Single cell analysis of the fibrotic landscape in Dupuytren's Disease



Dr Thomas Layton

St Hilda's College

University of Oxford

Supervisors

Professor Jagdeep Nanchahal

Professor Dominic Furniss

Trinity 2019

Thesis submitted for the degree of

Doctor of Philosophy in Molecular and Cellular Medicine

Table of Contents

Acknowledgements……

Abstract……

Statement of originality……..

Chapter I. Introduction……1

I.I. Fibrosis: a major unmet clinical need……2

I.I.I The complex cellular landscape in fibrosis……3

I.I.II The clinical landscape of human fibrosis……6

I.II The myofibroblast in fibrosis……7

I.II.I Myofibroblast origin……7

I.II.II Myofibroblast force generation……8

I.II.III Tools to measure cellular force generation……9

I.III Dupuytren's disease……13

I.III.I Emerging concepts in Dupuytren's disease pathogenesis……13

I.III.II The immunology of Dupuytren's disease……14

I.III.III The extracellular matrix in Dupuytren's disease……15

I.III.IV Emerging and existing treatments for Dupuytren's disease……18

I.IV Single cell genomics……20

I.IV.I Experimental considerations in single cell RNA-seq……21

I.IV.II Single cell profiling in fibrosis……23

I.V Thesis Hypothesis……26

I.VI Thesis Objectives……26

Chapter II. Materials and methods……28

II.I. Patient samples……28

II.II Tissue culture reagents……28

II.III Antibodies……30

II.IV Cell culture and tissue digestion……30

II.V Cell counting……31

II.VI Cell viability……31

II.VII Cell freezing and thawing……31

II.VIII Single cell isolation……32

II.IX Droplet-based single cell RNA-seq……32

II.X Computational analysis……33

II.XI Flow cytometry……37

II.XII FACS sorting……38

II.XIII Immunohistochemistry……38

II.XIV Immunofluorescence and confocal microscopy……39

II.XV Apoptosis assay……39

II.XVI Acid cleaning of coverslips……40

II.XVII Preparation and functionalization of polyacrylamide (PAA) gels ……40

II.XVIII Traction force microscopy……41

II.XIX Atomic force microscopy……42

II.XX Transfections……43

II.XXI Bulk RNA-seq library preparation……43

II.XXII Bulk RNA-seq data analysis……43

II.XXIII Cell cycle assay……44

Chapter III. Towards a single cell atlas of human fibrosis……45

III.I Introduction……46

III.II Hypothesis and aim……47

III.III Results……48

III.III.I Patient demographics and data quality control……48

III.III.II Feature selection in single cell RNA-seq……52

III.III.III Clustering cells in Dupuytren's nodules……56

III.III.IV Exploring the perivascular niche in human fibrosis……61

III.III.V The immune landscape in human fibrosis……65

III.IV Discussion……68

Chapter IV. Defining expression profiles of fibroblasts and myofibroblast in fibrosis……72

IV.I Introduction……73

IV.II Hypothesis and aims……75

IV.III Results……76

IV.III.I Elucidating transcriptomic signatures of fibroblasts and myofibroblasts……76

IV.III.II Defining three major subdivisions of fibroblasts in DD……81

IV.III.III Pseudotemporal ordering of fibroblasts in DD……86

IV.III.IV ICAM1^{+} fibroblasts are enriched in Dupuytren's nodules……89

IV.III.V Bulk RNA-seq of Dupuytren's nodule and matched palmar skin……91

IV.III.VI Integrating bulk transcriptomic signatures with cell type profiles……95

IV.IV Discussion……97

Chapter V. Classification of the myofibroblast in human fibrosis……103

V.I Introduction……104

V.II Hypothesis and aims……106

V.III Results……107

V.III.I Resolving distinct myofibroblast communities in DD……107

V.III.II The cycling stromal cell in DD is a myofibroblast……110

V.III.III Identifying a myofibroblast activation trajectory……114

V.III.IV Elucidating a novel myofibroblast marker CD82……119

V.III.V Expression of $CD82^{+}$ myofibroblast markers *in vivo*……122

V.III.VI CD82 function in myofibroblasts……124

V.III.VII $CD82^{+}$ stromal cells in murine fibrosis……129

V.VI Discussion……134

Chapter VI. Establishing an *in vitro* model to quantify myofibroblast force generation……138

VI.I Introduction……139

VI.II Hypothesis and Aims……140

VI.III Results……141

VI.III.I Experimental and computational workflows in TFM……141

VI.III.II Defining hydrogel thickness……142

VI.III.III Characterization of fiducial marker distribution……143

VI.III.IV Optimizing seeding time for TFM……145

VI.III.V Defining optimal gel stiffness……146

VI.III.VI Characterizing tissue mechanics of Dupuytren's nodules……..148

VI.III.VII Estimating the system resolution ……..151

VI.III.VIII Regularization factor for traction force reconstruction……153

VI.III.IX Tracking fluorescent marker beads with particle image velocimetry…….154

VI.IV Discussion……157

Chapter VII. Functional heterogeneity in human myofibroblasts at single cell resolution……161

VII.I Introduction……162

VII.II Hypothesis and aims……163

VII.III Results……164

VII.III.I Dissecting mechanical force in living human myofibroblasts……164

VII.III.II Spread myofibroblasts generate forcers in discrete foci……167

VII.III.III Dynamics of myofibroblast force generation during cell spreading……171

VII.III.IV Distinct biophysical communities in human myofibroblasts……175

VII.III.V Functional heterogeneity along myofibroblast activation continuum……178

VII.IV Discussion……181

Chapter VIII. Conclusion……186

VIII.I Dissecting the multicellular ecosystem of human fibrosis with single cell RNA-seq…187

VIII.II Fibroblasts in the fibrotic milieu……192

VIII.III A detailed classification of human myofibroblasts in fibrosis……196

VIII.IV The membrane tetraspanin CD82 in fibrosis……200

VIII.V What is a single cell? ……202

References……203

Acknowledgments

The past three years have been a truly amazing experience and I am thankful to all those who have made it possible. Since medical school I have looked for an opportunity to explore the basic science of human disease and feel blessed to have been given it here in Oxford. To be supervised by Jagdeep has been a privilege, who has instilled in me an ambition to always strive for more. His unwavering support, time and patience has produced not only a set of observations, I hope, that expand our knowledge of fibrosis, but also enabled me to develop the tools and broad scientific foundations to become an independent investigator. I will forever be indebted to Jagdeep, whose trust in me has ignited my desire for an academic career and I hope this is just the beginning of the journey.

During my foundation training, as I was wondering if or how I could begin my scientific career, a colleague in clinical medicine suggested I speak to Professor Furniss. His reputation was of a gifted teacher, an excellent scientist and talented clinician. All these qualities, and many more, embody Professor Furniss and I am thankful for his stewardship, insightful reviews and lively discussions during this DPhil. Dom has provided the perfect role model during the nascent stages of my research career.

Whilst at Oxford I have also been fortunate to meet some of the kindest, most dedicated and receptive individuals. I will be forever grateful for the continuous support, direction and patience of Lynn, Marisa and Fiona. Their encouragement and scientific knowledge have allowed me to

push through times of disappointment and frustration and opened my eyes to the wonders and intellectual rigors of basic science. Without them this thesis would not have been possible.

Lastly, I would like to thank my mum and brothers who have stood by me through the tumultuous past three years, and to Hannah, who remains the most brilliant, warmest and brightest person I have ever met. She represents all I aspire to be.

Abstract

Fibrotic disorders represent some of the most devastating and poorly treated conditions in the Western World. Despite this, effective therapeutic interventions have yet to be identified for many of these conditions. A major barrier to identification of novel targets and successful clinical translation is a poor understanding of the myofibroblast-rich human fibrotic microenvironment, compounded by the dynamic nature of this ecosystem. Here, I use Dupuytren's disease (DD), a localized fibrotic condition of the hand, as a model to construct a stromal cell atlas in human fibrosis at single cell resolution. I identify three fibroblast subsets in DD, including a novel ICAM1^{+} chemokine-enriched fibroblast. After this, I demonstrate heterogeneity in myofibroblasts and report discrete proliferative and CD82^{+} populations. Using a newly optimized functional assay, I show phenotypic diversity along modelled signature of myofibroblast activation and validate a novel cell surface marker of this cell type, CD82, Finally, I show CD82 functions to regulate cell cycle progression in DD myofibroblasts.

Statement of originality

I hereby declare that this submission is my own work and to the best of my knowledge contains no materials previously published or written by another person, or which have been accepted for the award of any degree or diploma at the University of Oxford, or any other educational institution. Any contributions made to the research by others, with whom I have worked, is explicitly acknowledged in this thesis.

I

For my father

II

Lend no part to any humanity that suffuses
you in its own light.
Be not chimera of morning,
Half-man, half-star.

(*Harmonium*, Wallace Stevens)

Chapter I. Introduction

I.I. Fibrosis: a major unmet clinical need

Fibrosis is defined as the accumulation of excess matrix proteins in a given organ(1-3). It can affect any tissue structure in the human body and represents a dysregulated wound healing response(4). The basic pathology of fibrosis is a continued production of matrix proteins that disrupts tissue architecture and ultimately compromises organ function. This stereotyped process occurs in many chronic inflammatory diseases of both infectious and sterile aetiology(5, 6). In addition, fibrosis is a major pathological feature of many autoimmune disorders including Crohn's disease and scleroderma(4, 7, 8). It is also seen during cancer progression, and within many tumours promotes metastatic dissemination by forming an aggressive desmoplastic reaction(9, 10).

There are numerous distinct triggers of fibrosis, including chronic autoimmune diseases, exposure to toxins and therapeutics, infections and sterile inflammatory disorders such as rheumatoid arthritis(4). Again, despite the plethora of fibrotic initiators, the basic pathology is highly conserved with the overproduction of matrix proteins(1). Nevertheless, fibrotic microenvironments are highly complex containing fibroblasts, myofibroblasts, immune, vascular and perivascular cells(8). It is increasingly recognized that interaction and communication between distinct cell types is crucial for the progression of fibrosis(11). Immune cells are now understood as important regulators of stromal cell activation and thus key mediators of fibrosis(5, 12-14).

As the underlying pathology of fibrosis involves the deposition of matrix proteins, much work has been undertaken to define the exact proteins involved(15-17). Major protein components of

fibrotic lesions are collagens and fibronectin(17), but many others are deposited in variable quantities(15). The major sources of matrix are fibroblasts and myofibroblasts which are equipped with developed rough endoplasmic reticulum and mature Golgi apparatus for protein production and secretion. During normal wound healing, this system secretes and remodels a nascent matrix scaffold to restore tissue architecture. In fibrosis, the excess matrix deposited amalgamates to tertiary protein structures that deform tissue morphology and mechanics(18). In response to this, further matrix is deposited by stromal cells(19, 20). Together, this vicious cycle of matrix production and contracture underpins the damaging effects of fibrosis on organ function(21).

I.I.I *The complex cellular landscape in fibrosis*

It is well established that inflammation plays a significant role in fibrosis. For example, in bleomycin-induced lung fibrosis, a common animal model of lung fibrosis, exposure to this agent induces epithelial cell necrosis whereby activating an inflammatory response(22, 23). Subsequently, infiltration and activation of immune cells induces an extensive fibrotic response characterized by excessive matrix deposition in the lung(24). Alongside acute reactions, chronic persistent inflammation is also common trigger of fibrosis(25, 26). One example, Dupuytren's disease (DD), is a localized fibrotic condition of the palm containing various immune cells and pro-inflammatory mediators that persist over a long period(27). In DD, disease progression exhibits great variation between patients and this may relate to temporal fluctuations in the inflammatory compartment(27).

As the interplay between inflammation and fibrosis is relevant across a wide spectrum of diseases(25, 28, 29), significant effort has been devoted to resolve the mechanisms promoting immune cell recruitment and activation(13, 14, 30-32). As a result, the entire gamut of the human immune system has been associated with fibrosis(29, 31, 33-35), suggesting that vastly differing upstream immunological mechanisms all terminate at a common final fibrotic pathway. With regards to immune cells in fibrosis, proposed mechanisms of activation can loosely be separated into innate or adaptive immune responses. One example of an adaptive response pertains to CD4 T cell subsets (Th1, Th2, Th17) and these have been shown to drive diverse immune reactions in fibrosis. Th2 T cells have emerged as central mediators of fibrosis in several organs, through production of cytokines including IL-4, IL-15 and IL-13(3, 36, 37). Specifically, IL-13 promotes fibrosis in systemic sclerosis and liver fibrosis(31, 38). Given the wide range of immune pathways involved, the focus on treating fibrosis has transitioned from removing a specific immunological trigger to elucidating key regulatory drivers of immune activation(31, 39). An alternative approach would be targeting final downstream effector mechanisms that act on myofibroblasts and their precursors(27, 40). This, hopefully, could facilitate development of an anti-fibrotic treatment efficacious across a range of disorders.

In response to tissue injury, inflammatory myeloid cells are the first recruited by chemokines(41). Myeloid cells then act via two broad pathways to promote fibrosis; exacerbation of tissue injury with further recruitment of immune cells or direct activation of collagen-producing stromal cells(5). One innate immune cell, the macrophage, has taken centre stage as a key regulator of tissue repair and fibrosis(5, 12-14). Macrophages display tremendous phenotypic plasticity, and

growing reports demonstrate that disturbances in macrophage function lead to aberrant wound healing(5, 42). Macrophage nomenclature describes classically activated (M1) and alternatively activated (M2) phenotypes based on cytokine profiles and pathogen signals and this dichotomy is of relevant in fibrosis(43). M1 macrophages are thought to predominate during the early phase of response to injury, releasing pro-inflammatory cytokines that if unchecked induce tissue damage. As the acute phase of inflammation subsides, the landscape shifts to one that promotes the polarization and recruitment of M2 macrophages(5). Despite previous reports demonstrating M2 macrophages as important for tissue repair and regeneration, recent studies have shown these cells to be an major source of pro-fibrotic cytokines and potentially central mediators of tissue destruction in fibrosis(44).

Over the past few years we have begun to integrate our understanding of both immune and stromal compartments in fibrosis(33, 45, 46). As in many other disease systems, interactions between distinct cell types is likely crucial and several soluble mediators of this process have been described. In particular, innate immune cells can express a range of soluble factors capable of driving stromal cell activation including TGF-β1, IL-6 and IL-8(25, 47, 48), with some of these molecules representing potential therapeutic targets. In addition to immune cell derived chemokines and cytokines, stromal cells also express multiple pro-fibrotic soluble factors. Specifically, stromal cells are a significant source of TGF-β1 and IL-6, as well as numerous chemokines including CXCL12, CXCL14 and CCL2(25). Therefore, when exploring the pathology of fibrosis, we must consider not only individual cellular phenotypes but also the multi-cellular niche in which these cells reside.

I.I.II *The clinical landscape of human fibrosis*

Despite the enormous body of research into the mechanisms of fibrosis, there are still no effective anti-fibrotic therapeutics. In idiopathic pulmonary fibrosis (IPF), a devastating lung disorder, a number of therapeutics have been recently approved, including perfenidone that acts to inhibit TGF-β1 induced collagen production(49). The efficacy of these compounds is limited to extending patient survival by only a few months and no currently approved treatment for IPF is able to halt disease progression(49). Unfortunately, this is a typical example of the unmet need in almost all other fibrotic disorders, including those affecting the liver, gut, hand and heart(4). Collectively, this unmet need reflects our still poor understanding of the biology of fibrosis. Potential reasons behind this are twofold. Firstly, our reliance on animal models that often fail to recapitulate many features of human disease(24, 50, 51). Secondly, the major challenge in obtaining human samples of early active stages of disease. IPF is a good example of both(52). The bleomycin lung model is the most commonly used murine system for lung fibrosis but involves an acute inflammatory insult with very different pathophysiology to that of IPF(52). Additionally, most studies using human tissue in IPF utilize small biopsy samples derived from end-stage fibrotic lungs obtained during a lung transplant procedure. The study of Dupuytren's disease addresses some of these issues as the Dupuytren's nodule is accessible and represents a fibrotic microenvironment rich in cells.

I.II. The myofibroblast in fibrosis

The key effector cell in all forms of fibrosis is the myofibroblast(53-55). Myofibroblasts were first described by Gabbiani who defined three key distinguishing features; fibrillar subcellular components reminiscent of smooth muscle cells, strong intercellular and cell-ECM attachments and nuclear deformations(56). In every organ undergoing fibrosis, myofibroblasts have been shown as a major driver of the disease process(53). At present, the modern definition of myofibroblasts has been generalized to the *de novo* expression of α-smooth muscle actin (α-SMA) and its incorporation into stress fibres(53). Other distinguishing features of myofibroblasts include intercellular adherens and gap junctions, large supermature focal adhesions (focal complexes) and a contractile phenotype(53). The defining protein of myofibroblasts, α-SMA, is the actin isoform typical of vascular smooth muscle cells and is coded by the ACTA2 gene. Currently, α-SMA is the most commonly used marker of myofibroblasts and several studies have shown this protein directly modulates cell phenotype(57-59).

I.II.I *Myofibroblast origin*

Myofibroblasts are crucial for the restoration of native tissue architecture during normal wound healing(54). In this context, myofibroblasts appear transiently and, after the secretion and remodelling of matrix proteins, are eliminated through apoptosis(45, 54, 56). In fibrosis, this process becomes dysregulated and myofibroblasts persist, ultimately contributing to the destructive cellular program. There are two major known stimuli for myofibroblast development; TGF-β1 and a stiff mechanical environment, which synergize to promote myofibroblast

differentiation(60). TGF-β1 is a profibrotic cytokine able to induce the expression of α-SMA, collagens and other key myofibroblast proteins. In addition, TGF-β1 regulates various biological processes in myofibroblasts including cell contraction, migration and matrix secretion(48, 61). These results highlighted TGF-β1 as a potential therapeutic target in fibrosis, but broad inhibition of this cytokine has not translated to the clinic(62). Current work aims to dissect the exact molecular pathways of TGF-β1 signalling in fibrosis, in the hope of developing a more bespoke and efficacious treatment regime(63).

The mechanical environment profoundly influences myofibroblast phenotype(64). Excessive matrix proteins stiffen the native tissue structure in fibrosis, providing a potent stimulus for stromal cell activation and further matrix secretion(54). This positive feedback loop in part explains how myofibroblast activation can be maintained in fibrosis. Indeed, several studies have shown that myofibroblasts are highly mechanosensitive, capable of sensing subtle changes in their environment(20, 64). An open question is whether these mechanotransduction pathways could be manipulated as part of a therapeutic strategy.

I.II.II *Myofibroblast force generation*

ECM remodelling in both physiology and disease is a force-dependent process(65). Myofibroblasts are defined by a contractile phenotype and α-SMA supports a robust cytoskeletal network forming a nexus for strong force generation(66). In normal tissue homeostasis, ECM remodelling requires the interaction between cell and environment through specialized contacts called focal adhesions(67, 68). Cellular traction forces are applied through these structures to sculpt and

organize the secreted matrix. In the context of fibrotic disorders, it is now recognized that stromal cell contraction is central to disease progression(64, 69).

Cellular force generation is dependent on a complex system of cytoskeletal proteins and regulators (70). In myofibroblasts, stress fibres are linear bundles of F-actin capable of generating, transmitting and sensing mechanical tension(54). Stress fibres are associated with another important structure, the focal adhesion complex, itself composed of a diverse array of proteins including paxillin, vinculin and talin(71, 72). Together, stress fibres and focal adhesions make up a force generating unit, connecting cell and extracellular matrix, under further regulation by non-muscle myosin and multiple actin-binding proteins(70). These structures enable the transmission of force required to remodel the macromolecular structure of secreted matrix proteins such as fibronectin and collagen(73). In addition, intracellular calcium signalling has been demonstrated as another important regulator of myofibroblast contraction(74). Given the role of force generation in myofibroblasts, developing methods to quantify this process has been an important goal of fibrosis research.

I.II.III *Tools to measure cellular force generation*

The first methods developed to measure cell generated forces relied on observations of the deformation of various elastic substrates following cell seeding(75). These pioneering studies were the first to demonstrate that cellular forces are important for a number of biological processes such as cell migration(75). Subsequently, improved techniques were developed to probe cellular forces in greater detail(76). Newer methods aimed to overcome limitations of initial techniques

that lacked true quantitative capacities. Currently, available techniques include collagen contraction assays, tissue pillar arrays, traction force microscopy and molecular techniques such as FRET-based biosensors(76). Each method has advantages and limitations and no gold standard exists. The most appropriate method is highly dependent on the underlying biological question.

Traction force microscopy (TFM) has become one of the most widely used techniques for measuring cellular forces. Key advantages of this method are an ease of implementation as well as providing the opportunity to quantify the spatiotemporal dynamics of force generation(77, 78). Indeed, TFM has been applied in many cellular contexts, including immune synapse formation(79), B cell activation(80) and fibroblast migration(70). The basic principle of TFM is based on the first methods developed to measure cellular forces, namely tracking deformations of an elastic substrate(75). The experimental setup first requires formation of an elastic polymer hydrogel, commonly synthesized from polyacrylamide or silicone. These substrates are impregnated with fluorescent beads that serve as fiducial markers for tracking substrate deformation in both space and time(81, 82). Often, the surface of the elastic substrate is functionalized with a protein that facilitates cell binding and for stromal cells this is usually either fibronectin or collagen(83). During data acquisition, the cell of interest adheres to the hydrogel surface and is located using optical microscopy. Following this, an image is acquired that captures the distribution of beads in the stressed state. Subsequently, the cell is removed from the gel surface and another image is acquired that determines the bead positions in the unstressed state. Finally, these two images or series of images are passed through computational algorithms that first quantify bead displacement and then reconstruct traction force. This final analysis step is reliant on *a priori*

knowledge of the hydrogel stiffness. As the beads tracked are smaller than the cell (40nm), TFM allows for the reconstruction of cellular forces with subcellular resolution, and has quantified the force dynamics of many distinct biological processes, from focal adhesion maturation to cell differentiation(84).

Alongside experimental developments in TFM, there has been a concerted effort to produce robust and generalizable computational algorithms for data analysis(85). The basic steps in data analysis first require quantification of bead displacement within the hydrogel. A number of approaches have been used to track beads including optical flow algorithms and techniques borrowed from fluid dynamics, such as particle image velocimetry (PIV)(86). PIV has emerged as the most commonly used technique for bead tracking in TFM and utilizes the cross-correlation of smaller regions of interest to obtain velocity measurements. These smaller regions, termed interrogation windows, enable PIV to be used with a high density of particles and avoids a reliance on tracking individual fiducial markers in the image. In TFM, the use of a high density of beads directly improves the resolution of force measurements as it facilitates capturing small deformations in the substrate(85).

Once bead displacement measurements have been obtained, the final step in data analysis is to reconstruct cellular forces(85). Diverse computational algorithms are available for this task and all rely on continuum elasticity theory that describes substrate deformations as a consequence of forces applied to the substrate surface(85). This means that the reconstruction of traction forces is inherently an inverse problem in which forces are recovered following measurement of

substrate deformation. With regards to force reconstruction, Fourier transform traction cytometry (FTTC) has been shown to be a sensitive and robust method to complete this task(87, 88). FTTC solves the problem of force reconstruction using the Fourier domain and benefits from not requiring any other information than bead displacements. This, alongside its accuracy, is the major reason FTTC is currently the most widely used traction reconstruction technique. One key principle in FTTC is the choice of regularization, which is weighted by a regularisation parameter λ. Regularization aims to balance contribution of noise in the image during force reconstruction with suppression of reconstructed force. Finally, all the analysis steps have been coded into multiple programs in several languages, including Java and MATLAB. These have facilitated more widespread use of TFM.

I.III. Dupuytren's disease

Dupuytren's disease (DD) is a common fibroproliferative disorder of the hand. It affects 12% among those aged 55 years, increasing to 29% of those aged 75 years in Western populations(89). The first clinical manifestation is the appearance of a firm nodule in the palm. After this, linear cord-like structures develop and can extend into the digits. During disease progression, cords condense and contract, leading to permanent flexion deformities. The rate and nature of disease progression is highly variable, but around 40% of patients eventually develop some degree of flexion deformity(90). We are only recently fully appreciating the significance of these flexion deformity with regards to activities of daily living, including self-care and employment(91).

Current guidelines for the management of DD recommend intervention when the digital flexion contractures limit hand function and/or the proximal interphalangeal joint is flexed to 30° or more(92-95). There are numerous treatment options for late stage disease, ranging in invasiveness and efficacy. The main therapeutic interventions include division of the cords using collagenase injection or needle fasciotomy, or surgical excision of the fibrotic tissue by limited fasciectomy and dermofasciectomy. Given the prevalence of DD and its associated morbidity there has been growing focus on improving our evidence base for appropriate management.

I.III.I *Emerging concepts in Dupuytren's disease pathogenesis*

The complexity of clinical manifestations in DD reflects an equally complex pathophysiology. We still lack a precise aetiological mechanism in DD, but it is probable that cellular, immunological,

genetic and environmental factors all play a part. What is established is a significant heritable component in DD, with one twin study from Denmark estimating the overall heritability as 80%(96). In addition, recent studies have begun to identify key genetic association and disease regulators. For example, a genome wide association study (GWAS) identified nine susceptibility genetic loci in DD, six of which harboured genes encoding proteins in the Wnt signalling pathway, including *WNT4*, *SRFP4* and *RSPO2*(97). Wnt signalling has been shown as a major pathogenic driver in a number of fibrotic diseases (98), and the GWAS data suggest an important role for this pathway in DD(99, 100).

Expanding upon these results, a second GWAS in DD, and the largest to date, elucidated many other significant genetic associations(101) and confirmed a close association between Wnt signalling and DD. A novel finding of this study was decreased secretion of *SFPR4*, a soluble Wnt-antagonist, in individuals homozygous for one high-risk allele. Together, these data link a genetic predisposition to DD and activation of Wnt signalling(97, 100-102). In other fibrotic disorders, SFPR4 has been shown to be a negative regulator of myofibroblast activation(103) and future work should aim to examine whether SRPF4 represents a novel therapeutic target in DD and if targeting this pathway could downregulate the myofibroblast phenotype.

I.III.II *The immunology of Dupuytren's Disease*

Stromal cell activation and fibrosis progression is intricately linked to inflammation in all forms of fibrosis(1). The inflammatory reaction can have a known trigger, such as in helminth-induced liver fibrosis, or be activated by some unknown mechanism such as in interstitial lung disease. The

entire spectrum of immune cell types has been studied in fibrosis and a number implicated directly in the pathogenesis of DD, including macrophages, lymphocytes(30) and dendritic cells(104). Despite this, the precise mechanism of immune cell activation in DD is not understood(105).

In DD nodules, around 10% of the cells comprise immune cells, with major populations being macrophages and T cells. In addition, a number of pro-inflammatory cytokines have been identified, including tumour necrosis factor (TNF), interleukin-6 (IL-6) and interleukin-8 (IL-8)(27) How these cells are recruited into nodules and what is the initial trigger remains to be known. One theory proposed DD as a T-cell-mediated autoimmune disorder, largely following the discovery of dense T-cell infiltrates in DD nodules(30). Subsequently, more studies have identified distinct T cell lineages in DD(87, 105). Although these studies point to a potential adaptive immune reaction in DD, no autoantigen has been discovered or validated. Indeed, given the lack of validated animal models for DD and variation in disease progression, discovering the initial trigger for immune infiltration may be challenging. Nonetheless, it is now accepted that nodules of DD are characterised by localized inflammation, regardless of the initial stimulus.

I.III.III *The extracellular matrix in Dupuytren's Disease*

The extracellular matrix (ECM) is a vital regulator of tissue homeostasis. In DD, composition of the ECM varies dramatically between disease states and stages(17, 55, 106). Cords are relatively acellular and composed of mature fibrillar collagen(17, 55). In cords, the main stromal cell population is fibroblasts that function to maintain the integrity and health of the cord structure(55). Nodules, in contrast, are highly cellular, with densely packed myofibroblasts(107,

108) arranged in an irregular pattern. Currently, we still have a poor understanding of how both these structures develop and their precise relationship to one another. In several studies, it has been proposed that nodules are the first structure to appear and act as precursors for cords(17, 55). Moreover, nodules are enriched with immune and proliferating cells and have been suggested to represent the active disease unit(17, 55) . Nevertheless, cord formation and shortening are necessary for the development of flexion contractures that ultimately compromises hand function in DD.

Several SNPs (single nucleotide polymorphisms) identified in GWAS studies in DD demonstrate an association with matrix remodelling genes, including discoid domain receptor (*DDR2*), matrix metalloprotease 14 (*MMP14*) and integrin alpha-11 (*ITGA11*)(101). ECM modulating proteins such as integrins have been extensively studied in the context of fibrosis and now are recognized as potential candidates for therapeutic inhibition. However, in DD we are still missing a complete understanding of key pathogenic ECM binding proteins. DDR2 is a membrane receptor tyrosine kinase whose ligands include type I and III fibrillar collagen, the most prominent ECM proteins in DD nodules and cords(106, 109). DDR2 has been shown to regulate fibrosis in the lung and liver, without a clearly defined function in DD(109, 110). In lung fibroblasts, DDR2 stimulates myofibroblast differentiation through modulating fibrillary collagen deposition and increasing the potency of TGF-β1 signalling(95,96). It is possible that in DD DDR2 acts via a similar mechanism to promote stromal cell activation and collagen secretion in the palmar fascia. *MMP14* (MT1-MMP) is another gene associated with a high-risk locus and is a type I transmembrane protein of the MMP family of proteases. MMP14 protein has been shown to be overexpressed in DD

nodules(111) and knockdown of *MMP14* in DD fibroblasts inhibited both cell contraction and MMP2 activation(112). In addition, data exists that report MMP gene expression in DD correlating with clinical outcomes, including recurrence following treatment(113). These observations mandate further work for a precise role of MMP14 in DD but suggest it may act to regulate myofibroblast function.

Tissue mechanics and cellular forces are important modulators of normal tissue function(114). Increasing tissue stiffness drives stromal cell activation and promotes the development and progression of fibrosis(64). In addition, myofibroblasts in various systems have been demonstrated as highly mechanosensitive, capable of responding to minute changes in their environment (16, 115). This phenomenon, termed mechanotransduction, defines the process of inducing cellular biochemical responses from mechanical energy and is tightly regulated in normal physiology. Two molecules that serve as master regulators of this process are transcriptional co-activators Yes-associated protein 1 (YAP1) and transcriptional coactivator with PDZ-binding motif (TAZ)(116). Both these proteins act to modulate profound changes in gene expression in response to tissue stiffness(116). In addition, YAP1 has been shown to regulate the myofibroblast phenotype and transduce signals from the mechanical environment in several fibrotic conditions(20, 64). Also, in an *in vitro* fibroblast spheroid system, TAZ activation directly increased cell contraction and ECM expression(19). More recently, YAP1 has been shown to be a crucial determinant of the myofibroblast phenotype in DD and silencing of *YAP1* in DD myofibroblasts demonstrated its key role in the expression of fibrotic genes and cell contraction in this system(20). Whether YAP1 and

TAZ are attractive therapeutic targets in DD remains to be confirmed, but collectively emerging work has highlighted the potential for targeting mechanosensitive pathways.

I.III.IV *Emerging and existing treatments for Dupuytren's Disease*

There is no cure for DD and current treatments for late stage disease aim to correct the flexion deformity of the finger and restore hand function. A number of non-surgical treatments are available, including pharmacological treatment with vitamin E or steroids, physical therapies and radiotherapy(117) and these are usually reserved for early stage disease. However, we currently have no data from appropriately blinded clinical trials providing efficacy for most of these treatments. For late stage disease, the gold standard remains surgical resection of the diseased tissue. Despite the vast number of clinical studies on the management of DD, findings have been limited to uncontrolled and unblinded studies, with no conclusive evidence for their efficacy(117). Ongoing work aims to explore the efficacy of TNF inhibition in DD, including promising results from a Phase 2A trial, providing hope that emerging medical treatments that could help some patients avoid surgery(40).

Collagenase histolyticum (CCH) injections(118) are now a commonly used treatment option for DD. CCH benefits from being less invasive than surgery with more rapid recovery and fewer complications(119-121). CCH acts to soften the cord through enzymatic digestion and several clinical trials have investigated this treatment modality in DD(118, 122-126). The results demonstrate that CCH can significantly reduce joint contracture and improve the range of joint motion compared to placebo(127, 128). Ongoing multi-centre clinical trials comparing the efficacy

and cost effectiveness of CCH with surgical excision should provide further evidence for or against its use in DD(129).

Despite the wealth of non-surgical interventions, the mainstay of treatment for late stage DD remains surgery(130). Several surgical procedures are available and include needle fasciotomy (also known as needle aponeurotomy, or PNF), limited fasciectomy and dermofasciectomy. These techniques vary in their invasiveness and have distinct advantages and limitations. More invasive procedures, such as dermofasciectomy, are associated with lower risk of recurrence but require longer post-operative rehabilitation(131-134). Although several studies have demonstrated improvement in flexion deformities with PNF, it carries a relatively high risk of recurrence of approximately 30% at 5 years compared to 6% for limited fasciectomy(135, 136).

More invasive surgical procedures include limited fasciectomy and dermofasciectomy. Limited fasciectomy (LF) involves excision of most of the diseased tissue whilst preserving the overlying palmar skin. In dermofasciectomy (DF) the excision is extended to include all subcutaneous fat and skin over lying the diseased tissue. Two major proposed advantages of these techniques are a reduced risk of recurrence as compared to minimally invasive procedures such as PNF and a better correction of flexion deformity(135). Theoretically, in DF clearance of the overlying dermis and subcutaneous fact could remove potential myofibroblast precursors and thus explain the reduced rate of recurrence. In support of this, a multi-centre cross-sectional study reported the reoperation rate following fasciectomy as 6% compared to 0% after dermofasciectomy at 5 years(137). Despite this, mechanisms to explain this disparity have yet to be described.

I.IV. Single cell Genomics

The expansion of genomic technologies has facilitated a rapid evolution of biomedical research(138). Moreover, in the quest for an improved understanding of genetics, ground-breaking technological advances have enabled researchers to ask different types of questions(139).The current exploration of genetics aims to resolve the governing factors of gene expression, dynamics and regulation(140, 141). Since the turn of the century, the transcriptomic revolution has catalysed the development of methodologies to profile gene expression in groups of cells in an unbiased manner(142). DNA microarrays were the first technique that provided this capability and utilised a hybridization based method to allow the study of differential gene expression in thousands of genes simultaneously(143). Many important insights were gained with this technology, but limitations restricted its application to many complex biological questions(143).

In an attempt to improve upon microarrays, RNA-sequencing (RNA-seq) was developed and for the first time allowed the measurement of absolute mRNA molecules across the entire genome(144). RNA-seq also eliminated some of the inherent bias of microarrays and reduced the amount of RNA input required for gene expression profiling(144). This technology greatly expanded our understanding of the human genome and has been applied to varied biological questions, including drug perturbations responses(144-146), gene expression dynamics during malignant transformation(147), and gene regulatory networks throughout embryonic development(148). Despite the wealth of information provided by RNA-seq, it still relies on data

generated from groups of cells. This means that changes in gene expression are averaged over a given population and potentially meaningful heterogeneity is lost. For example, RNA-seq cannot distinguish between 50% of cells expressing two RNA copies and 100% of cells expressing one RNA copy. However, as the complexity of cellular systems becomes fully appreciated, there is a recognition of the unique features of individual cells within populations once assumed to be homogeneous. Increasing effort has therefore been devoted to the development of methods capable of quantifying cellular features at single-cell resolution to probe biological phenomena in greater detail(149, 150). Previous inquiry into complex cellular networks, as found in the immune and haematopoietic systems, have employed single cell techniques such as flow cytometry with tremendous yield(151), forming a strong foundation for the burgeoning single-cell perspective. This has now expanded from the proteomic to genomic sphere, where single-cell technologies are becoming an established platform in scientific research.

I.IV.I *Experimental considerations in single cell RNAseq*

Single-cell genomics aims to provide a new framework to explore heterogeneity within a given organism or system, and by capturing multi-dimensional information in individual cells characterizes of cell structure and function in unparalleled detail(149). One great advantage of single cell sequencing over traditional methods such as flow cytometry is the avoidance of *a priori* defining which parameters to measure. This unbiased approach has facilitated the discovery of novel cell types and states in numerous contexts(152-154). Other key advantages are an increased resolution in heterogeneous cellular systems, the ability to evaluate rare cells and to identify transient and intermediate states in individual cells(152, 155-157). Nevertheless, the true extent

of transcriptional heterogeneity and the diversity of cell types in diseased or normal tissues remain largely undiscovered.

Single cell genomics is a general term for several distinct technologies(139, 150, 151). The first transcriptomic profiling in single cells was performed in 2009 using an adapted array-based protocol to measure over 1,000 transcripts in single oocytes(158). Subsequently, developments in sequencing chemistry provided the first techniques to isolate and sequence several thousand full length transcripts in single cells(159). Two examples, CEL-seq and SMART-seq, leveraged Nextera transposase reagent from bulk RNA-seq allowing individual cells to be tagged using cell-specific index pairs(160, 161). These methods can read up to 10,000 transcripts per cell and provided enlightening insights in several biological systems(153, 162). Two major factors limited the application of these methods. Firstly, given the low amount of starting material, significant cycles of PCR or linear amplification were required to generate enough material for sequencing. This process introduced technical artefact due to inconsistencies in the amplification process and affected interpretation of the data. The second major limitation of these initial technologies was a restricted throughput due to a dependence on plate-based protocols during library preparation. To overcome the second limitation, microfluidic and droplet platforms were developed that enabled the profiling of hundreds of thousands of cells with a significantly higher throughput(159). Several droplet-based single cell techniques exist, with the most commonly used platform currently being the Chromium system supplied by 10x Genomics(159). A second experimental development, also utilised in the Chromium system, is the use of barcodes termed unique molecular identifiers (UMIs) that aimed to overcome the first limitation of amplification bias. UMIs

are random barcodes between 6 to 10 bp assigned to individual transcripts before amplification. UMIs thus allow for sequencing reads to be assigned to individual transcript molecules, thereby allowing reads with the same UMI to be collapsed and in so doing help nullify amplification bias. UMI counts are then assembled to generate the expression matrix used in downstream analysis. Many droplet based techniques have been developed, including DropSeq(149), inDrop(163) and the 10x Platform(159).

I.IV.II *Single cell profiling in fibrosis*

As reviewed above, fibrotic lesions are inherently diverse and complex containing various immune and stromal cells, extracellular matrix proteins and humoral factors including cytokines, chemokines and growth factors(164). Under physiological conditions, this system of highly coordinated cells and mediators orchestrate critical processes such as wound healing, tissue repair and regeneration(165). However, in many inflammatory conditions, fibrosis becomes a final common pathway that ultimately leads to tissue injury and organ dysfunction. As inter-cellular communication is key in maintaining the fibrotic reaction, it is now clear that analysing transcriptional and epigenetic profiles of single cells will expand our understanding of the unique contributions of different cell populations to the disease process.

The cellular landscape of all fibrotic disorders spans a continuum of multiple different cell types(3, 34). Myofibroblasts represent the dominant mesenchymal cell but the fibrotic milieu also contains cells of the adaptive and innate immune system as well as parenchymal cells of the organ affected. One important advantage of single cell sequencing is the ability to characterize rare cells which

otherwise may be lost in bulk RNA-seq due to other more predominant populations. Here, "rare cells" is a loose term as the exact power to define individual cells is highly dependent on the sequencing technology and experimental design. Even so, all current single cell platforms are able to profile rare cells in complex biological structures and have allowed scientists to characterize rare cells amongst circulating tumour cells(166, 167), neurons(167) and individual sperm cells(158). Grun et al (2015) applied a modified version of CEL-seq to uncover the molecular signature of rare enteroendocrine cells in murine intestinal organoids. This work also introduced a novel algorithm, RaceID, whose development was focused on the identification of rare cell types in heterogeneous cellular ecosystems(168). The ability to profile rare cells may hold great potential in fibrosis research as multiple reports have described important cell populations at a low proportion. For example, Kramann et al (2015) described a Gli^{+} progenitor cell that occupied a perivascular niche in multiple murine models as an important source of myofibroblast during injury induced fibrosis(168). Despite their significance, these cells represented a small percentage of the total cell population, highlighting the unique potential for single cell sequencing to explore this population.

The fibrotic microenvironment houses not only cells from distinct lineages, such as immune, stromal and parenchymal, but also distinct populations of each(27, 32, 48, 52). Therefore, to fully appreciate the molecular diversity of all cell types in fibrosis, single cell sequencing must not only resolve distinct cellular states, but also cells with highly similar transcriptomes. In a seminal paper, Shekhar et al (2016) provided a molecular census of retinal bipolar neurons using massively parallel single cell RNA-seq(169). Here, new bipolar cell types were elucidated

and validated, illustrating the capacity of single cell technology to build a robust taxonomy composed of similar cell types, as well as uncovering transcriptional differences that distinguish cell types within a class(169). Li et al (2017) answered a similar question and used a newly developed single cell protocol and unsupervised machine learning algorithm, Iterative Clustering for Identifying Markers (ICIM), to classify Drosophila olfactory projection neurons(170). Given the dominant stromal cell population in fibrosis these studies provide compelling evidence that single cell RNA-seq could be leveraged to profile these cells and may help to define new cell types in fibrosis. Recent reports also highlight the presence of distinct myofibroblast populations in cancer associated desmoplasia(171), but this remains an unanswered question in human disease.

I.V. Thesis Hypothesis

This thesis aims to test the hypothesis that localised human fibrosis contains heterogeneous stromal cell populations that may contribute diverse pathogenic processes to the disease microenvironment.

I.VI. Thesis Objectives

I. In the third Chapter I aim to systemically dissect the multicellular landscape of Dupuytren's disease using single cell RNA-seq. Here, I seek to profile 12 Dupuytren's nodules and resolve major cell populations including immune cells, perivascular cells and endothelial cells.

II. In the fourth Chapter I aim to undertake a detailed characterization of fibroblasts and myofibroblasts in DD, and define key markers genes of each. In addition, I seek to investigate fibroblast subpopulations and examine their proportions within two Dupuytren's disease structures.

III. In the fifth Chapter, I aim to investigate whether distinct communities of myofibroblasts exist in fibrosis and ask whether single cell profiling can help elucidate novel cell surface markers in this population.

IV. In the sixth Chapter, I aim to establish an *in vitro* model to quantify myofibroblast contraction at single cell resolution, based on traction force microscopy. This will provide a platform to integrate functional readouts with observations made in the single cell expression data.

V. In the seventh Chapter, I aim to undertake detailed characterization of myofibroblast force generation using traction force microscopy. Finally, I aim to test whether transcriptomic profiles of myofibroblast communities are phenotypically distinct and integrate findings made in single cell expression data with biophysical readouts of myofibroblast contraction.

Chapter II. Materials and Methods

II.I. Patient samples

After approval by the local ethical review committee (REC 07/H0706/81), tissue samples were obtained after informed consent from patients. Dupuytren's nodular tissue, palmar skin (uninvolved skin overlying Dupuytren's tissue), and non-palmar skin (full-thickness skin harvested from the groin or medial aspect of arm) were obtained from individuals with Dupuytren's disease undergoing dermofasciectomy. Given the lack of well-defined controls in Dupuytren's disease, I initially focused on dissecting heterogeneity within nodules. These structures were identified during tissue dissection and separated for selected assays, including flow cytometry and immunohistochemistry. For single cell RNA-seq, samples were received ~4 hours following surgical resection and were processed immediately.

II.II. Tissue culture reagents

All reagents were of tissue culture grade and sterile. Reagents were purchased from the following sources: Dulbecco's modified Eagle's medium (DMEM) with high Glucose 4.5 g/l, L-Glutamine and Sodium Pyruvate with L-Glutamine, 100 X penicillin and streptomyocin (1000 U/ml penicillin and 10 mg/ml streptomycin) and 1xTrypsin-EDTA solution from PAA Laboratories Ltd. (Somerset, UK); sterile dimethyl sulphoxide (DMSO) from Sigma-Aldrich Ltd (Dorset, UK); Foetal Bovine Serum (FBS) from Gibco (Invitrogen, Paisley, UK); Type I collagen (Rat tail) from First Link UK; glass slides from Menzel-Glaser (Braunschweig, Germany); glass coverslips from VWR International; all plastic-ware for tissue culture from Invitrogen (Paisley, UK), BD Biosciences (Oxford, UK) and Corning BV (Amsterdam, The Netherlands) and Trypan Blue Solution from Sigma-Aldrich (Dorset, UK).

II.III. Antibodies

Antibodies used in flow cytometry: PE-ICAM1 (Biolegend HA58), PE/Cy7-CD34 (Biolegend 581-BL), APC-PDPN (Biolegend NC-08), APC-CD31 (Biolegend WM59), APC-CD146 (Biolegend SHM-57), FITC-CD146 (Abcam PIHI2), BV421-CD45 (Biolegend HI30), BV605-ki67 (Biolegend Ki67), APC-IL8 (eBioscience BCH), APC-IL6 (Biolegend MQ2-13A5), AF700-α-SMA (R&D Systems ICI420N), APC-CD29 (Biolegend 152/16), Pe-IgG2a, k isotype (Biolegend, 400211), Pe-Cy7-IgG1, k Isotype (Biolegend, 400251) and PE-CD82 (Biolegend ASL-24). Antibodies use for immunohistochemistry and Immunofluorescence: α-SMA (Abcam A5228), PLA2G2A (Novus Bio. 620501), CXCL14 (Abcam ab36622), C1R (Sigma-Aldrich P00736), ITG-β1 (Abcam EP1041Y), β-Catenin (Abcam ab16051), MMP14 (Abcam EP1264Y), FAPB5 (Abcam ab84028), IL-11RA (Abcam ab109697), MMP11 (Abcam ab53143), Rabbit Isotype (IgG1,2a-b cocktail, Abcam), Murine Isotype (IgG1,2a-b cocktail, Abcam) and CD82 (Abcam TS82b)

II.IV. Cell culture and tissue digestion

Cells from Dupuytren's disease were isolated from α-SMA rich nodules as described previously(27, 55). Briefly, tissue samples were dissected into small pieces and digested in DMEM (Lonza) with Type I collagenase (Worthington Biochemical Corporation) + DNase I (Roche Diagnostics) for up to 2 h at 37 °C. Cells were cultured in DMEM with 5% (vol/vol) FBS and 1% penicillin–streptomycin at 37 °C in a humidified incubator with 5% (vol/vol) CO_2. Cells before passage 2 were used for experiments and only freshly isolated cells were used for single cell RNA-seq and flow cytometry.

II.V. Cell counting

Cells were counted using a haemocytometer (Improved Neubaur by Hawksley) prior to use in experiments. The number of cells counted in the sixteen-square area of the haemocytometer was multiplied by $1x10^4$ to give the total number of cells per ml. This figure was then multiplied by the volume (ml) in the culture to give the total number of cells.

II.VI. Cell viability

Cell viability was checked using Trypan Blue Solution. Trypan Blue Solution is a dye excluded from viable cells but taken up by cells with compromised cell membrane permeability. Uptake of Trypan Blue Solution is verified by acquiring cells under the microscope (Olympus ULWCD 0.30, Japan). Cells were diluted 1:1 with Trypan Blue Solution and placed in a haemocytometer for counting. The final count calculated was multiplied by 2 to allow for dilution in Trypan Blue Solution.

II.VII. Cell freezing and thawing

Cells were frozen in FBS supplemented with 10% DMSO in 1ml cryovials. Approximately 1-3 $x10^6$ cells per ml were frozen in cryovials and placed in a Mr Freezy Container (Thermo Fisher) at -80°C for 24 hours. After 24 hours cryovials were transferred to long term liquid nitrogen stores. To thaw, cells were warmed up rapidly in a 37°C water bath and washed in culture media to remove DMSO. Cells were counted, and viability checked using Trypan Blue Solution.

II.VIII. Single cell isolation

Following surgical resection, Dupuytren's disease samples were rapidly transported to the research facility. Tissue specimens were transported in complete medium and on average arrived at the research facility 4 hours following resection. On arrival, samples were then rinsed in PBS and the nodule isolated using a scalpel. Each nodule was minced using a scalpel and transferred to 5ml of digestion medium described above. Dissociated cells were then washed in 5% FBS DMEM (Gibco) and passed through a 100 μm cell trainer. This single cell solution was slowly frozen using in a Mr Frosty Container (ThermoFisher). Single cell suspensions were thawed on the day of sequencing, diluted at a concentration of 1000 cells/μL in 0.04% BSA/PBS for loading into 10X Chromium Single Cell A Chips.

