fibroblast heterogeneity, and fibrosis. *Cell Stem Cell*. 2022;29(8):1161-1180. doi:10.1016/j.stem.2022.07.006

5. Cialdai F, Risaliti C, Monici M. Role of fibroblasts in wound healing and tissue remodeling on Earth and in space. *Front Bioeng Biotechnol*. 2022;10:1763. doi:10.3389/FBIOE.2022.958381/BIBTEX

6. Mhaidly R, Mechta-Grigoriou F. Fibroblast heterogeneity in tumor micro-environment: Role in immunosuppression and new therapies. *Semin Immunol*. 2020;48. doi:10.1016/j.smim.2020.101417

7. Lavie D, Ben-Shmuel A, Erez N, Scherz-Shouval R. Cancer-associated fibroblasts in the single-cell era. *Nat Cancer*. 2022;3(7):793-807. doi:10.1038/s43018-022-00411-z

8. McKayed KK, Simpson JC. Actin in Action: Imaging Approaches to Study Cytoskeleton Structure and Function. *Cells*. 2013;2(4):715. doi:10.3390/CELLS2040715

9. Xing F, Saidou J, Watabe K. Cancer associated fibroblasts (CAFs) in tumor microenvironment. *Front Biosci*. 2010;15(1):166. doi:10.2741/3613

10. Caligiuri G, Tuveson DA. Activated fibroblasts in cancer: Perspectives and challenges. *Cancer Cell*. 2023;41:434-449. doi:10.1016/j.ccell.2023.02.015

11. Ravikanth M, Soujanya P, Manjunath K, Saraswathi TR, Ramachandran CR. Heterogenecity of fibroblasts. *Journal of Oral and Maxillofacial Pathology*. 2011;15(2):247-250. doi:10.4103/0973-029X.84516

12. Irvine AF, Waise S, Green EW, Stuart B, Thomas GJ. Characterising cancer-associated fibroblast heterogeneity in non-small cell lung cancer: a systematic review and meta-analysis. *Scientific Reports 2021 11:1*. 2021;11(1):1-15. doi:10.1038/s41598-021-81796-2

13. Lebleu VS, Kalluri R. A peek into cancer-associated fibroblasts: origins, functions and translational impact. Published online 2018. doi:10.1242/dmm.029447

14. Phillip JM, Wu PH, Gilkes DM, et al. Biophysical and biomolecular determination of cellular age in humans. 2017;1(7). doi:10.1038/s41551-017-0093

15. Bartolozzi A, Viti F, De Stefano S, et al. Development of label-free biophysical markers in osteogenic maturation. Published online 2019. doi:10.1016/j.jmbbm.2019.103581

16. Liu Y, Mollaeian K, Ren J. An Image Recognition-Based Approach to Actin Cytoskeleton Quantification. *Electronics 2018, Vol 7, Page 443*. 2018;7(12):443. doi:10.3390/ELECTRONICS7120443

17. Flores LR, Keeling MC, Zhang X, Sliogeryte K, Gavara N. *Lifeact-TagGFP2 Alters F-Actin Organization, Cellular Morphology and Biophysical Behaviour*.

18. Gavara N, Chadwick RS. Relationship between cell stiffness and stress fiber amount, assessed by simultaneous atomic force microscopy and live-cell fluorescence imaging. *Biomech Model Mechanobiol*. 2016;15(3):511-523. doi:10.1007/S10237-015-0706-9

19. *MEMORIA CIENTÍFICO-TÉCNICA DE PROYECTOS INDIVIDUALES: Single-Cell Biophysical Biomarkers Coupled to Machine Learning for a Proof-of-Concept Drug Screen to Decrease the Activation State of Tumour-Associated Fibroblasts in Lung Cancer (BB-MLRforTAFS)*.

20. Alcaraz J, Ikemori R, Llorente A, Díaz-valdivia N, Reguart N, Vizoso M. Epigenetic Reprogramming of Tumor-Associated Fibroblasts in Lung Cancer: Therapeutic Opportunities. *Cancers (Basel)*. 2021;13(15). doi:10.3390/CANCERS13153782

21. Albrengues J, Bertero T, Grasset E, et al. Epigenetic switch drives the conversion of fibroblasts into proinvasive cancer-associated fibroblasts. *Nature Communications 2015 6:1*. 2015;6(1):1-15. doi:10.1038/ncomms10204

22. Yang P, Luo Q, Wang X, et al. Comprehensive Analysis of Fibroblast Activation Protein Expression in Interstitial Lung Diseases. *Am J Respir Crit Care Med*. 2023;207(2):160-172. doi:10.1164/rccm.202110-2414OC

23. Mahmoudi S, Mancini E, Xu L, et al. Heterogeneity in old fibroblasts is linked to variability in reprogramming and wound healing. *Nature*. 2019;574(7779):553. doi:10.1038/S41586-019-1658-5

24. Zeisberg EM, Zeisberg M. The role of promoter hypermethylation in fibroblast activation and fibrogenesis. *Journal of Pathology*. 2013;229(2):264-273. doi:10.1002/path.4120

25. Lambrechts D, Wauters E, Boeckx B, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*. 2018;24(8):1277-1289. doi:10.1038/s41591-018-0096-5

26. Kim N, Kim HK, Lee K, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-16164-1

27. Adams TS, Schupp JC, Poli S, et al. *Single-Cell RNA-Seq Reveals Ectopic and Aberrant Lung-Resident Cell Populations in Idiopathic Pulmonary Fibrosis*. Vol 6.; 2020. www.ipfcellatlas.com

28. Kieffer Y, Hocine HR, Gentric G, et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov*. 2020;10(9):1330-1351. doi:10.1158/2159-8290.CD-19-1384

29. Fernandes IR, Russo FB, Pignatari GC, et al. Fibroblast sources: Where can we get them? *Cytotechnology*. 2016;68(2):223. doi:10.1007/S10616-014-9771-7

30. Duch P, Díaz-Valdivia N, Ikemori R, et al. Aberrant TIMP-1 overexpression in tumor-associated fibroblasts drives tumor progression through CD63 in lung adenocarcinoma. *Matrix Biology*. 2022;111:207-225. doi:10.1016/j.matbio.2022.06.009

31. Farsani TM, Motevaseli E, Neyazi N, Khorramizadeh MR, Zafarvahedian E, Ghahremani MH. Effect of Passage Number and Culture Time on the Expression and Activity of Insulin-Degrading Enzyme in Caco-2 Cells. *Iranian Biomedial Journal*. 2018;22(1):70-75.

32. Strutz F, Zeisberg M, Renziehausen A, et al. TGF-β1 induces proliferation in human renal fibroblasts via induction of basic fibroblast growth factor (FGF-2). *Kidney Int*. 2001;59(2):579-592. doi:10.1046/j.1523-1755.2001.059002579.x

33. Midgley AC, Rogers M, Hallett MB, et al. Transforming Growth Factor-β1 (TGF-β1)-stimulated Fibroblast to Myofibroblast Differentiation Is Mediated by Hyaluronan (HA)-facilitated Epidermal Growth Factor Receptor (EGFR) and CD44 Co-localization in Lipid Rafts. *J Biol Chem*. 2013;288(21):14824. doi:10.1074/JBC.M113.451336

34. Ma Y, Iyer RP, Jung M, Czubryt MP, Lindsey ML. Cardiac Fibroblast Activation Post-Myocardial Infarction: Current Knowledge Gaps. *Trends Pharmacol Sci*. 2017;38(5):448. doi:10.1016/J.TIPS.2017.03.001

35. Zeisberg M, Strutz F, Müller GA. Role of fibroblast activation in inducing interstitial fibrosis. *J Nephrol*. 2000;13 Suppl 3(3 SUPPL.):S111-20. Accessed May 17, 2023. https://europepmc.org/article/med/11132027

36. Barnes JL, Gorin Y. Myofibroblast Differentiation During Fibrosis: Role of NAD(P)H Oxidases. *Kidney Int*. 2011;79(9):944. doi:10.1038/KI.2010.516

37. Juhl P, Bondesen S, Hawkins CL, et al. Dermal fibroblasts have different extracellular matrix profiles induced by TGF-β, PDGF and IL-6 in a model for skin fibrosis. *Scientific Reports 2020 10:1*. 2020;10(1):1-10. doi:10.1038/s41598-020-74179-6

38. Yang J, Velikoff M, Canalis E, Horowitz JC, Kim KK, Acti KK. Activated alveolar epithelial cells initiate fibrosis through autocrine and paracrine secretion of connective tissue growth factor. *Am J Physiol Lung Cell Mol Physiol*. 2014;306:786-796. doi:10.1152/ajplung.00243.2013.-Fibrogenesis

39. Takamura N, Renaud L, da Silveira WA, Feghali-Bostwick C. PDGF Promotes Dermal Fibroblast Activation via a Novel Mechanism Mediated by Signaling Through MCHR1. *Front Immunol*. 2021;12. doi:10.3389/FIMMU.2021.745308/FULL

40. Chen K, Chen J, Li D, Zhang X, Mehta JL. Angiotensin II Regulation of Collagen Type I Expression in Cardiac Fibroblasts Modulation by PPAR-Ligand Pioglitazone. Published online 2004. doi:10.1161/01.HYP.0000144400.49062.6b

41. Villee CA. Morphology - Dissection, Histochemistry, and Fluorescence Microscopy | Britannica. Accessed May 18, 2023. https://www.britannica.com/science/morphology-biology/Methods-in-morphology

42. Berdyyeva T, Woodworth CD, Sokolov I, Sokolov I. Visualization of Cytoskeletal Elements by the Atomic Force Microscope.

43. Webb DJ, Brown CM. Epi-Fluorescence Microscopy. *Methods Mol Biol*. 2013;931:29. doi:10.1007/978-1-62703-056-4_2

44. Cooper GM. *The Cell: A Molecular Approach*. Sinauer Associates; 2000. Accessed May 18, 2023. https://www.ncbi.nlm.nih.gov/books/NBK9908/

45. Ostrowska-Podhorodecka Z, Ding I, Norouzi M, McCulloch CA. Impact of Vimentin on Regulation of Cell Signaling and Matrix Remodeling. *Front Cell Dev Biol*. 2022;10:562. doi:10.3389/FCELL.2022.869069/BIBTEX

46. Kapuscinski J. DAPI: a DNA-specific fluorescent probe. *Biotech Histochem*. 1995;70(5):220-233. doi:10.3109/10520299509108199

47. Phalloidin Conjugates for Actin Staining - ES. Accessed June 5, 2023. https://www.thermofisher.com/es/es/home/life-science/cell-analysis/cell-structure/cytoskeleton/phalloidin-and-phalloidin-conjugates-for-staining-actin.html

48. ImageJ. Accessed June 5, 2023. https://imagej.nih.gov/ij/index.html

49. CellProfiler. Accessed June 5, 2023. https://cellprofiler.org/

50. MATLAB - MathWorks. Accessed June 5, 2023. https://ch.mathworks.com/products/matlab.html

51. Python. Accessed June 5, 2023. https://www.python.org/

52. Zhang X, Flores LR, Keeling MC, Sliogeryte K, Gavara N. Ezrin phosphorylation at T567 modulates cell migration, mechanical properties, and cytoskeletal organization. *Int J Mol Sci*. 2020;21(2). doi:10.3390/ijms21020435