II.IX. Droplet-Based single cell RNA-seq

Single cell libraries were prepared using the Chromium 3' v2 platform (10X Genomics, Pleasanton, CA) following the manufacturer's protocol. In brief, single cells were encapsulated into Gel beads in Emulsions (GEMs) in the GemCode instrument followed by cell lysis and barcoded reverse transcription of RNA, amplification, shearing and 3' adaptor and sample index attachment. Approximately 10,000 single cells were loaded on each channel and approximately 1,500-6,000 cells were recovered. Libraries were sequenced on the Illumina HiSeq 4000 (Paired end reads: Read 1, 26bp, Read 2, 98bp).

II.X. Computational analysis

Pre-processing

Sample de-multiplexing, alignment to the GRCh38 human transcriptome and UMI-collapsing were performed using the Cellranger toolkit (v2, 10X Genomics). Aspects of the downstream analysis were done in Seurat R package (Sajita Lab). I first performed quality control on each patient sample separately. I excluded poor quality cells that expressed fewer than 200 genes and those with over 10% of UMIs mapping to mitochondrial genes as well as cells that expressed over 4500 genes. After this, donors were aligned from the first and second batches. Alignment was implemented using the MergeSeurat() function. Batch correction was performed using Combat as implemented in the 'sva` R package using the default parametric adjustment mode(172). I employed a global-scaling normalization procedure as per(149) calling Seurat's LogNormalize() function. Briefly, the cell ranger UMI count values for each gene were divided by the sum of the total UMI counts per cell to normalize for differences in library size and then multiplied by a scaling factor that represented the median library complexity (10,000) producing TPM-like (Transcript-per-million) values. I then took the log transform of this procedure for downstream analysis.

Feature selection and dimensional reduction

Feature selection was first undertaken by defining highly variable genes using the FindVariableGenes() function *(n*=1,856 genes). This function provides an attractive visualization to select highly variable genes from a dispersion plot. Values were then centred and scaled before input to principal component analysis (PCA), which was implemented using the R function

'prcomp` from the "stats" R package. After PCA, significant PCs were identified using a permutation test implemented using the 'permutationPA' function from the 'jackstraw' R package(173). This test identified 20 significant PCs, and these were used as the input to further analysis. The jackstraw procedure is a permutation method that tests for association between the observed data and their systematic patterns of variations, constructing a null distribution for each principal component.

Clustering

To partition the data into clusters of transcriptionally related cells, I used unsupervised clustering based on the Louvain algorithm with the Jaccard correction implemented using the FindClusters() function. Initially, I used clustering to delineate major cell types and applied the Louvain algorithm on the *k*-NN graph of the entire dataset. Clusters were then annotated to know cell types using canonical marker genes and distinct cell types could be readily identified. In many of the patient samples and pooled dataset, I noted a small (<2%) population of proliferative stromal cells. After defining major cell types, I subsampled each (fibroblasts, myofibroblast, pericytes and endothelial cells) to identify potentially meaningful sub-clusters. This is important as it removes cell type variation from dominating the signal in reduced dimensional space. I applied the same steps described above to delineate sub-clusters including the selection of highly variable genes, PCA, defining significant PCs and graph-based clustering. As the average UMI count varied between distinct cell types, when analysing individual cell types (e.g., myofibroblasts), I performed a second round of quality control to remove doubles and contaminants. These were infrequent and represented < 1-0.5% of the total cell population. For doublet discrimination, I used a set of

canonical markers for each cell type and excluded those cells that shared markers of multiple cell types and those cells with a UMI count 3 standard deviations above the median cell-type UMI count. For visualization I used the "Barnes-hut" approximate version of t-distributed stochastic neighbour embedding (tSNE) using the 'Rtsne` R package. Although tSNE was not directly used to cluster the dataset, it provided a useful visualization tool.

Identifying the putative myofibroblast activation trajectory

On inspection of the PCA projection of the myofibroblast dataset I noted a strong association between A*CTA2 and* CD82^{+} myofibroblast marker expression with the first principal component (PC1). This was apparent in both pooled datasets and individual patient samples. To investigate this further, I extracted genes with top PC1 loading and again saw a strong negative correlation to ACTA2^{low} myofibroblast marker genes (e.g. *PLA2G2A* & *CXCL14*) and strong positive correlation to *ACTA2* expression and markers of the novel CD82^{+} population. Given this, I reasoned that this may represent the transcriptional profile of a putative differentiation or activation path. To visualize this, I defined a gene module of ACTA2^{low} myofibroblasts and CD82^{+} myofibroblast taking the top 100 genes correlated (Spearman) to *ACTA2* (as a marker for CD82^{+} myofibroblast as these showed high expression of *ACTA2*) and *CXCL14* (as a marker for ACTA2^{low} myofibroblasts as these showed high expression of *CXCL14*) based on average expression (UMI counts). Then, I plotted PC1 scores of each cell against the average expression of these gene modules that were adjusted with a control gene set and obtained the Pearson correlation coefficient. To account for differences in library complexity between cells I calculated a control gene set for each of these scores(174). This was selected by binning all genes in the dataset and randomly selecting 100 genes from each bin

containing every gene of the test gene set. Thus, the control gene set has a comparable distribution of expression levels to that of the *ACTA2* and *CXCL14* gene-sets and is 100-fold larger, such that its average expression is analogous to averaging over 100 randomly-selected gene-sets of the same size as the *ACTA2* and *CXCL14* gene-sets. This revealed a strong positive correlation to the *ACTA2* gene module (R=0.46) and a strong negative correlation to the fibroblast-like gene module (R=-0.71) along the first principal component, suggesting a strong transcriptome signature of this putative trajectory in the data.

Reconstructing the putative myofibroblast activation trajectory using diffusion maps

To assign myofibroblasts along the putative activation trajectory I first performed dimensional reduction using diffusion maps, implemented in the R package 'destiny`(175). Diffusion maps are a non-linear dimensional reduction method originally developed for signal processing in image analysis. Diffusion maps perform well in capturing biologically relevant structures in the data manifold and have been used extensively to explore potential developmental pathways or trajectories in single cell RNA-seq. The distance metric used was the Euclidean distance between pairs of cells in the reduced dimension space of the significant PCs (n=20), as defined above. In concordance with PCA, the first two diffusion components were highly associated with *ACTA2* expression. Therefore, I then fitted a principal curve (R package 'princurve', smoother ='lowess', f=1/3) through the first two diffusion co-ordinates(176). As the λ value of the curve reflects the arc-length from the beginning of the curve to each point I utilised this to assign each cell along a "Pseudotime" trajectory. Importantly, this modelled trajectory was focused on capturing myofibroblast differentiation associated with *ACTA2* expression but did not ascribe any

distinct cell type as a precursor. I used *ACTA2* expression to define a direction to the curve, as it is a directionless entity. In addition to this approach, I used the 'slingshot` R package to infer trajectories and align cells along developmental pseudo-time. Slingshot revealed a very similar trajectory structure with a path between $ACTA2^{low}$ and $CD82^{high}$ myofibroblasts. Finally, to model gene expression along the ordered pseudotime, gene expression was smoothed using generalised additive models implemented by the 'gam` function in the 'mgcv` R package. This function fits a *loess* regression to each gene along the pseudotime trajectory.

Gene Ontology

GO enrichment of cluster markers and differentially expressed genes was performed using the R package 'clusterProfiler'(177) with a Benjamini-Hochberg multiple testing correction and a false-discovery rate cut-off of 0.1, using all genes expressed in >3 cells as background. Visualization was performed using the R packages 'ggplot2` and 'igraph`. For each cell type I selected genes > 0.5 log fold change and an adjusted p-value < 0.1 unless otherwise stated. For gene ontology analysis I used Reactome and Hallmark pathways access in the 'clusterProfiler' package. For gene set enrichment (GSEA), I created a rank list using log fold change and implement the procedure using the 'fgsea' R package.

II.XI. Flow cytometry

After tissue dissociation into a single cell suspension, cells were first stained with a panel of fluorescently labelled antibodies to surface antigens, washed with FACS wash buffer (1% BSA, 0.01% NaN_3 in PBS) then fixed using CytoFix (eBiosciences Foxp3 staining buffer set # 00-5523-00)

for a minimum of 30 min at 4^{o}C. After three washes in perm wash (eBioscience Foxp3 staining buffer), intracellular antigens were stained with another panel of fluorescently labelled antibodies, and cells were then washed with perm wash and analysed by flow cytometry (BD LSR Fortessa X20) and FlowJo software. BD comp beads (anti-mouse Ig k #552843) were used to establish compensation settings. Dead cells were deselected using live/dead stain, added to surface staining panel prior to cell fixation (Life Technologies Live/Dead^{TM} Near IR fixable dead cell stain kit #L10119).

II.XII. FACS Sorting

Dupuytren's myofibroblasts were stained without prior magnetic bead enrichment into CD82^{high} and CD82^{low} based on the top and bottom 20% expressing cells, respectively. Samples were filtered through a 70µm strainer before sorting commenced. Single cell sorting was performed using a FACSAria (BD Biosciences). After doublet exclusion, isolated single cells were sorted into 5ml FACS tubes containing 1ml of complete medium (5%FBS DMEM + 1% pen/strep). Cell purity was assessed for each sorted population and was approximately 95%. Sorted stromal cell populations were seeded onto hydrogels (Collagen Type 1 coated, 2.47 kPA) for traction force microscopy after 24 hours in culture.

II.XIII. Immunohistochemistry

Dupuytren's tissue samples were fixed with 4% paraformaldehyde in PBS for 20 min, longitudinally bisected, and embedded in paraffin wax. 7µm sections were cut using a microtome and processed for immunohistochemistry(55). Sequential sections were stained with mouse monoclonal

antibodies. Antibodies were detected using a two-staged polymer enhancer system (Sigma). Murine or Rabbit IgG isotypes at the same protein concentration as the monoclonal antibody solution were used as a control. Images were acquired with an Olympus BX51 Microscope (Olympus).

II.XIV. Immunofluorescence and confocal microscopy

Dupuytren's tissue samples were fixed with 4% paraformaldehyde in PBS for 20 min, longitudinally bisected, and embedded in paraffin wax. 7µm sections were cut using a microtome and processed for immunofluorescence(55). Then, the tissue sections were stained with antibodies listed above followed by incubation with Phalloidin-AF488 (Life Technologies). Nuclei were counterstained with DAPI (4, 6-diamidino-2-phenylindole; Sigma-Aldrich) and mounted using Prolong ™ Gold anti-fade (Life Technologies). Fluorescent images were obtained with a confocal system (Zeiss LSM 710).

II.XV. Apoptosis assay

The Annexin-V-FITC apoptosis assay (eBiosciences) was performed as described by the manufacturer. Briefly, cells were washed in FACS buffer (PBS+1%FBS) and stained with Annexin-V-FITC for 30 minutes at room temperature. After this, cells were washed again in FACS buffer and stained with Propidium Iodine and run without washing on a BD Fortessa. Data analysis was performed using FlowJo software.

II.XVI. Acid cleaning of glass coverslips

This protocol involves hazardous chemicals and so masks and gloves were worn at all times. 20 ml of 99.99% sulfuric acid (Sigma 339741) was added to a 500ml glass beaker. To the same beaker, 10 ml of 30% (wt/wt) hydrogen peroxide solution (Sigma H1009) was added slowly. Coverslips were placed in this solution one by one, ensuring that all were completely submerged. After gentle agitation, the coverslips were incubated at room temperature (20 °C) for 1 h, agitating periodically. The coverslips were then washed three times in double distilled H20 (ddH20) and placed in a beaker that was covered with Parafilm. These could then be stored for up to two weeks.

II.XVII. Preparation and functionalization of polyacrylamide (PAA) gels

Ammonium percusulfate (APS) solution: 10 % APS solution was prepared by dissolving 1 g of APS (Sigma A3678) in 10 ml of ddH20. This was then frozen in 50 µl single use aliquots and store at -20 °C for up to 1 month.

Fluorescent bead solution: Dilute bead solutions were prepared by diluting the original manufactures solution in ddH20. For the 40 nm red fluorescent beads (Fluorospheres Carboxylate modified Microspheres, 40 nm, red fluorescent (580/605) F8793) a vol/vol dilution of 1 in 5,000 is adequate. To achieve this dilution, 1 in 100 dilution was made first by adding 10 µl of the concentrated fluorescent bead solution to 990 ml of ddH20. Subsequently a further 1 in 50 dilution

was performed and before each dilution step beads were sonicated for 10 minutes in an ultrasonic bath.

Polyacrylamide solution: The relative concentrations of acrylamide (Bio-Rad 1610140) and bisacryalmide (Bio-Rad 1610143) determine the gel stiffness. During my experiments I varied the concentrations of each to obtain the desired stiffness (0.67, 1.24, 2.55, 6.4, 8. 25KPa)

Sulfo-SANPAH crosslinker solution: Sulfo-SANPAH (Sigma 803332) was dissolved in 200 µl of DMSO, creating a 40 mg/ml solution. This was frozen as 40 µl single use aliquots and store at -80°C.

II.XVIII. Traction force microscopy

Polyacrylamide (PAA) hydrogels for traction force microscopy were prepared as previously described(178). Briefly, acid washed glass coverslips (18mm) were incubated with poly-L-lysine (10 µg/ml in H20) for 30 mins at 4^{o}C, then coated with 0.04 µm carboxylate-modified red FluoroSpheres (1:5000 in ddH20, Invitrogen). PAA gel formation was initiated with APS (10% solution in ddH20, Sigma) and TEMED (Sigma). Polymerized PAA gels were functionalized with sulfo-SANPAH (Invitrogen) and coated with Type 1 Collagen (200ug/ml Rat Tail, Thermo Fisher). The Young's modulus of the PAA gels were 2.55±0.5 kPa. Cells could adhere to the gel for 4-6 hours before image acquisition. Images were captured using an x40 Plan Apo objective on a Zeiss LSM 710 Confocal microscope and experiments were performed at 37 °C and 5% CO_2 in Phenol red free DMEM (Thermo Fisher) in a microscope stage incubation chamber. Bead positions were acquired before and after the addition of trypsin which removed cells from the gel surface. Bead

displacement was tracked with an ImageJ PIV plugin(179) and cellular forces reconstructed using a FTTC algorithm also implemented in ImageJ(88, 179).

Downstream analysis and visualization were performed using MATLAB (Mathworks) and R (R Version 3.5). Single cell force maps of $CD82^{high}$ and $CD82^{low}$ sorted cells were first normalized per cell area and scaled before input to PCA implemented in the "prcomp" function. Significant components were determined as described above for the single cell expression data using visualization of a scree plot and/or the Jackstraw procedure. Significant PCs were used as input for tSNE clustering using the Barnes-Hut implementation in Rtsne() function.

II.XIX. Atomic force microscopy

Mechanical measurements were obtained using an MFP-3D Asylum Research atomic force microscope (Oxford instruments). Fresh tissue samples from Dupuytren's disease were dissected to obtain an approximately 10 µm cube from the center of the nodule. This was then embedded in OCT and snap frozen in liquid nitrogen. Using a cryostat, 30 µm slices were cut and mounted on to a glass slide. Tissue stiffness was measured through micro-indentation using a 5 µm spherical tip probe (NovoScan, Oxford Instruments) and the cantilever spring constant was calibrated using the thermal fluctuation method. Measurements were taken from a 10 x 10 µm squares with sample in ddH20 at room temperature. Elastic modulus was calculated using Hertz model from force indentation profiles using MFP-3D software (Oxford Instruments).

II.XX. Transfections

For transfection experiments, cells were seeded at 2 x 10^5/6 well plate and the following day transfected with siRNA against CD82 (AM16708) or non-targeting control oligo at 20nM (Life Technologies). All transfections were performed with Dharmafect 1 (GE Healthcare) and Optimem™ (Life Technologies) according to the manufacturer's instructions. Media was replaced two hrs later with 5% FBS phenol red free DMEM (Gibco). Transfection was repeated after three days and experiments were performed at day six post transfection.

II.XXI. Bulk RNA-seq library preparation

Passaged 1 DD myofibroblasts from seven donors were transfected with siRNA against *CD82* (AM16708) or non-targeting control oligo at 20nM (Life Technologies). RNA was extracted using Direct-Zol™ RNA kit. Library preparation was started with the NEBNext® Poly(A) mRNA Magnetic Isolation Module followed by the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina with the NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) and (Index Primers Set 2). Concentrations of each library were determined using the NEBNext® Library Quant Kit for Illumina® and RNA quality assess using the High Sensitivity D1000 Screentape Bioanalyzer (Agilent).

II.XXII. Bulk RNA-seq data analysis

FASTQ files were assessed using FASTQC followed by the generation of TPM values with kallisto v0.42.4(180). TPM values were summed to obtain gene-level expression values using the 'tximport' R package and differential expression analysis was undertaken using the 'DeSEQ2' R

package(142). GO enrichment of differentially expressed genes was performed using the R package 'clusterProfiler'(177) with a Benjamini-Hochberg multiple testing adjustment and a false-discovery rate cut-off of 0.1. Visualization was performed using the R packages 'ggplot2` and 'igraph`.

II.XXIII Cell cycle assay

The cell cycle assay used was the Vibrant DyeCycle (ThermoFisher) Violet stain kit, performed as per the manufacturer's instructions. Briefly, cells were washed before staining with the violet dye (Vibrant DyeCycle Violet Stain) for 30 minutes and run on a BD Fortessa. Data analysis was performed in FlowJo software.

Chapter III. Towards a single cell atlas of human fibrosis

III.I. Introduction

Fibrotic lesions are complex and composed of diverse cell types(1, 25, 141). The proposed key effector cell in fibrosis is the myofibroblast, defined by the expression of α-SMA (45, 54, 181, 182). Alongside myofibroblasts, the fibrotic microenvironment contains a myriad of cell types, including pericytes(183, 184), immune cells(27, 165) and endothelial cells(185, 186). Each of these populations have been studied in fibrosis, but we are missing a precise molecular profile of a complete fibrotic ecosystem. Knowing the transcriptomic signatures of distinct cell types, and whether they house subpopulations, will be crucial to uncover fundamental mechanisms driving fibrosis.

As fibrosis can affect almost every organ in the human body there are a wide array of potential models to study this process. Dupuytren's nodules provide a valuable model for two main reasons. Firstly, nodules are a disease ecosystem rich in cells, containing high numbers of myofibroblasts in addition to pericytes, immune cells and endothelial cells(55). Secondly, surgical resection of nodules mandates an *en bloc* excision, avoiding the need to profile small biopsy samples that may fail to capture a representative population. Together, these two properties make Dupuytren's nodules an important system to study fibrosis using single cell RNA-seq.

III.II. Hypothesis and Aims

My hypothesis is that Dupuytren's nodules harbour diverse cell types and single cell sequencing will provide an appropriate tool to characterize this cellular heterogeneity.

In this chapter, I first aim to undertake broad phenotyping of the cellular landscape in Dupuytren's nodules using single cell RNA-seq. My initial goal was to build a single cell atlas of major cell types, outlining key computational approaches. These bioinformatic methods build upon those described in Chapter II.X and are vital for appropriate biological interpretation of the data. After building this single cell atlas, I sought to perform detailed analysis of each cell type alongside experimental validation. This chapter describes the start of this overarching goal and includes an investigation of immune, endothelial and perivascular cells in DD nodules.

III.III. Results

III.III.I *Patient demographics and data quality control*

In total, 12 nodules were processed in two batches of six patients as described in Chapter II.II.VIII. Surgery was performed between 2016-2017 at two sites located in Derby and Edinburgh. Before sample collection written informed consent was obtained from all patients. The mean age of patients was 63 years old (range 42-69); ten were male and two were female. Patient samples were transported to Oxford and processed immediately to reduce potential artefact from RNA degradation. During sample processing, nodules were located and excised from the diseased tissue. This was performed to isolate this structure for downstream processing and analysis (see Chapter II.II.VIII).

After formation of the single cell suspensions (see Chapter II.II.VIII), library preparation was undertaken at the Wellcome Trust Centre for Human Genetics, Oxford, as described in Chapter II.II.IX. Library preparation used the 10x Genomics Chromium platform to prepare single cell libraries and sequencing was performed on a HiSeq 4000 (Illumina). After upstream processing (see Chapter II.II.X), quality control was undertaken to identify low quality cells that may impact downstream analysis such as clustering and differential expression. Three crucial metrics are the number of genes (nGene), number of counts (nUMIs) and the percentage of reads mapping to mitochondrial genes per cell(187). These metrics were visualized using violin plots and mapped to clusters following tSNE embedding (Figure 1-2). After this, low quality cells were removed before downstream analysis (Chapter II.II.X).

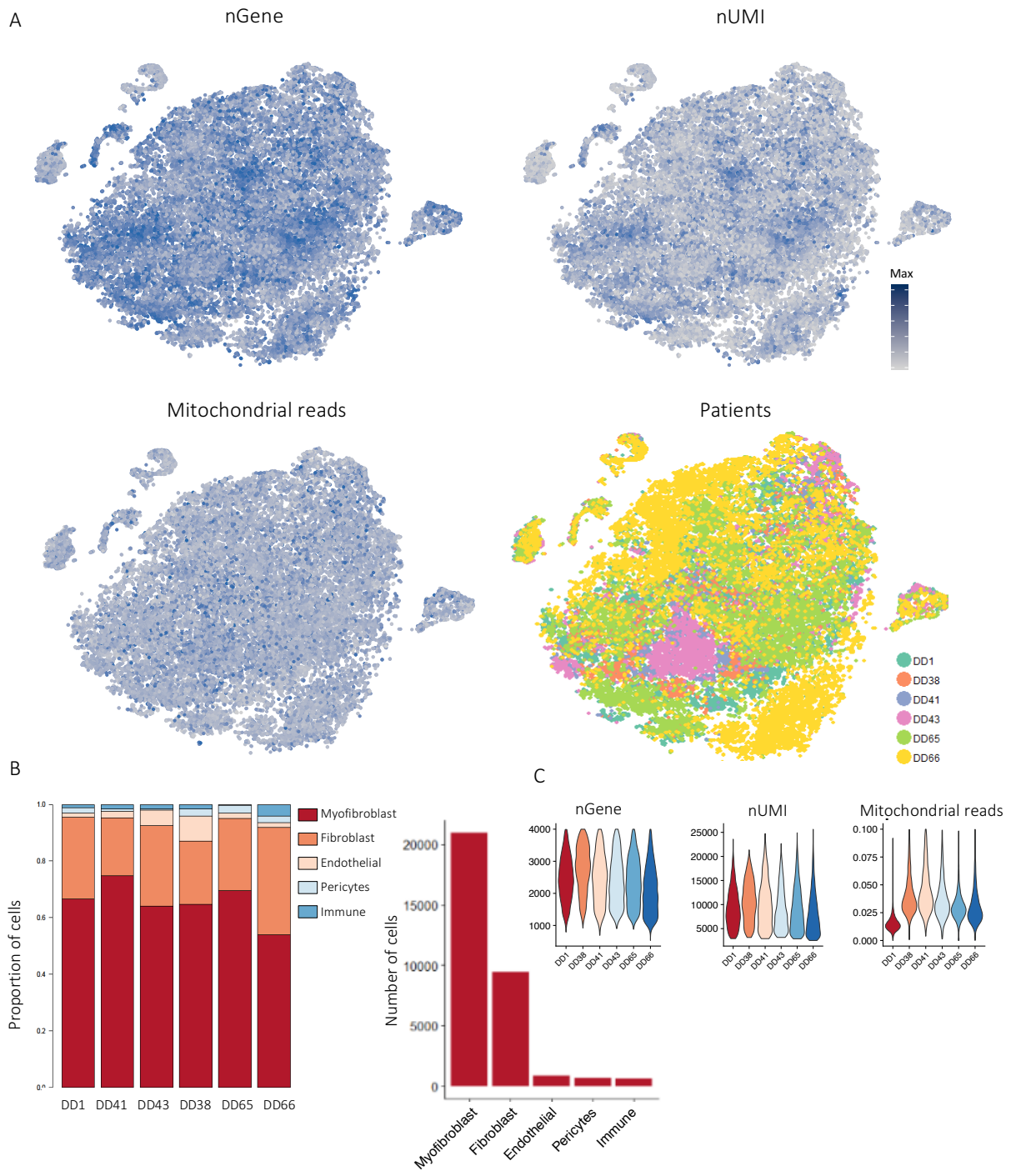


Figure 1. **Quality control for single cell RNA-seq of first patient batch**. A) tSNE projections showing the number of genes and UMIs, percentage of mitochondrial reads and patient donors for 36,864 single cells from the first six patients. B) Bar plots showing the proportion and number of each cell type from the first six patients. C) Violin plots showing quality control metrics in first six patients. X axis labels patient donor. Y axis in log(UMI+1) for nGene and nUMI and percentage of cells for mitochondrial reads. nUMI is number of UMI per cell, nGene is number of genes per cell and scale in (A) is z-scored values. DD is prefix for DD donor IDs (*n*=6). nUMI is number of unique molecular identifiers.



Figure 2. **Quality control for single cell RNA-seq of second patient batch**. A) tSNE projections showing the number of genes and UMIs, percentage of mitochondrial reads and patient donors for 7,332 single cells from the second six patients. B) Bar plots showing the proportion and number of each cell type from the second six patients. C) Violin plots showing quality control metrics in second six patients. X axis labels patient donor. Y axis in log(UMI+1) for nGene and nUMI and percentage of cells for mitochondrial reads. nUMI is number of UMI per cell, nGene is number of genes per cell and scale in (A) is z-scored values. DD is prefix for DD donor IDs (*n*=6). nUMI is number of unique molecular identifiers.

Each patient sample showed variation in quality control metrics, including percentage of mitochondrial reads (Figure 3). This could relate to technical artefact introduced during single cell isolation or library preparation or may reflect meaningful biological variation. In single cell RNA-seq, the percentage of mitochondrial genes provides a readout for dead or dying cells. I maintained a conservative filtering approach and removed cells expressing over 10% mitochondrial reads (see Chapter II.II.X). In total, I collected high quality profiles of 44,196 cells from 12 patient donors.

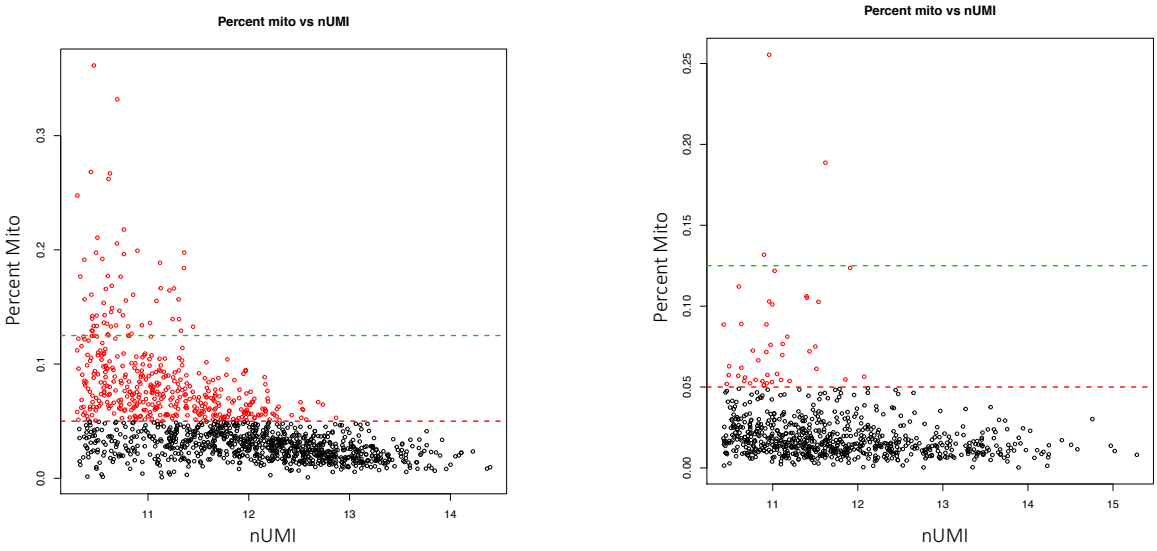


Figure 3. **Reads mapping to mitochondrial genes in Dupuytren's patients.** Scatterplots for number of UMIs (nUMI) in $\log_2$ scale and percentage of mitochondrial reads (Percent Mito) in two representative Dupuytren's patients (DD38 &DD47). Red line at 5% and green line at 12.5% mitochondrial reads. Red points indicate cells expressing > 5% mitochondrial reads. nUMI is number of unique molecular identifiers in log(UMI+1).

III.III.II *Feature selection in single cell RNA-seq*

High dimensional datasets suffer from the so-called "curse of dimensionality", representing computational challenges faced in extracting representative information in high-dimensional space(188). Single cell RNA-seq measures the gene expression of thousands of genes in tens of thousands of cells and thus presents many obstacles for appropriate data analysis. In machine learning, feature selection describes the process of extracting key features in a dataset that will contribute most to the predicted output of interest(189). In single cell RNA-seq, we select a subset of features (i.e. genes) exhibiting high variability across cells that likely hold major sources of biologically meaningful information. Highly variable genes are then used as input for downstream analysis, including dimensional reduction and clustering (Figure 4). Initially, I apply these steps before identifying major cell populations in DD.

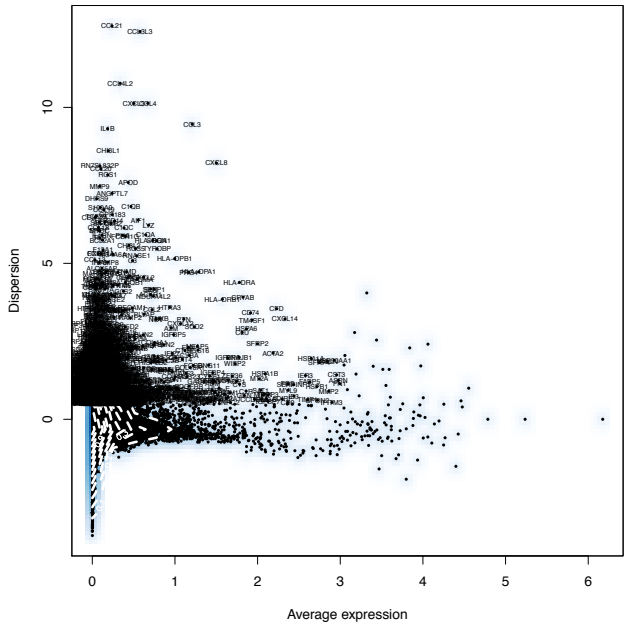


Figure 4. **Selecting highly variable genes in single cell RNA-seq.** Dispersion plot illustrating highly variable genes in Dupuytren's dataset. For each gene the average z-score expression is plotted against a calculated dispersion metric. Dispersion represents the mean-stabilized variance of each gene (z-score value). Each point represents a gene and gene labels are included if the gene is selected as a highly variable. Inspection of the plot allows visualization of the expression and dispersion characteristics of genes selected.

Principal component analysis (PCA) is a commonly used linear dimensional reduction method that aims to capture prominent sources of orthogonal variation in a dataset(190). In single cell biology, PCA provides a useful visualization of the data in two or three dimensions. In addition, it can help identify genes driving major sources of variation that may have been missed during variable gene selection.

PCA demonstrated that major sources of variation represented the difference between cell types in DD (Figure 5). For example, along the first two principal components (PC), immune and endothelial cells were highly distinct as compared to other stromal cells (Figure 5). Moreover, endothelial and immune cell marker genes dominated the gene loadings for the first few principal components (Figure 6). These results highlight the use of PCA in distinguishing transcriptomically distinct cell populations in complex systems, but also show why each cell type must be analysed individually to prevent cell-cell differences dominating the signal (see Chapter II.II.X). In other words, clustering the entire dataset based on PCA scores (see Chapter II.II.X) would abrogate the ability to discern subpopulations as the signal input would be dominated by cell to cell variation.

Despite the widespread use of PCA, its limitations make application to single cell clustering challenging(191). As a linear dimensional reduction technique, it often fails to capture complex non-linear structures in the data. Given this, other methods such as graph-based clustering have gained prominence in partitioning single cell expression profiles into discrete communities(157, 192). Before applying graph-based clustering, I selected significant PCs for downstream analysis

using visualization of a scree plot and a permutation method (Figure 7-8) described in Chapter II.II.X.

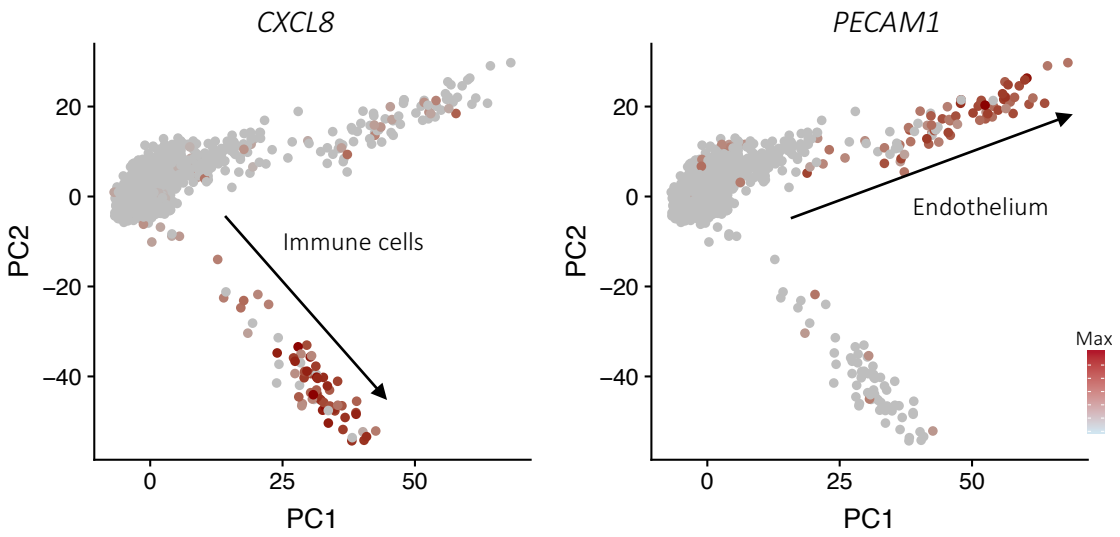


Figure 5. **Principal component analysis delineates major cell types in Dupuytren's nodules.** Scatterplots of PCA projections from representative DD patient (DD65) along first two principal components coloured by the expression of *CXCL8* and *PECAM1*. PC = principal component. Scale bar represents scaled expression in log(UMI+1).

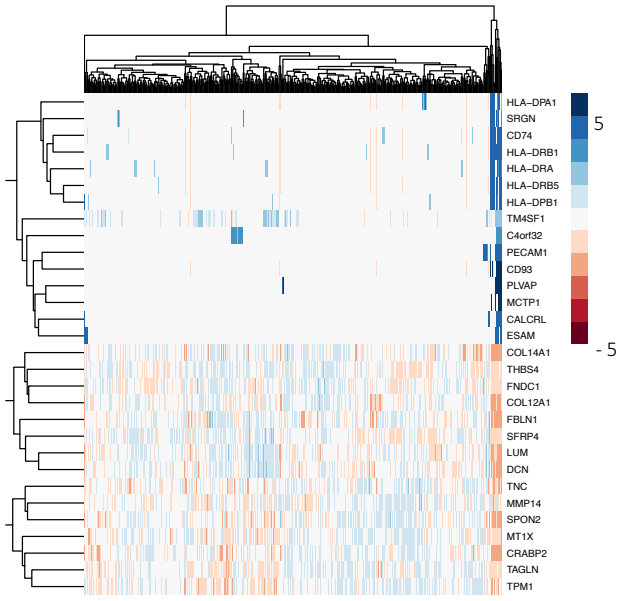


Figure 6. **Principal component 1 captures immune and endothelial signatures in Dupuytren's nodules.** Heatmap of z-score mean expression of top 30 genes along the first principal component. Scale bar represents z-score expression values.

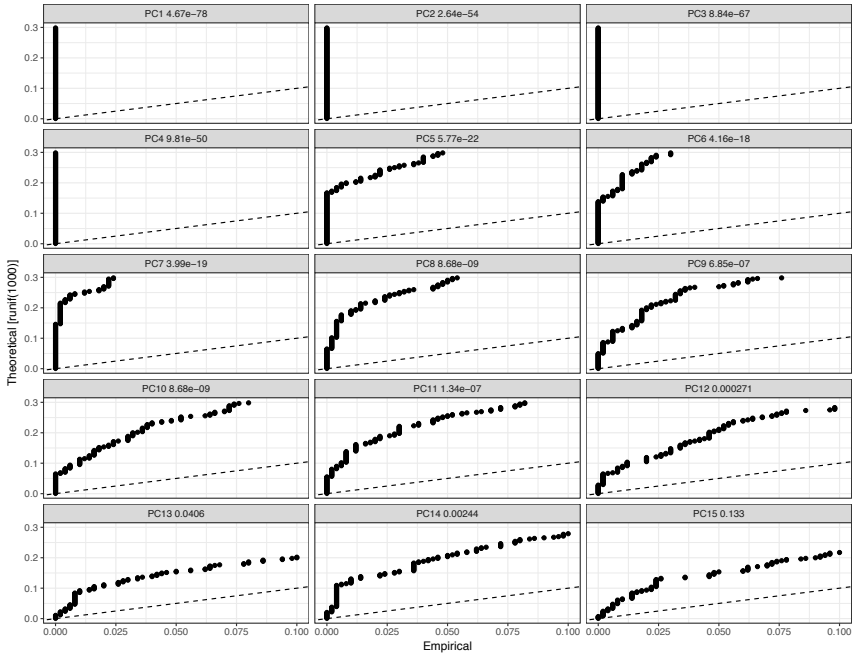


Figure 7. **JackStraw procedure for defining significant principal components.** Scatterplots of JackStraw permutation method for defining significant principal components by constructing a null distribution for each component. PC = principal component. Empirical represents null distribution of PC and theoretical is the measured variance explained by each PC in the single cell RNA-seq data.

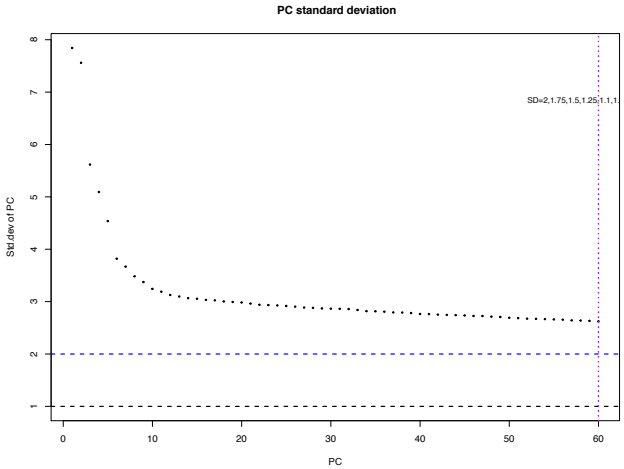


Figure 8. **Scree plot for principal components.** Scatterplot showing eigenvalues (percentage of variance explained) for each principal component. SD = standard deviation and PC = principal component.

III.III.III *Clustering cells in Dupuytren's nodules*

After defining significant principal components using the Jackstraw procedure and inspecting the scree plot (Figure 7-8), these were used as inputs for clustering. Building PCA based meta-signatures improves clustering by simultaneously encoding within each cell major sources of variation whilst reducing noise. I employed a popular graph-based clustering routine as described in Chapter II.II.X(193). First, PCA meta-signatures was assembled into a *k*-Nearest Neighbour (*k*-NN) graph. This graph was then refined using the shared local neighbourhoods between cells (Jaccard Distance). Finally, this graph was partitioned using the Louvain algorithm that aims to optimize a modularity parameter in the graph structure(194) (Figure 9).

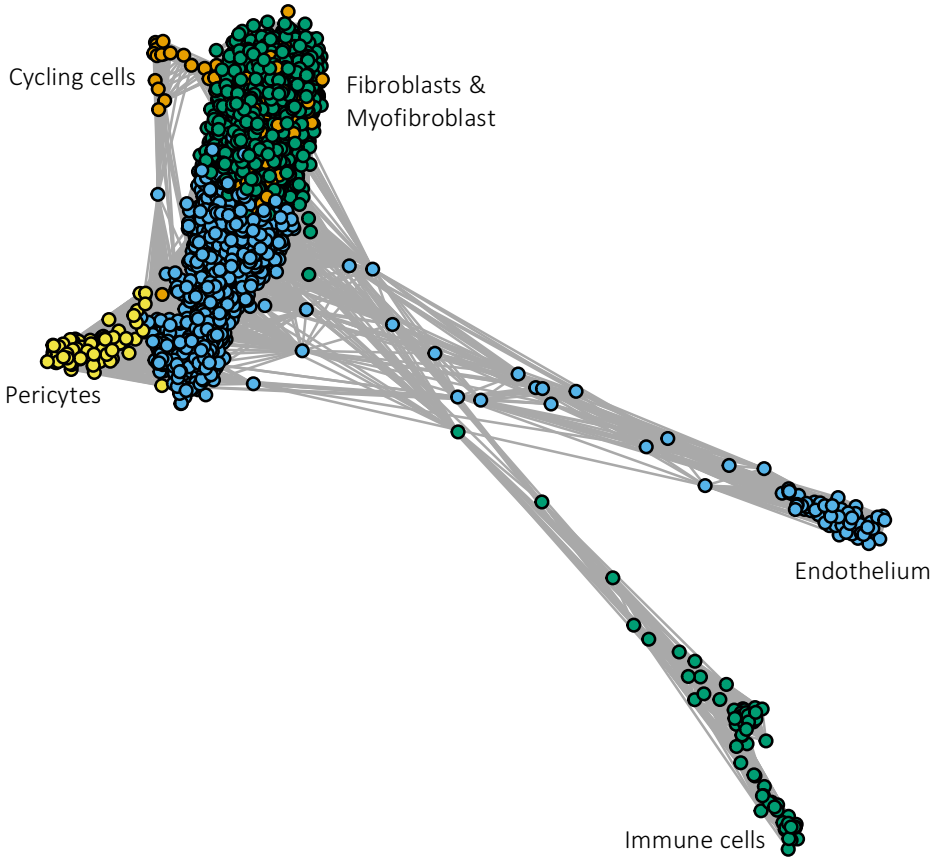


Figure 9. **Shared nearest neighbour graph of cell types in Dupuytren's nodules**. Shared nearest neighbour graph of entire dataset from twelve DD patients, coloured by assigned Louvain clusters. *k*=5,000 cells (Down sampled for visualization only).

To investigate heterogeneity within DD nodules, I applied unsupervised graph-based clustering to the entire dataset. This partitioned nodular cells into five clusters representing distinct cell types (Figure 10). Cluster 1 was assigned as fibroblasts and myofibroblast and showed high expression of genes associated with matrix protein homeostasis, including *COL1A1, COL3A1* and *FN1.* In addition, several genes involved in myofibroblast biology, such as *ACTA2* and *TPM1*, were also highly expressed in this cluster. This was the major population, making up ~80% of total cells in the nodules.

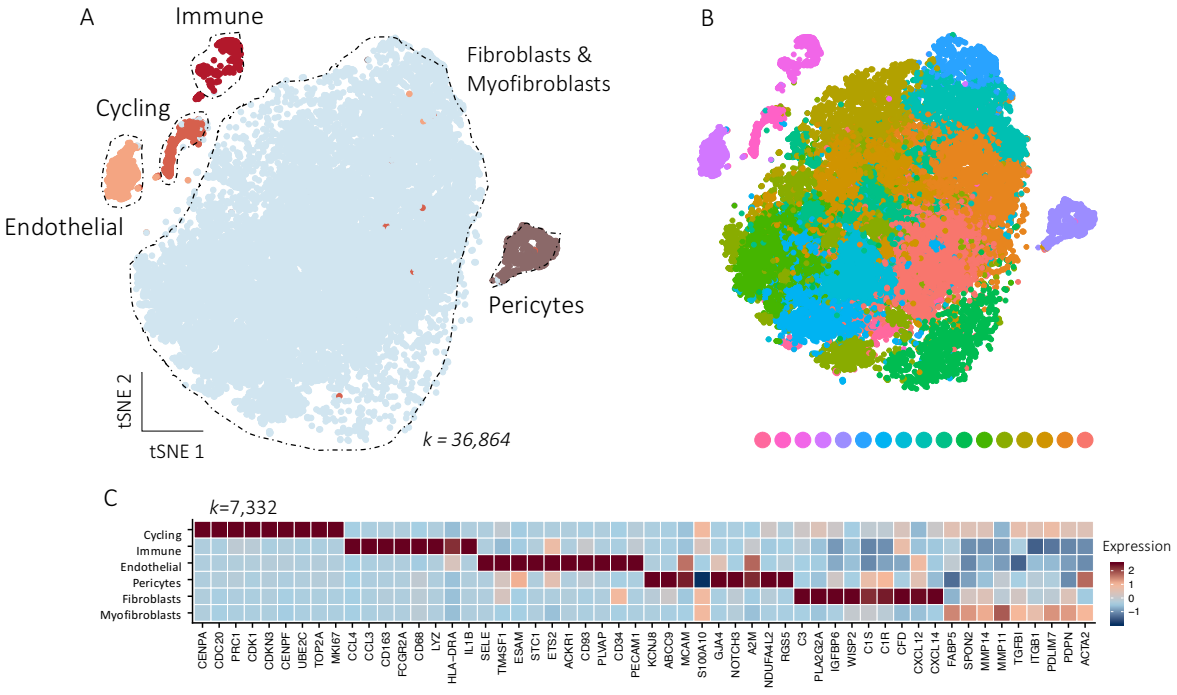


Figure 10. **Single cell RNA-seq resolves the complex ecosystem in Dupuytren's nodules**. A) tSNE projections for 36,864 single cells in Dupuytren's nodules from six DD patients (first batch) showing major cell types (A) and Louvain clusters (B). C) Heatmap of z-score normalized mean expression of selected cell type marker genes for 7,332 single cells from 6 patients (second batch).

The second largest cluster represented perivascular cells. This population showed high expression of many general stromal cell markers, such as *PDGRB*, but also had a unique set of specific genes that included *RGS5* and *NDUFA4L2. RGS5* encodes the regulator of G protein signalling 5 (RSG5) protein that functions to modulate GPCR signalling(195). RSG5 has previously been shown to be expressed by pericytes and may function to regulate endothelial-pericyte communication *in vivo*(196).

The third largest cluster represented endothelial cells, marked by high expression of *PECAM1*. In each patient sample there was a discrete endothelial cluster and other markers of this population included *ICAM1* and *TIE2*. Sub-clustering the endothelial population resulted in two transcriptomically similar subsets (Figure 11), with few differentially expressed genes between the two clusters. One cluster showed slightly higher expression of chemokines (e.g. *CXCL8*) and markers of endothelial activation (e.g. *HLA-DRA*) (Figure 11).

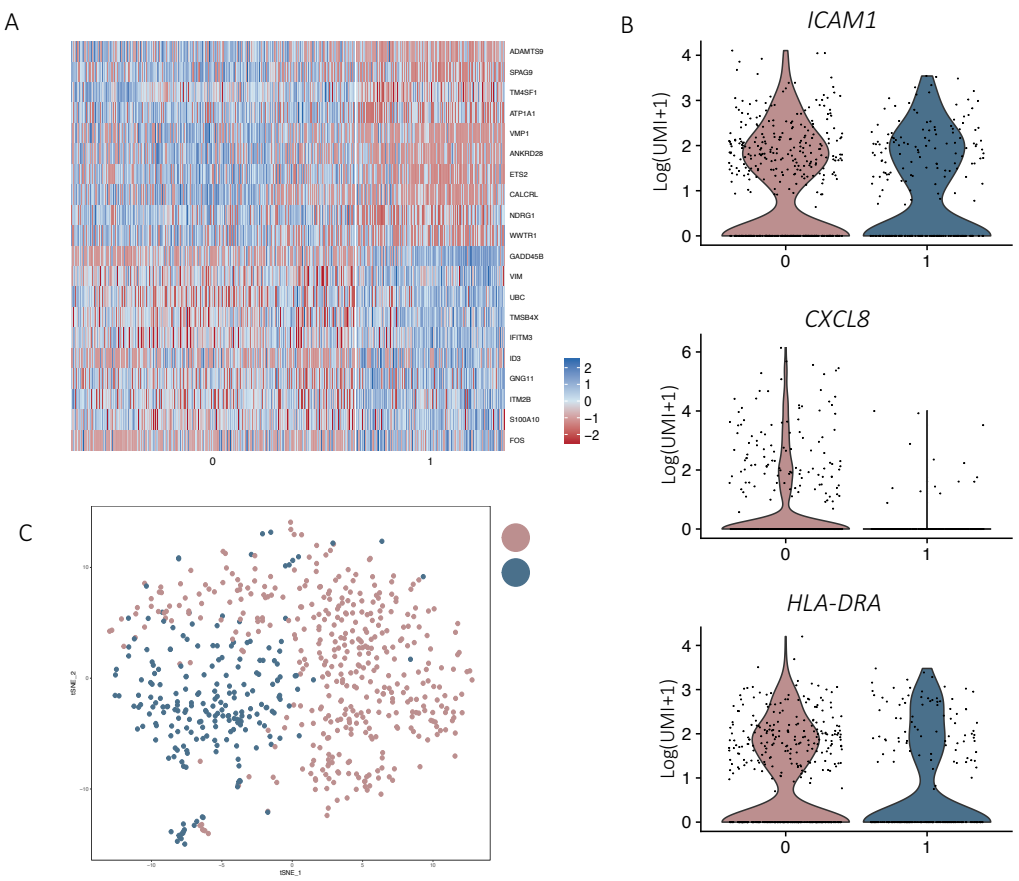


Figure 11. **The endothelial compartment in Dupuytren's nodules.** A) Heatmap of z-score mean expression of marker genes for two endothelial cell clusters (0-1). Scale represents z-score expression. B) Violin plots showing the expression of selected marker genes in two endothelial clusters (0-1). Expression in in log(UMI+1). C) tSNE projection of endothelial clusters showing two loosely demarcated clusters. Colour correspond to cluster ID in (B). *k*=2,432 endothelial cells and *n*=12 DD patients.