53. Meng H, Chowdhury TT, Gavara N. The Mechanical Interplay Between Differentiating Mesenchymal Stem Cells and Gelatin-Based Substrates Measured by Atomic Force Microscopy. *Front Cell Dev Biol*. 2021;9. doi:10.3389/FCELL.2021.697525/FULL

54. R: The R Project for Statistical Computing. Accessed June 5, 2023. https://www.r-project.org/

55. Home - GraphPad. Accessed June 5, 2023. https://www.graphpad.com/

56. sklearn.linear_model.LogisticRegression — scikit-learn 1.2.2 documentation. Accessed June 5, 2023. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

57. Charrad M, Ghazzali N, Boiteau V, Maintainer AN. Package "NbClust." Published online 2022.

58. Sandbo N, Dulin N. The actin cytoskeleton in myofibroblast differentiation: Ultrastructure defining form and driving function. *Translational Research*. 2011;158(4):181. doi:10.1016/J.TRSL.2011.05.004

59. Pham TX, Lee J, Guan J, et al. Transcriptional analysis of lung fibroblasts identifies PIM1 signaling as a driver of aging-associated persistent fibrosis. *JCI Insight*. 2022;7(6). doi:10.1172/JCI.INSIGHT.153672

60. Shani O, Raz Y, Monteran L, et al. Evolution of fibroblasts in the lung metastatic microenvironment is driven by stage-specific transcriptional plasticity. *Elife*. 2021;10. doi:10.7554/ELIFE.60745

61. Dey S, Dutta A, Toledo JI, Ghosh SK, Llados J, Pal U. SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification. Published online July 7, 2017. Accessed June 7, 2023. https://arxiv.org/abs/1707.02131v2

62. Schroff F, Kalenichenko D, Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2015;07-12-June-2015:815-823. doi:10.1109/CVPR.2015.7298682

63. Mhaidly R, Mechta-Grigoriou F. Role of cancer-associated fibroblast subpopulations in immune infiltration, as a new means of treatment in cancer. *Immunol Rev*. 2021;302(1):259-272. doi:10.1111/IMR.12978

64. Nath Kundu A, Dougan CE, Mahmoud S, et al. Tenascin-C activation of lung fibroblasts in a 3D synthetic lung extracellular matrix mimic. doi:10.1101/2023.02.24.529926

65. Tarifes. *Consell Social*. Published online 2022.

66. *Codi d'Integritat En La Recerca de La Universitat de Barcelona*.; 2020. Accessed May 22, 2023. http://hdl.handle.net/2445/166917

67. Casado M, Patrão Neves M do C, Leucona Ramírez I de, Carvalho AS, Araújo J. Declaració sobre integritat científica en recerca i innovació responsable. Published 2016. Accessed May 22, 2023. http://hdl.handle.net/2445/103268

68. Hayes B, Kuyumdzhieva A. *Ethics and Data Protection*.; 2021.

69. Stages of Cancer | Cancer.Net. Accessed May 16, 2023. https://www.cancer.net/navigating-cancer-care/diagnosing-cancer/stages-cancer

70. Zhang X, Flores LR, Keeling MC, Sliogeryte K, Gavara N. *Ezrin Phosphorylation at T567 Modulates Cell Migration, Mechanical Properties and Cytoskeletal Organization*. Vol 20.; 2019. www.mdpi.com/journal/ijms

12. Annexes

12.1. Supplementary information regarding the obtained cells from Duch et al. (2022)

Duch et al.[30] study used different immortalized cell lines from healthy lung tissue (which was used as a control FC) and tumoral lung tissue from the same patient. The cells were obtained from a total of twenty patients, who gave their informed consent on the procedure. When working with human samples, it is crucial to ensure compliance with ethical guidelines and obtain the necessary approvals from relevant ethics committees or institutional review boards (IRBs). The "Comitè d'Ètica de l'*Hospital Clínic de Barcelona*" and "Comitè d'Ètica de l'*Universitat de Barcelona" 's* (ethical committees from Hospital Clínic de Barcelona and Universitat de Barcelona) approved protocols were followed. The clinical characteristics of said patients can be seen in **table A1**: the patient reference identification, age, histological subtype of the tumor (SCC, squamous cell carcinoma or ADC, adenocarcinoma), TNM staging to describe the tumor (pT T1-T4 describes the size and location of the tumor on a scale of 1 to 4 with a lowercase letter; and pN describes the affectation of the cancer on the lymph nodes with a number from 0 to 3[69]) and the cancer staging from I to IV, depending on the spread of the tumor.

Table A1. Summary of patient's clinical from Duch et al.[30]

Patient ref.	Age (y.o.)	Histologic subtype	pT	pN	Stage
#2	82	SCC	T3	N2	IIIA
#4	73	SCC	T2b	N1*[1]	IIB
#5	69	SCC	T2b	N2	IIIA
#6	65	SCC	T1a	N0	IA
#7	61	ADC	T1b	N0	IA
#10	76	ADC	T1a	N0	IA
#12	70	ADC	T3	N0	IIB
#13	59	ADC	T2b	N2	IIIA
#15	73	ADC	T1a	N0	IA
#16	64	SCC	T2a	N1*[1]	-
#18	83	SCC	T2a	N0	IB
#20	78	SCC	T1b	N0	IA
#22	76	SCC	T2a	N1	IIA
#27	71	ADC	T1a	N0	IA
#28	80	ADC	T1a	N0	IA
#31	62	SCC	T3	N0	IIB
#33	72	SCC	T2a	N0	IB
#34	76	SCC	T2b	N0	IIA
#35	60	SCC	T2a	N1	IIA
#37	58	ADC	T1b	N0	IA

[1] N*: direct hilar ganglionar infiltration by the tumor

12.2. Procedure for single cell analysis of immunofluorescence images

In this section the pipeline for analyzing single cells from immunofluorescence images is described as it was explained in Zhang et al. (2019)[70]. It should be noted that some modifications may have been made to the pipeline since then.

To analyze single cells from immunofluorescence images, the initial step involves selecting and cropping individual cells within the images. Because images were taken with a 20x objective and cell sizes are small, there are normally many cells in one single image which must be selected. Preferably cells that are isolated from others in the image will be selected with a rectangular area. Once all desired cells are cropped, the cell outline are manually drawn along the cell's perimeter. Then the cell mask is generated automatically by detecting the difference between signals inside and outside the cells' outline.

Following the identification of the cells' outline, the next step is to perform an initial segmentation of the fibers. This is achieved by convolving the cell image with elongated Laplace of Gaussian (eLoG) kernels. At each pixel location within the cell, a small window with a size of 21x21 pixel acts as a fiber template filter. The window is rotated 180º with a step size of 6º gradually. By maximizing the image cross-correlation signals, a map of putative fibers is generated.

This map of putative fiber undergoes fiber refinement, which is carried out by using a coherence enhancing diffusion filter. This process involves extending and interconnecting interrupted fibers. Then, the values of newly inserted pixels are examined by comparing them with the average orientation of neighboring pixels within the same fiber, using a 9x9 pixel window. If the difference is larger than pre-defined threshold, the pixel is discarded. Additionally, this step corrects artifacts such as bright dots. These fiber-enhancement and trimming steps are iterated until the algorithm converges.

At last, to eliminate fluorescent signals of non-specific binding that do not belong to the fibers, a background subtraction process is employed. A background fluorescence map is generated by computing the median signal intensity within a 21 x 21 window surrounding each non-fiber pixel near fiber edges. This results in a smoothly changing intensity map, with fiber pixels replaced by the median of non-fiber pixels. This background map is then subtracted from the original image, and the pixels that obtained negative values that belong to the fiber map are removed. This process is iterated until convergence is achieved, ensuring that only pixels truly belonging to a fiber are included in the final fiber map.

In the final step of the pipeline, 19 parameters which characterize the cytoskeleton and nuclear properties are computed. Compiling the results, a total of 21 features are obtained. **Table A2** shows the aforementioned features.

Table A2. Features extracted from the images.

Compiled_results.txt		
1. Area	Sum of pixels contained within the mask of cell shape	Pixels^2
2. Aspect ratio	Ratio between major and minor axis of cell shape.	1-∞

3. F-actin	Level of fluorescence of what is in the shape of a filament. Sum of pixel intensities for all pixels identified as belonging to a fibre.	Arbitrary units
4. Fibre thickness	Average of pixel intensities for all pixels identified as belonging to a fibre. Given that in our imaging conditions that pixel size is larger than the diffraction limit and the thickness of an actin filament, pixel intensity constitutes a good surrogate measure to estimate the number of individual fluorophores bound to a filament and the number of filaments making up a stress fibre.	Arbitrary units
5. CoV fibre thickness	Variability in the distribution of all pixel intensities for all pixels identified as belonging to a fibre.	
6. Fibre alignment	Using the previously computed orientations for all pixels identified as belonging to a fibre, alignment is defined as 1-circular variance, computed using directional statistics of the distribution of angles. Angle variance. Values close to 0 indicate that the majority of fibres are oriented in the same direction, whereas values close to 1 indicate random orientation of fibres	0-1
7. Curvature of fibres	Similar to alignment but computed as circular variances for all pixels within a single fibre. Values close to 0 indicate that the fibres are curvier, whereas the closer the values to 1 the more straight the fibres are.	0-1
8. Peak fibres	Preferent radial location of the fibres. A value of 1 indicates the cell edge (closer to the cell periphery), whereas a value of 0 indicates the cell's centroid (closer to the centre of the cell)	0-1
9. Spread fibres	Variance associated with location of fibres in the radial position A lower value (closer to 0) means that fibres are preferentially concentrated at a certain point of the cell (at a single radial position), whereas values closer to 1 indicates that cells are more dispersed through the cell diameter.	0-1
10. Convexity/concavity of cell outline	Measurement of the convexity and concavity of cell area.	
11. Fibre length	Average length of fibres in a cell.	Pixels
12. CoV fibre length	Variance associated with the length of the fibres in a cell.	
13. Chirality	Once the centroid of the cells is identified, the fiber orientation map is converted to compute the relative orientation of each fiber with respect to the cells centroid. Chirality indicates if the fibers point towards the center of the cell or tangentially. A value close to 0º indicates that the fibers are preferentially pointing in the radial direction (towards the center of the cell) whereas a value closer to 90º indicates that fibers are preferentially pointing in the circumferential direction (in parallel to the cell edge).	0º-90º
14. Variance of chirality	Variance associated with chirality measurements.	
15. Nuclear brightness	Normalized value, does not provide any information	
16. Chromatin condensation	Level of chromatin condensation calculated through DAPI intensity.	AU
17. Nuclear volume	Extrapolation of the relative volume of the nucleus if cells were found in a 3D space.	No units.
18. Nuclear Poisson's ratio	Behaviour of the nucleus under compression in one direction. Nuclear volume is not constant, so the expected value is negative.	0.5 to -∞
19. Nuclear stiffness	Indirect measurement of the nucleus stiffness derived from the obtained number of fibres and nuclear deformation.	AU
20. Nuclear centrality	Indicates whether the nucleus is near the centre of the cell or far from it.	0-1
21. Nucleus/fibre alignment	Indicates whether nucleus elongation is aligned with cell direction	0-1
all_res_csk.txt or all_res_vim.txt		
1. Area (pixels2)		
2. Aspect ratio		