Within most DD patients I also noted two further smaller clusters, one composed of cycling stromal cells and the other immune cells. The immune cells will be further characterized in the final stages of this chapter and the cycling stromal cells in Chapter V. Both represented minor populations in the data making up around 1-2% of total nodular cells.

To complement the gene expression data, I performed flow cytometry (see Chapter II.XI) on freshly disaggregated DD nodular cells to validate major cell populations, alongside fibroblasts and myofibroblasts. In concordance with the single cell RNA-seq data, the major cell populations delineated were pericytes ($CD146^+$), endothelial cells ($CD31^+$) and immune cells ($CD45^+$) (Figure 12).

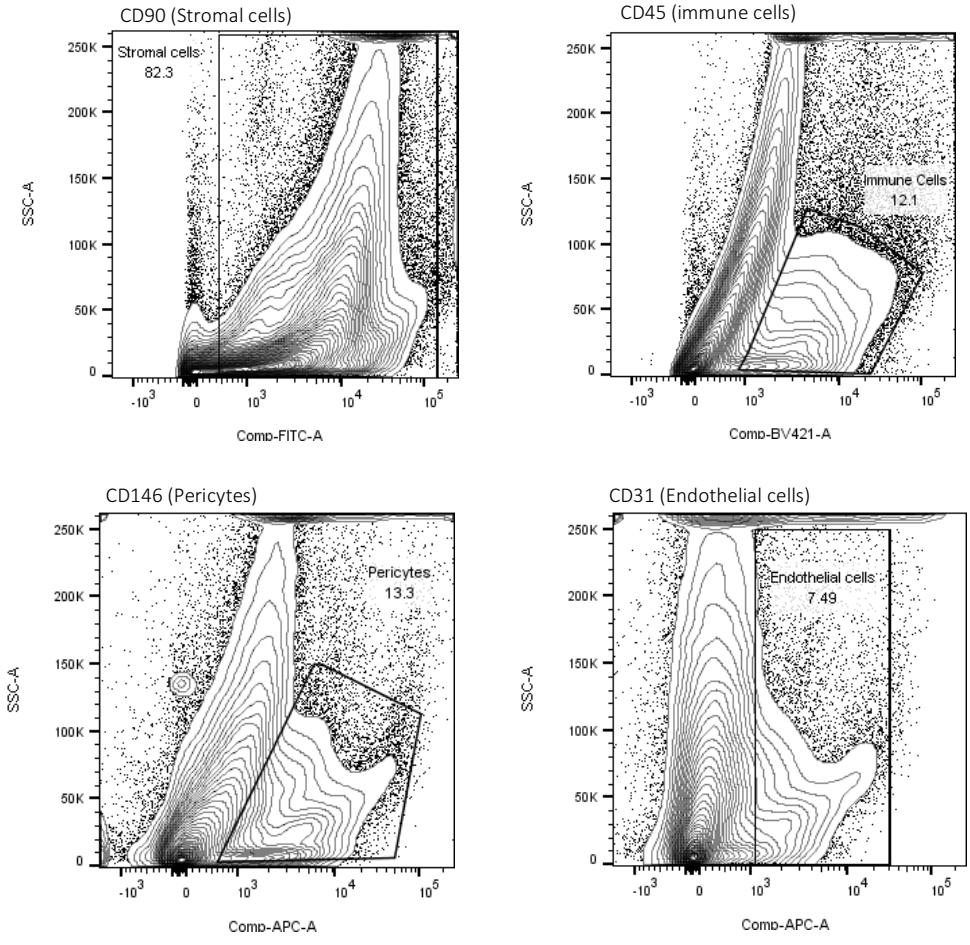


Figure 12. **Validation of major cell types in freshly disaggregated Dupuytren's nodules.** Representative density plots of flow cytometry analysis of freshly disaggregated cells in nodules. Stromal and immune populations are labelled, and data is representative of eight independent DD donors. Markers for each population were as follows; CD90 for stromal cells, CD45 for immune cells, CD146 for pericytes and CD31 for endothelial cells. Comp = compensated.

III.III.IV *Exploring the perivascular niche in human fibrosis*

Pericytes have been extensively studied to analyse their contribution to myofibroblast formation and fibrosis(197-200), but questions remain with regards to what differentiates pericytes from resident fibroblasts and myofibroblasts. In DD, I wanted to explore this population in greater detail, first performing differentially expression analysis. Significant markers of the pericyte cluster were *RGS5, NDUFA4L2, COL18A1, SLN, MCAM* and *NOTCH3* (Figure 13), with many of these genes associated with pericytes in several organs(201). Of note, pericytes also showed high expression of genes encoding contractile proteins, such as *TPM2, ACTA2* and *MYH11*. These findings support the role of pericytes modulating vascular homeostasis through direct cell-cell communication and the transduction of contractile force.

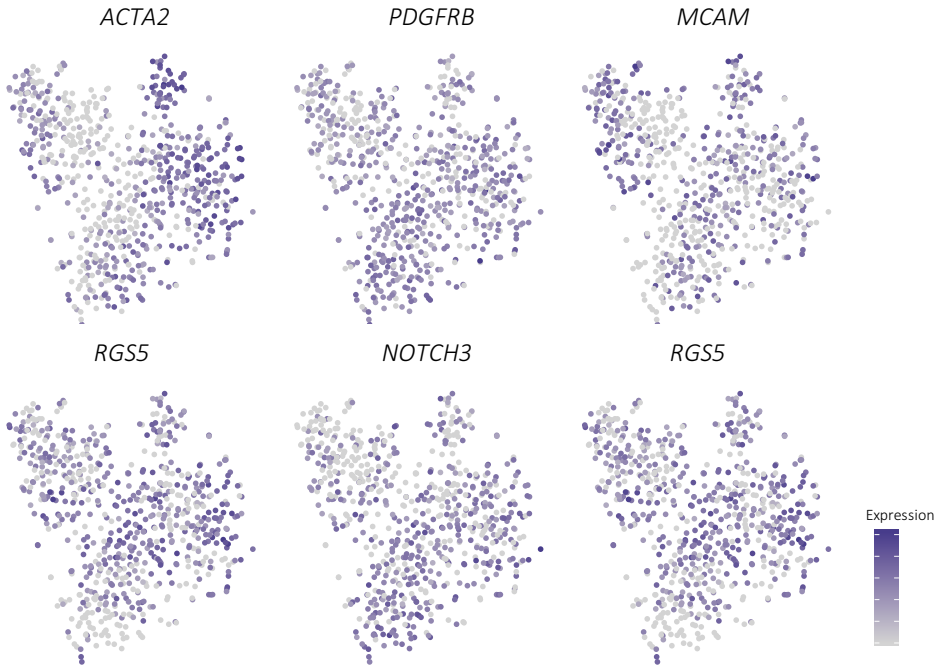


Figure 13. **Gene markers of pericytes in Dupuytren's nodules.** tSNE projections for selected pericyte markers. Expression in scaled log(UMI+1). *MCAM* was widely expressed throughout pericytes and was used to isolate this population during flow cytometry. *k*=*7,323* pericytes and *n*=12 DD patients.

Next, unsupervised clustering of pericytes revealed four distinct clusters (Figure 14A-B). Many clusters shared markers, but each expressed a unique set of genes. Cluster 1 (pericyte 1) showed high expression of *THY1* in addition to genes encoding collagens such as *COL1A1* and *COL1A2*. Cluster 2 (pericyte 2) expressed high levels of *LDHA, EIF1B* and *CYCS*. These markers suggest a metabolically active pericyte population. Cluster 3 (pericyte 3) shared several marker genes with Cluster 1, but was distinguished by *SLN*, *FN1* and collagen encoding genes, including *COL6A3*. *SLN* encodes sarcolipin, a Ca^{2+}-ATPase modulator first described in smooth muscle cells(202). The smallest cluster (pericyte 4) demonstrated high expression of many genes encoding contractile proteins, including *ACTA2,* with other markers including *MYL11* and *JUN*. Finally, I validated a discrete perivascular population in freshly disaggregated DD nodules using flow cytometry, showing that pericytes made up ~6% (2.4-8.4%) of total cells isolated (Figure 14D).

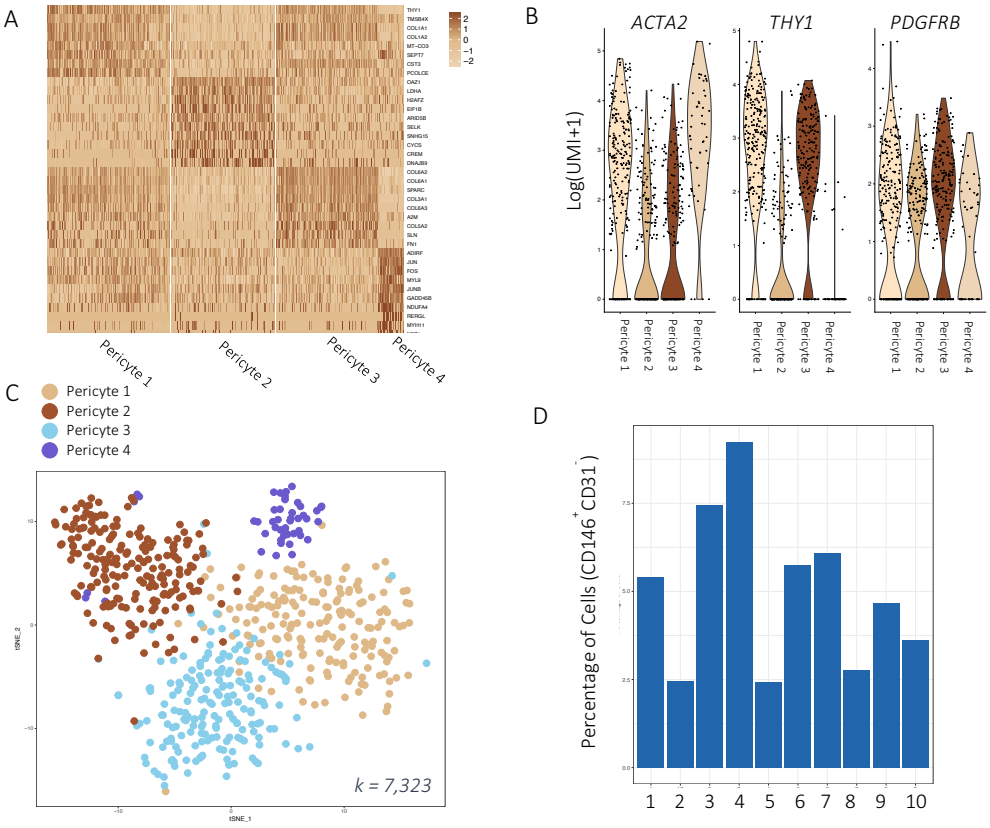


Figure 14. **Distinct perivascular niche in Dupuytren's nodules.** A) Heatmap showing the z-score expression of marker genes in four perivascular clusters. Scale bar in z-score expression. B) Violin plots showing expression of genes in four pericyte clusters (pericyte 1-4) in log(UMI+1). C) tSNE projection of pericytes showing Louvain cluster assignment. (pericyte 1-4). *k*=*7,323* pericytes and *n*=12 DD patients. D) Barplot of flow cytometry analysis of the percentage of perivascular cells ($CD146^+CD31^-$) in freshly disaggregated nodular cells as a total of all live cells. X axis represents nodules derived from independent DD patients (*n*=10 DD patients).

After this, I wanted to explore potential relationships between pericyte clusters and so projected the data into low-dimensional space using diffusion maps (see Chapter II.II.X) (Figure 15A). Diffusion maps also allow cells to be ordered along pseudotime, uncovering prospective differentiation trajectories. Visualization of the data along the first two diffusion components (DM-DM2) revealed an arching topology containing two curved branches, moving from the pericyte 1 and pericyte 2 towards the pericyte 3, before ending at the rare and contractile pericyte 4. This structure could suggest that pericyte 4 signifies an intermediate cellular state between a pericyte and a myofibroblast. In support of this, pericyte 4 showed a high expression of *ACTA2,* and other myofibroblast markers*,* and was transcriptomically the most distinct pericyte cluster containing the highest number of differentially expressed genes (*n*=1,022).

Finally, I was interested in the spatial distribution of pericytes in DD and focused on the α-SMA^+ pericyte 4. Using α-SMA protein staining in DD tissue sections, perivascular cells could be readily identified by a fusiform morphology surrounding a vessel lumen as shown in Figure 15B. Figure 15B also illustrates that α-SMA^+ pericytes were highest in areas of less densely packed myofibroblasts. In addition, Figure 15B shows their size and distribution varied, with some surrounding small vessels and others extending across large tortuous vessels. Together, these results provide evidence for a distinct perivascular niche in DD nodules.

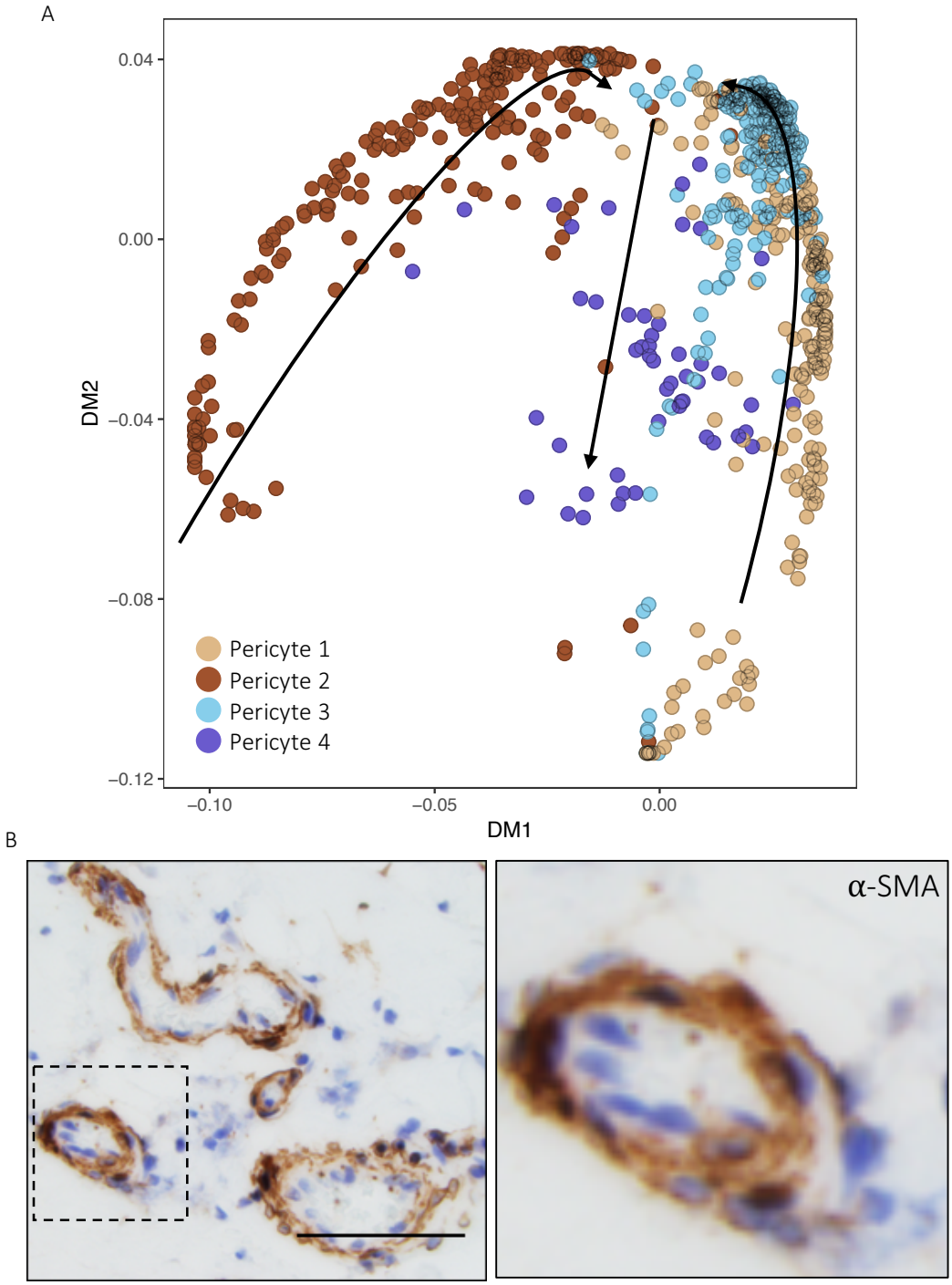


Figure 15. **Distinct perivascular niche in Dupuytren's nodules.** A) Diffusion map embedding of pericytes in DD coloured by Louvain clusters. *k*=*7,323* pericytes and *n*=12 DD patients. DM = Diffusion map coordinate. B) Microscopy image of immunohistochemistry showing α-SMA^{+} perivascular cells in Dupuytren's nodule. Scale bar 40 µm. Left panel represents zoomed area highlighted by black square.

III.III.V *The immune landscape in human fibrosis*

Immune cells (*k*=1033) in nodules were partitioned using unsupervised clustering into five major clusters (Figure 16A). This revealed a complex immune landscape with predominant cell types being macrophages, T cells and dendritic cells. The five immune cell clusters showed two macrophage rich clusters (*CD68*$^{+}$), two T cell clusters (*IL7R*$^{+}$*CD3D*$^{+}$) and one cluster enriched for dendritic cells (*FCER1A*$^{+}$). Within the macrophage population, I noted one cluster enriched for markers of M2-macrophages (*CD163*$^{+}$*)* and another showed expression of more classical M1-macrophage markers, including *IL1B, CCR7 and KYN.* Next, I visualised the data using PCA, which demonstrated a linear trajectory between the two macrophage enriched clusters (Figure 16B). This could indicate a differentiation path from a monocyte to a macrophage. Supporting this, along the trajectory there was a gradual decrease in monocyte markers, such as *CD14,* and an increase in macrophage markers including *CD68*.

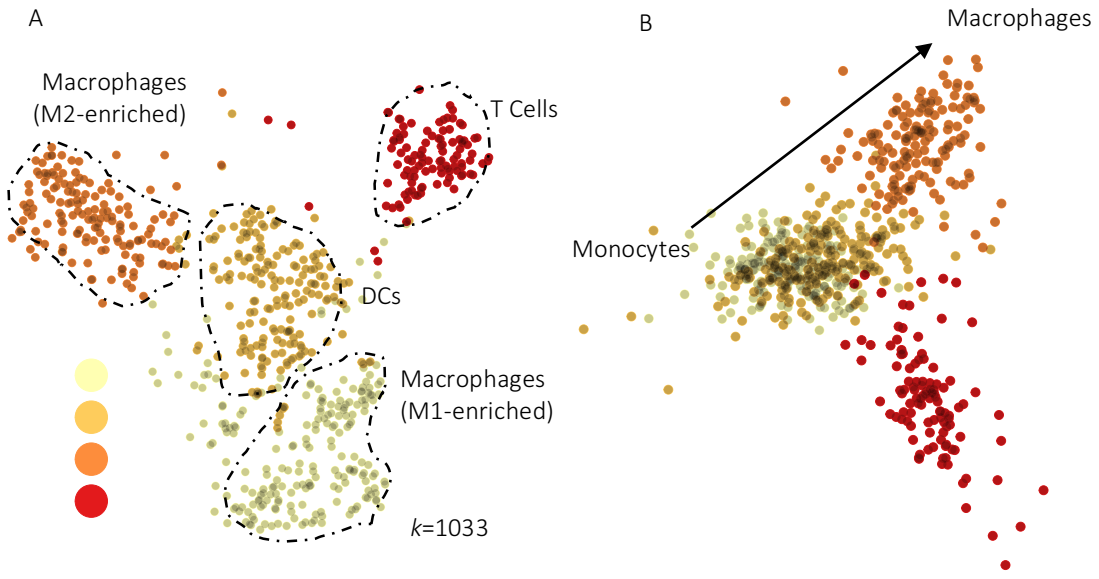


Figure 16. **The immune landscape in Dupuytren's nodules.** A-B) tSNE embedding (A) and PCA projection (B) of single cell RNA-seq of *1,033* immune cells isolated from Dupuytren's Nodules. Cells are coloured by Louvain clusters with annotated cell types. *n*= 12 DD patients. DCs = dendritic cells.

The trajectory noted in the PCA projection did not discriminate between M1 and M2-like macrophages but rather appeared as a general monocyte to macrophage transition. One potential reason for this poor resolution could relate to the small number of immune cells sampled. After this, effector cytokines were mapped to immune cell clusters and provided further confidence in identifying distinct populations (Figure 17A). *CCL5* was predominantly expressed in T cell clusters; in contrast, mononuclear cell clusters expressed high levels of *CXCL8, CXCL3, CXCL5* and *CXCL3* (Figure 17B). These two populations (monocytes/macrophages & macrophages) were also distinguished by the expression of *CCL2* and *CCL18*. Interestingly, I found the chemokine receptor *CXCR4* was highly expressed on all immune cell clusters (Figure 17B).

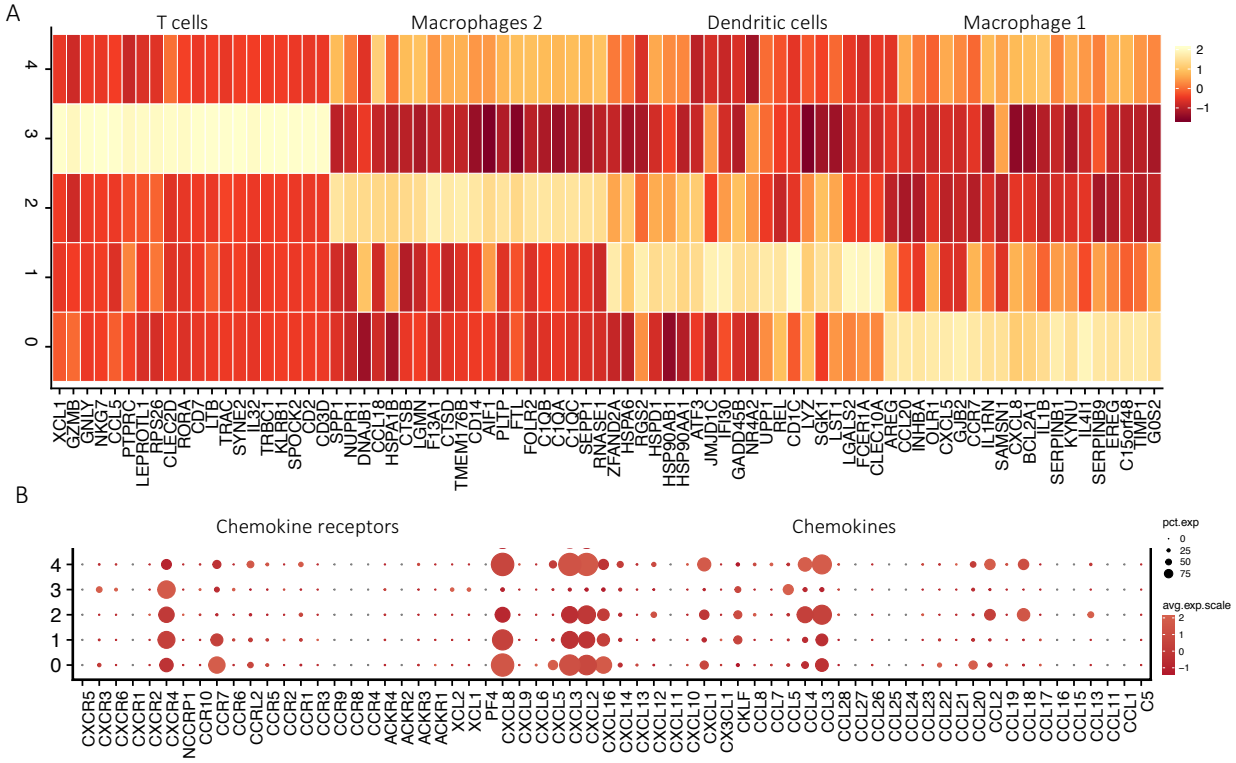


Figure 17. **Marker genes of immune cell clusters in Dupuytren's nodules.** A) Heatmap of markers genes of four immune cell types displaying the mean row scaled expression in log(UMI+1). Y axis 0-4 labels represent immune cell clusters. B) Dot plot mapping genes coding for chemokines and chemokine receptors across immune cell clusters. 0-4 on Y axis represent immune cell clusters. Cluster 0=macrophages, cluster 1=dendritic cells, cluster 2 & 4=macrophages and cluster 3=T cells. pct.exp = Percentage of cells in cluster expressing gene and avg.exp.scale represents z-score mean expression in log(UMI+1).

One interesting observation was the differential expression of *IL32* between the two T cell clusters (Figure 18A), I term *IL32*$^{+}$ and *IL32*$^{-}$ T cells. *IL32*$^{-}$ T cells were also distinguished by the high expression of *IL1B* and *IL1RN*. Given the small size of the sampled T cell population it was difficult to delineate true lineages such as CD4 T cells, CD4 T cells or regulatory T cells (Treg). However, I did observe a small number of CD4 T helper cells (*CD3D*$^{+}$) but could not identify any Treg (*FOXP3*$^{+}$) or CD8 T cells (*CD8*$^{+}$) in the data (Figure 18B).

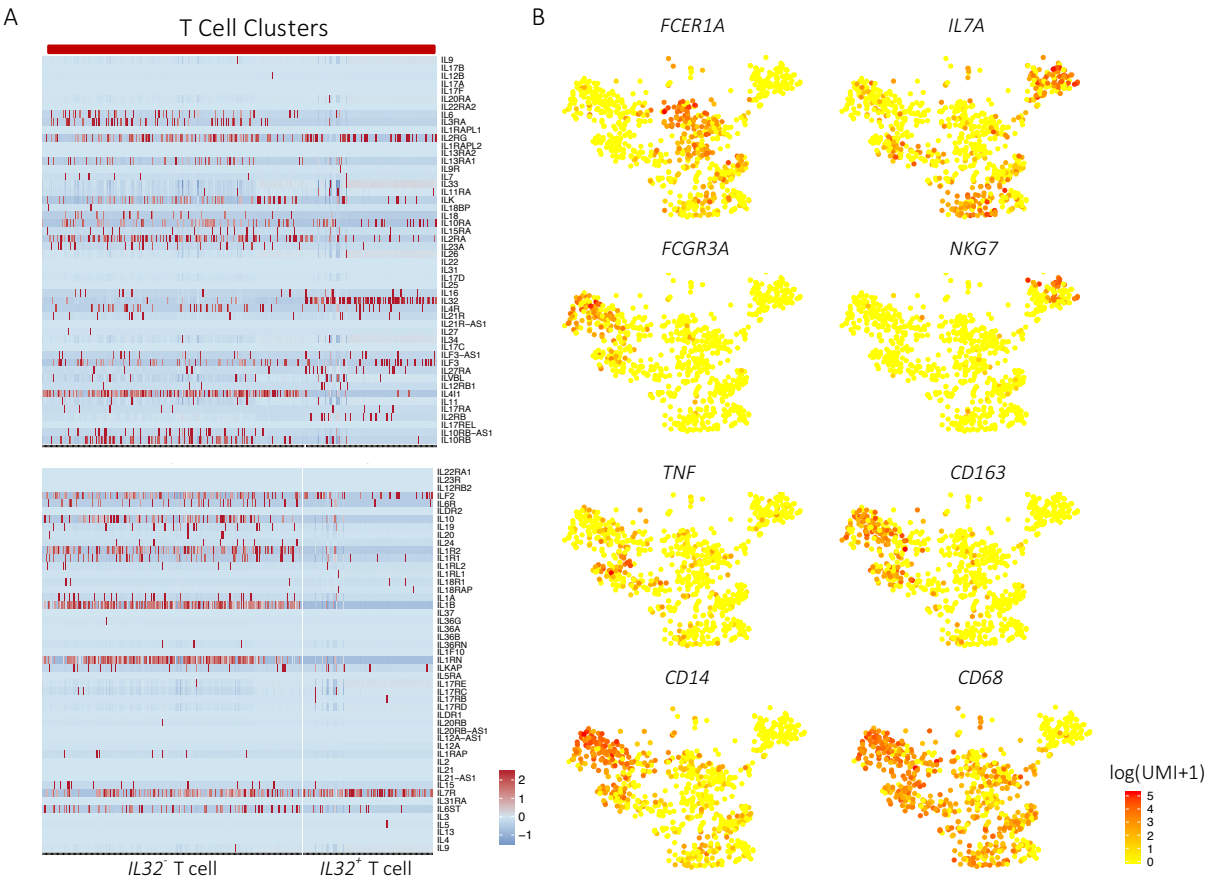


Figure 18. **Marker genes of immune cell clusters in Dupuytren's nodules.** A) Heatmap of genes coding for cytokines across two T cell enriched clusters displaying (Scale bar) average row scaled expression in log(UMI+1). B) tSNE projections for immune cell marker genes in log(UMI+1) in entire immune cell dataset. *k*=1,033 immune cells and *n*=12 DD patients.

III.IV. Discussion

The cellular landscape of fibrosis includes a spectrum of different cell types(28, 164). Myofibroblasts are the most studied mesenchymal cell, but the fibrotic milieu also contains cells of immune system, perivascular and vascular cells, as well as parenchymal cells of the organ effected. To unravel this complexity and inform the development of effective therapeutics, we must first gain a detailed understanding of the states, subtypes and functional properties of fibrotic stromal cells in humans. As described in Chapter III.I, DD is an exciting model to explore a cellular and complete fibrotic ecosystem in humans(203). In this chapter, major cell types in DD nodules were identified using single cell RNA-seq, with fibroblasts and myofibroblasts making up the major cell populations. For each cell type in DD nodules I report key marker genes. Moving forward, we must extend our descriptive and atomized understanding of the fibrotic microenvironment and investigate how the interaction between cell types or states may drive disease progression. Bridging this knowledge will enable a better understanding of whether targeting specific key regulators or pathways may alter the entire disease ecosystem.

After delineating major cell types in DD, another goal of this chapter was to dissect pericytes in human fibrosis at single cell resolution. My description of pericytes support the use of established markers such as *MCAM* and *RGS5,* shown to distinguish perivascular cells in DD*(204-206)*. Using single cell RNA-seq, I uncover four distinct subsets in the perivascular niche (pericyte 1-4) and one significance of this finding relates to informing future studies in fibrosis. Previously, established markers (e.g. *PDGRFB*) have been used to examine the origin, dynamics and phenotype of

pericytes(201, 207, 208). Often, only one marker is ostensibly used to label the entire pericyte population. However, the results in this chapter demonstrate that these markers are not evenly distributed across all pericyte subsets. Therefore, using one marker to universally tag a supposed pure population may be inherently bias towards distinct subsets. For example, pericyte 4 showed a lower expression of *PDGFRB* as compared to pericyte 1 and 3. Thus, if *PDGFRB* alone was used to isolate pericytes it may skew the population away from pericyte 4 towards a more pericyte 1/3 phenotype. Importantly, the diffusion map projections and marker genes imply that pericyte 4 could represent an intermediate cellular state between a perivascular cell and myofibroblast. Thus, losing pericyte 4 during an isolation protocol could greatly impact conclusions drawn from the data.

Nevertheless, the perivascular niche does express general markers suitable for labelling the entire cell population, including *MCAM* (CD146). In this Chapter, I extend observations made in the single cell RNA-seq data, validating a $CD146^{+}$ pericyte population in freshly disaggregated nodular cells using flow cytometry. Furthermore, I confirm a α-SMA^{+} perivascular niche in nodules using immunohistochemistry. Important questions remain with regards to whether these pericyte subsets carry distinct functional properties and if pericyte 4 represents a proto-myofibroblast. Future work should look to answer these questions, and may resolve the first myofibroblast precursor in DD.

Inflammation is intricately linked to fibrosis. In fibrotic disorders, the whole gamut of immune cells have been associated with distinct mechanisms driving stromal cell activation(165). In DD, the

immune compartment has been previously examined using supervised techniques such as flow cytometry and immunohistochemistry(27, 105). The results presented in this Chapter correlate with previous findings of major immune cells in DD nodules as macrophages and T cells(27). However, several rarer immune cell types previously described in DD, such as mast cells, were not captured in the single cell RNA-seq data. This may be due to the low overall number of immune cells sampled or bias introduced during tissue transportation and digestion. This highlights one potential limitation of the experimental protocol I applied (see Chapter II). In my methodology, after tissue processing, I avoided any steps that could influence cell health or integrity and thus decided not to sort or purify populations. To further explore the immune landscape in DD with greater resolution, I would FACS sort $CD45^{+}$ cells in freshly disaggregated nodular cells to enrich this population and run this sample individually on a sequencing lane. I would then capture a much larger pure population providing greater detail to discern specific lineages such as CD4 T helper cells, Tregs and myeloid cells.

Mapping chemokines genes to immune cells provides a perspective on the compartmentalization of key soluble factors in DD nodules. Chemokines are potent chemoattractants for leucocytes and play an important role in driving fibrosis by recruiting macrophages, myofibroblasts and other key effectors cells to the site of tissue injury(25). The entire spectrum of chemokines and chemokine receptors has been studied in fibrosis, but CXC receptor and CC chemokine families exhibit key pathogenic functions(209-211). In particular, CCL2 (monocyte chemoattractant protein-1) is recognised to promote fibrosis acting via the recruitment of mononuclear phagocytes(212). The specific expression of *CCL2* in mononuclear cells in DD imply a potential role in promoting

inflammation, but the results presented here are only descriptive and require further functional validation. Other highly expressed chemokines included *CXCL10* and *CCL26*, known chemokines for mast cells, in addition to *CCL4* and *CXCL8,* consistent with the presence of macrophages.

Lastly, another interesting observation was the high expression of *CXCR4* in all immune cell clusters. The natural ligand for CXCR4 is CXCL12, which is a specific soluble factor expressed by fibroblasts in DD (see Chapter IV.III.I). This could provide a fibroblast-immune cross talk supporting immune cell recruitment into DD nodules.

In summary, the results presented in this chapter construct a detailed single cell atlas of major cell types in DD nodules, including fibroblasts, myofibroblasts, endothelial cells, pericytes and immune cells. In addition, new subsets of transcriptomically distinct perivascular cells are identified and a description of the immune landscape in DD highlights the diversity of this compartment in DD nodules.

In the next chapter, I will undertake a detailed examination of fibroblasts and myofibroblasts and a focused investigation of fibroblast expression profiles in DD nodules.

Chapter IV. Defining expression profiles of fibroblasts and myofibroblasts in fibrosis

IV.I. Introduction

Accurate classification of cell type and function has been a major focus of scientific discovery since its inception. Indeed, we now have a detailed understanding of key parenchymal cells and how they change in normal physiology or disease. Despite this, traditional methods have been limited to measurements of bulk cell populations without the ability to resolve single cell states. Single cell genomics now offers the capacity to obtain global gene expression profiles in individual cells(139, 150, 213).

Cellular heterogeneity in fibrotic disorders provides an exciting model to refine the classification of stromal cells(1, 141). As fibrosis signifies a complex multicellular ecosystem, profiling single cells in these disorders may help identify key molecular signatures of pathogenic stromal cells(1, 141). Single cell transcriptomics will also facilitate a robust comparison between myofibroblasts and fibroblast. Pertinent questions relate to key marker genes and whether each has unique functional properties. This information will be invaluable to guide future therapeutic strategies, enabling mechanistic insights of treatments to be combined with an in-depth knowledge of their cellular context.

Since its first description, the myofibroblast has taken centre stage as key to all forms of fibrosis(4, 5, 214, 215). Although there is no doubt myofibroblasts are important during the progression of fibrosis, it has meant that other stromal cells have been partially overlooked. Specifically, α-SMA$^-$ fibroblasts are poorly understood in the context of fibrosis(216) and their significance in the

disease microenvironment and precise molecular function remains unknown. Also, we are still missing clear evidence for distinct sub-populations of fibroblasts in human fibrosis. In contrast, fibroblast subsets are increasingly reported in other disease systems, including cancer(217), providing support for a similar disease landscape in fibrosis.

IV.II. Hypothesis and aims

My hypothesis is that myofibroblasts and fibroblasts have distinct transcriptomic signatures in DD and fibroblasts house discrete subpopulations.

In this chapter, I leverage the single atlas generated in Chapter III to undertake a focused investigation of fibroblasts and myofibroblasts in DD. A major goal is to elucidate robust gene markers of each population and explore associations with biological processes using gene ontology. After this, I aim to test whether distinct subpopulations of fibroblasts can be resolved and how their proportions may change between Dupuytren's cords and nodules.

IV.III. Results

IV.III.I *Elucidating transcriptomic signatures of fibroblasts and myofibroblasts*

In each patient I first defined the proportion of fibroblasts and myofibroblast using *ACTA2* expression. For this, I selected clusters with >20% of cells expressing *ACTA2* and defined these as myofibroblasts. *ACTA2*$^{+}$ myofibroblasts were the predominant cell in nodules, representing ~60% (40-78%) of the total population. In contrast, *ACTA2*$^{-}$ fibroblasts comprised ~40% (13-60%), with significant variation between patients (Figure 1).

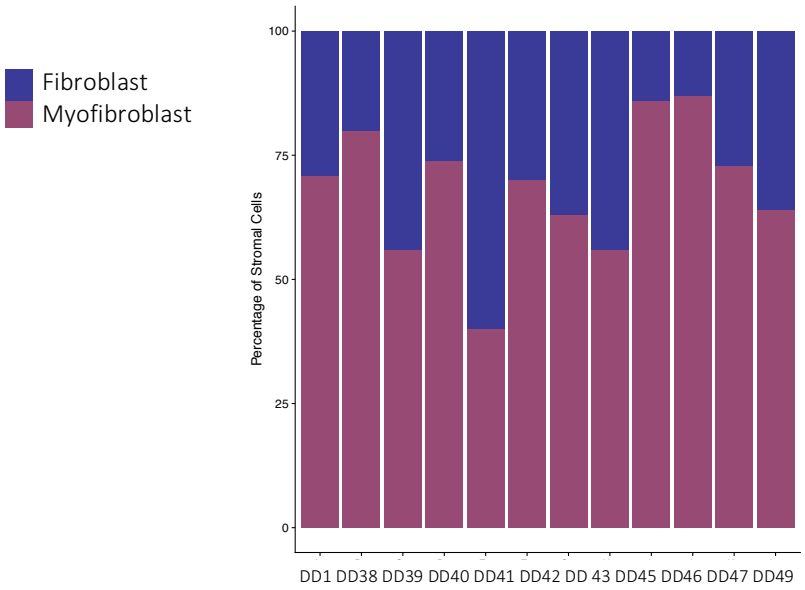


Figure 1. **Proportion of fibroblasts and myofibroblasts in Dupuytren's nodules.** Bar plot showing the proportion of fibroblasts to myofibroblasts within each patient donor (*n*=12 DD patients). Fibroblasts were defined as *ACTA2*$^{-}$ and myofibroblasts as *ACTA2*$^{+}$, with each patient showing distinct clusters separating cell types. X axis lists DD patients and DD1-DD65 represents individual patient IDs.

Next, I used differential expression analysis to define key marker genes for each population (see Chapter II.II.X). This defined many significant myofibroblast specific markers, including *FABP5, POSTN, SPON2 and LGALS1,* as well as the established marker *ACTA2* (Figure 2A). Many myofibroblast specific genes function to regulate cell contraction, such as *MYL6, PDLIM7, ACTA2*

and *TPM1,* highlighting this biological process as a key feature of this cell. After unsupervised clustering, I also noted a visible distinction between fibroblasts and myofibroblasts on the partitioned *k*-NN graph projected using tSNE. This was apparent in merged datasets and within individual patient samples. (Figure 2B).

Interestingly, many significant fibroblast marker genes were associated with immune pathways, including *CXCL14*, *PLA2G2A* and *C1R* (Figure 2A). C1R is a cell surface complement receptor expressed in wide variety of cells and is a strong stimulus for inflammation(218, 219).

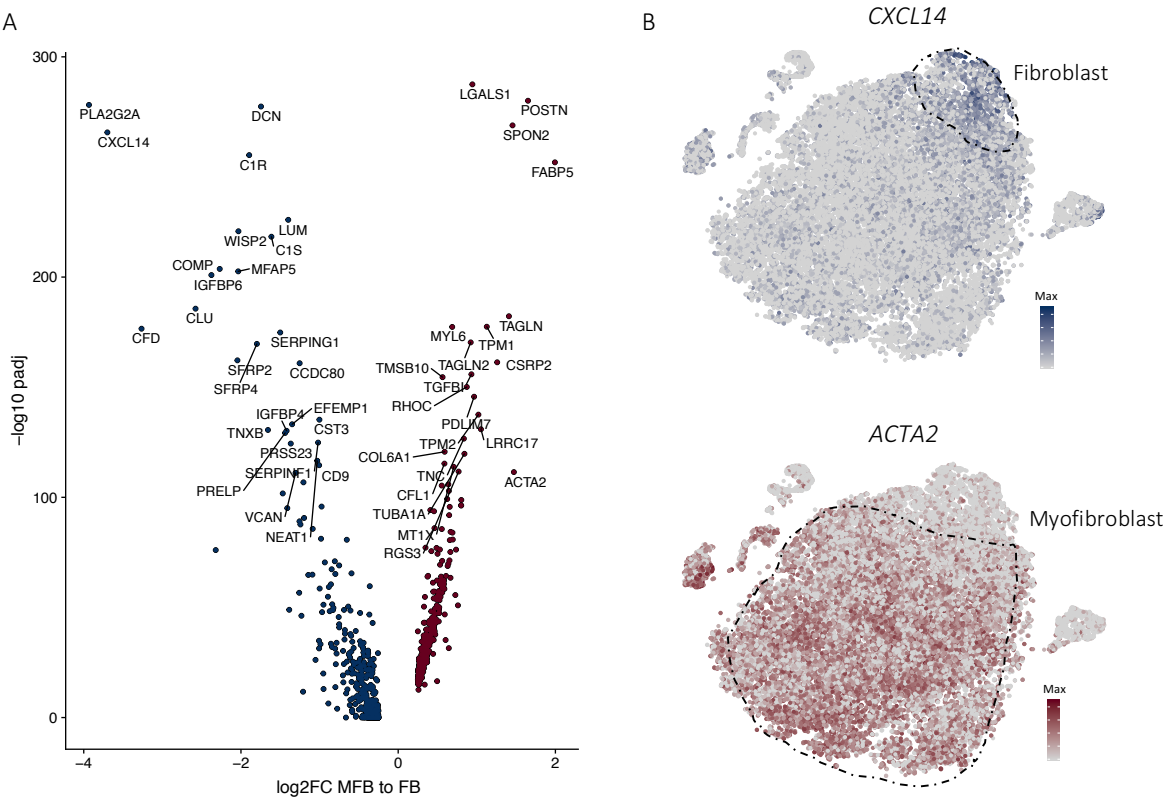


Figure 2. **Fibroblasts and myofibroblasts have distinct transcriptomic signatures in Dupuytren's nodules.** A) Volcano plot showing differentially expressed genes (Wilcoxon Rank Sum Test) between myofibroblast and fibroblast clusters with FDR corrected adjusted p-values (BH correction). B) tSNE projections from first patient batch (*n*=6 DD patients) coloured by the expression of fibroblast marker gene (*CXCL14*) and myofibroblast marker gene (*ACTA2*). Expression in scaled log(UMI+1). MFB = myofibroblast and FB = fibroblast.

PLA2G2A is a membrane phospholipid implicated in promoting inflammation via the arachidonic pathway(220). CXCL14 is a chemokine and a member of the CXC superfamily. It is a chemoattractant for varied immune cells including dendritic cells and monocytes, and is expressed by both immune and stromal cells(221).

Next, I looked for the protein expression of two fibroblast markers *in vivo* using immunohistochemistry (see Chapter II.II.XIII). PLA2G2A was selectively expressed in cells with a fibroblast-like morphology, located in areas lacking densely packed cells (Figure 3). This suggests a specific expression on fibroblasts in nodules. In contrast, when staining for C1R, I observed more widespread staining, on fusiform fibroblast-like cells in addition to densely packed cells that represent myofibroblasts (Figure 3).

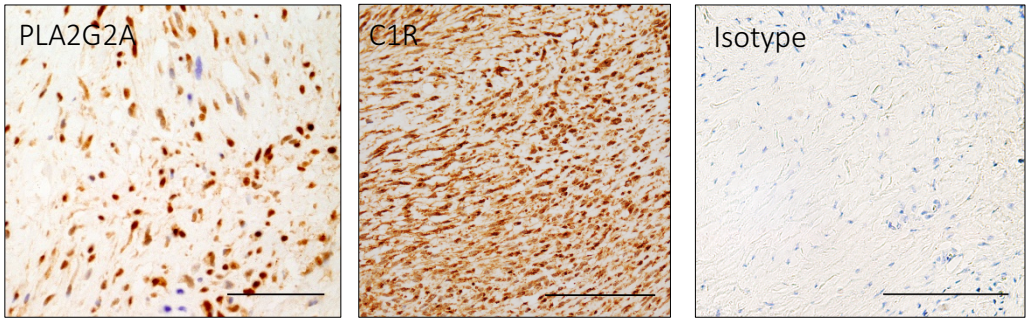


Figure 3. **Expression of fibroblast markers in Dupuytren's nodules.** Microscopy images of Immunohistochemistry validation of fibroblast marker expression in Dupuytren's nodules with isotype control (Rabbit IgG). Scale bar 40µm.

After identifying fibroblast marker genes, I performed functional annotation using gene ontology to explore associations with annotated biological processes(222). To obtain genes for this, I used differentially expressed genes between myofibroblasts and fibroblast and selected those above 1

log fold change with an adjusted p value < 0.05 (Wilcoxon Rank Sum Test). This demonstrated a specific enrichment for immune-related pathways including “leucocyte chemotaxis” and “regulation of the immune response” (Figure 4B) in fibroblasts as compared to myofibroblasts.

Next, I performed principal component analysis (PCA) on immune cells, myofibroblasts and fibroblasts to visualize potential relationships between them. This demonstrated a strong association of principal component 4 (PC4) with a putative immune-signature, ordering cells from myofibroblasts to fibroblasts, terminating at immune cells (Figure 4C). Together, these results suggest α-SMA$^-$ fibroblast could adopt an immune regulatory function in DD.

In the pathway analysis, myofibroblasts showed a strong enrichment for pathways pertaining to cell contraction and ECM remodelling (Figure 4B). Example pathways included “collagen biosynthetic process” and ‘regulation of muscle contraction”. Given this, I looked for the gene expression of key matrix protein coding genes in myofibroblasts and fibroblasts. Myofibroblasts showed higher expression of matrix protein coding genes (*COL1A1, COL3A1* and *FN1*) compared to the fibroblasts (Figure 4A). Together, these data support myofibroblasts as a major contributor to matrix expression in DD nodules.

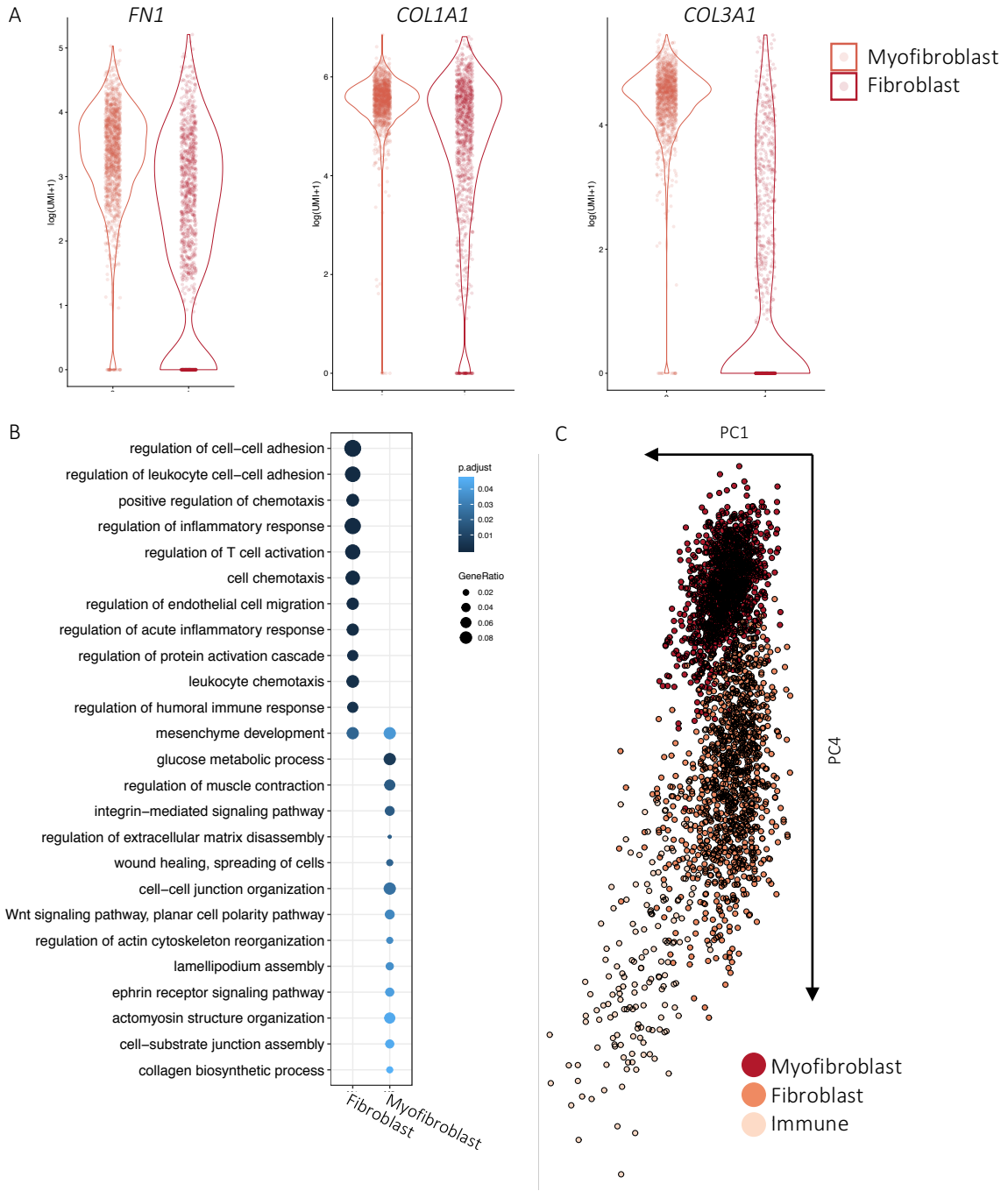


Figure 4. **Dupuytren's fibroblasts show an association with immune-regulatory properties.** A) Violin plots showing the expression of matrix protein genes in myofibroblasts and fibroblasts. Expression in scaled log(UMI+1). B) Dot plot showing the pathways enriched in fibroblasts and myofibroblasts. Fibroblasts show a specific enrichment for several immune related pathways. Gene ratio represents the number of genes found in listed pathways and p.adjust represents adjusted p-value (BH correction). C) PCA projection of stromal cells (myofibroblasts & fibroblasts) and immune cells showing PC4 aligning cells along a putative immune-signature. PC = principal component. *n*=12 DD patients.

IV.III.II *Defining three major subdivisions of fibroblasts in DD*

As α-SMA$^-$ fibroblasts are relatively understudied in fibrosis, I analysed this population further. Unsupervised clustering defined three major subsets; *CD34*$^+$, *PDPN*$^+$ and *ICAM1*$^+$ fibroblasts (Figure 5A-B). *ICAM1*$^+$ fibroblasts could be further divided into *ICAM1*$^+$*IL6*$^+$ and *ICAM1*$^+$*IL6*$^-$.

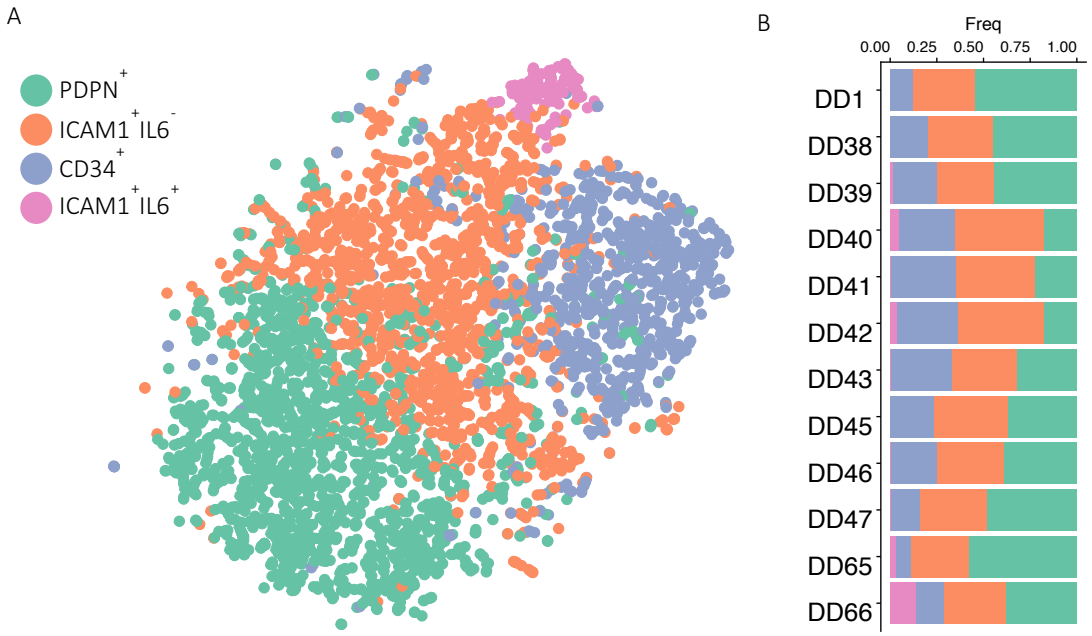


Figure 5. **Conserved fibroblast subsets in Dupuytren's nodules.** A) tSNE projection of Dupuytren's disease fibroblasts coloured by assigned Louvain cluster. B) Bar plot showing the proportion of fibroblast subsets in 12 Dupuytren's patients as a total of the entire fibroblast population. Colours correspond to Louvain clusters in (A). Freq = Frequency. Y axis lists DD patients with DD1-DD66 representing patient IDs. *k*=4,323 fibroblasts from 12 DD patients.

Key marker genes for *CD34*$^+$ fibroblasts were *CD34*, *CD55* and *CD248* (Figure 6). In contrast, *PDPN*$^+$ fibroblasts expressed several myofibroblast specific genes, including low levels of *TAGLN, PDLIM3* and *TPM1* (Figure 7A). Other gene markers for *PDPN*$^+$ fibroblasts included *COL11A1* and *IL32*

(Figure 6). The *ICAM1*$^{+}$ fibroblast was more diffuse and could be partitioned into *ICAM1*$^{+}$*IL6*^{high} and *ICAM1*$^{+}$*IL6*^{low} subsets by the expression of *IL6* and other chemokines (Figure 6). Initially, I was concerned *ICAM1*$^{+}$*IL6*^{high} fibroblasts may represent a doublet or contaminated population. However, I noted its conservation in the larger patient samples and it lacked the expression of any clear immunophenotypic markers such as *CD3D* or *PTPRC* (CD45). In addition to high expression of *CXCL6*, the *ICAM1*$^{+}$*IL6*^{high} fibroblast also expressed other chemokine coding genes, including *CXCL8, CCL3* and *CCL5* (Figure 7A-B).

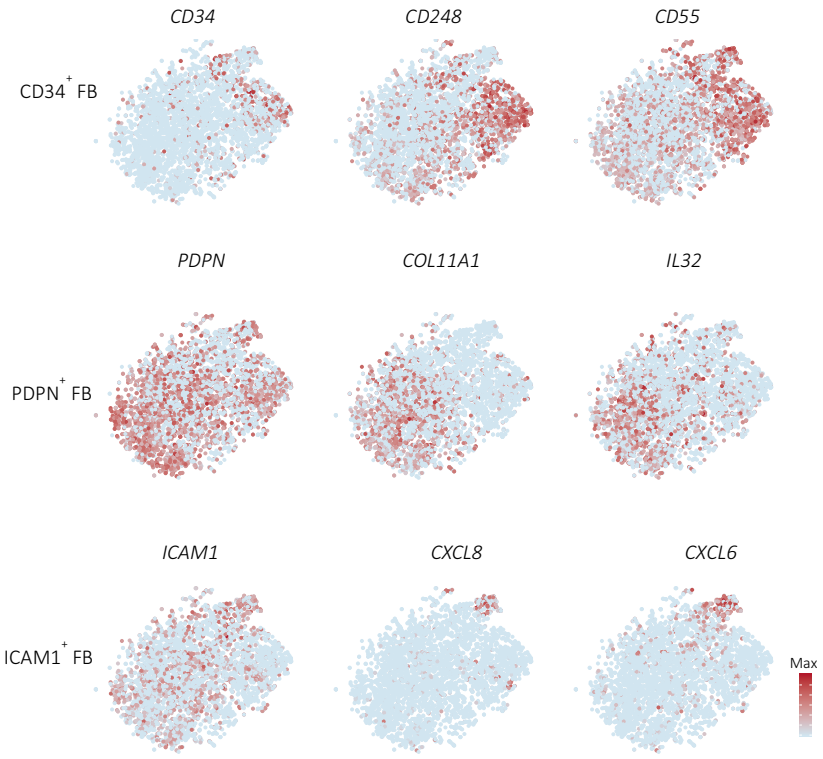


Figure 6. **Three subsets of Dupuytren's disease fibroblasts.** tSNE projections coloured by the expression of genes showing three fibroblast subsets (*CD34*$^{+}$, *PDPN*$^{+}$, *ICAM1*$^{+}$). Fb=Fibroblast. Expression (Scale bar) in z-score expression in log(UMI+1). *k*=4,323 fibroblasts from 12 DD patients.

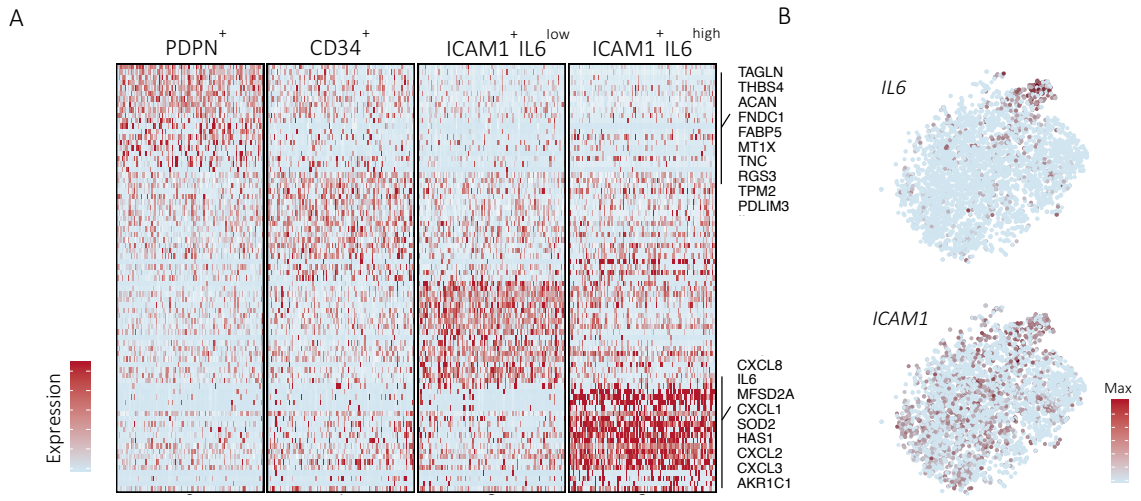


Figure 7. **Four clusters of fibroblasts in Dupuytren's nodules housing a chemokine-rich ICAM1^{+} subset.** A) Heatmap of the z-score expression of top 10 marker genes in fibroblast clusters. B) tSNE projections of entire fibroblast population coloured by *ICAM1* and *IL6* expression. Expression (Scale bar) in z-score expression in log(UMI+1). *k*=4,323 fibroblasts from 12 DD patients.

After this, using freshly disaggregated cells from DD nodules, I validated the protein expression of three main markers of the fibroblast subsets (CD34, PDPN and ICAM1) using flow cytometry (see Chapter II.II.IX) (Figure 8). To isolate a pure fibroblast population, I first negatively gated on immune cells (CD45^{+}), endothelial cells (CD31^{+}), dead cells (LIVE/DEAD^{-}), perivascular cells (CD146^{+}) and myofibroblasts (ITG.β1^{high}).

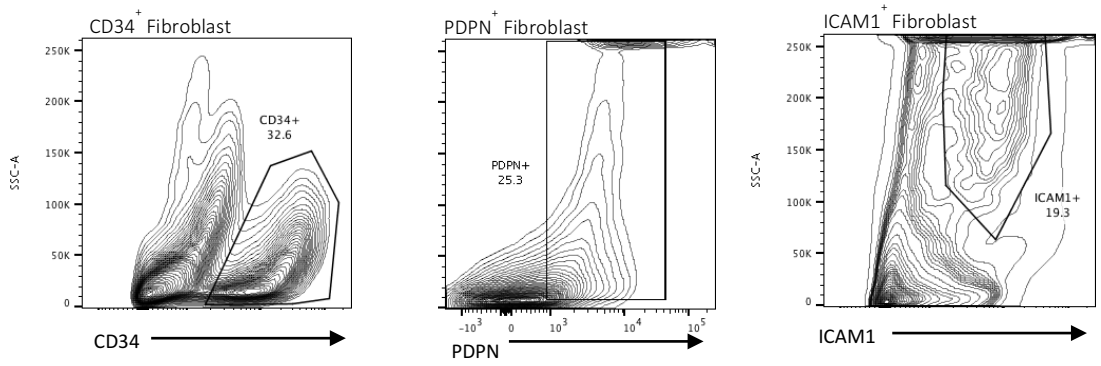


Figure 8. **Flow cytometry of three major fibroblasts in Dupuytren's nodules.** Representative density plots of flow cytometry analysis of three fibroblast (CD146^{-}CD45^{-}CD31^{-}ITG-β1^{low}LIVE/DEAD^{-}) populations in Dupuytren's nodules. Representative of seven DD patients.

After identifying gene markers of fibroblasts subsets, I tested for an association with known biological processes using pathway analysis (Figure 9). To obtain genes for this, I used differentially expressed genes between fibroblast subsets and selected those above 0.5 log fold change with an adjusted p value < 0.05 (Wilcoxon Rank Sum Test) (see Chapter II.II.X). In addition, gene set enrichment analysis (GSEA) was undertaken using ranked gene lists of fibroblast subset markers. Functional annotation of marker genes for each fibroblast subset showed an enrichment for more general stromal cell pathways, including "ECM Homeostasis" and "Matrix Remodelling" in two of the three subsets ($CD34^{+}$ & $PDPN^{+}$). However, the relatively small $ICAM1^{+}IL6^{high}$ fibroblast displayed specific enrichment for several chemotactic and immune responsive signatures driven by the expression of various chemokines (*IL6*, *IL8*, *CXCL12*, *CCL2*) (Figure 9B). This was demonstrated by GSEA that showed core enrichment of immune pathways included *IL6, IL8* and *CXCL12*.

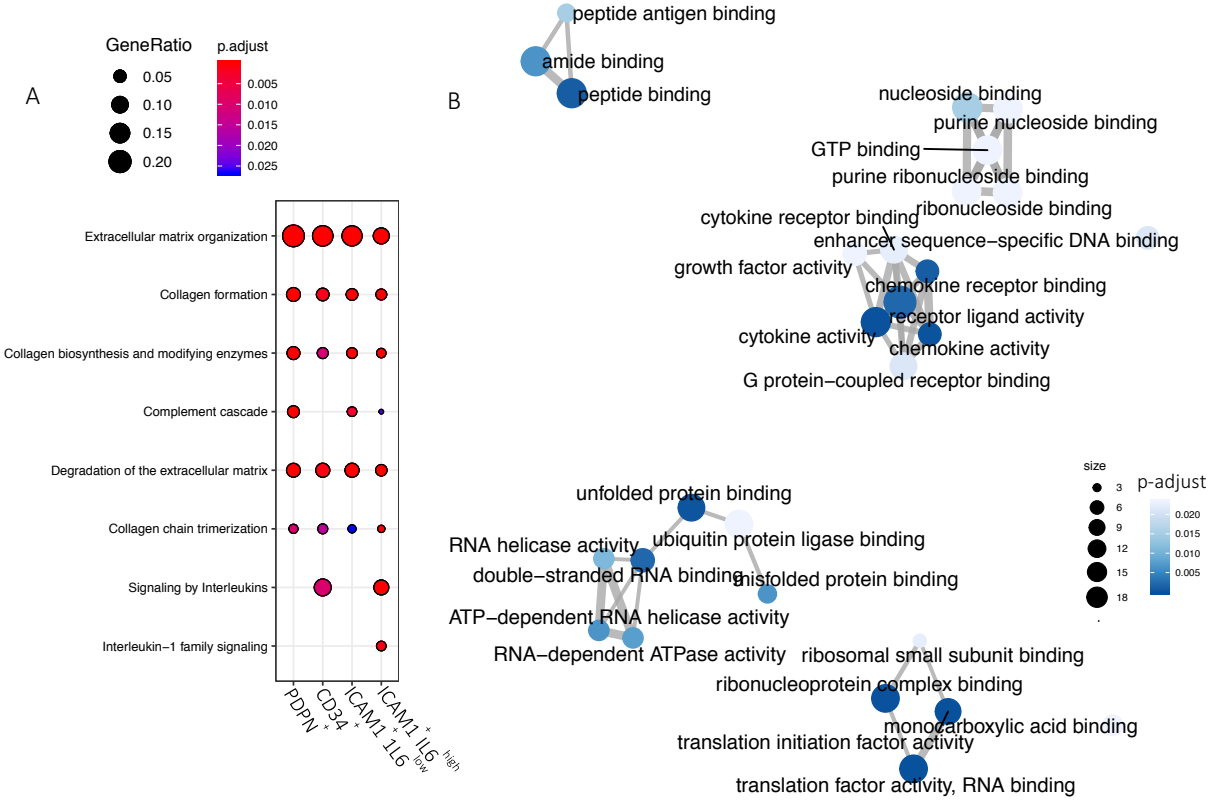


Figure 9. **Pathway analysis in Dupuytren's disease fibroblasts.** A) Dot plot showing the enrichment of gene ontology pathways in fibroblast subsets (X axis). Gene ratio represents the number of genes detected in annotated gene set used for pathway analysis and p.adjust represents the adjusted p-value (BH correction). B) Network plot of pathways upregulated in *ICAM1*$^{+}$ *IL6*^{high} fibroblast. Here, size represents the number of genes detected in annotated gene set used for pathway analysis.

As ICAM1^{+} fibroblasts carried an immune regulatory signature, I wanted to validate the protein expression of genes driving this signature. For this, I collected freshly disaggregated nodular cells and performed flow cytometry using an intracellular staining protocol (Figure 10) (see Chapter II.II.XII). Often these protocols utilise a stimulation step with a protein transport inhibitor to augment staining. However, I decided to avoid this step as I was interested in defining the native expression of these chemokines without any perturbations. I used a FACS strategy described previously (see Figure 8) to isolate fibroblasts and then co-stained for IL-6 and IL-8. This demonstrated an increased chemokine protein expression in the ICAM1^{+} subset.

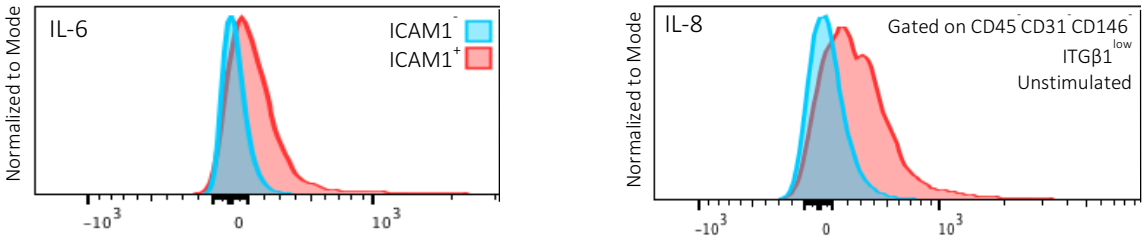


Figure 10. **Chemokine expression in ICAM1^{+} fibroblast in Dupuytren's nodules.** Histograms of flow cytometry analysis of intracellular staining in freshly isolated unstimulated nodular cells. Fibroblasts (CD31^{-}CD45^{-}CD146^{-} ITGβ1^{low}) were gated for ICAM1$^{+/-}$ populations and co-stained for IL-6 and IL-8. Representative of three DD patients.

IV.III.III *Pseudotemporal ordering of fibroblasts in DD*

The presence of discrete transcriptomic profiles of fibroblast clusters suggests functional differences between them. How this heterogeneity emerges and the developmental process underlying it are unclear. Single cell expression profiling enables the inference of cell-cell relationships and putative transitional states along differentiation trajectories. Computationally, these methods are collectively known as pseudotemporal or pseudotime analysis, with one popular method being diffusion maps (see Chapter II.II.X)(223). Diffusion maps were used in Chapter IIII.III.IV to study the perivascular compartment, and I applied the same method to fibroblasts in DD to explore relationships between subsets (Figure 11).

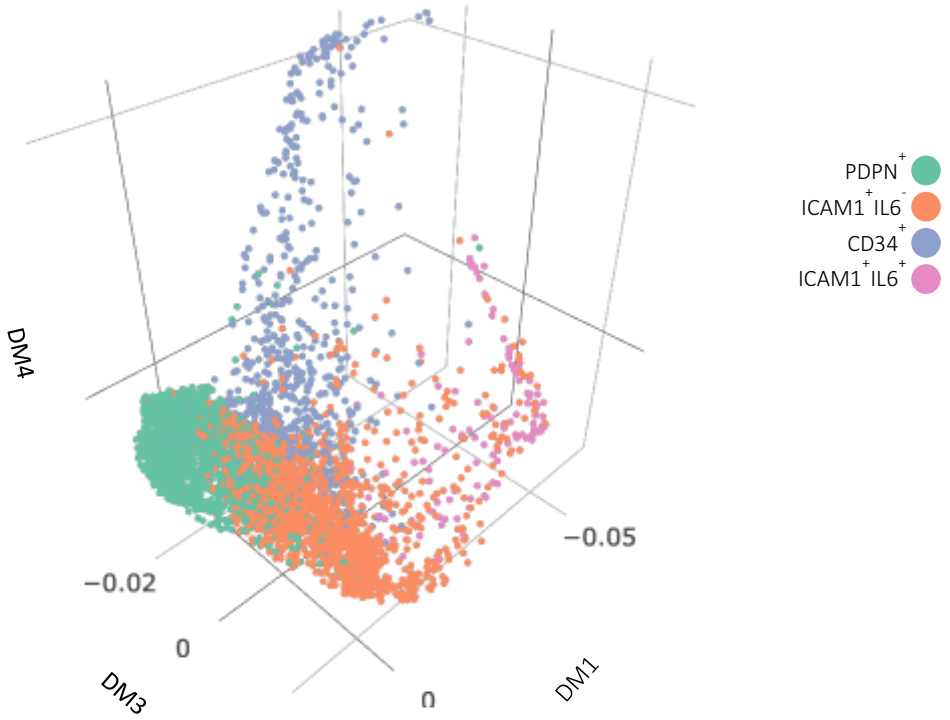


Figure 11. **Pseudotemporal analysis of fibroblasts in Dupuytren's nodules.** Diffusion map projection of 4,323 fibroblasts (*n*=12 DD patients) coloured by Louvain clusters. DM1-3 represents diffusion component coordinates

Visualizing the data along the first three diffusion components revealed a complex structure containing two discrete trajectories (Figure 11). The first moved from $CD34^{+}$ to $ICAM1^{+}IL6^{-}$ fibroblasts terminating at the $PDPN^{+}$ fibroblast. The second began at $ICAM1^{+}IL6^{+}$ fibroblasts, transitioned through $ICAM1^{+}IL6^{-}$ fibroblasts and ended at $PDPN^{+}$ fibroblasts. These results suggest either $CD34^{+}$ or $ICAM1^{+}IL6^{+}$ fibroblasts may represent precursor populations. Each trajectory shared common features, including $ICAM1^{+}IL6^{-}$ as an intermediate cell state and $PDPN^{+}$ fibroblasts marking a terminal phase of the developmental path. In accordance with this observation, building a phylogenetic tree of fibroblast subsets showed $ICAM1^{+}IL6^{-}$ and $PDPN^{+}$ fibroblasts were more transcriptomically similar as compared to $CD34^{+}$ and $ICAM1^{+}IL6^{+}$ fibroblasts (Figure 12). Here, the phylogenetic tree builds a hierarchical relationship based on inter-cell distance and provides a complementary metric to the diffusion map projection.

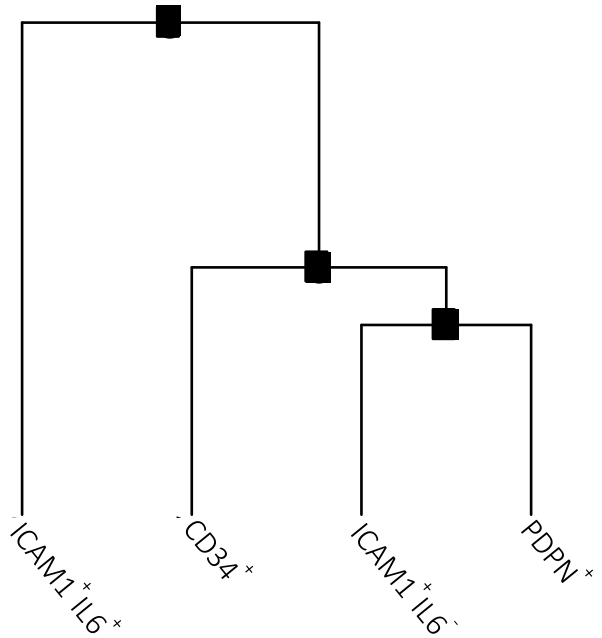


Figure 12. **Phylogenetic tree of fibroblasts in Dupuytren's disease.** Phylogenetic tree of fibroblast subsets representing the inter-cell disances between the average cell for each cluster in gene expression space. Fibroblast subsets clustered using hierarchical clustering in the hclust() function.

The results of the pseudotime analysis suggest PDPN^{+} fibroblasts may be a terminal cell state representing the end of a final common pathway from two precursors. In this regard, several PDPN^{+} fibroblast markers overlapped with myofibroblast markers including *ACTA2, TPM1* and *FAPB5* (Figure 7A). Together, these results may suggest PDPN^{+} fibroblasts are proto-myofibroblasts, capable of differentiating into a mature and activated myofibroblast following appropriate stimulation(60). A detailed annotation of PDPN^{+} fibroblast markers using gene ontology also showed close association with myofibroblast related pathways, including "EPH-Ephrin Signalling" and "Extracellular Matrix Organization" (Figure 13). To obtain genes for this, I used differentially expressed genes between fibroblast subsets and selected those above 0.5 log fold change with an adjusted p value < 0.05 (Wilcoxon Rank Sum Test).

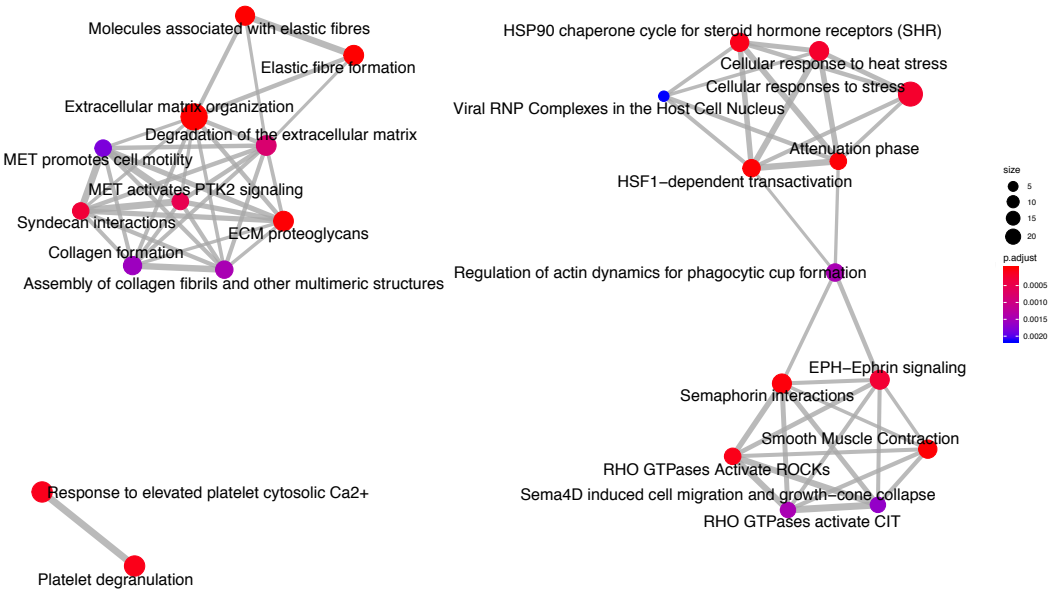


Figure 13. **Pathways enriched in PDPN^{+} fibroblasts.** Network plot showing enriched pathways in PDPN^{+} fibroblast. Size represents the number of genes detected in annotated gene set used for pathway analysis and p.adjust represents the adjusted p-value (BH correction).

IV.III.IV *ICAM1^{+} fibroblasts are enriched in Dupuytren's nodules*

After identifying the fibroblast subsets in DD, I sought to explore their temporal dynamics during fibrosis pathogenesis. To assess this, I used flow cytometry to quantify their proportions within two distinct Dupuytren's structures, the myofibroblast-rich nodule and matrix-rich cord(55, 224) (Figure 14-15). This provided a perspective on how fibroblast subsets differ in distinct fibrotic microenvironments. Moreover, as DD nodules harbour a greater proportion of inflammatory cells and myofibroblasts, it allowed me to investigate how fibroblast subsets may correlate with other cell types in fibrosis(17, 105). Using flow cytometry of freshly disaggregated cells, I observed a greater proportion of ICAM1^{+} fibroblasts in DD nodules. In contrast, CD34^{+} and PDPN^{+} fibroblasts had stable proportions in DD nodules and cords (Figure 14).

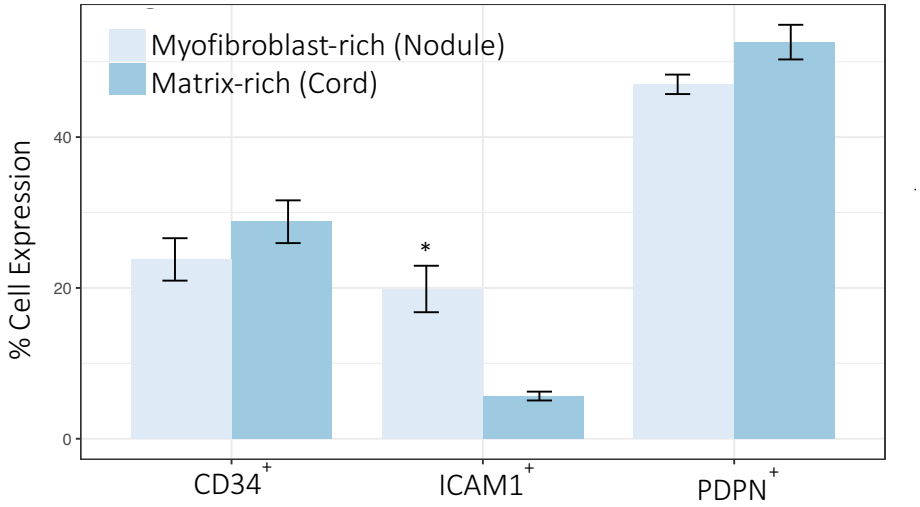


Figure 14. **Dynamic ICAM1^{+} fibroblast in Dupuytren's disease.** Bar plot showing the percentage of fibroblast subsets in freshly disaggregated Dupuytren's cords and nodules. p-values calculated by Wilcoxon rank Sum test. %Cell Expression represents the percentage of fibroblasts expressing protein markers as determined by flow cytometry. Mean +/- standard error plotted. * is p-value < 0.05. *n*= 8 independent DD nodules. X axis marks the three fibroblast subsets (ICAM1^{+}, CD34^{+} and PDPN^{+}).

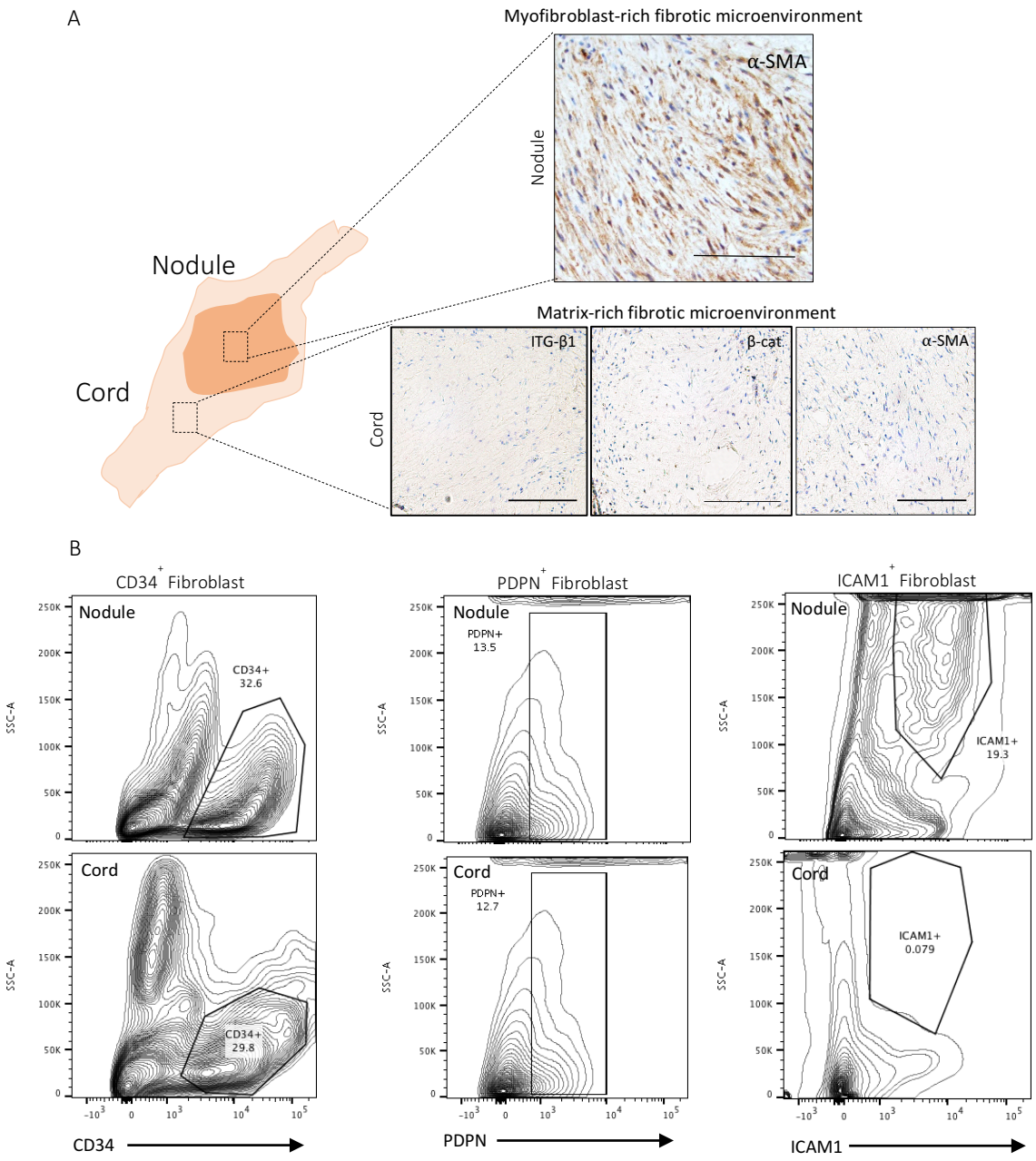


Figure 15. **ICAM1$^+$ fibroblasts are specifically enriched in Dupuytren's nodules.** A) Schematic of Dupuytren's cord and nodule with images of immunohistochemistry staining for myofibroblast markers. Scale bar 50 µm. B) Representative density plots for flow cytometry analysis of three major fibroblast populations in a myofibroblast-rich nodule and matrix-rich cord. Representative of 8 DD patients.

IV.III.V *Bulk RNA-seq of Dupuytren's nodule and matched palmar skin*

After defining key marker genes for myofibroblasts and fibroblasts, I finally wanted to integrate these findings with bulk transcriptomic profiles of fibrotic tissue (Figure 16). This may help gain an insight into how the fibrotic microenvironment could influence stromal cell phenotypes. To this regard, I performed bulk RNA-sequencing to construct transcriptomic signatures of fibrotic (DD nodule) and healthy (dermal fibroblasts) stromal cells. After this, I sought to use cell type expression profiles of myofibroblasts and fibroblasts to deconvolute the bulk transcriptomic signatures with the aim of discovering potential phenotypic markers.

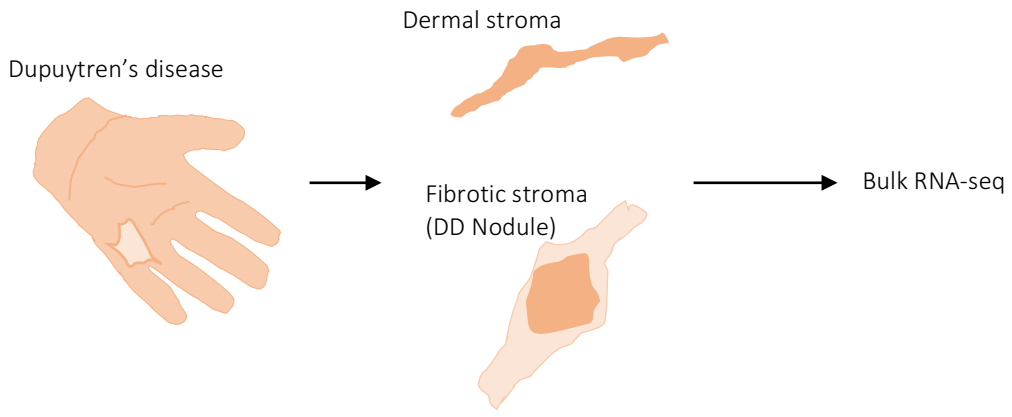


Figure 16. **Bulk RNA-sequencing of matched skin and Dupuytren's nodular cells.** Schematic showing the experimental strategy for mRNA sequencing of match Dupuytren's nodule tissue, non-palmar dermal and palmar dermal fibroblasts isolated following dermofasciectomy.

As there is no accepted and accessible control for DD, I used dermal (palmar and non-palmar) skin as a control for the bulk RNA-sequencing as DD is confined to the palm of genetically susceptible individuals. I collected eight matched samples and performed mRNA-sequencing to systematically

uncover genome-wide expression profiles marking bulk stromal cell populations, with data analysis performed as described in Chapter II.II.XXII.

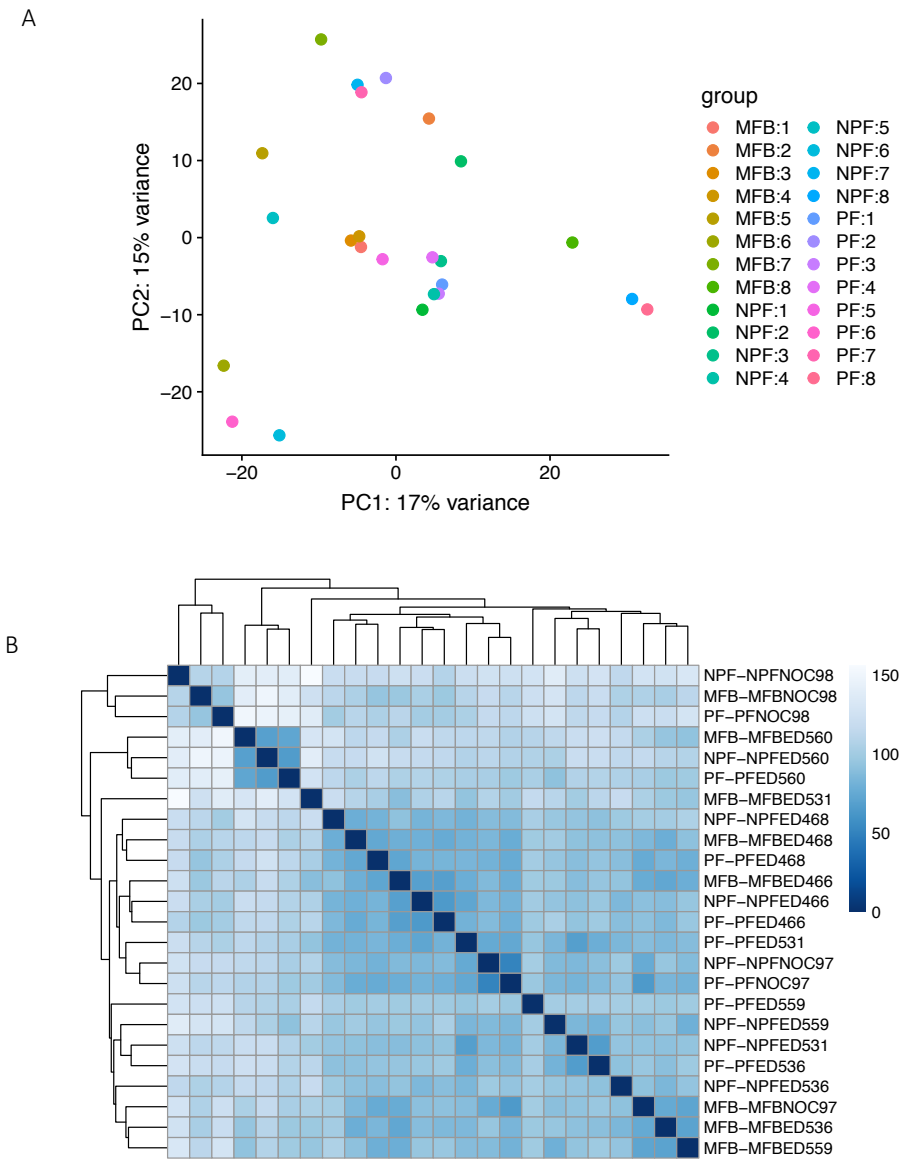


Figure 17. **Quality control for bulk RNA-sequencing of matched skin and Dupuytren's nodular cells.** A) PCA projection of patient donors from bulk RNA-seq. Points represent individual DD patients. B) Heatmap of sample clustering of patient donors from bulk RNA-seq. Distance metric is Pearson correlation between regularized log transformed values. Samples cluster mainly by patient donor and cell type. Scale bar represent Euclidean distance between samples. Suffix following cell type ID represents DD patient ID (*n*=8 DD patients). MFB is myofibroblast, NPF is non-palmar dermal fibroblast and PF is palmar dermal fibroblasts.

As expected, *ACTA2* demonstrated higher expression in DD nodular cells as compared to dermal fibroblasts. Other specific DD nodular cell markers included *MMP1, RSPO3* and *RSPO4.* RSPO3 and RSPO4 belong to the R-spondin protein family, associated with fibrosis in the lung acting via Wnt signalling(225). Exemplar non-palmar dermal fibroblast marker genes included *SFRP4, MMP9* and *COL14A1.* In total, there were 443 genes differentially expressed (logFC </> 1, p-adjust < 0.05) between DD nodular cells and non-palmar dermal fibroblasts, and 132 between DD nodular cells and palmar dermal fibroblasts (Figure 18-19).

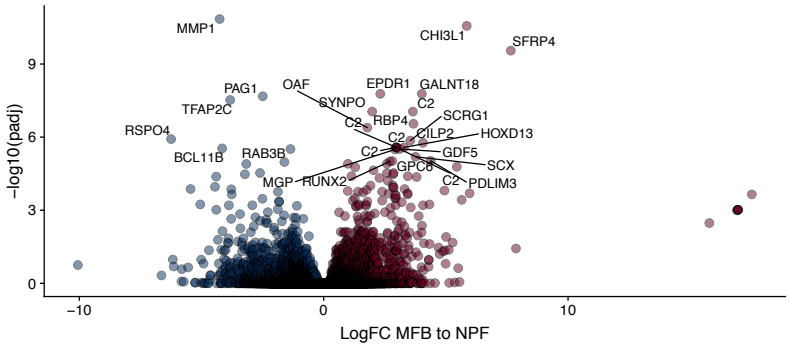


Figure 18. **Differentially expressed genes between non-palmar dermal fibroblasts and Dupuytren's nodular cells.** Volcano plot showing differentially expressed genes between DD nodular cells (MFB) and non-palmar dermal fibroblasts (NPF). FDR-corrected p-values (BH correction).

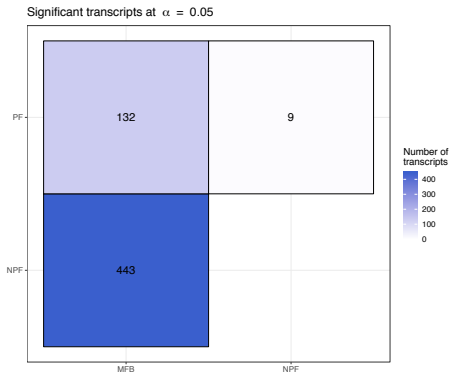


Figure 19. **Palmar dermal fibroblasts are transcriptionally closer to Dupuytren's nodular cells than non-palmar dermal fibroblasts.** Heatmap showing the number of differentially expressed genes (logFC </> 1, p-adjust < 0.05) between each group (PF=palmar dermal fibroblasts, NPF=non-palmar dermal fibroblasts and MFB=DD nodular cells). Colour represents number of transcripts differentially expressed.

One interesting observation from the RNA-seq was a greater transcriptomic similarity between palmar dermal fibroblasts (P-DF) and DD nodular cells as compared to non-palmar dermal fibroblasts (NP-DF). NP-DF had almost four-fold more differentially expressed genes as compared to P-DF (Figure 19B). This may be significant, as it has been proposed that dermal fibroblasts in the palm may be a precursor of myofibroblasts in DD(20, 226). Although only descriptive, my results suggest P-DF could represent an intermediate transcriptomic state between fibroblasts from non-palmar dermis and fibrotic myofibroblasts in DD. Finally, I was interested in applying functional annotation to DD nodular cells and dermal fibroblast marker genes (Figure 20). This was performed using gene ontology taking differentially expressed genes between nodular and dermal cells (non-palmar dermis) and selecting those with a log fold change > 1 and a p-adjust < 0.1. DD nodular cells showed a more significant enrichment in pathways pertaining to cell contraction including “contractile fiber part” and “sarcomere”.

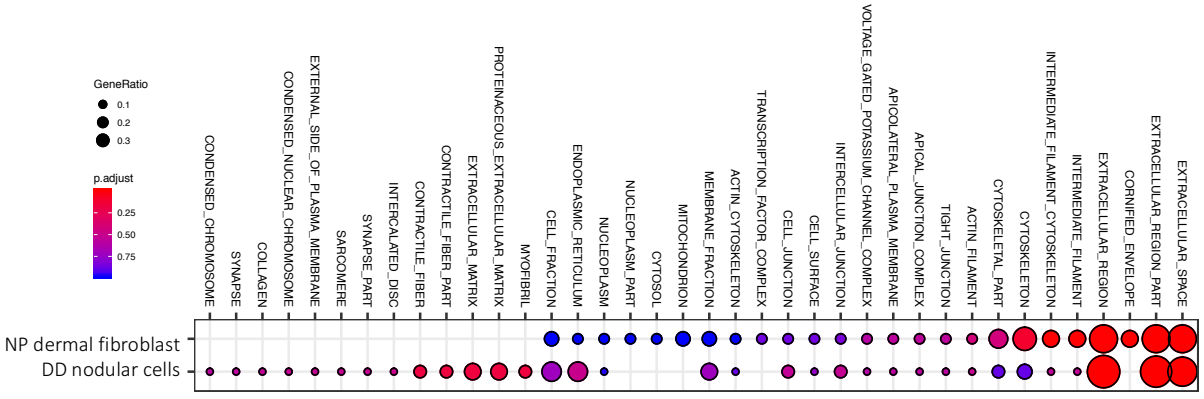


Figure 20. **Dupuytren’s nodular cells are enriched for cell contraction pathways.** Dot plot showing pathways enriched in DD nodular and fibroblast (NP-DF) marker genes identified in bulk RNA-seq. Gene ratio represents the number of genes detected in the annotated gene set used for pathway analysis and p.adjust represents the adjusted p-value (BH correction).