3. Total fluorescence
4. F-actin
5. Fibre thickness
6. CoV fibre thickness
7. Chirality
8. Fibre alignment
9. Preference direction of fibres
10. Curvature
11. Variance of chirality
12. Peak fibres (0-1)
13. Spread fibres
14. Convexity/concavity of cell outline
15. Fibre length
16. CoV fibre length
18. Centre of mass of cells in the X direction
19. Centre of mass of cells in the Y direction

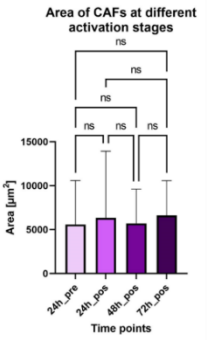
Additionally, a parameter indicating the level of activation of the fibroblasts was obtained thorough the information in the green channel of αSMA, as indicated in

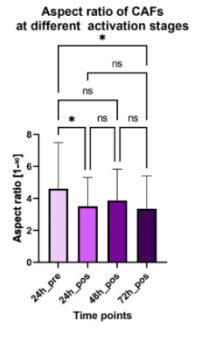
$$\alpha SMA = col(3) - 0.1569 * col(1) - 26983 \qquad Eq.2$$

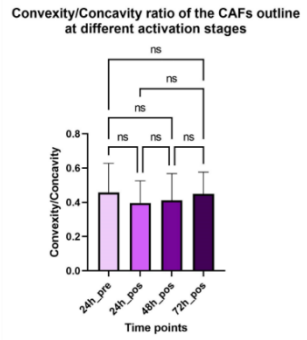
where col(x) indicated the column in the all_res_vim.txt file.

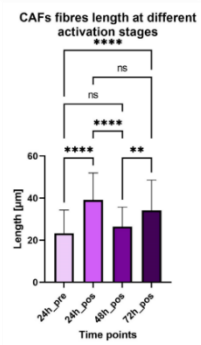
12.3. All figures for the obtained compiled results of tumoral and non-tumoral fibroblasts

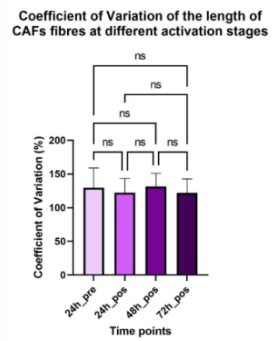
This annex shows all the statistical analysis done for each previously specified parameter (from gross morphology, individual fibers, organization of fibers and nucleus) as graphs, for both CAFs and NF.

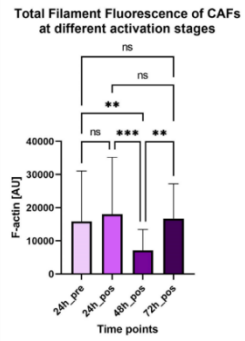


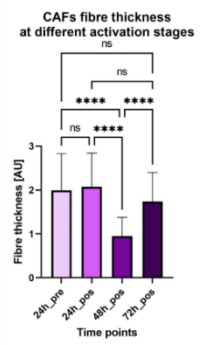


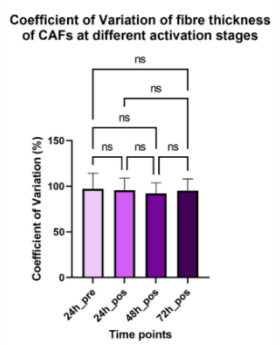


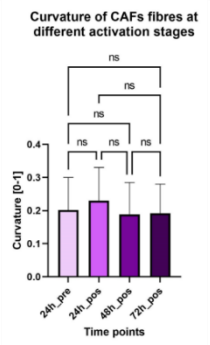


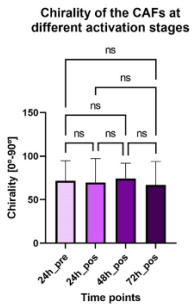


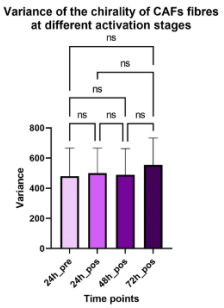


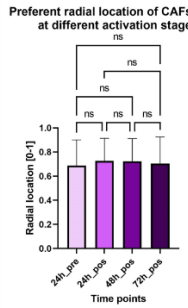


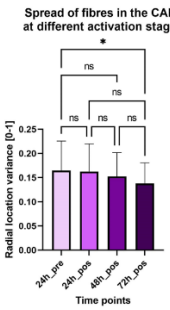


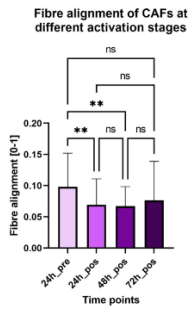












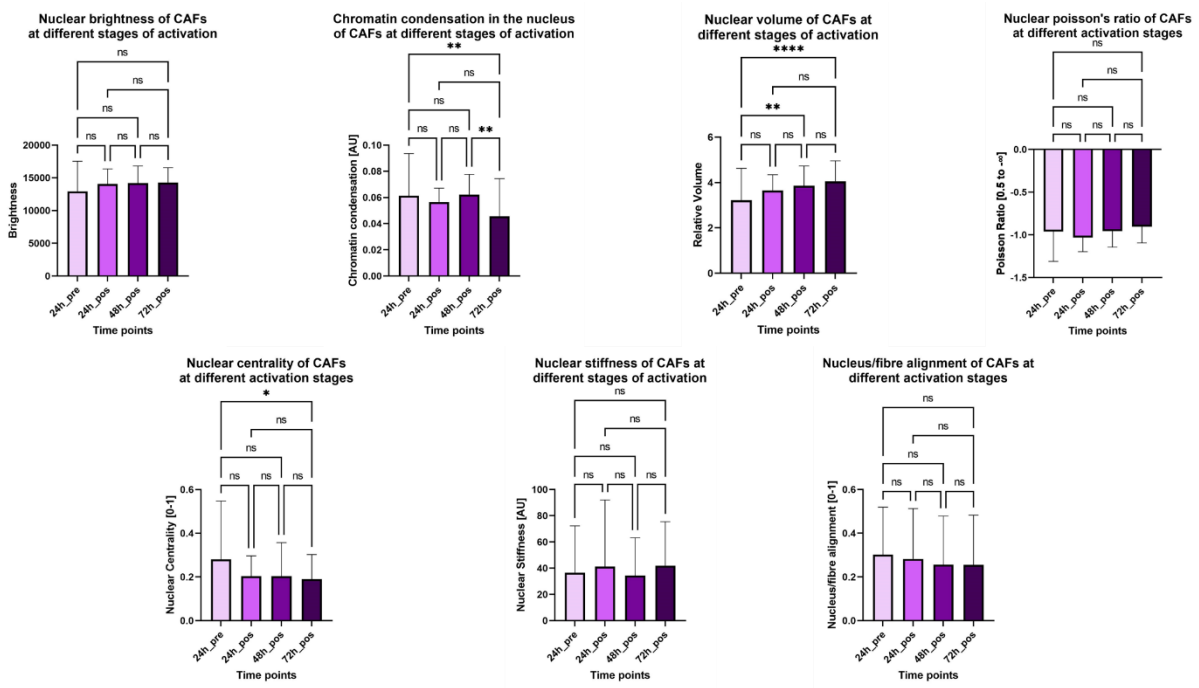
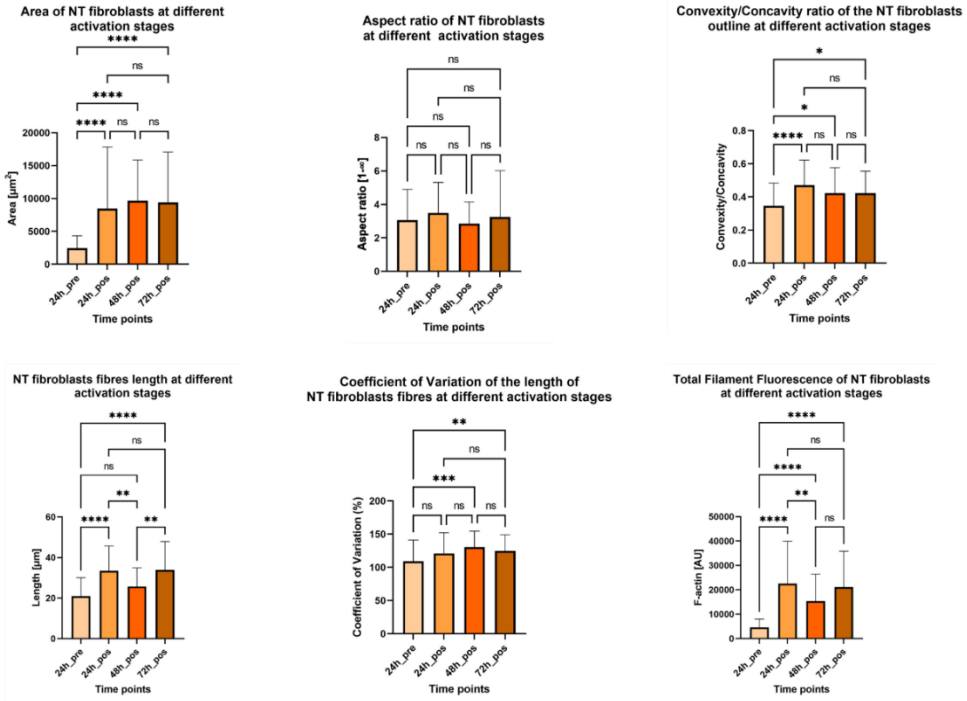
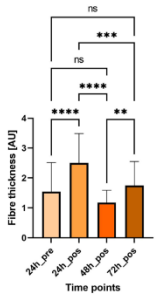


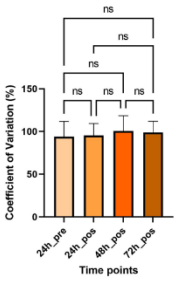
Figure A1-A21. Graphs for CAFs with the statistical analysis



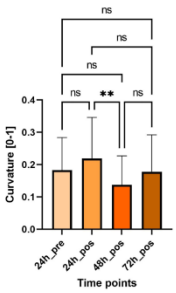
Fibre thickness of NT fibroblasts at different activation stages



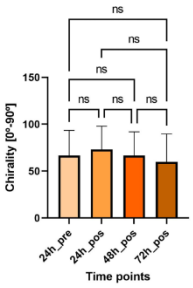
Coefficient of Variation of fibre thickness of NT fibroblasts at different activation stages



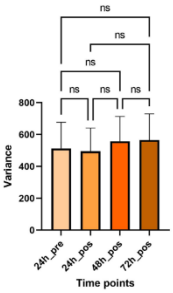
Curvature of NT fibroblasts fibres at different activation stages



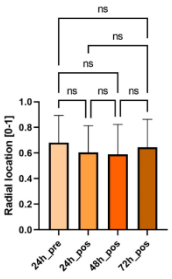
Chirality of NT fibroblasts at different activation stages