IV.III.VI *Integrating bulk transcriptomic signatures with cell type profiles*

From the bulk RNA-seq I could define broad transcriptomic programs that distinguish healthy (dermal) and fibrotic (DD nodular) stromal cells. Subsequently, I used the single-cell expression profiles to deconvolute these programs and assembled meta-signatures for comparing fibrotic and dermal fibroblasts. This provides insights into conserved phenotypic genes and allows for the identification of genes that may be upregulated within the fibrotic microenvironment (Figure 21).

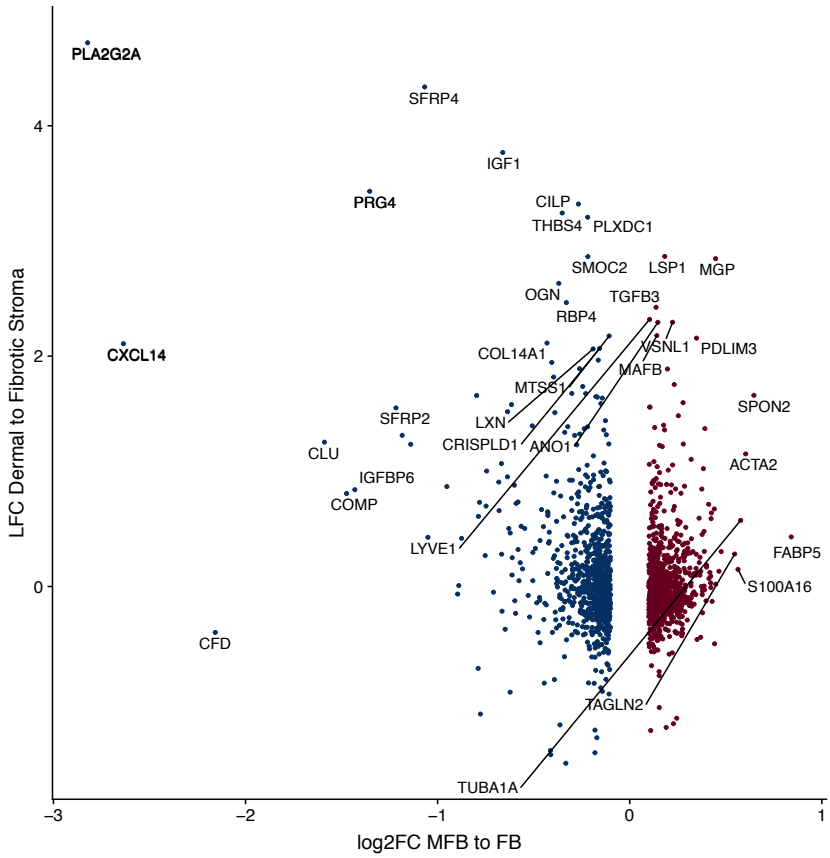


Figure 21. **Deconvolution of bulk expression profiles using single cell transcriptomics.** Volcano plot showing intersection of differentially expressed genes between DD myofibroblasts and fibroblasts (single cell RNA-seq) and fibrotic (DD nodular) and dermal (NP-DF) stromal cells (bulk RNA-seq). X axis represents log fold change between myofibroblast and fibroblast clusters in single cell RNA-seq. Positive values reflect genes showing higher expression in myofibroblast cluster. MFB = myofibroblast and FB = fibroblast. Y axis represents log fold change between non-palmar dermal fibroblasts and DD nodular cells (fibrotic stromal cells) in bulk RNA-seq data. Positive values reflect genes showing higher expression in DD nodular cells (fibrotic stromal cells).

Fibroblast meta-signatures were assembled by intersecting fibroblast marker genes identified in single cell RNA-seq with genes differentially expressed between dermal and fibrotic (DD nodular) stromal cells in bulk RNA-seq. One of the most striking observations was the specific upregulation for immune regulatory genes in the fibrotic fibroblasts meta-signatures (Figure 21-22). For example, *CXCL14, PLA2G2A, CXCL3* and *CXCL8* were all significantly upregulated in DD nodular cells compared to non-palmar dermal fibroblasts. In addition, these genes were differentially expressed between fibroblasts and myofibroblast clusters in single cell RNA-seq. I also noted upregulation of many genes coding for chemokines in DD nodular cells as compared to dermal stromal fibroblasts (Figure 22A).

Finally, I performed functional annotation of markers genes for the pro-fibrotic fibroblast meta-signature (>1 log fold change & adjusted p-value < 0.1). This showed enrichment for multiple immune related pathways including "Regulation of chemotaxis" and "neutrophil activation" (Figure 22B). There was also enrichment for other pathways including cell growth, proliferation and matrix remodelling. In summary, these data could suggest that fibroblasts in fibrosis adopt an immune regulatory role.

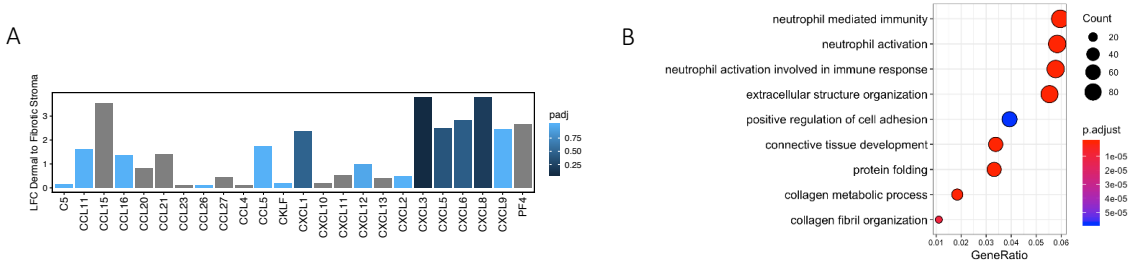


Figure 22. **Fibroblasts have an immune regulatory signature in fibrosis.** A) Barplot showing log fold change of selected chemokines between fibrotic and dermal stroma. p.adj = adjusted p-value (BH correction). LFC = log fold change. B) Dot plot of enriched pathways in the fibrotic fibroblast meta-signature. Gene ratio and count represent the ratio and absolute number of genes that appear in the gene sets used for pathways analysis. p.adjust = adjusted p-value (BH correction).

IV.IV. Discussion

Myofibroblasts are crucial mediators of fibrosis and this chapter provides evidence that they make up a significant proportion of DD nodules. Myofibroblasts have been characterized in DD and reports have suggested an association between digital contracture and myofibroblast distribution(55). In this Chapter, I use single cell RNA-seq to confirm myofibroblasts make up ~60% of the total population in nodules. Unsupervised clustering and differentially expression analysis revealed a distinct expression profile of myofibroblasts and showed these cells to have the highest expression of genes encoding major ECM proteins (*COL1A1*, *FN1* & *COL3A1*).

One limitation of the data is that the patients profiled are relatively homogeneous, without clear stratification based on clinical parameters such as rate of disease progression or degree of contracture. Moving forward, we should aim to profile DD nodules along a spectrum of disease severity using single cell RNA-seq. Although DD nodules are obtained following surgical resection and therefore represent a later disease stage, patients could still be stratified on clinical parameters such as rate of disease progression. This would enable testing for a correlation between clinical parameters in DD and the expression profiles of myofibroblasts.

Another aim of this Chapter was to present the first detailed molecular description of myofibroblasts at single cell resolution and here I identify a clear expression profile that distinguished myofibroblasts from fibroblasts. Ranking differentially expressed genes (Wilcoxon Rank Sum Test) specific to myofibroblasts by fold change showed the top five were; *FABP5* (LogFC

1.99, p-adjust $8.06e^{-253}$), *POSTN* (LogFC 1.65, p-adjust $8.57e^{-281}$), *ACTA2* (LogFC 1.47, p-adjust $3.51e^{-112}$), *SPON2* (LogFC 1.45 p-adjust $1.33e^{-269}$) and *TAGLN* (LogFC 1.41, p-adjust $6.13e^{-183}$). *TAGLN* encodes transgrelin, an actin-binding protein belonging to the calponin family(227). A recent publication based on a murine model of inflammatory bowel disease reported that *TALGN* marked an *ACTA2* positive myofibroblast population, suggesting conservation across species and anatomical sites(228). *POSTN* encodes periostin, a secreted extracellular matrix protein. Periostin has been associated with many fibrotic conditions including idiopathic pulmonary fibrosis, in which it has been shown to promote myofibroblast activation. In addition, periostin has been demonstrated to regulate myofibroblasts isolated from Dupuytren's patients, influencing apoptosis, proliferation and expression of α-SMA(229). A direct role for periostin in DD mandates further investigation, but the results presented in this chapter highlight several myofibroblast markers, including *TALGN*, that may be regulated by this protein.

FABP5 encodes fatty-acid binding protein 5 (FAPB5), an intracellular lipid carrier and transporter of hydrophobic fatty acids. First described in epidermal cells, FABP5 has recently been reported as upregulated in radiation induced human skin fibrosis(230). Interestingly, in this study overexpression of FAPB5 in human WS1 fibroblasts resulted in the nuclear translocation of SMAD2 and activation of TGF-β1 signalling. As this is the only published report of FAPB5 expression in human fibrosis, an important future direction would be to explore an exact role of FABP5 in DD. Although only descriptive, the data in this chapter support *FABP5* as a specific myofibroblast marker and in the future we should look to define its exact molecular function and whether it may represent a therapeutic target.

In normal physiology, fibroblasts exhibit enormous heterogeneity(231, 232). Their functions extend from providing structural support(233), guiding wound healing(234) to directing immune responses through both antigen presentation and the secretion of cytokines(235). It is therefore unsurprising that fibroblasts have been implicated in a diverse range of diseases in which they perform highly variable functions(231). For example, in rheumatoid arthritis (RA), fibroblasts within the joint are key mediators of inflammation, promoting joint destruction(154). In cancer, fibroblasts are capable of directly binding to and pulling cancer cells, supporting metastatic dissemination(236). In fibrosis, the exact role of resident fibroblasts remains obscure, but the results presented in this chapter suggest these cells could regulate immune cell recruitment. Evidence for this proposed function are threefold. Firstly, the putative immune signature visualized in PCA. Secondly, several key fibroblast marker genes associate with immune pathways including *PLA2G2A, CXCL14* and *C1R*. Finally, ICAM1^{+} fibroblasts were shown to be the major source of stromal-cell derived chemokines, including IL-8 and IL-6, and specifically upregulated in DD nodules. However supportive, these results do not allow for a direct causation of fibroblasts driving inflammation in DD. We must now look to better characterize this intriguing ICAM1^{+} fibroblast and ask, alongside IL-6 and Il-8, which soluble factors they secrete. This, in turn, may reveal key soluble factors driving their function. Considering this, an important next step would be to isolate individual fibroblast subsets (CD34^{+}, PDPN^{+} & ICAM1^{+}) and measure a full panel of cytokines and chemokines and correlate these observations with their transcriptomic profiles. This may uncover highly contrasting protein and gene markers of fibroblasts in DD.

In this chapter I describe three major subsets of fibroblasts in DD and parallels can be drawn with fibroblasts in RA, where distinct subsets appear to contribute diverse pathological processes(154). One contrast to RA, where $CD34^+$ fibroblasts are the major source, is the chemokine-enriched $ICAM1^+$ fibroblasts in DD(154). With increasing single cell datasets on stromal cells, it will be important to accurately integrate and compare molecular signatures of fibroblasts. In many diseases, including fibrosis and cancer, core biological processes such as inflammation and proliferation are paramount in disease progression. Single cell biology will enable mapping of these crucial processes to distinct cell subsets.

What must be remembered is that fibroblasts may adopt a panoply of functions in fibrosis, as in RA. Although my results suggest $ICAM1^+$ fibroblasts may be an important driver of immune cell recruitment they may also have a completely undiscovered but vital role in DD. In line with this, pathway analysis of $ICAM1^+$ fibroblast marker genes included immune pathways, but also several other associations including cell migration and ECM remodelling. Here, it is pertinent to consider limitations of techniques such as pathway analysis that are composed of annotated gene sets initially conceived in completely different biological systems. Thus, alongside a focused exploration of $ICAM1^+$ fibroblasts role in inflammation, other key pathogenic processes need to be investigated to uncover possible hidden functions.

To conclude this chapter, I sought to integrate bulk RNA-seq signatures of fibrotic (DD nodular) and physiological (dermal fibroblasts) stromal cells with cell type specific profiles identified using single cell RNA-seq. Importantly, there are limitations that must be considered when drawing

conclusions from these observations. Firstly, there remains no universally accepted control tissue for Dupuytren's disease; dermal fibroblasts are not the true healthy counterpart to Dupuytren's nodules. Secondly, bulk RNA-sequencing was performed on early passage (Passage 1) cells to obtain enough material for sequencing. With consideration to these factors, bulk RNA-sequencing did identify genes upregulated between dermal and fibrotic stromal (DD nodular) cells that could be deconvoluted using the single cell RNA-seq data. Of note, *ACTA2*, the canonical marker of myofibroblasts, was a shared marker of myofibroblasts (scRNA-seq profiles) and upregulated in fibrotic stromal cells (bulk RNA-seq), providing some support for the robustness of the meta-signatures constructed. Other genes associated with this fibrotic myofibroblast meta-signature include *MAFB* and *FAPB5*. *MAFB* encodes Mafb, a basic leucine zipper transcription factor. Interestingly, a recent genome-wide association study showed *MAFB* to be associated with a single nucleotide polymorphism in DD(237). In addition, *MAFB* was reported as a significantly upregulated gene in Dupuytren's cords as compared to palmar fascia in a microarray(238). More work is required to tease apart a role of this transcription factor in DD, but previous GWAS and microarray data, in concert with findings in my single cell RNA-seq, suggest it may be an important transcriptional regulator of myofibroblasts in this disease.

The bulk RNA-sequencing results demonstrated that many chemokine and pro-inflammatory genes showed higher expression in fibrotic stromal cells (DD nodules) as compared to dermal fibroblasts. Pathway analysis also showed an enrichment for immune regulatory pathways in fibroblasts derived from fibrotic tissue. One speculation from this result could be that the fibrotic microenvironment primes stromal cells to adopt a pro-inflammatory signature, increasing the

expression of chemokines. My results provide only indirect evidence for this but do show that chemokines are differentially expressed between the two cellular types. To directly test this hypothesis, one could isolate primary human fibroblasts and perform perturbation experiments. Altering the mechanical and biochemical environment *in vitro* may offer answers to how transformations in native tissue ecosystems specifically influences fibroblast phenotype. One final question is how this immune regulatory phenotype may arise. One possibility is that fibroblasts are inherently primed to adopt this phenotype in fibrosis, and perhaps is the initial trigger of the disease. An alternative is that fibroblasts are primed with this signature as the fibrosis develops, perpetuating disease progression.

In summary, the results presented here identify significant marker genes of fibroblasts and myofibroblasts in DD nodules and resolve three major subsets of fibroblasts. Using flow cytometry, I show that one subset, $ICAM1^{+}$ fibroblasts, exhibit different proportions in two distinct fibrotic structures. Finally, I use bulk RNA-seq to inform potential key phenotypic markers of fibroblasts and myofibroblasts in DD and deconvolute these signatures to highlight genes that may be upregulated within the fibrotic microenvironment.

In the next chapter I aim to undertake a focus investigation of myofibroblasts in DD. My goal is to isolate this population and apply unsupervised clustering and dimensional reduction to the dataset in the hope of illuminating and validating novel cell surface markers and distinct sub-populations.

Chapter V. Classification of the myofibroblast in human fibrosis

V.I. Introduction

The cellular landscape of fibrosis spans a continuum of multiple cell types(1, 28, 165). Myofibroblasts represent the predominant mesenchymal cell, but the fibrotic milieu also contains immune, endothelial and perivascular cells, as described in Chapter III. Single cell genomics can resolve not just distinct cell types, but also discrete cell states along developmental pathways and uncover genes driving phenotypic transitions in the data(156, 239, 240).

Molecular descriptions of myofibroblasts loosely define an "activation" paradigm along a differentiation path towards a fully pathogenic phenotype(64, 215). Here, a precursor population termed a "proto-myofibroblast" is identified as a cell primed to adopt a mature myofibroblast phenotype(60, 74). The precise features of proto-myofibroblasts are unclear, but it is characterised by the upregulation of α-SMA that is yet to be incorporated in F-actin stress fibres(8, 60, 214). Adding to the complexity of myofibroblast ontogeny, almost every cell type has been reported as myofibroblast precursors, including endothelial, epithelial, myeloid and resident stromal cells(54, 60, 182, 241, 242). In DD, the exact signature of myofibroblast activation is unknown.

Despite their role as key mediators of fibrosis, myofibroblasts remain a poorly described and enigmatic population. We have a collection of putative markers for this cell type but are lacking a robust and comprehensive description of their molecular phenotype in human disease. Moreover,

we currently do not know if distinct myofibroblast communities exist and how this diversity may relate to fibrosis pathogenesis.

V.II. Hypothesis and Aims

My hypothesis is that the fibrotic microenvironment houses distinct communities of myofibroblasts.

In this chapter I aim to profile myofibroblasts at single cell resolution using the fibrotic atlas generated in Chapter III. My broad goal was to identify novel markers of myofibroblasts and ask whether they represent a heterogeneous population harbouring distinct communities. To achieve these aims, I integrated single cell RNA-seq and multi-parameter flow cytometry to elucidate a novel myofibroblast state, describe a transcriptomic signature of myofibroblast activation, and validate my findings experimentally. In addition, I uncover a novel myofibroblast cell surface marker and map a distinct myofibroblast subset on to a murine model of lung fibrosis.

V.III. Results

V.III.I *Resolving distinct myofibroblast communities in DD*

Myofibroblasts are central mediators of the dysregulated wound-healing program that defines fibrosis(45, 243) and I sought to study this population in detail. First, I performed the same computational workflow as applied to delineate major cell types in DD nodules (see Chapter III.III.III). This involved selecting out the myofibroblast cluster in the data and performing quality control to remove low quality cells. Next, I undertook feature selection by defining highly variable genes (*n*=1,865) and significant principal components (PCA) and used these as input for graph-based clustering. Unsupervised clustering of the SNN graph using the Louvain algorithm (*k*=30, resolution=0.3) resolved four clusters including a proliferative population (Figure 1A-B). One cluster was characterized by lower expression of *ACTA2* and intermediate expression of fibroblast marker genes (*CXCL14, PLA2G2A, C1R*) that I term the $ACTA2^{low}$ myofibroblast. A second cluster was a highly conserved $CD82^{+}$ population (*OX40L, MMP11, MMP14, COL6A1* and *CD82*) (Figure 2-3). Finally, there was a third less distinct community that I designated the intermediate myofibroblast. Visualizing the top 10 gene markers readily demonstrated that the first three clusters ($ACTA2^{low}$, $CD82^{+}$ and proliferative) where highly distinct, in contrast to the intermediate myofibroblasts, which appeared more heterogeneous (Figure 1C). I noted that these distinct communities were prominent in all 12 patient samples and aligned datasets, suggesting a core molecular signature of myofibroblasts in DD (Figure 1B).

The intermediate myofibroblasts represented 62.0% (51.7-71.7%), whilst the CD82^{+} subset constituted 24.2% (13.1-31.5%) and the ACTA2^{low} subset made up 26.7% (26.7-41.3%).

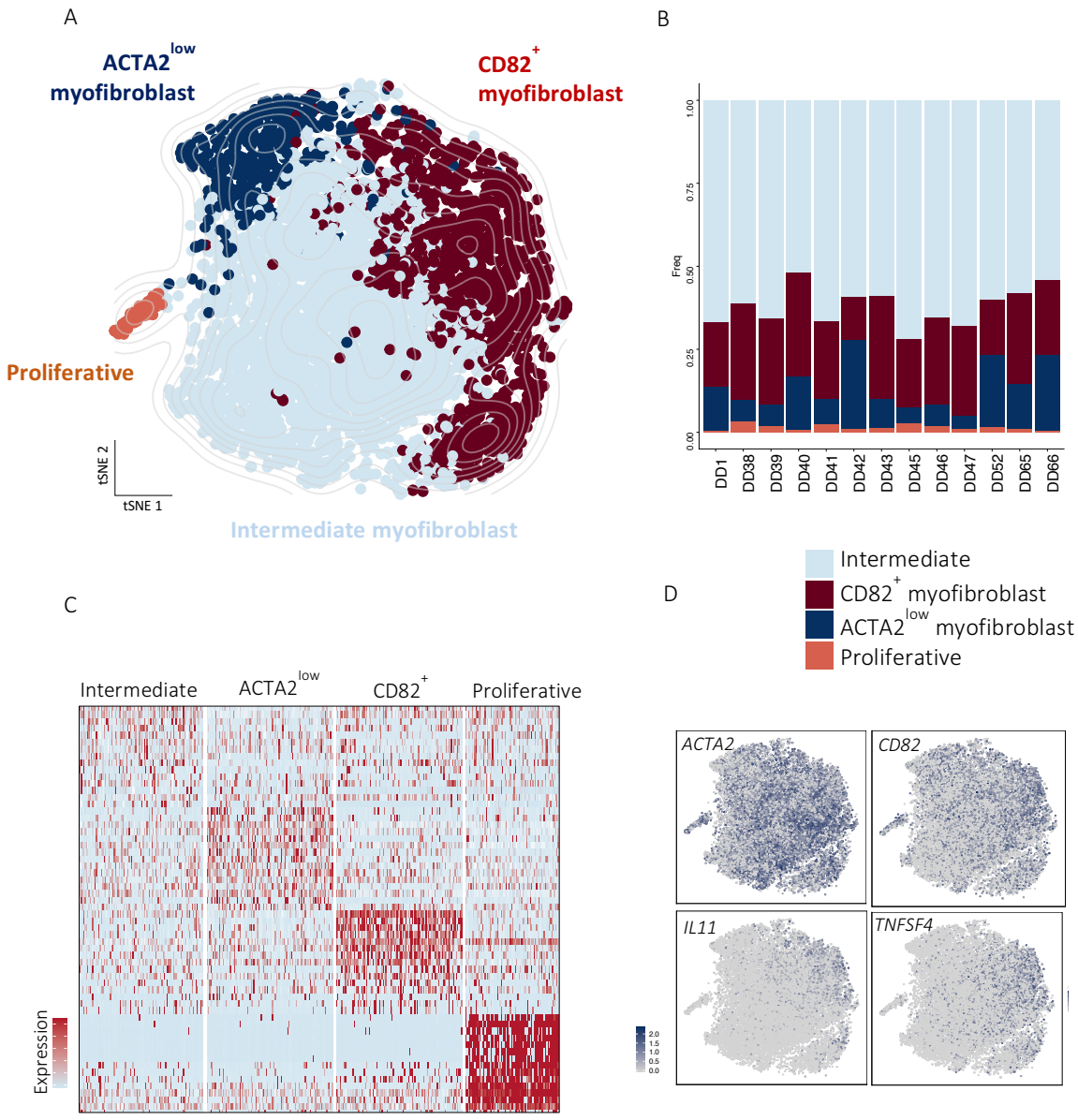


Figure 1. **Four transcriptional communities of human myofibroblasts.** A) tSNE projection of human myofibroblasts showing four communities coloured by Louvain cluster assignment. B) Barplot showing the proportions of each myofibroblast community in DD patients. C) Heatmap showing the z-score expression of the top 10 marker genes of each myofibroblast community. D) tSNE projections of myofibroblasts coloured by the expression CD82^{+} myofibroblast markers. Expressed in scaled log(UMI+1). DD is a prefix for DD patient IDs (DD1-DD65). Freq = Frequency (Proportion of all cells sampled). *n*=12 DD patients.

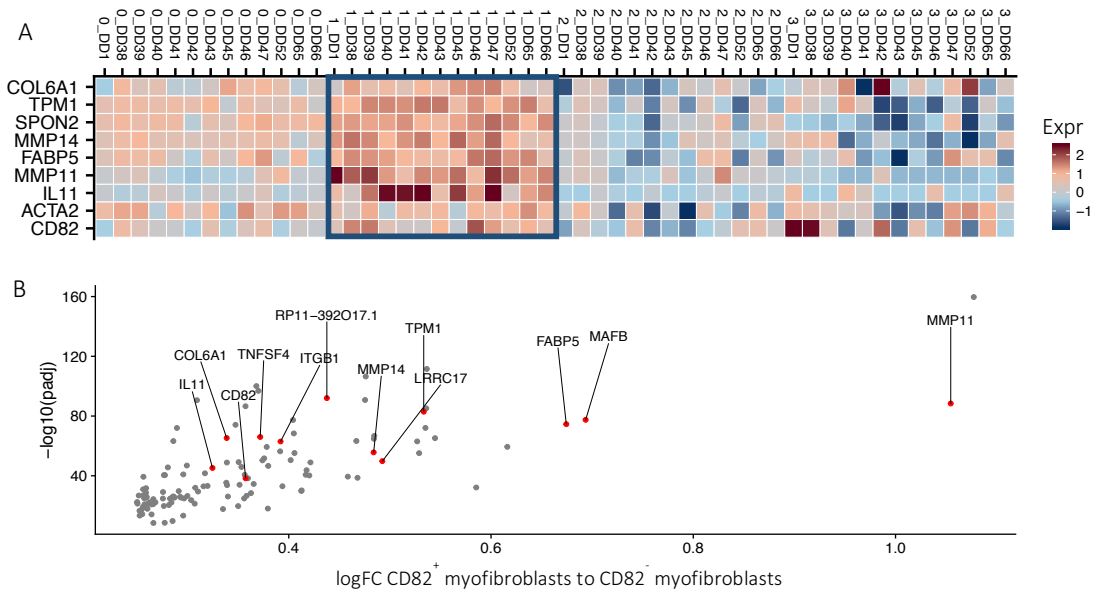


Figure 2. **Gene markers of novel CD82$^+$ myofibroblast.** A) Heatmap showing expression of selected CD82$^+$ myofibroblast markers demonstrating conservation of markers in the same cluster (1) across patient samples. Expression in z-score scaled log(UMI+1). At top, prefix 0-3 denotes cluster and suffix DD1-DD65 represents DD patient IDs. B) Scatterplot showing top positive markers for CD82$^+$ myofibroblast population. FDR-adjusted p values (BH correction) and log fold change plotted. Expr = Expression.

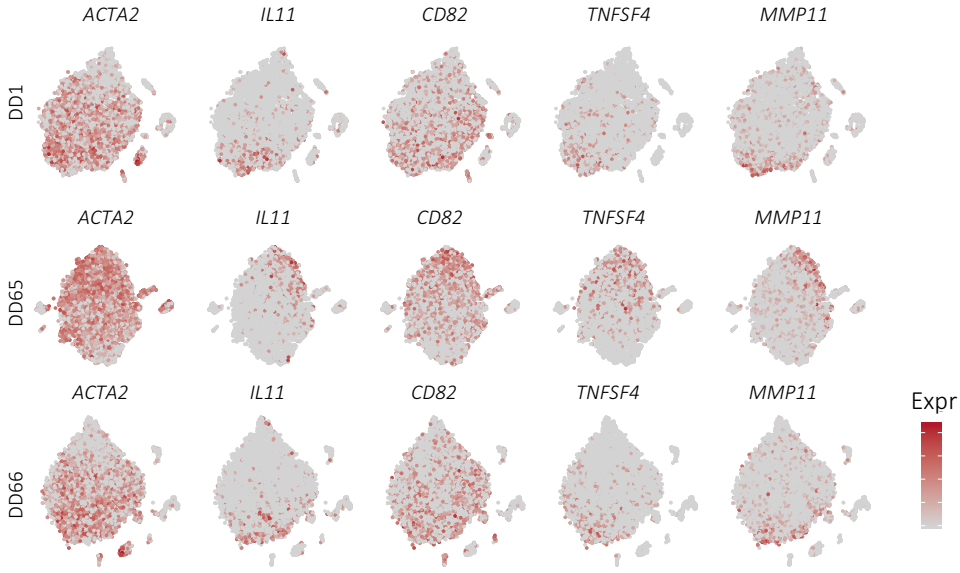


Figure 3. **Highly conserved CD82$^+$ myofibroblast in Dupuytren's nodules.** tSNE projections for selected markers of CD82$^+$ myofibroblast in three representative DD patients (DD1, DD65, DD66). Expression in scaled log(UMI+1). Expr = Expression.

V.III.II *The cycling stromal cell in DD is a myofibroblast*

On closer inspection, I confirmed a small proliferative cluster (Mean=1.5%,3.27-4.17%) in the myofibroblast compartment (Figure 4). This subset expressed pure myofibroblast markers such as *ACTA2* and *ITGB1*, but lacked expression of pericyte markers (e.g. *MCAM, RGS5*). After finding this proliferative population, I sought to investigate it further.

The small cycling cluster demonstrated high expression of many cell cycle genes, including *MKI67,* alongside several classical myofibroblast markers. Looking within the fibroblast compartment, I could not identify any similar proliferative cluster. Next, I calculated a cell cycle phase score of each myofibroblast cluster based on the expression of canonical cell cycle specific genes. This was corrected for differences in gene expression between cells using a control set as described in Chapter II.II.X. Then, I projected the S Phase (Proliferative) score for each cluster along a smoothed density plot, whereby computing and drawing a kernel density estimate (Figure 4D). This demonstrated that the small proliferative cluster dominated the overall signal for a proliferative (S Phase) signature within myofibroblasts, suggesting that within DD nodules the cycling population is discrete with other cells remaining quiescent.

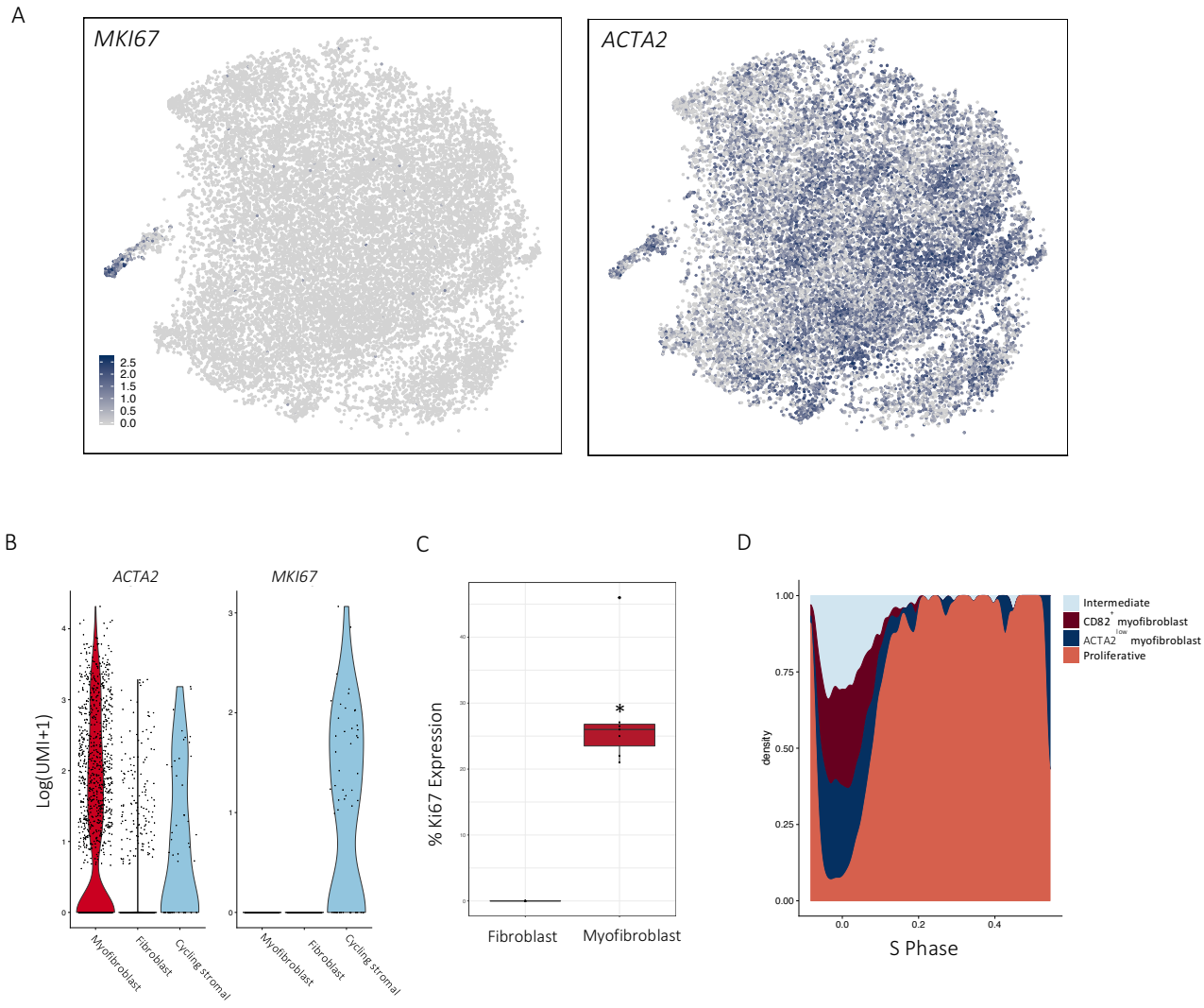


Figure 4. **Distinct cycling myofibroblast in Dupuytren's nodules.** A) tSNE projections of human myofibroblasts coloured by expression of *MKI67* and *ACTA2* highlighting distinct cycling cluster. Expression in scaled log(UMI+1). *n*=12 DD patients. B) Violin plots showing expression of *MKI67* and *ACTA2* in myofibroblasts, fibroblasts and cycling stromal cell clusters. Expression in log(UMI+1). *n*=12 DD patients. C) Box and whisker plot of flow cytometry analysis showing Ki67 protein expression in myofibroblast (ITG-$\beta1^{high}$) and fibroblasts (ITG-$\beta1^{low}$). %Ki67 Expression = percentage of stromal cells expressing Ki67 as determined by flow cytometry. D) Conditional density plot of Gaussian Kernel density of S phase score in myofibroblast clusters. Small cycling cluster dominants the overall signal. * = p-value < 0.001 (Paired t-test).

Subsequently, I sought to validate the identity of this cycling stromal cell. Using flow cytometry, I isolated major cell types in freshly disaggregated nodules using canonical protein markers (CD45, CD31, CD146, ITG-β1) as described in Chapter II.II.I. ITG-β1 was a specfic marker for myofibroblast in the single cell expression data and was highly expressed in freshly isolated cells, therefore, I used this protein to define myofibroblasts (ITG-$\beta1^{high}$) and fibroblasts (ITG-β^{low}) co-staining for Ki67 (Figure 5). In accordance with the single cell RNA-seq, the $KI67^{high}$ cell population was in the myofibroblast compartment, expressing high protein levels of ITG-β1 and negative for CD45, CD31 and CD146. In contrast, I did not observe a $KI67^{high}$ population in the $CD34^{+}$ fibroblast used as a comparison. In summary, these results suggest that in this archetypal localised fibrotic disease, proliferating stromal cells ($Ki67^{+}$) are housed within the ITG-$\beta1^{high}$ myofibroblast compartment. I noted a variation in Ki67 protein (~20%) and gene (~1-2%) expression in the myofibroblasts, which may relate to the low sequencing depth of droplet based single cell RNA-seq missing lowly expressed nuclear genes.

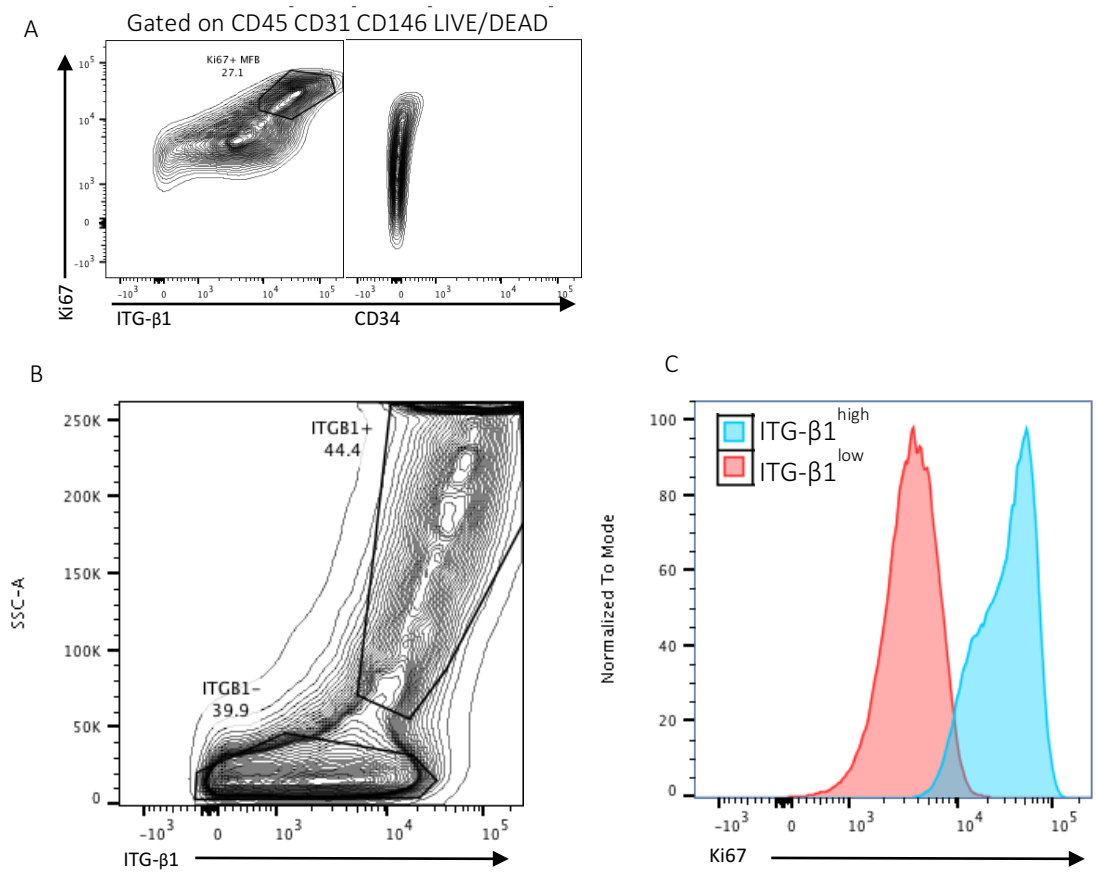


Figure 5. **Flow cytometry localizes cycling stromal cell to the myofibroblast compartment in Dupuytren's nodules.** A) Representative density plot of flow cytometry analysis showing Ki67 protein expression in myofibroblast (ITG-$\beta1^{high}CD34^{-}$) and fibroblast (ITG-$\beta1^{low}CD34^{+}$) compartments. MFB = myofibroblast. B) Density plot of ITG-β1 surface expression showing gating strategy for identifying myofibroblasts (ITG-$\beta1^{high}$) and fibroblasts (ITG-$\beta1^{low}$). C) Histogram for Ki67 protein expression in gated populations from (B) showing higher expression in ITG-$\beta1^{high}$ myofibroblast.

V.III.III *Identifying a myofibroblast activation trajectory*

The overall topography of myofibroblast clusters suggested an overarching trajectory in which the two distinct cell communities ($ACTA2^{low}$ & $CD82^{+}$) represent diverging ends of a continuum separated by an intermediate cellular state (intermediate myofibroblast). This concept reflects a central principle in myofibroblast biology, which describes an "activation" paradigm along a differentiation path towards a fully pathogenic phenotype(53, 241, 242). $CD82^{+}$ myofibroblasts had a distinct transcriptomic signature and showed high expression of *ACTA2*. $ACTA2^{low}$ myofibroblast were also well-defined but in contrast showed a low expression of *ACTA2* and relatively high expression of fibroblast marker genes such as *CXCL14*. (Figure 1C). Finally, intermediate myofibroblasts were more heterogeneous and shared markers of both more discrete subsets. This again fits with a model in which the intermediate myofibroblast represents a cell undergoing some form of phenotypic transition. Given this, I interrogated myofibroblasts in DD to gain a better understanding of any overarching structure or developmental path.

Projection of the myofibroblast data across principal components revealed a strong association between the first principal component (PC1) and *ACTA2* expression. Looking closer, it appeared that PC1 captured a putative myofibroblast differentiation or activation trajectory moving from the $ACTA2^{low}$ myofibroblasts towards $CD82^{+}$ myofibroblasts (Figure 6). After this, I defined gene modules of the two myofibroblasts subsets ($CD82^{+}$ & $ACTA2^{low}$) as described in Chapter II.II.X, which demonstrated a positive (R=0.46) and negative correlation (R=-0.71) to PC1, respectively. Together, PCA points to myofibroblasts existing along a linear or branched continuum in fibrosis.

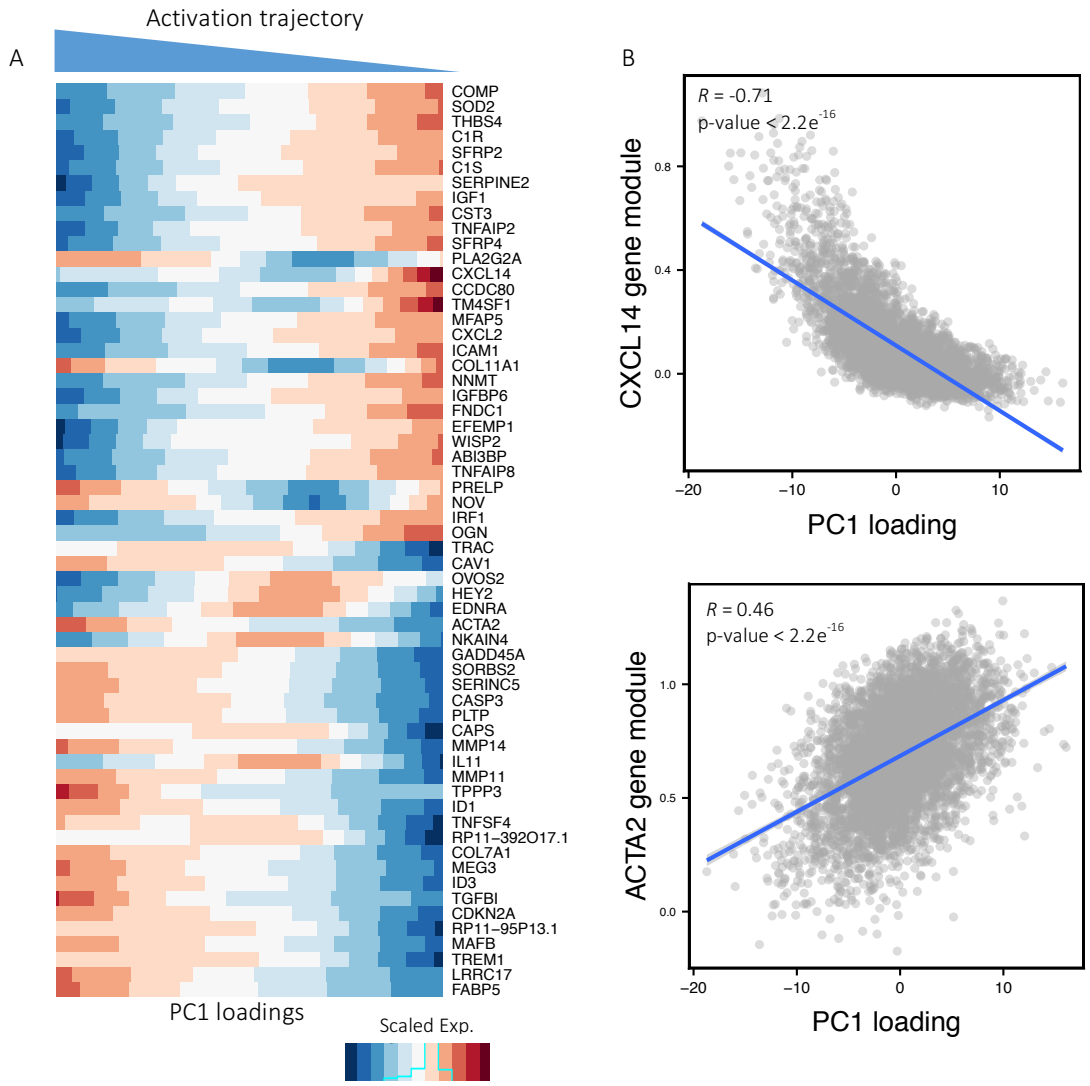


Figure 6. **Principal component analysis captures a putative myofibroblast differentiation path.** A) Heatmap of scaled z-score expression of 50 genes with top PC1 loadings. Colour represents scaled z-score expression and gene expression has been loess smoothed along rows. B) Scatterplot with linear regression line showing correlation between PC1 loading and *ACTA2* gene module (myofibroblast differentiation program) and *CXCL14* gene module (fibroblast differentiation program). PC = principal component. *R* = Pearson correlation coefficient.

To better understand this putative myofibroblast activation signature, I next used diffusion maps to reconstruct a potential developmental path (Figure 7). In accordance with PCA, the first two

diffusion map components(223, 244) captured a robust trajectory associated with increasing *ACTA2* expression (Figure 7).

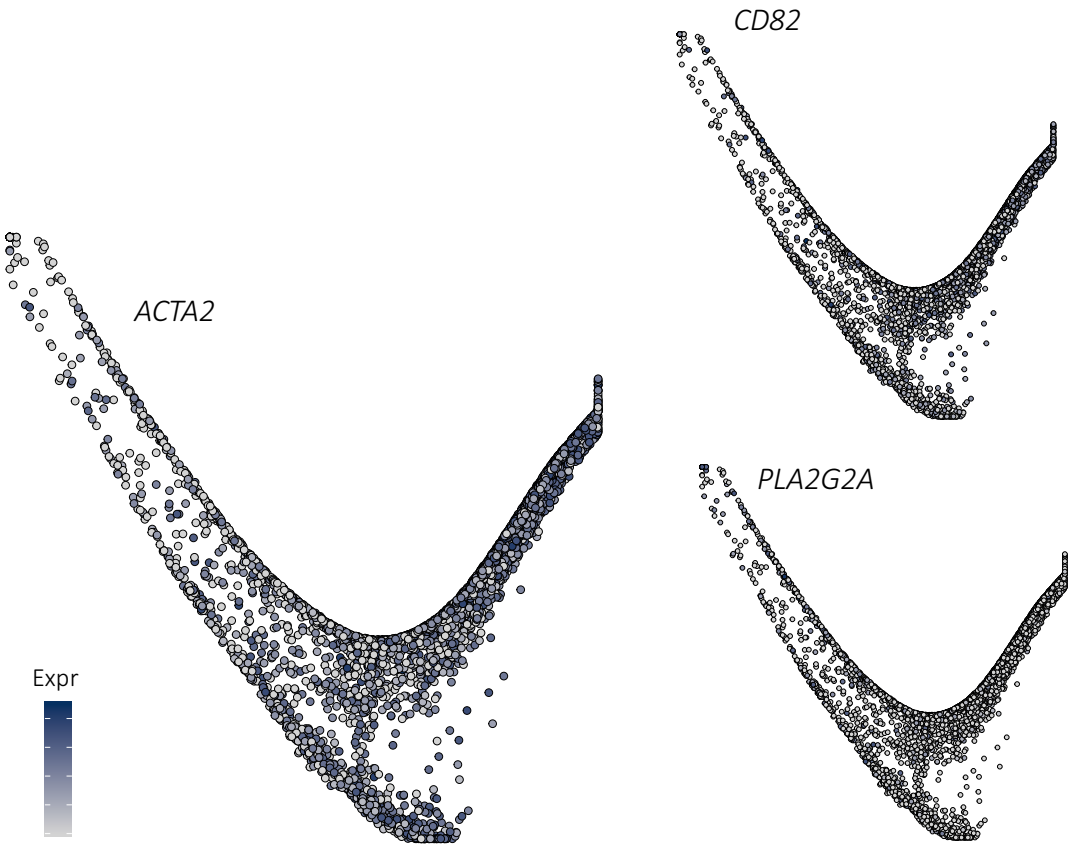


Figure 7. **Diffusion maps reconstruct myofibroblast differentiation trajectory.** Diffusion map embeddings for myofibroblasts coloured by expression of listed genes. Gene expression in scaled log(UMI+1). Diffusion map projection represents first and second diffusion coordinates. Expr = expression. Down sampled to 5,000 cells for visualization only.

Along the first two diffusion components, I observed a gradual increase in the expression of $CD82^{+}$ myofibroblast markers (e.g. *CD82*) and a corresponding decrease in the expression of $ACTA2^{low}$ myofibroblast markers (e.g. *PLA2G2A*). Fitting a principal curve(150) through the first two diffusion

components enabled me to systematically define a dynamic gene module along this continuum (see Chapter II.II.X). Principal curves represent a non-linear smoothing through a *n*-reduced dimension dataset. Again, along the modelled pseudotime from the principal curve, I observed a decrease in ACTA2^{low} myofibroblast markers and gradual increase in CD82^{+} myofibroblast markers (Figure 9). Finally, constructing a *k*-NN graph of cell-cell correlations (Spearman), whereby removing a dependence on PCA meta-signatures for clustering, on to a force directed graph (Fruchterman-Reingold layout) demonstrated a clear developmental trajectory (Figure 8).

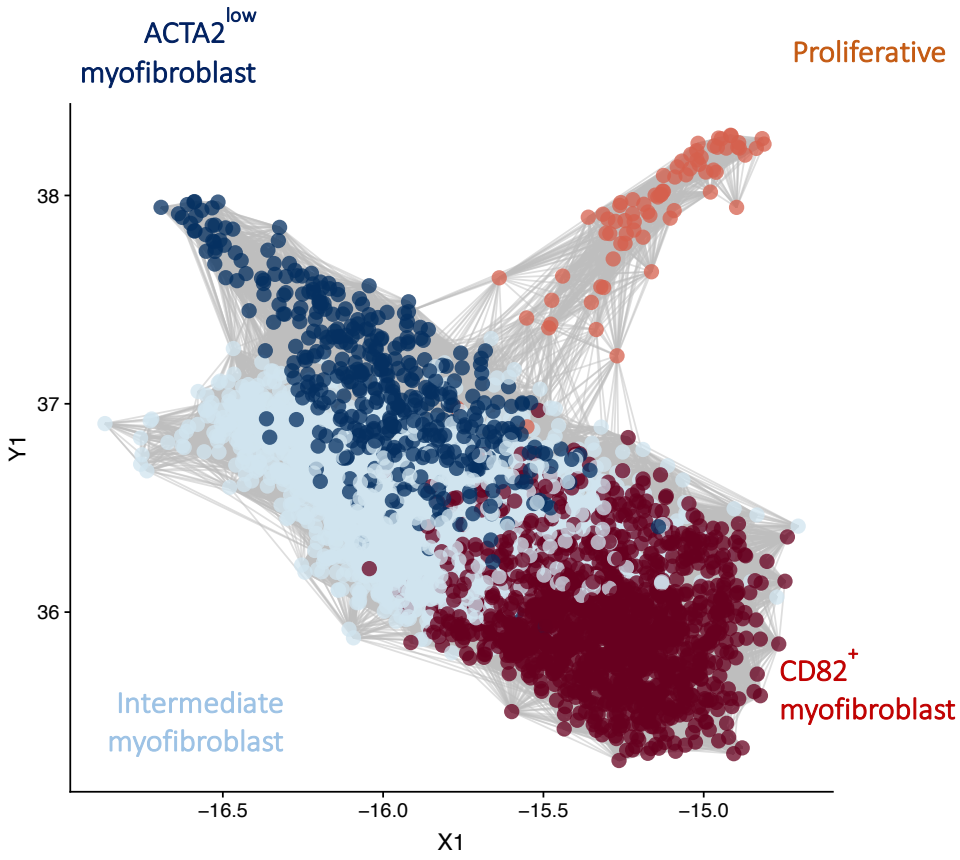


Figure 8. **Force directed graph of human myofibroblasts.** Force directed graph (Fruchterman-Reingold layout) of *SNN* graph of ~30,000 myofibroblasts from 12 DD patients coloured by myofibroblast communities (Louvain clusters). Down sampled to 5,000 cells for visualization only.

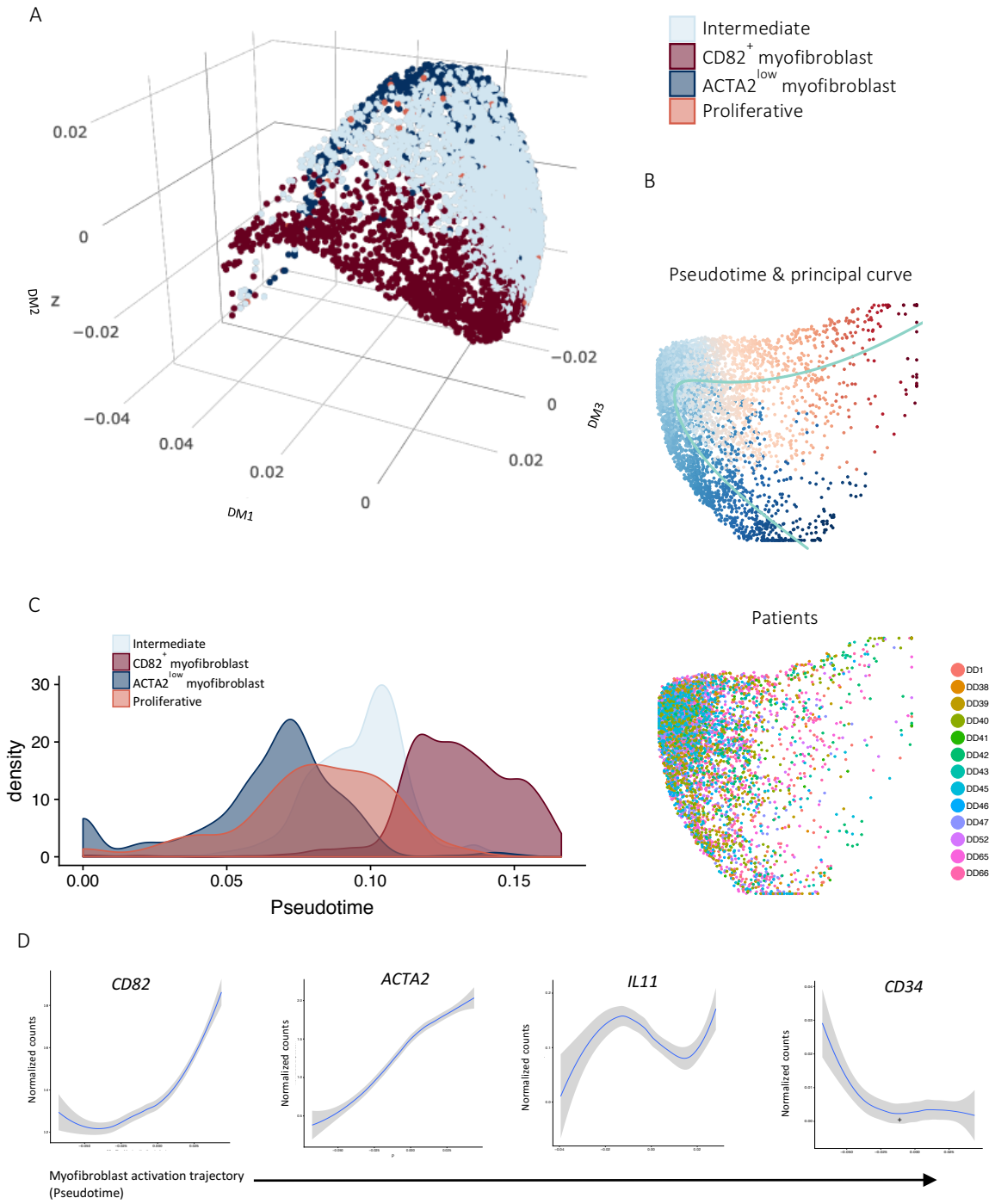


Figure 9. **Pseudotemporal analysis of human myofibroblasts using diffusion maps.** A) Diffusion map embedding of human myofibroblasts coloured by Louvain clusters. DM= diffusion map coordinates. B) Diffusion map embedding of human myofibroblasts coloured by patient donors (*n*=12 DD patients*)* and pseudotime with principal curve fit. Down sampled for visualization only. (*k*=5,000 cells). C) Density plot of myofibroblast clusters along pseudotime moving towards CD82^{+} myofibroblasts. Data represent Gaussian kernel probability densities. D) Line plots of *loess* smoothed expression and standard error in normalized counts of selected genes along pseudotime. DD1-DD65 represent DD patient IDs (*n*=12).

V.III.IV *Elucidating a novel myofibroblast marker, CD82*

The membrane tetraspanin *CD82* was highly associated with the putative myofibroblast activation trajectory and *ACTA2* expression in the myofibroblast dataset. Given this, I aimed to validate CD82 as cell surface marker of this population. First, I performed immunofluorescence co-staining for established myofibroblast markers (α-SMA & ITG-β1) and CD82 in DD nodule sections. This revealed a tight co-expression of these proteins, with highest expression of CD82 in myofibroblast rich foci (Figure 10-11).

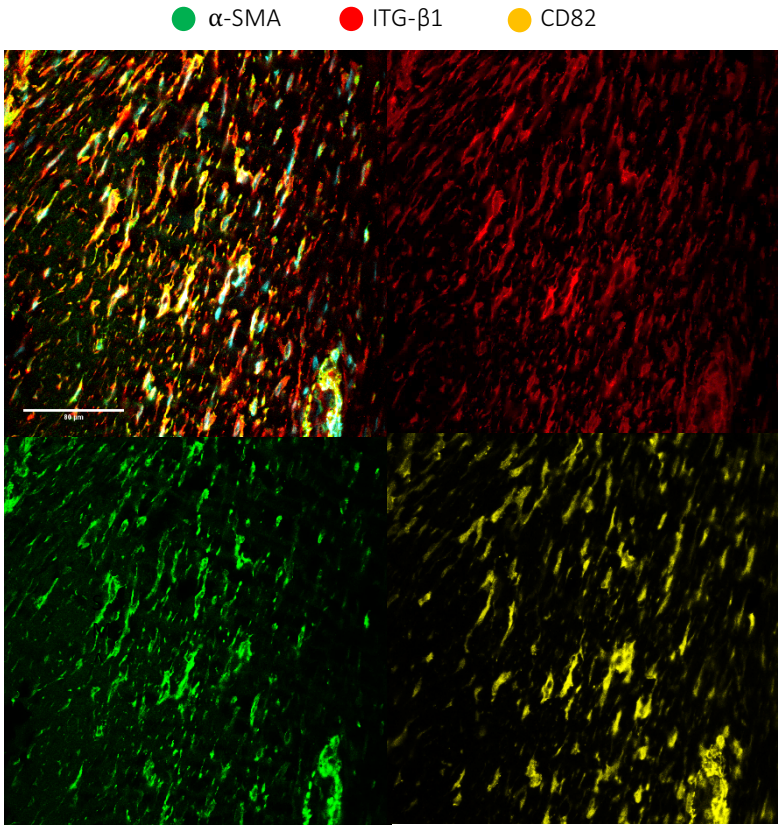


Figure 10. **CD82 expression in Dupuytren's nodules.** Confocal image of immunofluorescence validation of co-expression of CD82, α-SMA and ITG-β1 in Dupuytren's nodule sections. Scale bar 20 µm.

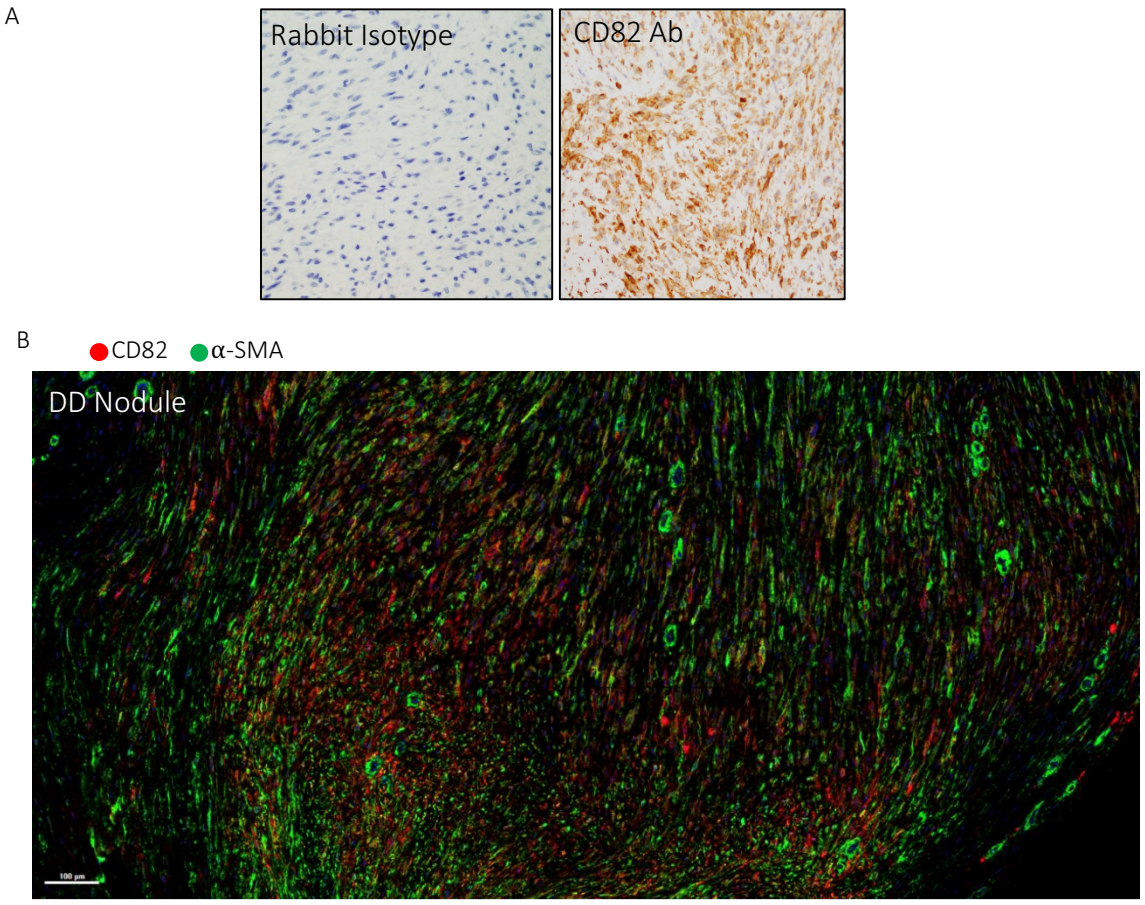


Figure 11. **CD82 Expression in Dupuytren's nodules.** A) Validation of CD82 Rabbit Ab (a50909) with rabbit isotype control using immunohistochemistry staining of Dupuytren's nodule section. x40 magnification. B) Confocal image of immunofluorescence validation for co-expression of CD82 and α-SMA in Dupuytren's nodule sections. Scale bar 100 µm.

After identifying CD82 as a myofibroblast marker *in situ,* I explored whether there was a correlation between the protein expression of this tetraspanin and established myofibroblast markers, as suggested by the pseudotemporal analysis in Chapter V.III.III. Here, I found *CD82* to be highly correlated with *ACTA2* and *ITGB1* gene expression. Proving this at the protein level would provide further support for CD82 as a marker of myofibroblast differentiation in DD. In this regard, I performed flow cytometry in freshly disaggregated DD nodule cells, co-staining for CD82 and

myofibroblasts markers α-SMA and ITG-β1, with an example fibroblast marker (CD34) used as a comparator (Figure 12).

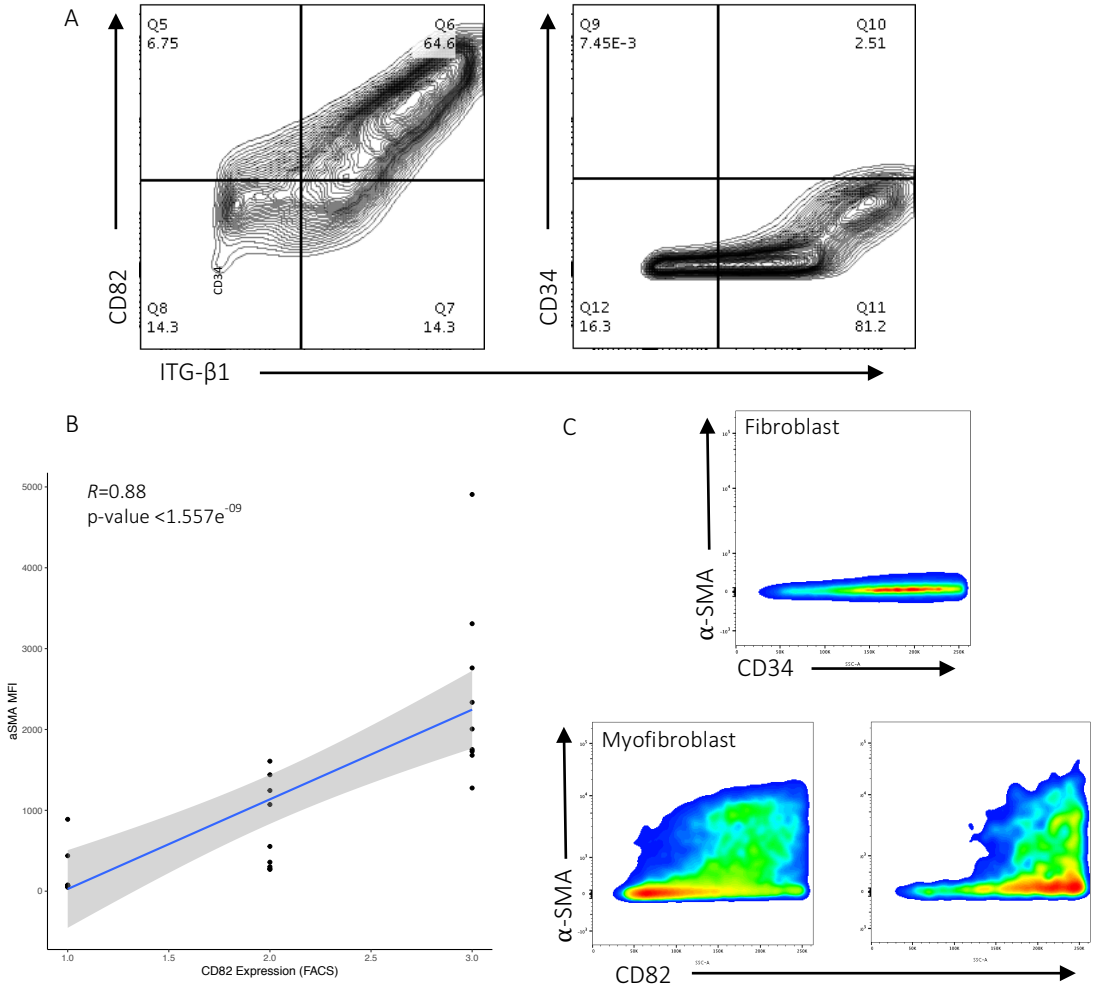


Figure 12. **CD82 protein expression is highly correlated with established myofibroblast markers.** A) Representative density plots of flow cytometry analysis showing positive correlation between CD82 and ITG-β1, but not CD34. B) Scatter plot with linear regression line +/- standard of flow cytometry analysis showing strong positive correlation of CD82 and α-SMA. Data points represent mean fluorescence intensity (MFI) of α-SMA in $CD82^{low}$, $CD82^{mid}$ and $CD82^{high}$ myofibroblasts. These were selected as the top, middle and bottom 20% of cells based of CD82 expression (MFI), respectively. *n*=8 DD patients. *R*=Pearson correlation coefficient. C) Representative density plots of flow cytometry analysis showing co-expression of α-SMA and CD82 in DD myofibroblasts. CD34 used an example fibroblast marker.

From the flow cytometry analysis, mean fluorescent intensity of CD82 and α-SMA were highly correlated (*R*=0.88), indicating that CD82 could be a key differentiation marker in DD myofibroblasts, in accordance with the observations made in the transcriptomic data. Together, the immunohistochemistry and flow cytometry results provide direct evidence for CD82 as a novel myofibroblast marker in DD.

V.III.V *Expression of CD82$^+$ subset markers in vivo*

After characterising the novel myofibroblast marker and subsets I was interested in spatially mapping the distinct CD82$^+$ population *in situ.* DD offers a unique lens to study *en bloc* resection samples without a reliance on small biopsy specimens. Histologically, DD nodules are heterogeneous structures with irregular clusters of myofibroblasts arranged in densely packed foci. These could be readily identified when staining for markers of myofibroblast activation such as β-catenin, ITG-β1 and α-SMA (Figure 13). Using immunohistochemistry, I then looked for the protein expression of selected CD82$^+$ myofibroblast markers (Figure 14).

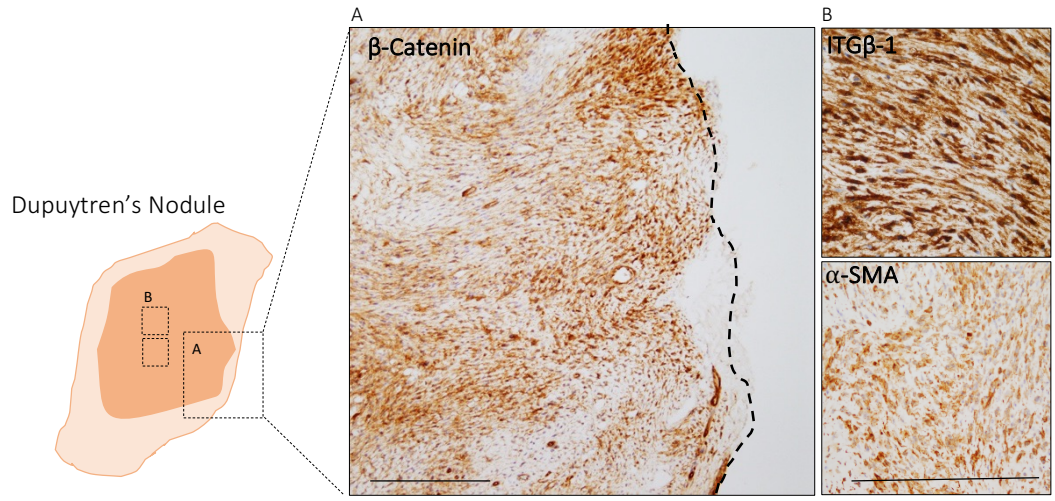


Figure 13. **Dupuytren's nodule cellular architecture.** Schematic of Dupuytren's nodule with representative immunohistochemistry staining for established myofibroblast markers illustrating the distribution of myofibroblasts in nodules. Scale bar is 50 µm.

I found CD82^{+} myofibroblast markers (e.g. MMP11, CD82, MMP14) show high expression in densely packed, α-SMA^{+} myofibroblast-rich foci (Figure 14).

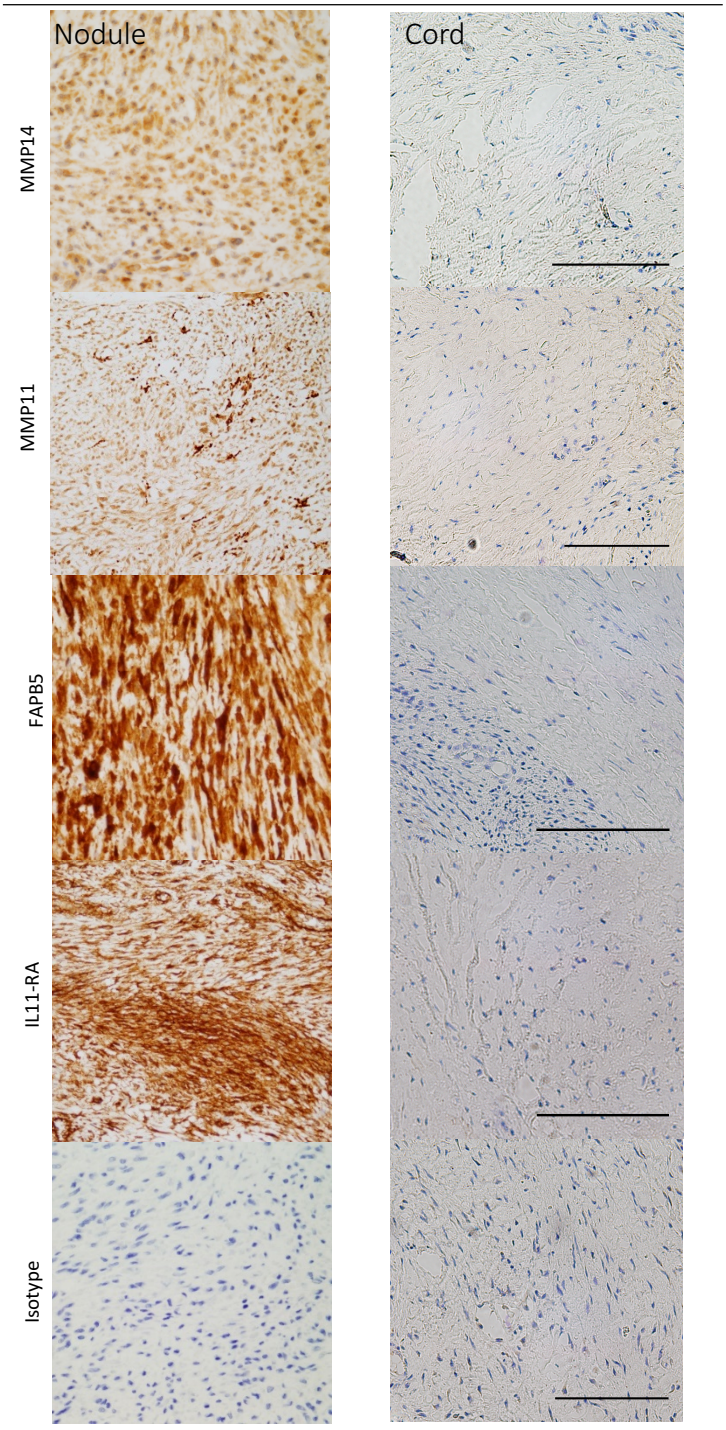


Figure 14. **CD82^{+} myofibroblast markers are highly expressed within Dupuytren's nodules.** Immunohistochemistry staining of Dupuytren's nodule and cord for selected markers of novel CD82^{+} myofibroblast population with isotype control (Rabbit IgG). Scale bar 20µm.

V.III.VI *CD82 function in myofibroblasts*

Together, the previous results in this chapter support the role of CD82 as a marker of myofibroblast differentiation, with increased expression along a developmental path towards a fully activated phenotype. However, these descriptive observations fail to define a molecular function for CD82 in DD. Therefore, I next investigated how CD82 may regulate the myofibroblast phenotype in DD. I collected myofibroblasts from seven DD donors up to passage 2 to limit artefactual changes introduced during *in vitro* culture. I then performed siRNA mediated knockdown of *CD82* and profiled gene expression using bulk RNA-sequencing. Firstly, I validated downregulated of CD82 protein expression following siRNA knockdown and confirmed significant (p-value $1e10^{-16}$) downregulation of *CD82* gene expression (Figure 15). Given the stable nature of some mRNA molecules, I decided on a two-step siRNA-mediated knockdown and collected cells for RNA extraction at day six post first transfection (see Chapter II.II.XX).

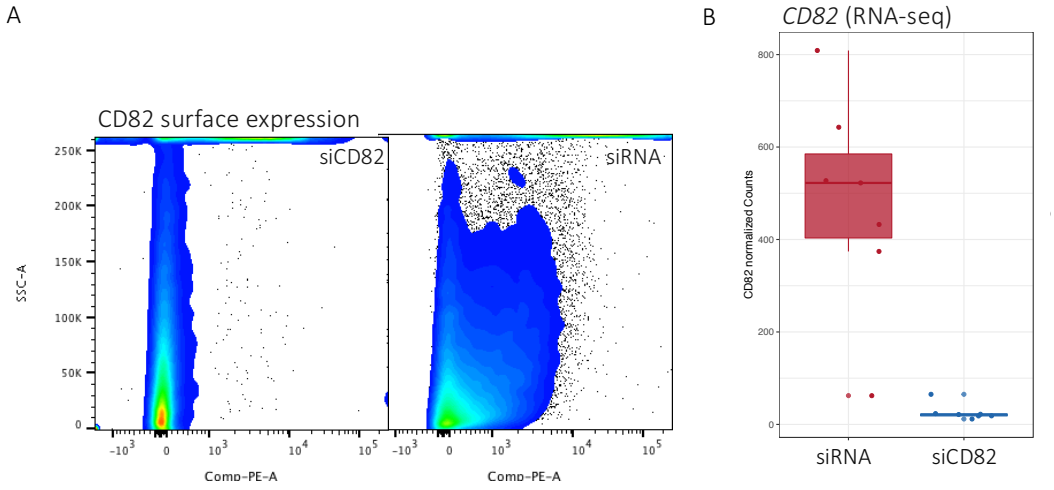


Figure 15. **Validation of CD82 knockdown in primary human myofibroblasts**. A) Representative scatterplot of flow cytometry validation of CD82 surface expression in human myofibroblasts following siCD82 transfection (siRNA used as a control). Comp-PE = CD82. B) Box and whisker plot of *CD82* normalized counts in RNA-seq for siCD82 and siRNA transfections (*n*=7 DD patients).

The RNA-seq dataset was analysed using the DESeq2 model as described in Chapter II.II.X (Figure 16). This demonstrated just 22 genes significantly upregulated following *CD82* knockdown (p-adjust < 0.05 & LFC > 1) and 453 significantly downregulated (p-adjust < 0.05 & LFC < 1).

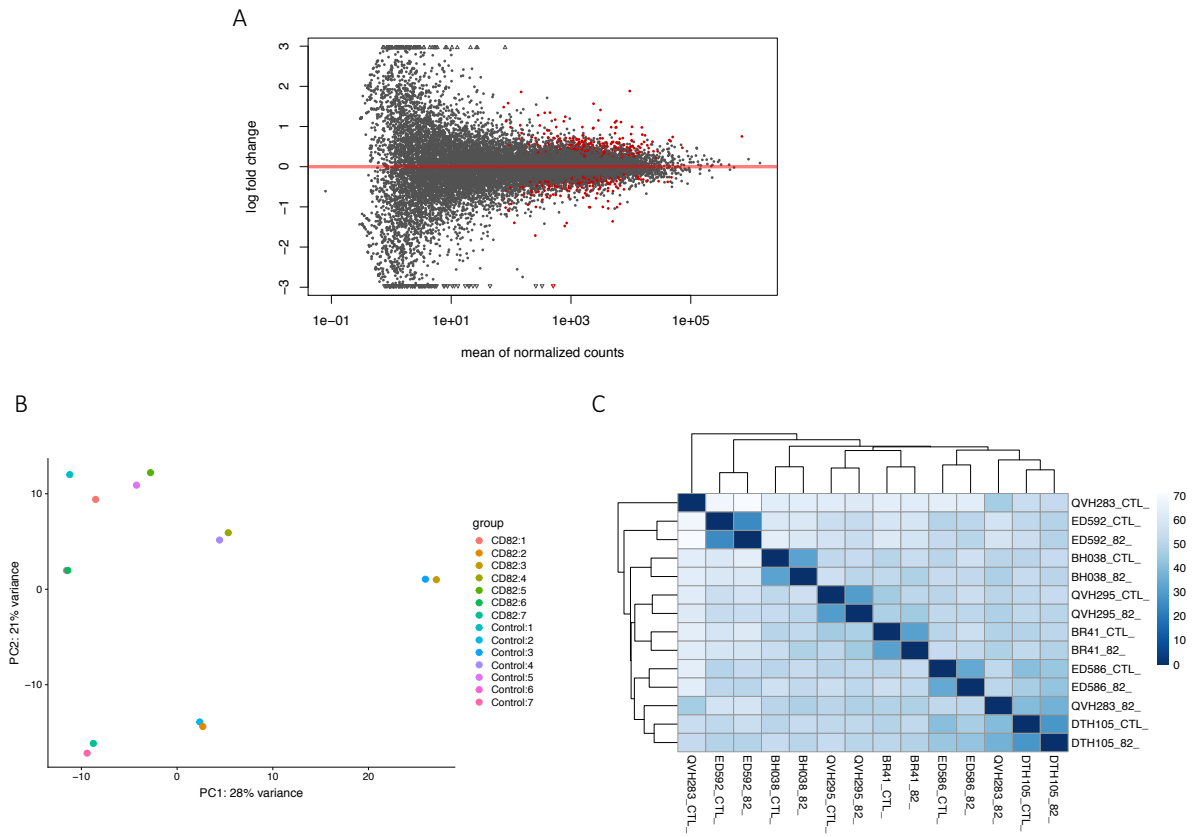


Figure 16. **Differential gene expression and sample clustering in CD82 knockdown**. A) MA plot for differentially expressed genes between siRNA and siCD82 in human myofibroblasts (*n*=7 DD patients). Points represent genes and red points genes with adjusted p-values < 0.05. B) PCA projection of patient donors in RNA-seq (*n*=7 DD patients). CD82 = siCD82 and Control = siRNA. PC = principal component. B) Heatmap of sample clustering of transformed count matrix showing sample to sample distance matrix. Distance metric is Euclidean distance between samples after regularized log transformation. Prefix represents patient ID and suffix is condition (CTL = siRNA and 82 = siCD82).

Whilst CD82 regulated a wide array of myofibroblast genes associated with diverse biological processes, six genes were highly downregulated and were *LAMTOR1, FLOT1, LENG1, CTD-2545G14.7, AL513523.1* and *ARHGAP11A* (Figure 17A-B). Pathway analysis of *CD82* regulated genes using gene ontology demonstrated an enrichment for multiple pathways including cell cycle, protein modification and metabolic signalling (Figure 17C).

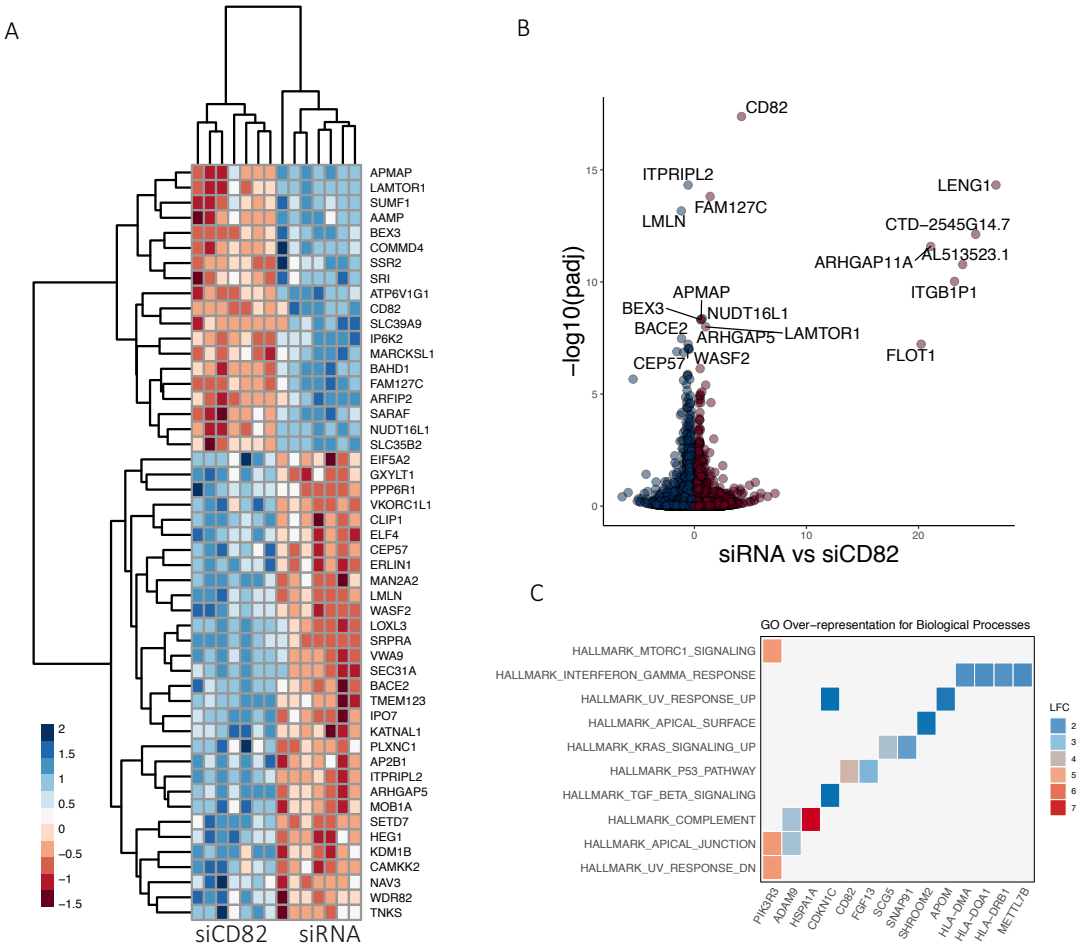


Figure 17. **RNA-seq of CD82 knockdown in Dupuytren's myofibroblasts.** A) Heatmap of regularized-log transformed values from RNA-seq of differentially expressed genes for siCD82 and siRNA transfection in human myofibroblasts (*n*=7 DD patients). Scale bar shows expression in z-score row scaled expression. X axis = log fold change and p-values adjusted with BH correction. B) Volcano plot for differentially expressed genes between siCD82 and siRNA transfection in human myofibroblasts. C) Heatmap for pathways enriched in differentially expressed genes. Colour represents log fold change of genes driving pathways. LFC = log fold change.

The results of the RNA-seq uncovered many cell cycle genes regulated by CD82 and this was also reflected in the pathway analysis. In addition, I noted that *CD82* was expressed by the proliferative stromal cell cluster in DD. Given this, as apoptosis resistance is a major factor leading to myofibroblast accumulation in fibrosis, I asked whether CD82 effected this cellular process. Using an Annexin-V flow cytometry assay, I saw no effect of *CD82* knockdown on myofibroblast apoptosis (Figure 18).

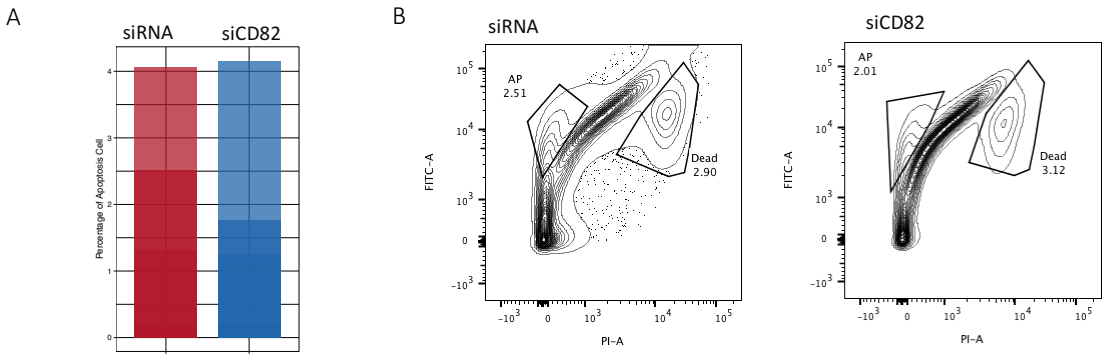


Figure 18. **CD82 does not influence Dupuytren's myofibroblast apoptosis.** A) Barplot showing the percentage of apoptotic DD myofibroblasts following siRNA and siCD82 knockdown (*n*=8 DD patients). Mean plotted. B) Representative density plots of flow cytometry analysis showing gating strategy for siRNA and siCD82 knockdown in Annexin-V apoptosis assay. FITC-A is Annexin-V channel and PI-A is Propidium iodide. AP = apoptotic.

Next, I explored whether CD82 directly regulated the cell cycle, especially considering the RNA-seq data. I used a DNA staining protocol to visualise distinct phases of the myofibroblast cell cycle. This showed that in DD myofibroblasts, siRNA knockdown of *CD82* inhibited progression of cell cycle and limited the proportion of cells moving into G2-M phase (Figure 19). Together, these results suggest CD82 is a marker of myofibroblast differentiation in DD and acts to promote cell cycle progression (Figure 19).

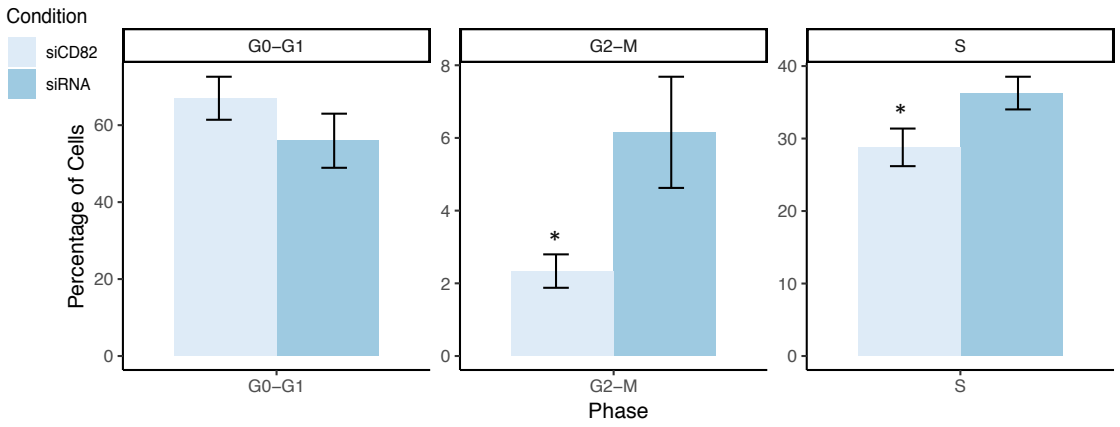


Figure 19. **CD82 regulates Dupuytren's myofibroblast cell cycle progression.** Bar plot showing the proportion of each cell cycle phase in DD myofibroblasts following siRNA and siCD82 knockdown. *n*=8 DD patients. Mean +/- standard error plotted. * = p-value < 0.05 (Paired t-test).

V.III.VII *CD82^{+} stromal cells in murine fibrosis*

Finally, I sought to explore my newly described CD82^{+} myofibroblast in a murine model of pulmonary fibrosis. Here, I analysed a public dataset that used droplet-based single cell RNA-seq (10x Genomics) to profile CD45^{-} stromal cells before and after installation of the chemotherapy agent bleomycin (*n*=2, *k*=7,346). Despite its limitations, this is the most commonly used murine model of human lung fibrosis and provides an opportunity to map the CD82^{+} subset in a contrasting disease system(22). Firstly, I performed the same computational workflow as applied to the DD data to delineate major cell clusters (see Chapter II.II.X). Next, using canonical cell type markers, I identified the stromal cell clusters, that contained a *Col1a1^{+}Acta2^{+}* myofibroblast cluster and two *Col1a1^{+}Acta2^{-}* fibroblast clusters (Figure 20). These clusters represented only a minor proportion of the murine fibrotic lung atlas, with the other major cell types being vascular, perivascular and immune cells.

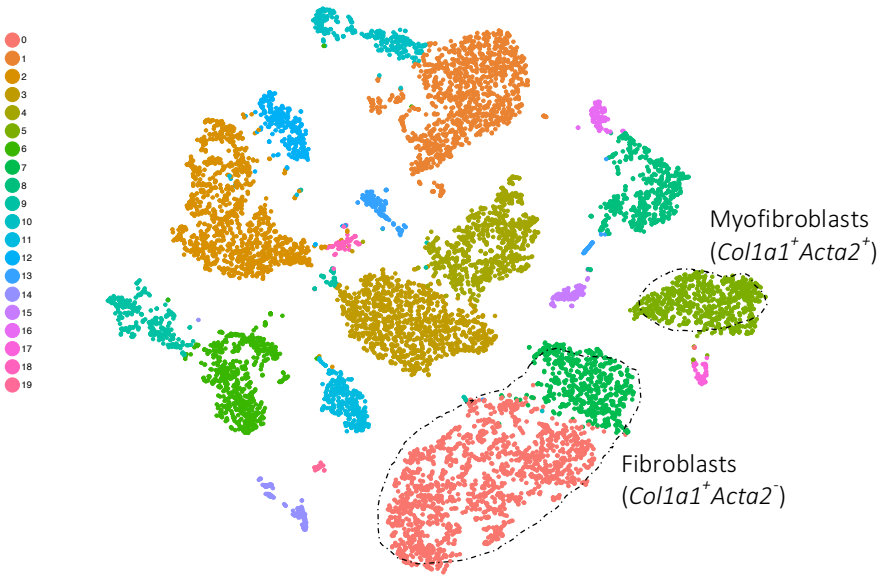


Figure 20. **Single cell atlas of murine pulmonary fibrosis.** tSNE projection of murine bleomycin pulmonary fibrosis model coloured by Louvain clusters (0-19) with fibroblasts and myofibroblasts outlined (*n*=2 independent mice).

After identifying the murine stromal cells, I selected these cells and performed a second round of quality control, feature selection and clustering to attempt to resolve potential sub-populations (Figure 21). Sub-clustering revealed two distinct clusters of fibroblasts separated by the expression of *Col13a1* and *Col14a1,* as well as a smaller intermediate cell population (Figure 21). In addition, there was a clear *Acta2*$^{+}$*Tagln*$^{+}$ myofibroblast cluster. I then looked for the expression of *Cd82* in this dataset. Firstly, I confirmed that murine and human *CD82* were one-to-one paralogs via the Ensembl annotation and then visualized its expression within the stromal cell clusters. Remarkably, I found *Cd82* marked the distinct *Col13a1*$^{+}$ fibroblast. Whilst this subset did not express *Acta2*, it did express other putative human myofibroblast markers, including *Tagln* and *Tgfb1* (Figure 21A).

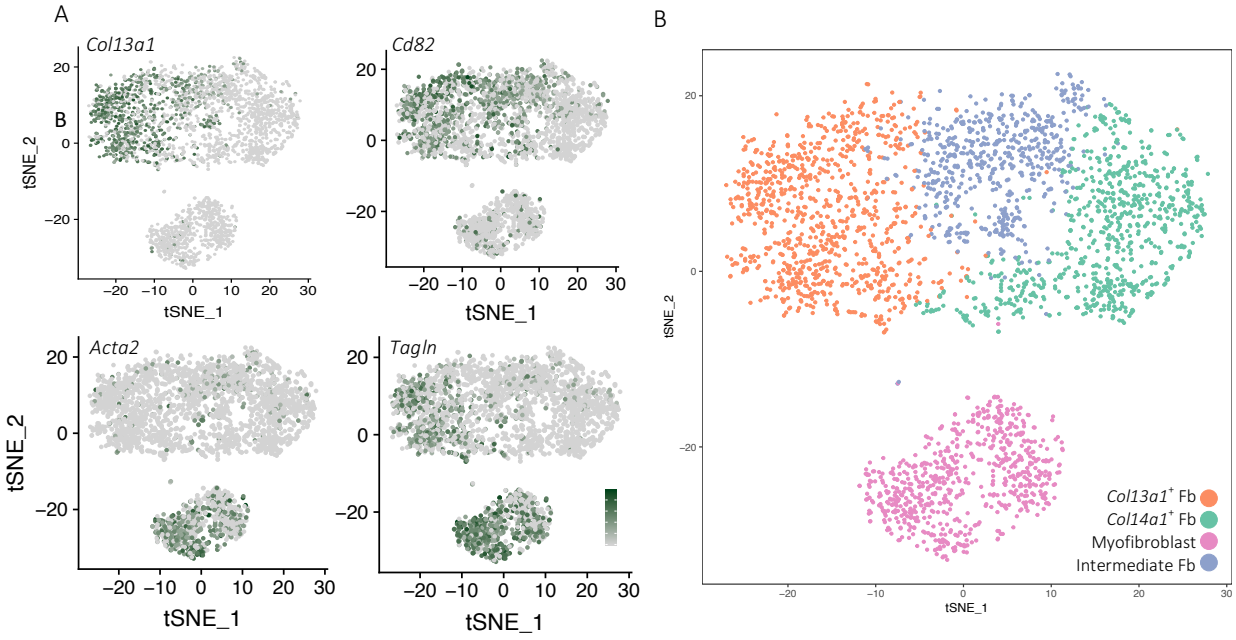


Figure 21. **Three major divisions of pro-fibrotic murine fibroblasts.** A) tSNE projections of fibroblasts and myofibroblasts from bleomycin murine model marked by the expression of selected genes including *Cd82* (*n*=2 independent mice). Expression in scaled log(UMI+1). B) tSNE projection of murine fibroblasts and myofibroblasts in bleomycin model coloured by Louvain clusters (*n*=2 independent mice). Fb = Fibroblast.

Interestingly, in this murine model of fibrosis, the *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts were the most upregulated stromal cell population during fibrosis progression (Figure 22).