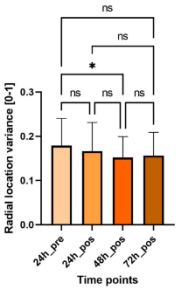
Variance of the chirality of NT fibroblasts fibres at different activation stages



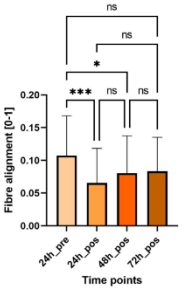
Preferent radial location of NT fibroblasts fibers at different activation stages



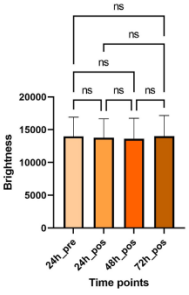
Spread of fibres in the NT fibroblasts at different activation stages



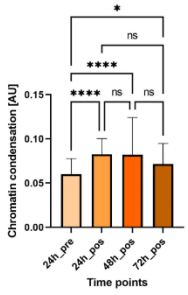
Fibre alignment of NT fibroblasts at different activation stages



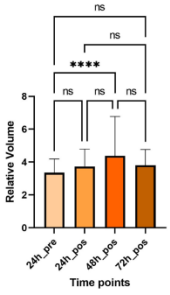
Nuclear brightness of NT fibroblasts at different stages of activation



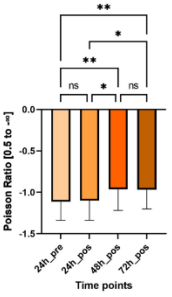
Chromatin condensation in the nucleus of NT fibroblasts at different stages of activation



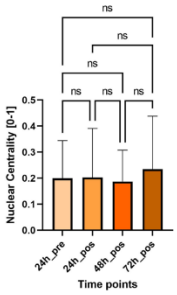
Nuclear volume of NT fibroblasts at different stages of activation



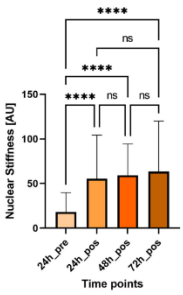
Nuclear poisson's ratio of NT fibroblasts at different activation stages



Nuclear centrality of NT fibroblasts at different activation stages



Nuclear stiffness of NT fibroblasts at different stages of activation



Nucleus/fibre alignment of NT fibroblasts at different activation stages

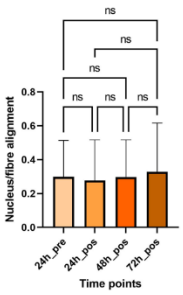


Figure A22-A42. Graphs for NAFs with the statistical analysis

12.4. Machine Learning application

The code used to apply Logistic Regression algorithms can be seen below. Although this code is specific for one of the conditions, the same process has been followed for the other.

```
#We import the files
import numpy as np
import pandas as pd
from os import chdir
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import AffinityPropagation, AgglomerativeClustering, KMeans
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, ConfusionMatrixDisplay
from scipy.stats import zscore

chdir(r'C:\Users\lourd\OneDrive\Escriptori\Lourdes\Enginyeria Biomèdica\4EBM\2SEM\tfg\.EXCELS DADES\NO TUMORAL\PhaVimD') #We select the file

#We obtain the different datasets
df1 = pd.read_csv(r'1compiled_results_Pre24hNT_PhaVimD20x.csv', delimiter = ';', decimal = ',')
df2 = pd.read_csv(r'2compiled_results_Post24hactivNT_PhaVimD20x.csv', delimiter = ';', decimal = ',')
df3 = pd.read_csv(r'3compiled_results_Post48hactivNT_PhaVimD20x.csv', delimiter = ';', decimal = ',')
df4 = pd.read_csv(r'4compiled_results_Post72h_activNT_PhaVimD20x.csv', delimiter = ';', decimal = ',')

#the datasets are joined into one
frames = [df1, df2, df3, df4]
df = pd.concat(frames)

#the target label is defined
vector = np.concatenate([np.repeat(0, df1.shape[0]),
                         np.repeat(24, df2.shape[0]),
                         np.repeat(48, df3.shape[0]),
                         np.repeat(72, df4.shape[0])])

df['Target'] = vector

#NaN and Inf rows are dropped
df = df[~np.isinf(df)]
df = df.dropna()
```

```
#We apply Z-score transformations
df.iloc[:, 0:21] = df.iloc[:, 0:21].apply(zscore)


#We built the correlation matrix
correlation = df.corr()

plt.figure(1)
sns.heatmap(correlation, cmap = 'seismic')

#We define features and outcome

X = np.array(df.iloc[:, 0:21]) #features
y = np.array(df.iloc[:,21]) #outcome

#%% Logistic regression
repeat = 30
acc_list = []
for i in range(30):
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=i) #multiple random train-test splits
    model = LogisticRegression(random_state=i, solver = 'sag', multi_class= 'multinomial') #we define the model

    fit_model = model.fit(X_train, y_train) #fit the model to the data

    pred = fit_model.predict(X_test) #obtain the predicted labels

    acc = accuracy_score(y_test, pred) #computed accuracy score
    acc_list.append(acc)
    print(acc)

    #We can plot the confusion matrix
    conf = confusion_matrix(y_test, pred)

    disp = ConfusionMatrixDisplay(conf)

    disp.plot()
    plt.show()

# Different accuracy scores
plt.figure()
plt.plot(list(range(30)), acc_list, '-')
```

```
plt.title('Accuracy score of NT cells on different activation stages regression')
plt.ylabel('Accuracy score')
plt.xlabel('Iterations (#)')
props = dict(boxstyle='round', facecolor='wheat', alpha=0.5)
plt.text(15, 0.715, 'Average Accuracy score = {:.2f}'.format(np.mean(acc_list)), bbox = props)
plt.show()
```

The plots of the average scores obtained for different train-test partitions on the different LogisticRegression algorithms can be seen in **Figure A43**.

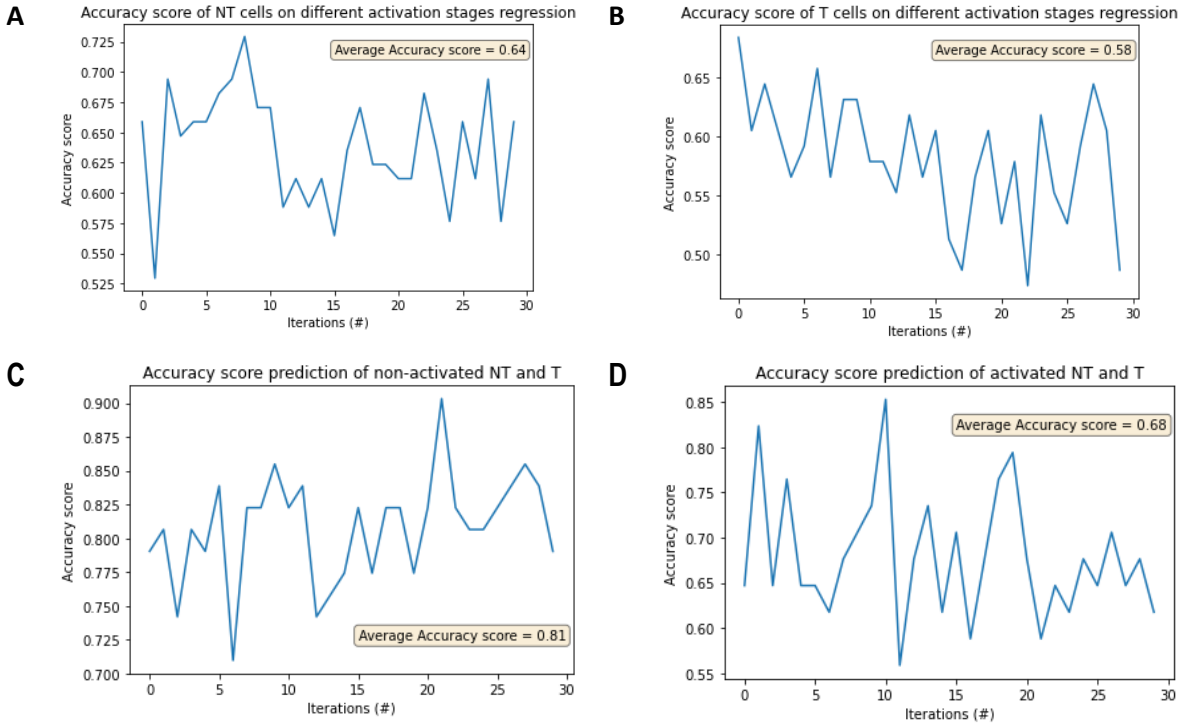


Figure A43. Accuracy scores obtained from the application of Logistic Regression for (A) prediction between the different time points within the non-tumoral fibroblasts; (B) prediction between the different timepoints within the tumoral fibroblasts; (C) prediction between tumoral or non-tumoral fibroblasts of non-activated populations; (D) prediction between the tumoral or non-tumoral fibroblasts of activated populations.

In the same manner the overall structure of the code used in R to perform clustering can be seen below. The needed libraries must be downloaded.

```
cels <- read.csv (file = "C:\\Users\\lourd\\OneDrive\\Escriptori\\Lourdes\\Enginyeria Biomèdica\\4EBM\\2SEM\\tfg\\.EXCELS DADES\\TUMORAL\\PhaVimD\\4compiled_results_Post72hactivT_PhaVim20x.csv", sep = ';', dec = ',') #we import the data

library(matrixStats)
dfN <- cels[1:54,] #define dataset
dfN <- do.call(data.frame,                # Replace Inf in data by NA
         lapply(dfN,
                function(x) replace(x, is.infinite(x), NA)))
data <- as.data.frame(dfN) #data.frame
data = as.data.frame(data)
```

```
data <- data[!(is.na(data$`Nuclear.stiffness`)), ] #we drop NaN data

set.seed(123)
#We obtain the optimal number of clusters
library("NbClust")
optimo.cluster <- data %>%
  scale() %>%
  NbClust(distance = "euclidean",
          min.nc = 2, max.nc = 10,
          method = "complete", index ="all")
# The optimal number of clusters is obtained
library(factoextra)

#We apply the zscore to the data
dfN <- data#copia data
zscore = (dfN - rowMeans(dfN))/(rowSds(as.matrix(dfN)))[row(dfN)]
# We perform kmeans
km.zscore <- kmeans(zscore, 2, nstart = 50) #kmeans
fviz_cluster(km.zscore, data = zscore,
             ellipse.type = "convex",
             palette = "rainbow",
             ggtheme = theme_light()) #representación

km.res$size #magnitude of the clusters
km.res$cluster


data_cluster = data.frame(data,cluster = km.zscore$cluster) #add assigned cluster number to each of the cells
data_cluster <- data_cluster %>%
  group_by(cluster) %>%
  arrange(cluster)
# we import data
write.csv2(data_cluster,                "C:                \\Users\\lourd\\OneDrive\\Escriptori\\Lourdes\\Enginyeria
Biomèdica\\4EBM\\2SEM\\tfg\\.EXCELSDADES\\TUMORAL\\PhaVimD\\pos72h_clustering.csv",
row.names=FALSE)
```