After this, I integrated fibroblast and myofibroblast markers between human and mouse to explore any species dependent differences in their expression. Of note, *TALGN* and *ACTA2*, two highly studied myofibroblast markers, were common myofibroblast markers between species. In addition, *MAFB, MYL6, MYL9* and *TGFBI* were all conserved myofibroblasts markers between mouse and human. In contrast, *CD82* was a myofibroblast marker in humans (DD), but a fibroblast marker in mouse (Lung fibrosis). As this was the specific population expanded in murine lung fibrosis, it raises the possibility that *CD82* marks pathogenic stromal cells, but on different populations in the various fibrotic diseases.

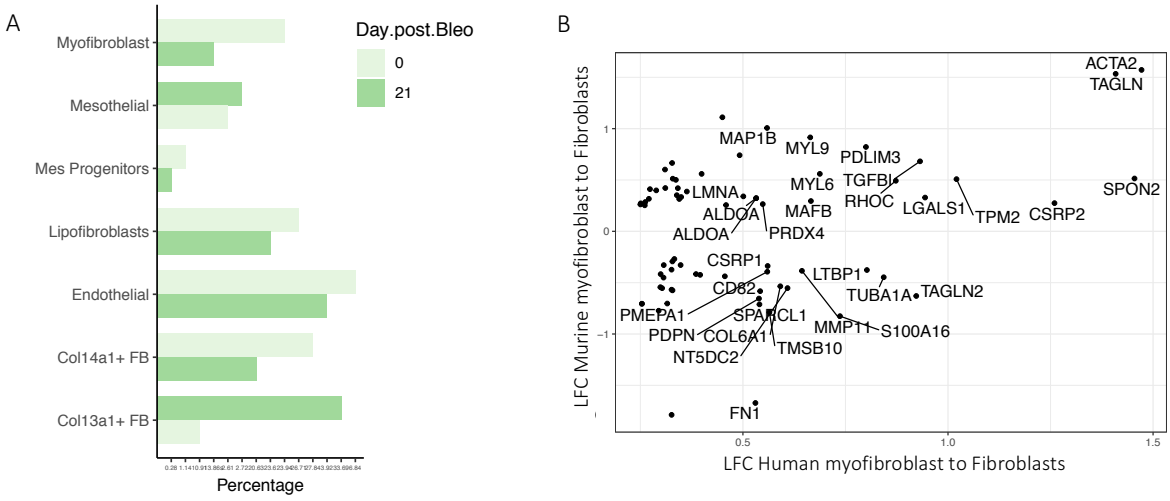


Figure 22. ***CD82*$^{+}$*Col13a1*$^{+}$ fibroblasts are expanded in murine pulmonary fibrosis.*** A) Barplot showing the percentage of stromal cell types before and after the installation of bleomycin in murine fibrosis model. Day.post.bleo = Days after bleomycin installation (0 & 21). Mes Progenitors = Mesothelial progenitors. B) Scatterplot integrating differentially expressed genes (Wilcoxon Rank Sum Test) in murine and human myofibroblasts as compared to fibroblast populations. LFC = log fold change. Murine genes selected by comparing fibroblast (*Col13a1*$^{+}$) and myofibroblast clusters and human genes obtained by comparing fibroblast and myofibroblast clusters.

To explore potential relationships among murine fibroblasts in fibrosis, I used diffusion maps to align cells along pseudotime, excluding myofibroblasts. The resulting three-dimensional diffusion map had one main curved branch. The fibroblast groupings on this branch were consistent with the Louvain clusters, extending between *Col14a1*$^{+}$ and *Col13a1*$^{+}$ fibroblasts (Figure 23). This structure suggests that *Col13a1*$^{+}$ fibroblasts could be derived from *Col14a1*$^{+}$ fibroblasts and expand during the progression of fibrosis. Consistent with this hypothesis, *Col13a1*$^{+}$ fibroblasts expressed several myofibroblast markers, including *Tgfb1* and *Talgn,* and therefore could represent a proto-myofibroblast, providing a precursor pool for mature myofibroblasts.

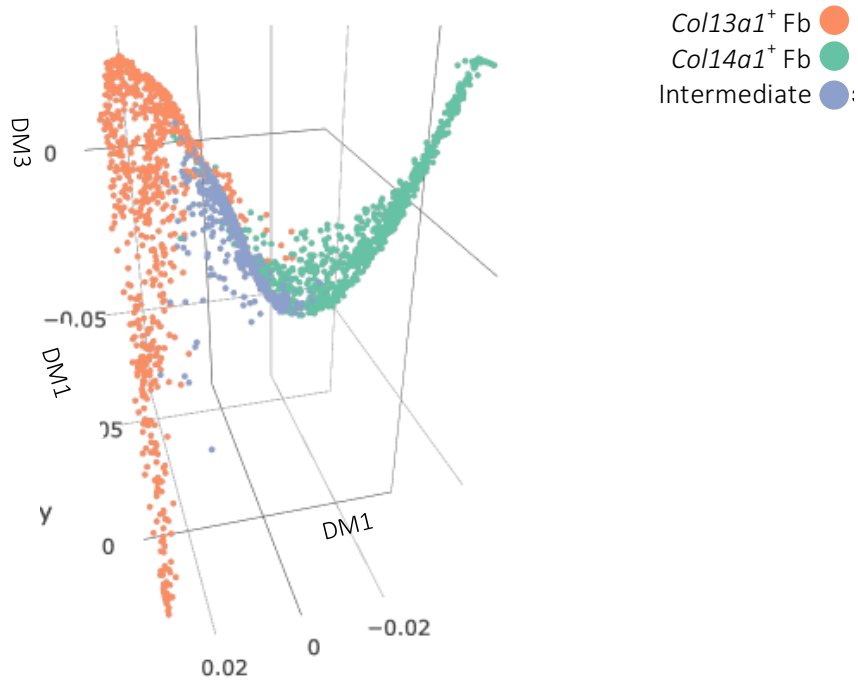


Figure 23. **Diffusion map embedding of murine fibroblasts in bleomycin lung fibrosis model.** Diffusion map embedding of murine fibroblasts in bleomycin model showing trajectory moving between *Col13a1*$^{+}$ and *Col14a1*$^{+}$ fibroblasts. DM= diffusion map coordinates (*n*= 2 independent mice). Fb = Fibroblast.

Based on these findings, I looked for CD82 protein expression in human pulmonary fibrosis (IPF). This was facilitated by an international collaboration between Kennedy Institute, Oxford and Celgene Corporation, with staining of samples undertaken in San Diego, USA. Co-staining for α-SMA and CD82 was performed in tissue sections of IPF lung. Here, we found CD82 expression within fibrotic foci in IPF, not co-localized with α-SMA, mirroring observations in the murine model (Figure 24). Together, these data show that CD82 marks a fibroblast population in murine and human pulmonary fibrosis.

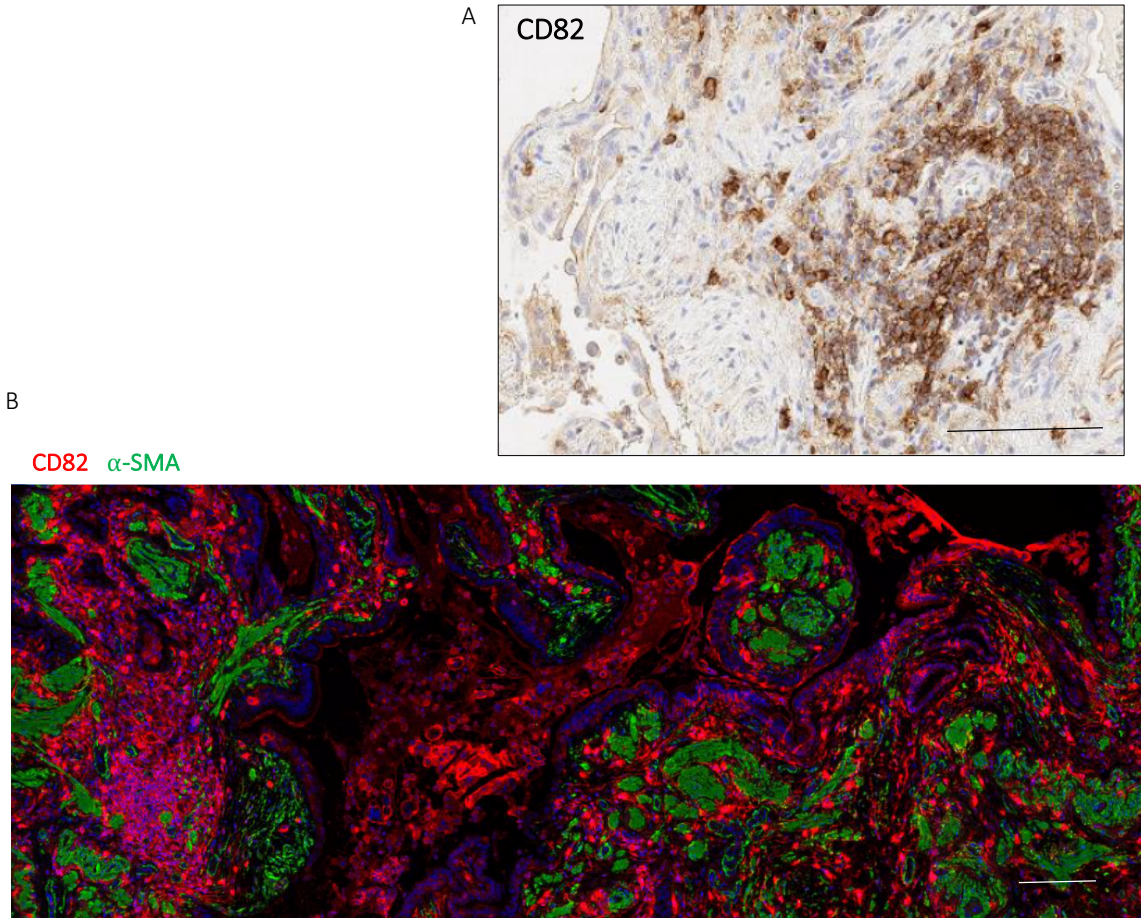


Figure 24. **CD82 expression in human idiopathic pulmonary fibrosis.** A) Immunohistochemistry staining of CD82 expression in human IPF samples. Scale bar is 30 µm. B) Confocal image of immunofluorescence staining showing expression of CD82 and α-SMA in human IPF. Scale bar is 100 µm. Staining by Weilin et al.

V.IV. Discussion

In this Chapter, I uncover profound phenotypic diversity in human myofibroblasts. These results correspond to a recent description of stromal cells in a murine model of wound healing and skin fibrosis(45, 245). Here, major stromal subsets were reported as $CD29^{high}$, $CD29^{low}$, adipogenic progenitors and $CD34^{+}$(45). In DD, CD29 (ITG-β1) is also a highly specific marker of myofibroblasts and in agreement with published findings(45), distinct DD myofibroblast subsets had overlapping but discrete expression profiles. In another study, using a murine model of lung fibrosis, distinct α-SMA^{+} populations were identified using only a few markers(246). It is reasonable to assume that myofibroblast heterogeneity will continue to be reported across fibrotic conditions and it will be intriguing to see whether distinct disease processes or anatomical sites imprint unique gene markers in this cell.

The overall structure of myofibroblast clusters could imply a plasticity supporting the transition between $ACTA2^{low}$ and $CD82^{+}$ myofibroblasts. Although I initially conceptualised the topography of the myofibroblast dataset as containing a linear differentiation or activation pathway, this transcriptomic profile may instead signify a reversible dynamic process that provides a flexibility to instruct the development of either $CD82^{+}$ myofibroblasts or $ACTA2^{low}$ myofibroblasts. Indeed, this paradigm would allow functionally distinct populations to be spatially and temporally tuned during different stages of the disease. For example, in response to continued inflammatory signals, stromal cells could be primed to adopt a myofibroblast phenotype to support the expression of matrix proteins. This would allow different pathological processes to exist in equilibrium and foster

crucial processes (Inflammation & ECM remodelling) required to maintain fibrosis. In support of this, in the bleomycin model of pulmonary fibrosis, α-SMA positive cells are most abundant at day 14 post exposure as compared to day 21(23). This finding is consistent with reports in both rodent and human lung fibrosis suggesting that the expression of α-SMA in myofibroblasts is dynamic, with higher levels during early stages of disease(2, 23, 247, 248). Connecting these observations with the results presented in this Chapter imply that *ACTA2* expression is finely balanced in concert with the disease ecosystem. The transcriptomic signature I describe could in fact signify this equilibrium paradigm, with cells moving in both directions between $CD82^{+}$ myofibroblasts and $ACTA2^{low}$ myofibroblasts.

Having a modelled and experimentally validated signature of myofibroblast differentiation, it will be important to identify key molecular regulators of this process. My results identify many transcription factors, including *MAFB*, that precisely follow the emergence of $CD82^{+}$ myofibroblasts. Future work should examine these genes functionally and could help determine a master transcriptional switch controlling myofibroblast differentiation in DD. We are now well poised to decode regulatory networks governing myofibroblast development. This will be facilitated by advances in computational methods, enabling more accurate modelling of biological features. Once these processes have been systemically characterised, our final goal will be to integrate this information with perturbation experiments to investigate how disease-promoting signatures can be manipulated, helping to inform therapeutic development.

Identifying a proliferative stromal cell in DD has important implications. Although I noted a discrepancy between the protein (~20% cells) and gene expression (~1-2%) of Ki67, both readouts localised the cycling population within the ITG-$\beta1^{high}$ myofibroblast community. As a highly proliferative population, this subset could be a pool of cells *in vivo,* providing a continuous source of precursors to maintain myofibroblasts. However, given the high cell density and tight packing of myofibroblasts in DD(55), my results could support other factors in promoting myofibroblast survival in localised human fibrosis. In other forms of fibrosis, apoptosis resistance is reported as the major mechanism through which stromal cells accumulate(249-252). In DD, myofibroblasts may accumulate primarily due to apoptosis resistance, secondarily supported by the cycling population(17, 253). We should now aim to further characterize the cycling stromal cell in DD with a logical next step exploring genes that may regulate its transition towards a mature myofibroblast. Learning the cell cycle pathways controlling this proliferative myofibroblast could ascertain mechanisms that govern this process.

$CD82^+$ stromal cells in DD and pulmonary fibrosis had contrasting expression profiles suggesting distinct roles at each site. In the murine model, *Cd82*$^+$*Col13a1*$^+$ fibroblasts were also accurately delineated by *Itga8*, *Cxcl14* and *Npnt*. In addition, transcription factors *Lbh*, *Nr2f2*, *Tbx2*, and *Meox2* strongly marked this fibroblast. This differs to DD, where *CXCL14* and *CD82* marked distinct populations of fibroblasts and myofibroblast, respectively. Interestingly, in a recent study of murine lung fibrosis, *Itga8* was shown to regulate *Col1a1* expression in $PDGFR\beta^+$ stromal cells *in vitro*, without a clear function *in vivo*(254). High expression of *Itga8* and its regulator, *Npnt*, in *Cd82*$^+$*Col13a1*$^+$ fibroblasts suggests a significant role of this pathway in this population(141). The

study described above, is also a prime example of how investigating genes without knowledge of their cellular context may impair an understanding of their function *in vivo*. Here, the *in vivo* role of *Itga8* was tested by inhibiting expression of this gene only in PDGFRβ^{+} stromal cells. However, the results presented in this Chapter demonstrate that *Itga8* is highly expressed in a PDGFRβ^{-} and *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblast, showing how this study may have failed to define a function of *Itga8 in vivo*. Therefore, we must now look to integrate known or emerging molecular mechanisms of fibrosis with their cellular context to help inform strategies to test or validate their function. Finally, in addition to a better understanding of *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts in mouse models, we should characterize this population across human fibrotic diseases. Just as in the lung and hand, it may mark contrasting cell populations at different anatomical sites or may be a conserved myofibroblast marker in several diseases. One limitation to my investigation of CD82^{+} stromal cells in the lung was that we only co-stained for alpha smooth muscle actin and CD82. In future, using markers of the distinct *Col13a1*$^{+}$ subset (e.g., *ITGA81*, *PDGRFA*) would be vital to see if this stromal cell is conserved in human fibrotic lung disease.

In summary, in this Chapter I characterize previously unknown sub-populations of myofibroblasts and elucidates a transcriptomic signature of myofibroblast activation towards a novel cell state. Finally, I identify and validate CD82 as a myofibroblast cell surface marker.

Chapter VI. Establishing an *in vitro* model to quantify myofibroblast force generation

VI.I. Introduction

There are myriad tools to measure mechanical force in living cells, but one that has gained increasing popularity is traction force microscopy (TFM)(255). Major reasons behind this are an ease of implementation and the ability to monitor the spatiotemporal dynamics of cellular forces. The basic design of a TFM experiment is one in which hydrogels are formulated with fluorescent beads that serve as fiducial markers(82). Often, the surface of the gel is functionalized, usually with some form of extracellular matrix protein, to promote cell adherence (see Chapter II.II.XVII). Cells are then seeded on to the surface, allowed to adhere and contract against the hydrogel producing deformations in its structure. Using fluorescence microscopy, cells on the surface of the gel are located and subsequently removed by the addition of trypsin. As cells are removed, correction of gel deformation caused by cell contraction can be visualized by movement of the fluorescent beads. Bead displacements can then be tracked and through knowledge of the mechanical properties of the gel, forces exerted by cells quantified. The measurement of cellular forces requires a stereotyped workflow involving several separate steps.

VI.II. Hypothesis and Aims

My hypothesis is that optimizing a traction force microscopy assay will enable probing of myofibroblast force generation at single cell resolution.

Here, I outline experimental and computational methods encompassing traction force microscopy and seek to validate severe inherent assumptions in the assay. I systemically test whether key experimental assumptions are met to build a robust assay to quantify myofibroblast force generation.

VI.III. Results

VI.III.I *Experimental and computational workflows in TFM*

The figure below (Figure 1-2) demonstrates the experimental and computational methods encompassing TFM. For my system I used hydrogels between 2.55 to 4.47 kPa and 40 nm fluorescent marker beads to track deformation in the substrate surface (see Chapter II.XVII).

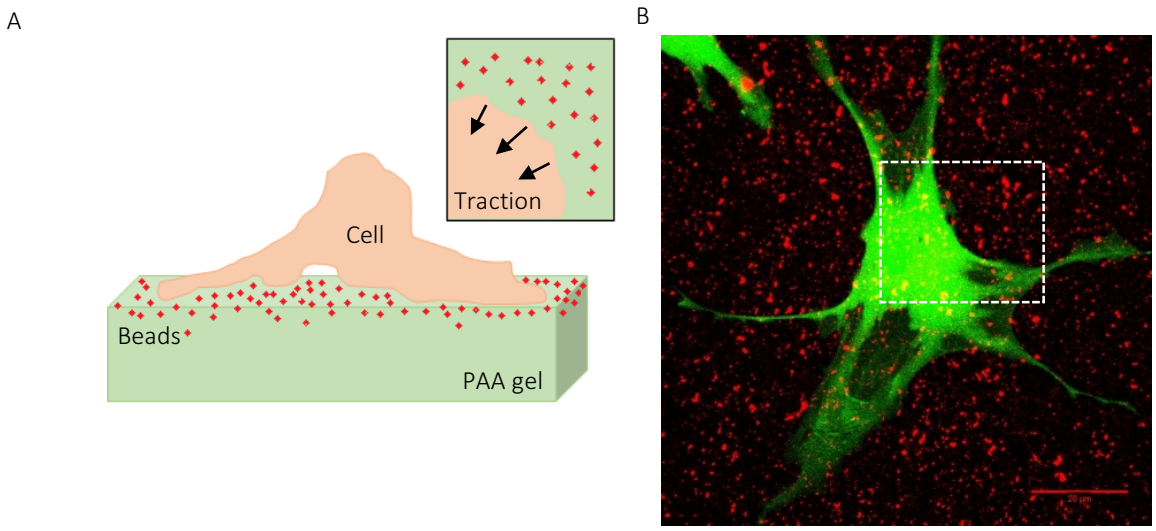


Figure 1. **Traction force microscopy (TFM) of Dupuytren's disease myofibroblasts**. A) Schematic for the experimental setup in TFM with cell seeded on surface of polyacrylamide (PAA) hydrogel with 40 nm fluorescent marker beads (red points) in surface. Square demonstrates schematic of cell traction forces that lead to bead displacement. B) Representative confocal image of myofibroblast tagged with Calcein AM (Green) on 2.47 kPa PAA gel with fluorescent beads (Red) on the surface. Scale bar 10 µm.

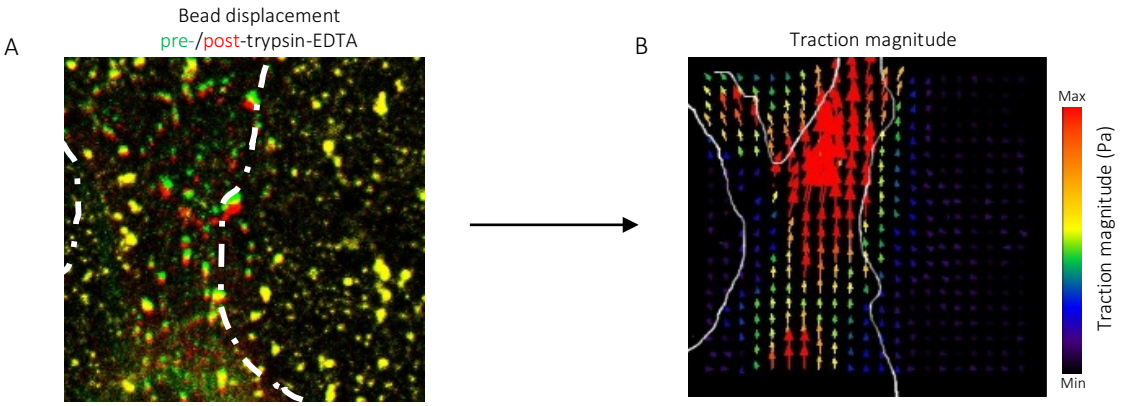


Figure 2. **Computational workflow of traction force microscopy (TFM)**. A) Confocal image from white square in Figure 1B demonstrating bead displacement on gel surface before and after addition of trypsin. Bead displacement tracked using particle image velocimetry. B) Corresponding force vector plot with scale bar. Force vectors are reconstructed using FTTC algorithm implemented as an ImageJ plugin. White lines represent cell boundary. Pa= Pascal.

Calculation of cellular traction forces relies on several assumptions relating to certain theoretical and experimental parameters. In the next section I describe optimisation of the TFM assay through defining these critical parameters. Table 1 lists pertinent information for each stage of the TFM workflow.

Experimental Step	*Critical parameters*
Hydrogel preparation	Gel stiffness, thickness, bead density
Cell culture	Seeding time, cell density
Image acquisition and processing	Microscope resolution
Bead displacement	Displacement magnitude
Force reconstruction	Regularization factor

Table 1: Key experimental aspects of TFM and important parameters to be defined. Each step is in sequential order with force reconstruction representing the final stage.

VI.III.II *Defining hydrogel thickness*

TFM was pioneered by Harris et al (1980), who first demonstrated the mechanical activity of fibroblasts through visualizing wrinkles on an elastic silicone substrate(75). However informative, this technique suffered from poor quantification of forces due to non-linear elastic properties of the silicone substrate and poor spatial resolution(75). Consequently, TFM substrates are now most commonly made from polyacrylamide (PAA) which has several important advantages(76). Firstly, it is linearly elastic, meaning cell induced deformation is directly proportional to the applied force.

Secondly, even minute elastic deformations can be described in a quantitative manner that allows cellular traction forces to be recovered with high precision(76). Finally, polyacrylamide is optically clear and transparent, facilitating image acquisition through the gel material.

Calculation of cellular forces relies on several experimental assumptions pertaining to the PAA gels that must be validated. Firstly, PAA gels are assumed to be a semi-infinite half space with minimal contribution of the glass coverslip to the applied force(85). Practically, this means gels should be at least 30 µm thick and so I measured the thickness of my own gels using a z-stack imaging mode (Figure 3)(85). The mean thickness of the PAA gels was 36.45µm (31.7-45.1µm).



Figure 3. **TFM PAA gel characterization.** A) Confocal z-stack image of PAA gel thickness showing red marker beads at surface. Yellow line labels gel thickness. Representative of three independent experiments and *n*=9 gels. Z = Z axis and X = X axis. B) Scatterplot showing measured gel thickness of PAA hydrogels. *n*=9. Mean +/- standard error plotted. Gel thickness measured from three independent experiments and *n*= 9 gels. PAA = polyacrylamide.

VI.III.III *Characterization of fiducial marker distribution*

Force induced gel deformations are tracked through the displacement of fluorescent beads that serve as fiducial markers(85). The density of beads must be sufficiently high to prevent information being lost when areas of the gel are deformed, but insufficient beads are present to be displaced. This must be balanced with not exceeding a density that will impede bead tracking algorithms.

Through several iterations I found a 1:5,000 bead dilution (in ddH20) produced a density of ~8 μm^{-1} (Figure 4). Another important consideration is that only beads on the surface of the gel are useful for measuring forces. Therefore, I followed a protocol described in Colin-York et al (2017) to localize fluorescent beads on gel surface (see Chapter II.II.XVII)(81, 82). Figure 4 shows three representative images of fluorescent beads on the surface of the gel.

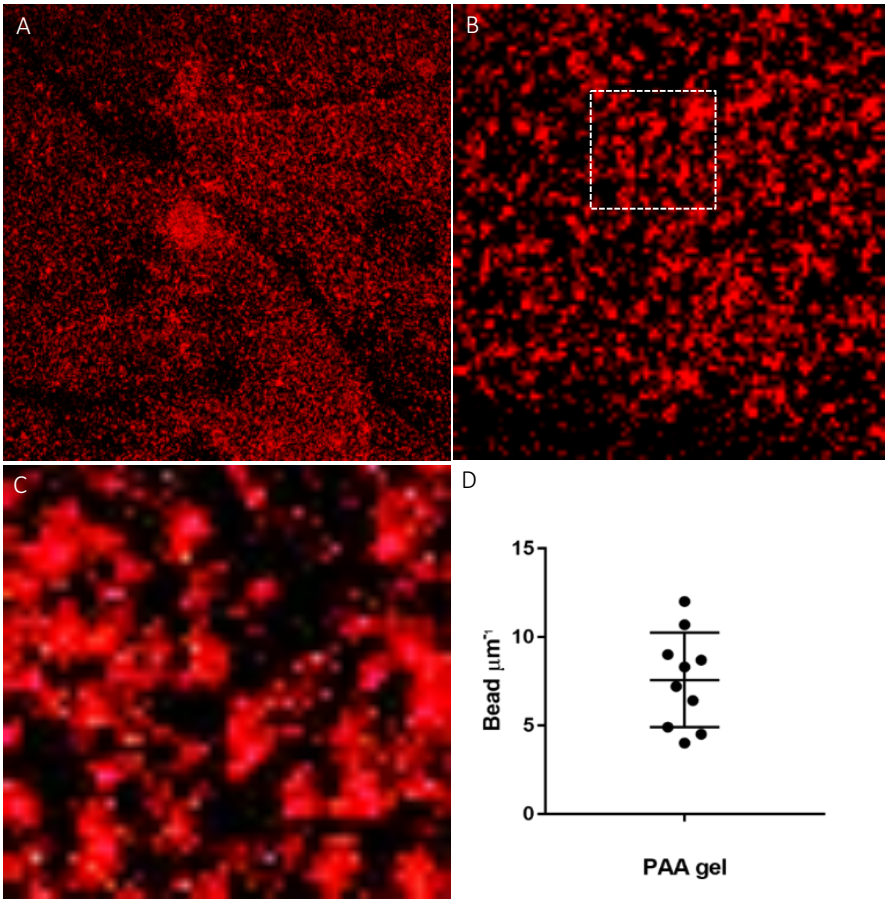


Figure 4. **Fiducial marker (fluorescent bead) density on PAA gel surface**. A-C) Confocal images of 40nm marker beads (Red) on gel surface. Images at x63 magnification. C) Confocal image of white square in (B). D) Scatterplot showing bead density (beads per micrometre) on gel surface. *n*=10 gels from three independent experiments. Mean +/- standard error plotted. PAA = polyacrylamide.

VI.III.IV *Optimizing seeding time for TFM*

Interestingly, a parameter I discovered as crucial for success of the assay was seeding time. This relates to the time between seeding cells and image acquisition. The clear majority of TFM protocols describe seeding cells overnight to facilitate adhesion(64). However, when I attempted this, movement of beads following cell removal was too small to quantify. In addition, cells required large volumes of concentrated trypsin to release from the gel. Through perusal of the literature I identified that one group(256) that restricted seeding time to only 4-6 hours. Their reasoning was that stromal cells can remodel ECM including the protein coating on the gel surface and this can prevent robust force measurement. I tried this protocol and found that bead displacement was far more consistent and established a timeframe of 4-6 hours. Considering this result, as actin stress fibres are required for cellular traction forces, I then ascertained whether Dupuytren's stromal cells formed actin stress fibres within this timeframe on soft 2.55kPa hydrogels. Phalloidin staining was used to visualize F-actin stress fibres and revealed most cells had developed stress fibres within 6 hours (Figure 5). These experiments also allowed me to confirm the seeding density of cells during the TFM procedure. This is important as it allows measurements from individual cells to be taken, avoiding cell-cell interactions, which may influence the results(76).

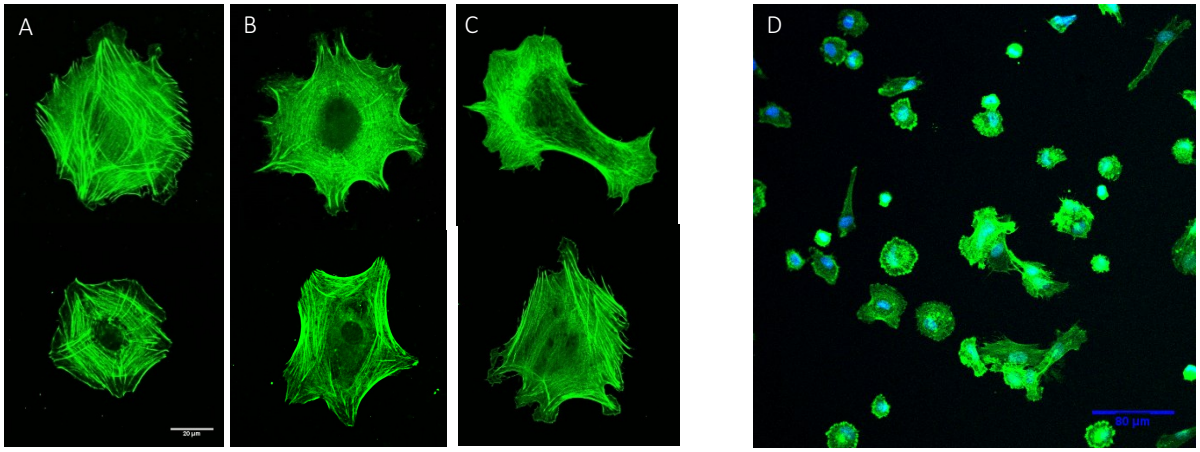


Figure 5. **Stress fibre formation on 2.55kPa PAA hydrogel.** A-C) Confocal images of DD myofibroblasts on the surface of 2.55kPa hydrogel stained with Phalloidin-AF488 (Green) labelling F-actin. Cells were seeded on to hydrogels 6 hours prior to staining to confirm stress fibre formation on softer environments. Scale bar is 20 µm. D) Low magnification confocal image (x20 Magnification) of same protocol in (A) showing myofibroblasts on 2.55 kPA gel surface. Scale bar is 80 µm. PAA = polyacrylamide.

VI.III.V *Defining optimal gel stiffness*

Traction force measurements are valid if the polyacrylamide gel behaves as a linearly elastic solid (85). In practice, this means bead displacement should be small, linear and in the range of 1 µm. This must be balanced with the ability to detect bead displacements consistently, which becomes challenging if too small. Therefore, I performed a series of optimisation experiments to select a gel stiffness that enabled enough bead displacement for robust downstream analysis. I found that a 2.55kPa hydrogel produced the most reliable results (Figure 6). Although a proportion (37%) of bead displacements were above 1 µm, when I attempted to use gels of higher stiffness (eg 8kPa), bead displacements became far less consistent and meant that many of the gels could not be used for analysis.

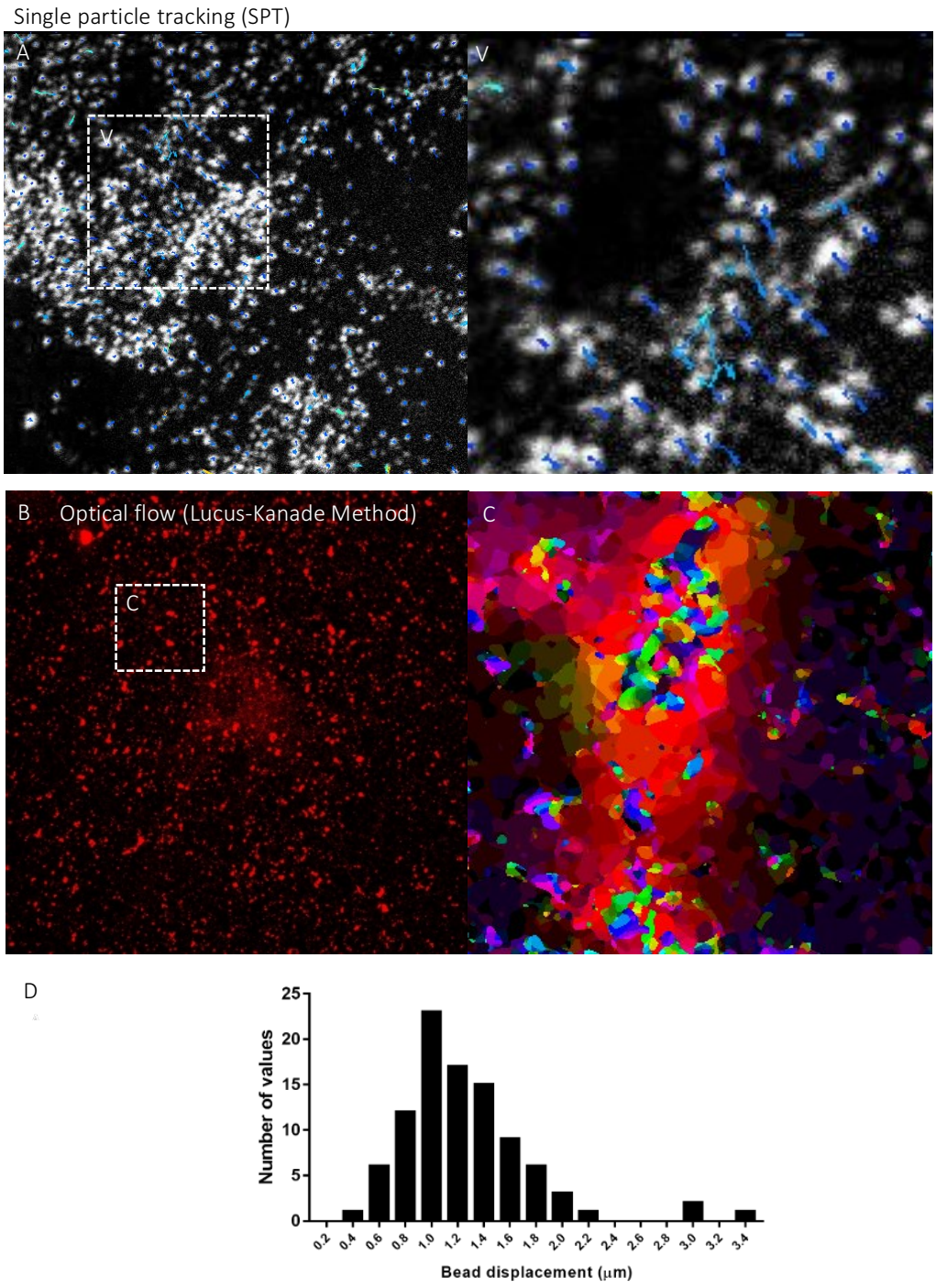


Figure 6. **Defining optimal gel stiffness and quantifying bead displacement.** A) Single particle tracking (SPT) of fiducial markers on PAA gel surface. Magnification x63. White points represented beads and tracks signify movement after removing cell from gel surface. B-C) Optical flow (Lucus-Kanade Method) of bead displacement (C) in the white square in (B) demonstrating localized area of bead displacement. Colours represent areas of bead displacement. D) Histogram of bead displacements from > 50 cells from three independent experiments with cells seeded on 2.55kPa hydrogels. kPa = kilopascal.

VI.III.VI *Characterising tissue mechanics of Dupuytren's nodules*

The mechanical environment of fibrotic tissue is increasingly recognized as crucial in driving fibrosis(64). Myofibroblasts exist within a dense ECM that progressively becomes less compliant in disease progression(64). During development of the traction force assay, I set out to quantify the tissue mechanics of Dupuytren's nodules to better understand the mechanical milieu whilst gaining information to formulate hydrogels of a stiffness that recapitulates, as closely as possible, the *in vivo* environment. In collaboration with Dr Cameron Brown, group leader in Bioengineering, I used atomic force microscopy (AFM) to probe freshly prepared Dupuytren's nodules (Figure 7, see Chapter II.II.XIX). The reasoning behind this method is that the AFM probe is 5 micrometres and therefore allows measurements to be taken at cell-relevant scales, gaining a perspective of the mechanical signals transduced through stromal cells(64). Our measurements showed that Dupuytren's nodules had a mean Young's Modulus of 8.47kPa. Young's modulus, or elastic modulus, is a common metric used to describe stiffness of solid materials(257). Our results were comparable to other fibrotic pathologies, that have a modulus of 10 -20kPa(257).

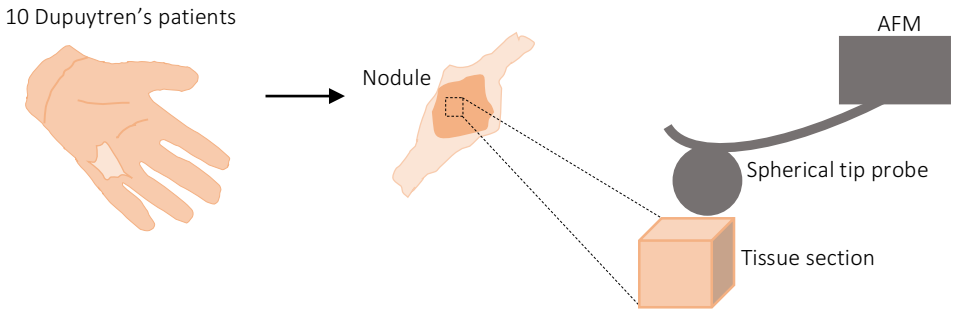


Figure 7. **Measuring native tissue stiffness in Dupuytren's nodules using atomic force microscopy.** Schematic demonstrating experimental workflow for obtaining measurements of tissue stiffness. Central nodular tissue was dissected to create cube like structures and these were then sectioned in 30µm slices. The tissue was then probed using micro-indentation with a 5µm sphere tip probe mounted on the AFM (atomic force microscopy) cantilever.

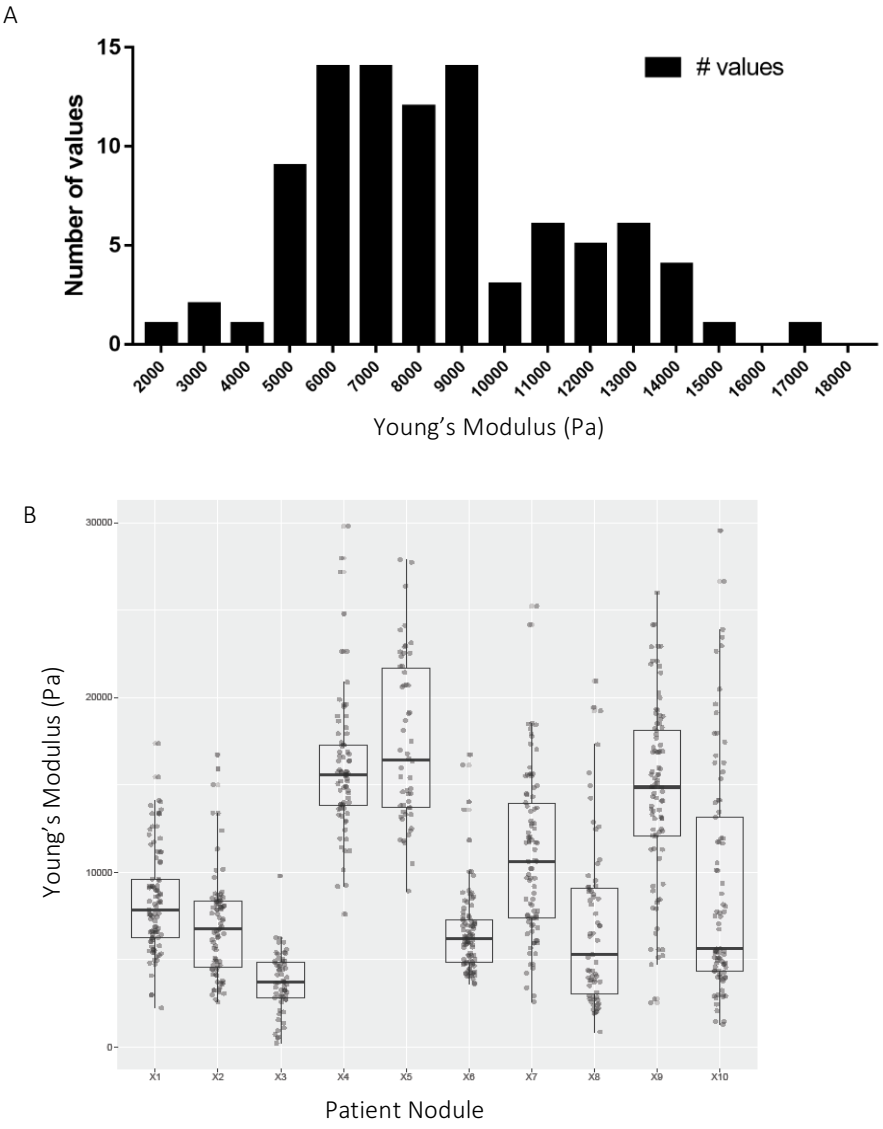


Figure 8. **Defining the Young's Modulus of Dupuytren's nodules**. A) Histogram showing the frequency of stiffness measurements in representative nodule. X-axis in Young's' Modulus (Pa) and Y axis is frequency of observations. Each sample was measured using mechanical maps of 10 x 10 µm with 100 individual measurements taken in each map. Three maps were taken per sample (DD nodule). B) Box and whisker plot with overlay scatterplot for Young's modulus of 10 Dupuytren's ndoules. Line represents mean value. Each point represents one micro-indentation measurement (*n*=300 per DD nodule). X1-X10 represent nodule section from 10 independent DD patients. Pa = Pascal.

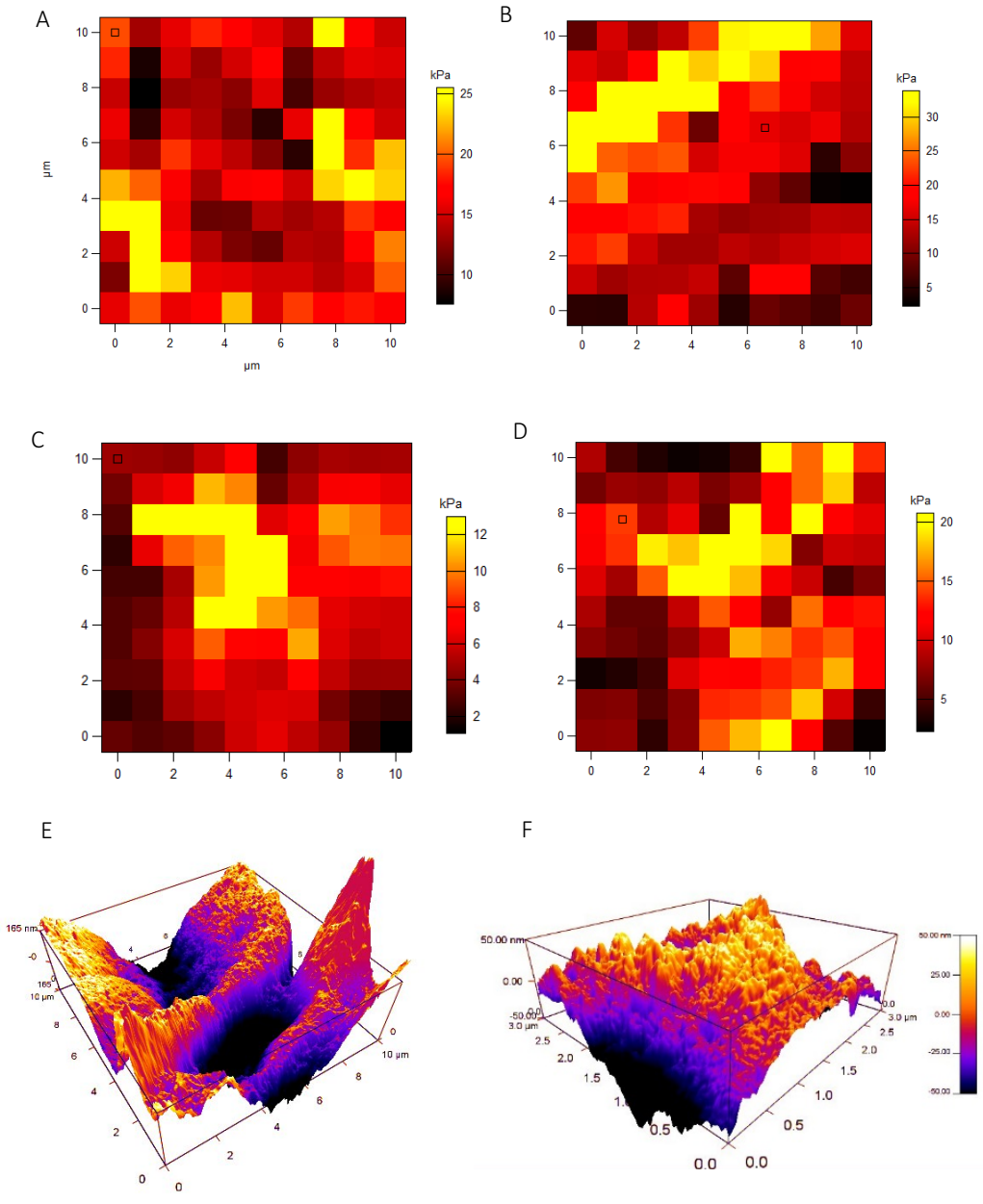


Figure 9. **Dupuytren's nodule tissue mechanics.** A-D) Mechanical map profiles obtained through atomic force microscopy. Each map is derived from a different DD patient nodule and represents a 100-point force profile. Scale is in kPa with yellow areas signifying highest stiffness. X and Y axis represent µm. E-F) Surface profile of Dupuytren's nodules. Normalized scale bar is in nanometres and figures show the surface topography of nodular tissue with linear bundle structures of individual collagen fibres. kPa = kilopascal.

VI.III.VII *Estimating the system resolution*

An important consideration in TFM is the ability to extract bead displacement for further analysis. In this regard, I wanted to estimate the resolution of my system as bead tracking is an important step in recovering cellular forces and directly relates to resolution of the microscope(81). One common method for estimating resolution is full width at half maximum (FWHM) of the point spread function (PSF)(81). As each marker bead in the gel represents the finite point spread function of the image, beads can be examined to define this parameter. To do this, I plotted line profiles of each fluorescent bead, applied a Gaussian fit to the fluorescence intensity, and calculated the FWHM to be approximately 300 nm (Figure 10).

According to Nyquist criterion, the minimal sampling density to capture all information in an image is twice the highest frequency expected(81). This is important, as if each marker bead is defined by a 300 nm PSF, the optimal pixel size for my system would be 150 nm. Applying Nyquist criterion in TFM provides a framework to enable the most accurate recovery of traction forces. Throughout my experiments, I chose a pixel size of 100 – 200 nm as this balanced accurate bead tracking with the ability to capture the whole cell at higher magnification.

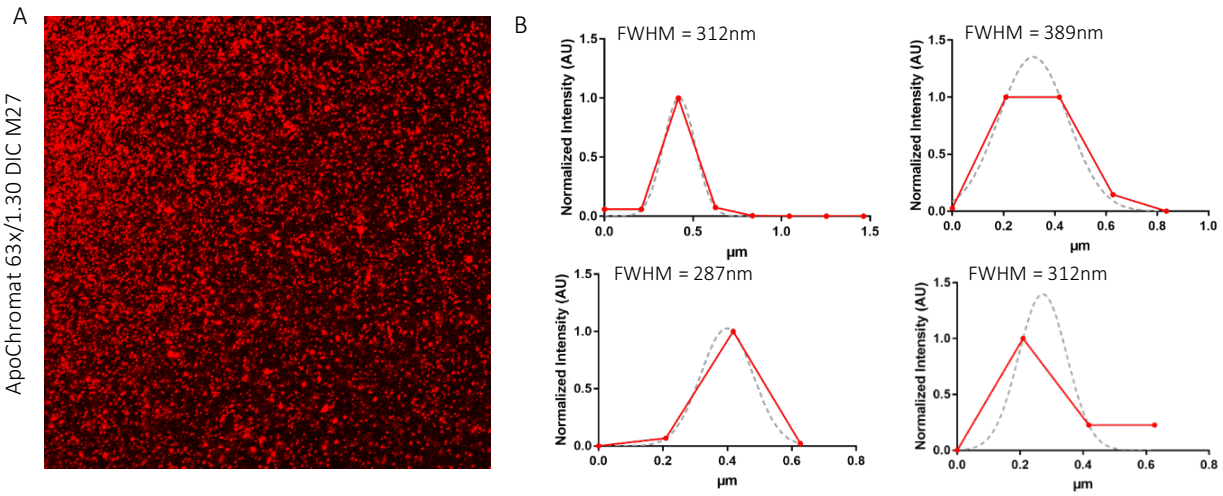


Figure 10. **Estimation of TFM resolution and defining optimal pixel size**. A) Typical bead image of PAA gel at x63 magnification using ApoChromat 1.40 NA objective. Red points represent fluorescent marker beads (40nm). B) Representative line profiles of fluorescent intensity across individual beads illustrating a resolution of 300 nm. Red line is the measured fluorescent intensity across individual beads with overlay of Gaussian Fit (grey dashed line). FWHM = Full width at half maximum. AU = Arbitrary Units.

VI.III.VIII *Regularization factor for traction force reconstruction*

A plethora of methods are available for calculating cellular forces, but one that has gained prominence is Fourier Transform Traction Cytometry (FTTC)(88). This method implements the procedure in the Fourier domain and is a computationally efficient means of calculating traction forces from the measured displacement field. Traction force reconstruction is inherently an inverse problem; it requires calculation of cellular force from a measured displacement field. Given this, FTTC employs a Tiknohov regularization scheme defined by the scalar product λ. This product essentially balances accurate reconstruction of traction forces with smoothing noise in the experimental image. An increasing value of λ reducing the recovery of traction forces. Considering this, I performed a regularization sweep and defined a value of $1x10^{-10}$ that optimized traction force recovery with noise suppression and kept this value constant throughout all experiments (Figure 11). This value is comparable with previous reports implementing FTTC in traction force microscopy(258).

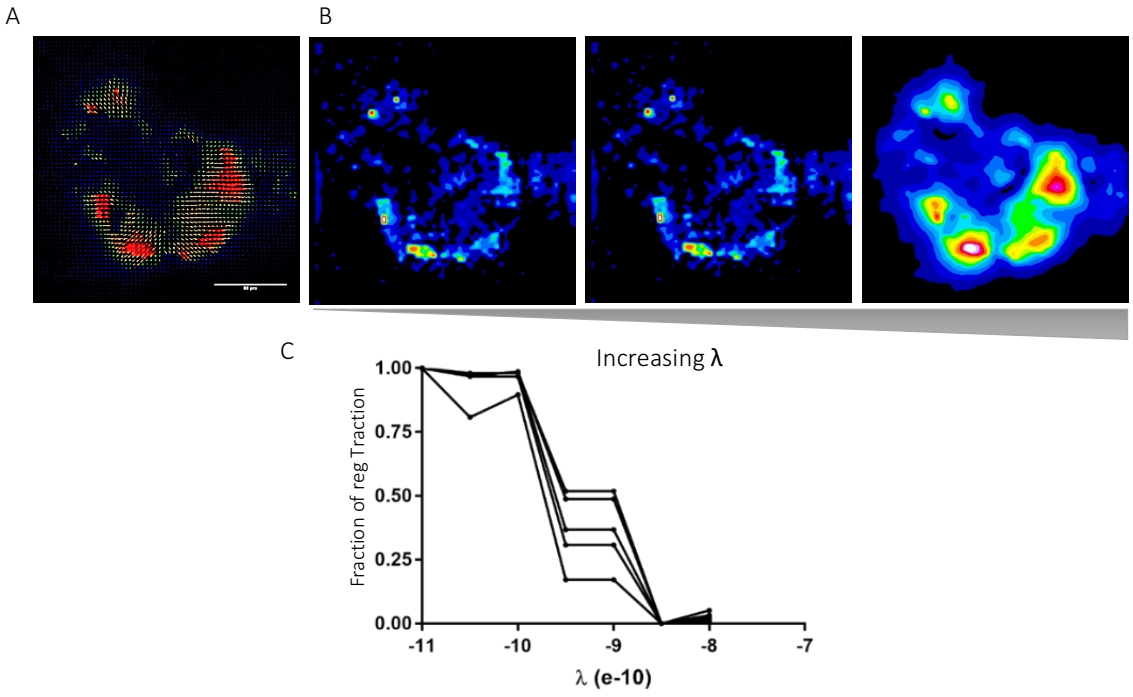


Figure 11. **Determining a value for FTTC regularization**. A) Vector plot of displacement field calculated through ImageJ PIV plugin. Vector plot depicts magnitude of highest vectors in red. Scale bar is 20µm B) Regularization parameter was changed from $1e^{-11}$ to $1e^{-8}$ and normalized traction force calculated for cell in (A). Figures represent heatmaps of traction force for regularization factor $1e^{-11}$, $1e^{-10}$ and $1e^{-9}$, respectively. Red areas represent highest force magnitude. C) Line plot illustrating the force recovered during iterations of increasing the regularization factor, normalized to the cellular force measured at $1e^{-11}$. λ = Regularization factor.

VI.III.IX *Tracking fluorescent marker beads with particle image velocimetry*

Particle image velocimetry (PIV) is an optical method of flow visualization originally developed to obtain velocity measurements in fluid dynamics, but is now the most popular method of quantifying bead displacement in traction force microscopy(85). PIV is a pattern-based tracking procedure that divides a pair or series of images into smaller regions termed interrogation windows, with each interrogation window spatially correlated between frames to assess the degree of movement. In my protocol, I applied a particle image velocimetry (PIV) plugin to quantify bead displacements (see Chapter II.II.XVII). Specifically, I used a cross-correlation algorithm with an iterative window size of 64 pixels and 32 pixels yielding displacement vectors 16 pixels apart with a final resolution of 2.4 µm (Figure 12). Post-processing of PIV vectors was undertaken using a normalized median test (noise level of 0.2 and threshold of 5.0) to filter and replace erroneous displacement vectors with a median value from nearby vectors. This process compares each vector with its 30 nearest neighbours and corrects for inaccurate displacement vectors in the image.

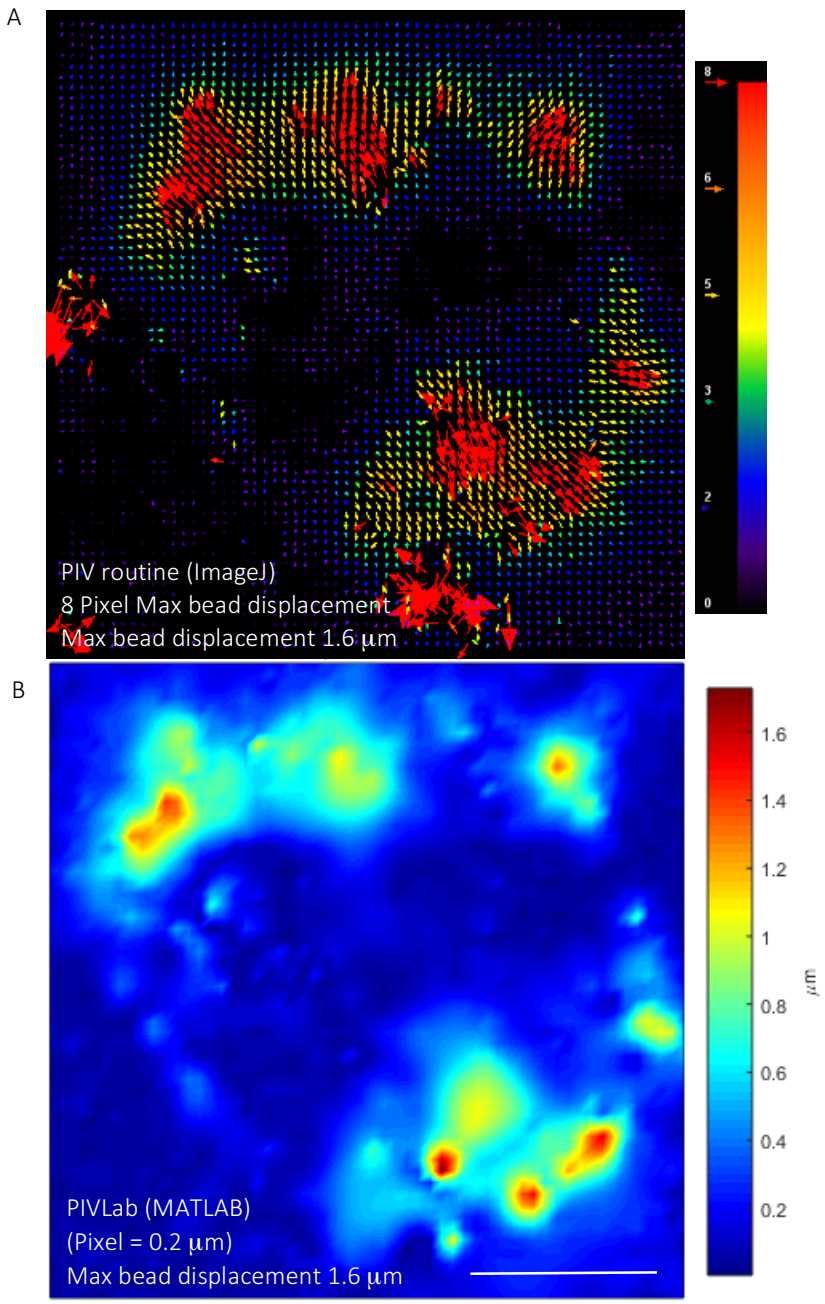


Figure 12. **Validating bead tracking using particle image velocimetry**. A) Vector plot of bead displacement (in pixels) with red vectors representing vectors of highest magnitude. Bead tracking performed by ImageJ plugin. B) Heatmap of bead displacement in the same cell as depicted in (A) showing PIV implementation using MATLAB software (PIVLab). Scale bar is 20µm.

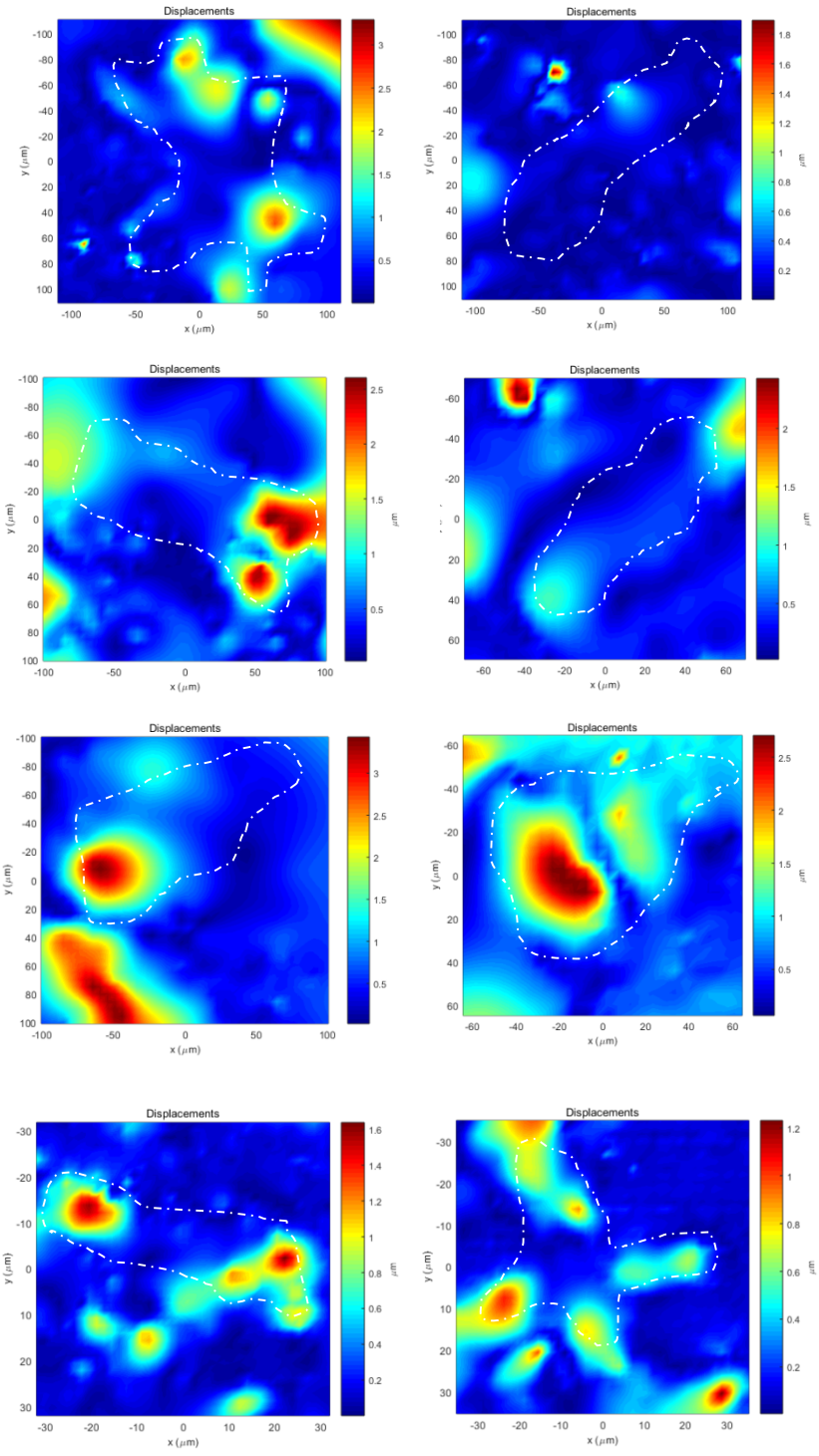


Figure 13. **Traction force microscopy of human myofibroblasts.** Heatmaps of bead displacement (in micrometres) induced by human myofibroblasts on 2.47 kPa PAA hydrogels quantified using particle image velocimetry. White dashed lines represent cell boundaries.

VI.IV. Discussion

The mechanical environment profoundly influences cell phenotype(16, 64, 224). Myofibroblasts are highly mechanosensitive, capable of sensing and responding to minute fluctuations in tissue mechanics. Increasingly, it is recognized that in fibrosis tissue mechanics play an important role in promoting myofibroblast activation and disease progression(20, 21). Indeed, studies have elegantly shown that targeting mechanotransduction pathways may signify an effective therapeutic strategy in fibrosis(64). At present, pathways that enable cells to sense, integrate and respond to changes in their mechanical environment are becoming better defined and two principle molecules are Yes-associated protein (YAP1) and Transcriptional Coactivator With PDZ-Binding Motif (TAZ)(64). These molecules allow cells to transduce mechanical stimuli into biochemical cellular responses and directly regulate myofibroblast activation(64). Given the significance of tissue mechanics in fibrosis, I first set out to map the mechanical landscape of Dupuytren's nodules, and in this chapter provide the first AFM measurements in this disease. The overall stiffness of nodules was around 5-10kPa, comparable to fibrotic skin and lung(249) and in contrast, the stiffness of normal dermis ~ 1kPa(257). These results provide an important reference for future mechanical studies in this disease. Although the observations presented here remain descriptive, knowing the true mechanical environment in DD could help inform future studies. Testing known or emerging therapeutic targets on cells spanning a mechanical spectrum found in DD nodules could elucidate which modulate mechanotransduction pathways. Before these mechanical measurements, DD stromal cells have been largely studied *in vitro* on hard plastic that

imparts profound effects on cell phenotype. It will be fascinating to see how our perspectives or preconceptions change if we begin to study DD cells in their native mechanical environment.

ECM remodelling in normal and diseased tissues is a force-dependent process(255). Myofibroblasts are defined by their contractile phenotype, with the canonical protein marker of these cells, alpha-smooth muscle actin, incorporated into linear actin structures called stress fibres(56, 60). In normal tissue homeostasis, ECM remodelling requires stromal cells to interact with their environment using focal adhesions, and subsequently apply traction forces to sculpt and organize the secreted matrix(56, 60). In the context of fibrotic disorders, stromal cell contraction is central in promoting disease progression and thus developing methods that allow us to quantify this processwill be crucial to further our understanding of their pathobiology *in vivo(259, 260)*. Here, I have optimized a live cell imaging assay that provides the opportunity to measure force generation in myofibroblasts. Moving forward, this assay could provide a functional readout for testing therapeutic targets or molecular regulators in myofibroblasts. In addition, it would be a vital next step to accurately define the force signature in myofibroblasts to build a perspective on how this may be manipulated when developing therapeutics.

The nature of the traction force assay requires certain experimental and computational parameters to be optimized. One of the most important is PAA hydrogel stiffness, as the mechanical milieu exerts significant effects on cellular force generation, generally increasing force on less compliant environments(259). As I found the native tissue stiffness of DD to be 5-10 kPa, I initially synthesize gels at 10kPa and titrated down the stiffness until I could obtain quantifiable bead displacements. My overall aim was to formulate gels of comparable stiffness to DD nodules

to build an assay that reflects, as closely as possible, the native disease environment. However, I found optimal bead displacement occurred at 2.55 kPa, slightly lower than the measured stiffness of the DD nodules (5-10kPa). At this gel stiffness, almost every cell deformed the matrix sufficiently to calculate traction forces. Thus, if I set out to use this assay for perturbation experiments, for example testing effects of inhibiting soluble factors, I would select this gel stiffness as it would enable me to collect a greater number of cell measurements. Fortunately, during optimization of the TFM assay I did observe less reliable, but consistent, bead displacements on 4.47kPa gels, closer to the measured stiffness in DD. Therefore, I selected a 4.47kPA gel to probe myofibroblast force generation as shown in the next chapter, to balance obtaining measurements on representative mechanical environment with optimizing throughput of the assay.

During optimization of the TFM assay, I noted that a critical parameter was seeding time of the cells. If this was extended beyond a few hours, or even overnight, bead displacements could not be quantified even if the cell moved over the surface of the gel. Explanations for this observation could be twofold. Firstly, it may be that after time myofibroblasts remodel the matrix beneath their surface, lifting them from the hydrogel and creating a surface over which they can glide when released from the gel. A second possibility could be that as matrix proteins on the surface are remodelled, local viscoelastic properties of the gel are altered. This process could thus create localized areas under the cell which are far stiffer that the fabricated PAA gel solution. If this were to occur, the gel may become too stiff to produce quantifiable deformations. Although only speculative, finding seeding time as a critical parameter in TFM does support the role of DD

myofibroblasts as efficient at matrix remodelling, able to initiate this process within a few hours after seeding.

In summary, in this chapter I detail the validation steps during optimization of a TFM assay and provide the first AFM measurements of the stiffness of DD nodules.

In the next chapter I aim to undertake a detailed characterization of force generation in human myofibroblasts, tracking the spatiotemporal dynamics of this process using the TFM assay developed here. Finally, I aim to integrate expression profiles of myofibroblasts discovered in the single cell RNA-seq (see Chapter V) with biophysical readouts of cell contraction.

Chapter VII. Functional heterogeneity in human myofibroblasts at single cell resolution

VII.I. Introduction

Myofibroblasts are crucial mediators of normal wound healing but contribute excessive matrix during fibrosis(4, 17). Crucial to myofibroblast function in health and disease is traction force as this provides the foundation for matrix adhesion and remodelling(261). Despite several studies quantifying cellular traction force(26, 55, 259, 262), we still lack an understanding of the biophysical profiles of human myofibroblasts. In addition, we are yet to obtain subcellular measurements of myofibroblast traction force on mechanical environments that reflect pathological fibrotic tissue.

Cellular forces have been measured in many different systems, including immune cells(80, 263), neurons(263) and fibroblasts(255). Interest in uncovering biophysical properties in distinct cell types is in response to the growing recognition that these parameters are central to cell function(114). Cellular traction forces underpin diverse biological processes ranging from immune synapse formation, intracellular vesicle trafficking to cell-cell communication(264). Forces are also important in embryology and allow for precise spatiotemporal patterning of all developing tissues(264). Thus, dissecting mechanical forces in living cells will help improve our understanding of the mechanics behind cell function. Much remains to be discovered with regards to mechanical forces generated by myofibroblasts.

VII.II. Hypothesis and aims

My hypothesis is that myofibroblast force generation is spatially and temporal dynamic. In addition, I predict that myofibroblasts house distinct biophysical communities and exhibit functional heterogeneity along the modelled transcriptomic differentiation signature.

To test this hypothesis, myofibroblast force generation was quantified at subcellular resolution (~2µm). I utilized the optimized TFM workflow described in Chapter VI to quantify force generation in myofibroblasts isolated from Dupuytren's nodules and only used cells up to passage 2 as extended culture can have a profound impact on myofibroblast phenotype(55, 224). Importantly, I focused on observing the dynamics of myofibroblast force generation in mechanical environments that emulate the stiffness of the DD nodules (4.47kPa). Finally, I aimed to integrate single cell expression profiles with biophysical readouts in myofibroblasts to probe functional heterogeneity.

VII.III. Results

VII.III.I *Dissecting mechanical force in living human myofibroblasts*

After optimising the traction force assay I sought to probe cellular forces in myofibroblasts isolated from DD nodules. I first quantified the magnitude of force on polyacrylamide (PAA) hydrogels as illustrated in Chapter VI. Six hours after seeding cells, the majority had begun to or were fully spread. Myofibroblast morphology was similar to their appearance when cultured on plastic, with spreading dendritic processes radiating from a central cell body (Figure 1-2). However, in comparison to *in vitro* culture on plastic, fewer cells were fully spread at the six hour time point. In fully spread myofibroblasts, force generation was focused to discrete oval areas that likely represent focal complex structures previously described in this population(214) (Figure 1A).

TFM revealed average traction forces were ~125 Pa on 4.47 kPa hydrogels (Figure 3), similar to some reports of stromal cells(73, 236), but slightly lower than other TFM measurements of fibroblasts(265). One explanation for this could be the use of fibroblast cell lines by some authors. My results in Dupuytren's myofibroblasts may reflect the use of human cells, with these being more sensitive to changes in their physical and mechanical environment during TFM. After quantifying average traction forces, I next examined the maxima of myofibroblast force generation. Figure 3 demonstrates the mean and maximum values of cellular traction forces in human myofibroblasts. Peak traction forces were ~634 Pa, with some cells able to produce forces in the kPa range (Figure 3B).

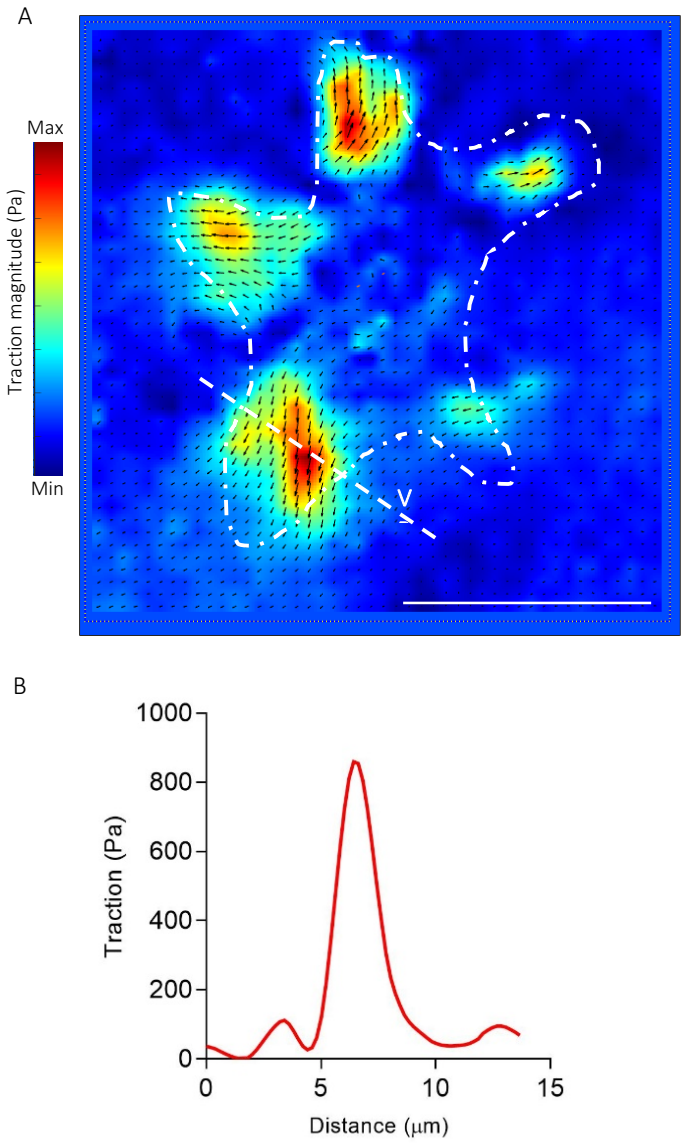


Figure 1. **Peak Traction force generation in a Dupuytren's myofibroblast.** A) Heatmap with white cell outline illustrating the distribution and magnitude of cellular forces across one representative human myofibroblast. Traction force scale bar to 1067 Pa. Scale bar represents 30 μm. B) Line profile of Line V in Figure 1A showing linear profile of peak traction force generation from one human myofibroblast. Pa = Pascal.

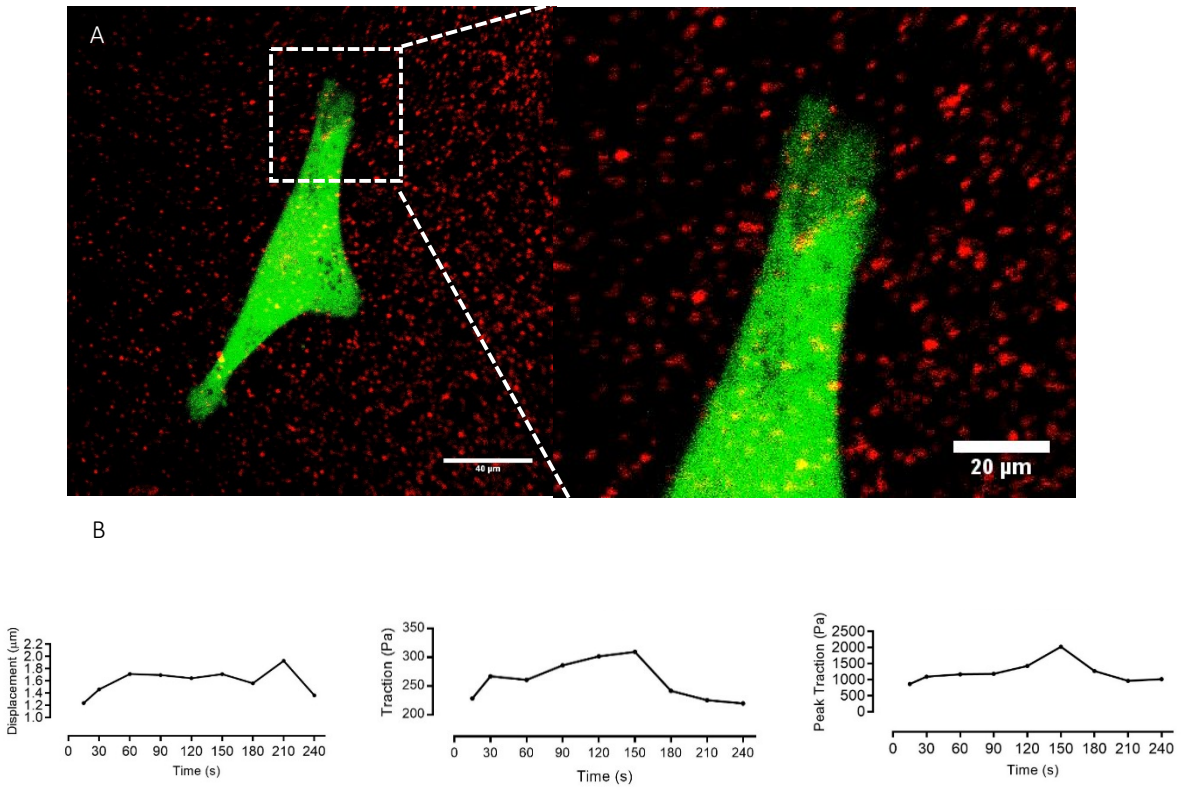


Figure 2. **Traction force microscopy defines human myofibroblast force generation**. A) Confocal image of Calcein AM tagged human myofibroblast (Green) seeded on a 4.47kPa PAA hydrogel with marker beads (Red). Scale bars at 40 and 20 µm. B) Line plots showing series of time-lapse recording over 240 seconds of cellular forces with peak bead displacement, average traction over cell area and peak traction force. Representative data from one myofibroblast. S = Seconds. Pa = Pascal.

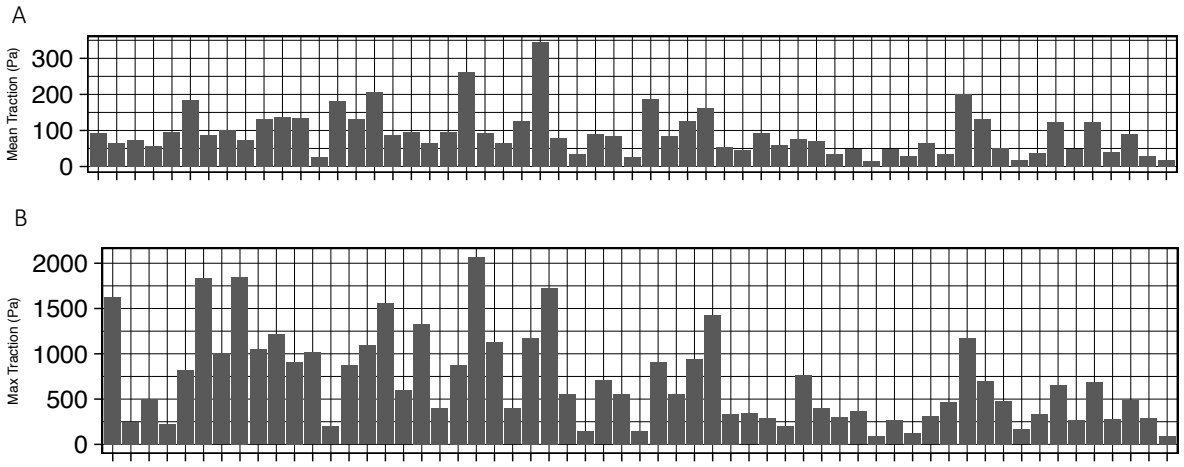


Figure 3. **Cellular traction forces of human myofibroblasts.** A-B) Bar plots demonstrating average normalized (A) and peak (B) traction forces in myofibroblasts on 4.47kPa PAA hydrogels (*n*=46 cells from >10 independent donors). X axis represents individual myofibroblasts. Pa = Pascal.

VII.III.II *Myofibroblasts generate traction force in discrete foci*

The distribution of cellular forces in fully spread myofibroblasts may relate to their underlying function*(214, 261)*. Oval discrete force foci could provide mechanical support for cells to bind and remodel ECM proteins(73, 259). Therefore, I next undertook detailed characterization of these structures. The overall topography of the force foci was highly conserved, with localized areas of high force magnitude encompassed by a more smoothed distribution of lower force (Figure 4). These structures were largely absent at earlier stages of cell spreading but matured as cells elongated and adopted a fusiform morphology (Figure 4-5).

Figure 4 illustrates the localized, discrete nature of myofibroblast traction forces, positioned along individual dendritic process (Figure 4A, 5, 6A). The surface plots also highlight the morphology of these structures with localized areas of high force, and the histograms (Figure 4C) show the distribution of cellular forces exerted by single myofibroblast. These demonstrate that there is a low frequency of high force magnitude across single cells as visualized in Figure 4A.

Next, I isolated and quantified traction forces within these focal areas. This showed forces followed a Gaussian distribution (Figure 6B). Finally, segmentation of the force foci revealed that areas of the cell outside these structures generated little force above noise level (Figure 6C). Thus, these data confirm that these focal adhesions must be responsible for the cell binding to the matrix on the surface of the hydrogel.

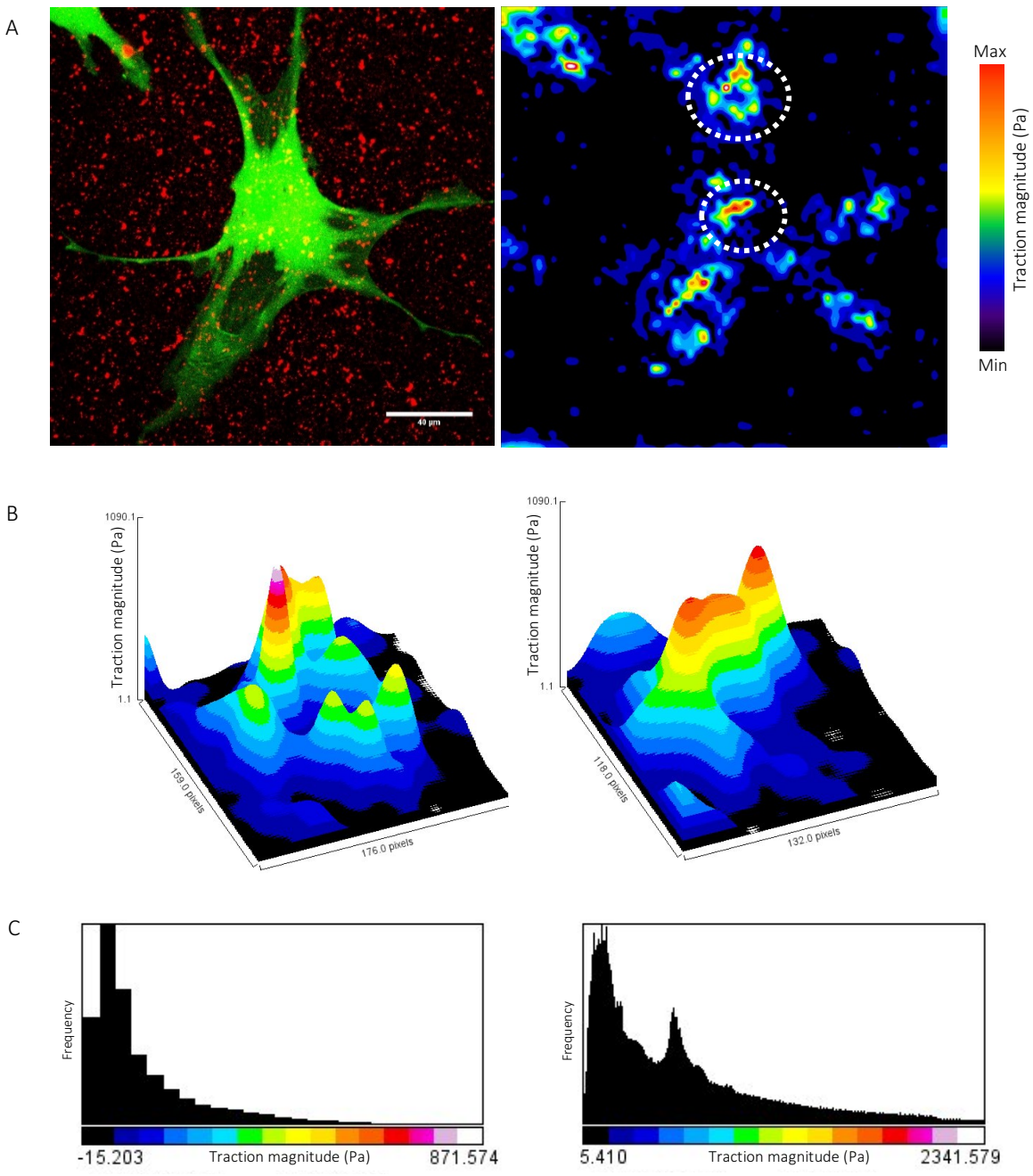


Figure 4. **Myofibroblast traction forces are discrete and localized**. A) Confocal image of Calcein AM tagged myofibroblast (Green) on 4.47 kPa PAA gel with marker beads (Red) and corresponding traction stress heatmap showing localized area of high traction force in red. Scale bar is 40µm. B) 3D surface plots of regions indicated by two white circles in Figure 4A, with Z axis representing magnitude of traction forces in Pascals. X and Y axis represent cell area in pixels. C) Histograms of traction forces over two representative myofibroblasts demonstrating few values of high force over entire cell area. Scale bar below in Pa. Pa=Pascals.

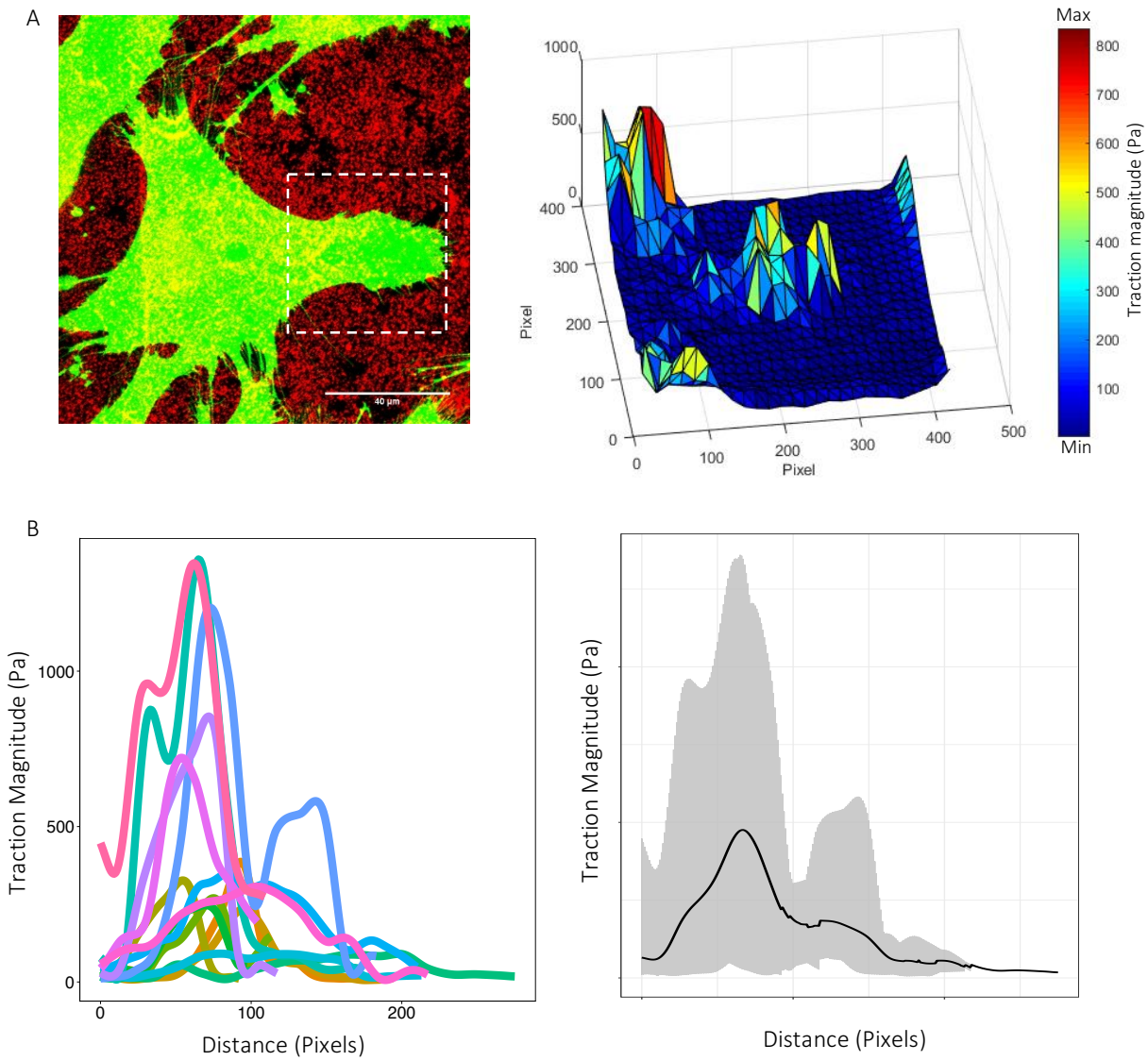


Figure 5. **Spread myofibroblasts exert traction forces in discrete localized focal structures**. A) Confocal image of Calcein AM tagged myofibroblast (Green) on 4.47 kPa PAA gel with marker beads (Red) and corresponding 3D traction stress surface plot showing localized area of high traction force in red. Scale bar is 40µm. Pa = Pascal. B) Line plots showing the distribution of forces along discrete foci in human myofibroblasts. Left panel demonstrates exact profile and right the mean force +/- range. *n*=10 force profiles taken from 10 individual myofibroblasts.

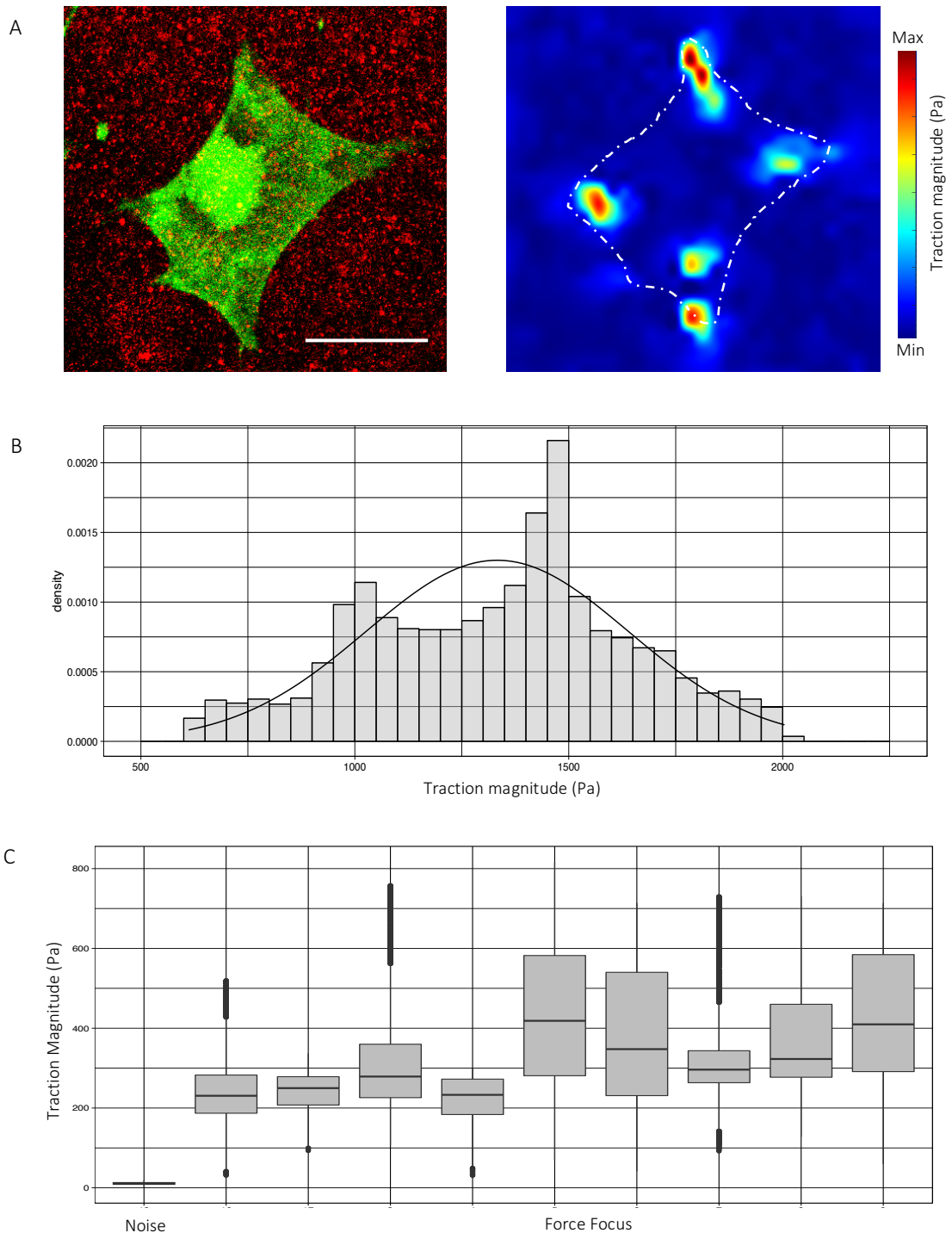


Figure 6. **Mapping the biophysical profile of force foci in human myofibroblasts.** A) Confocal image of Calcein AM tagged myofibroblast (Green) on 4.47 kPa PAA gel with marker beads (Red) and corresponding traction stress heatmap showing localized area of high traction force in red. Scale bar is 40µm. B) Histogram showing the distribution of forces in segmented force focus in human myofibroblasts. *n*=10 force foci from three independent experiments. C) Box and whisker plot showing the distribution of forces in segmented force foci in human myofibroblasts. *n*=10 force foci from three independent experiments.

VII.III.III *Dynamics of myofibroblast force generation during cell spreading*

After this, I used single cell force measurements to explore the spatiotemporal dynamics of myofibroblast force generation during cell spreading (Figure 7). I noted the distribution of forces evolved during this process. On initial contact with the PAA gels, myofibroblasts were circular and traction force localized to the lamellipodium, a discrete peripheral cell structure characterized by dendritic actin networks (Figure 9C). However, during spreading, the location of high cellular force moved to the discrete oval regions described previously. Based on these data, I performed correlation analysis to examine the relationship between total cell area and cellular forces. I found no clear correlation between cellular area and force generation (Figure 8).

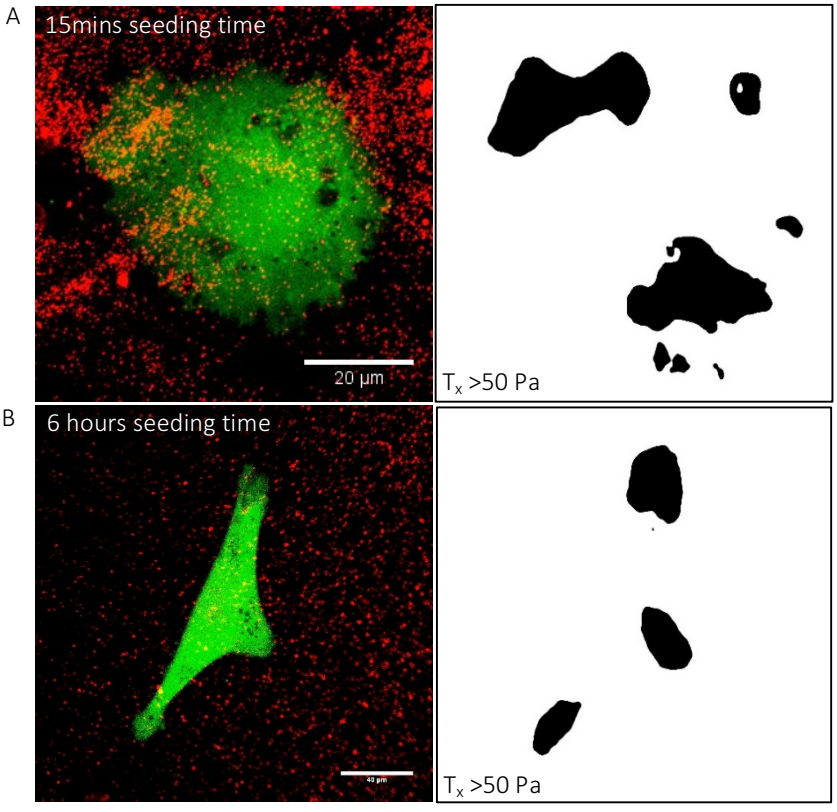


Figure 7. **Spatiotemporal dynamics of cellular forces in human myofibroblasts**. A-B) Confocal image of Calcein AM tagged myofibroblasts (green) seeded on 4.47 kPa PAA hydrogel 15 minutes after seeding (A) and 6 hours (B) after seeding with corresponding binary plot of traction forces above background level. Black regions represent areas of cellular traction force above background level (50Pa). Pa = Pascal. T_x = Traction. Scale bars at 20µm (A) and 40µm (B).

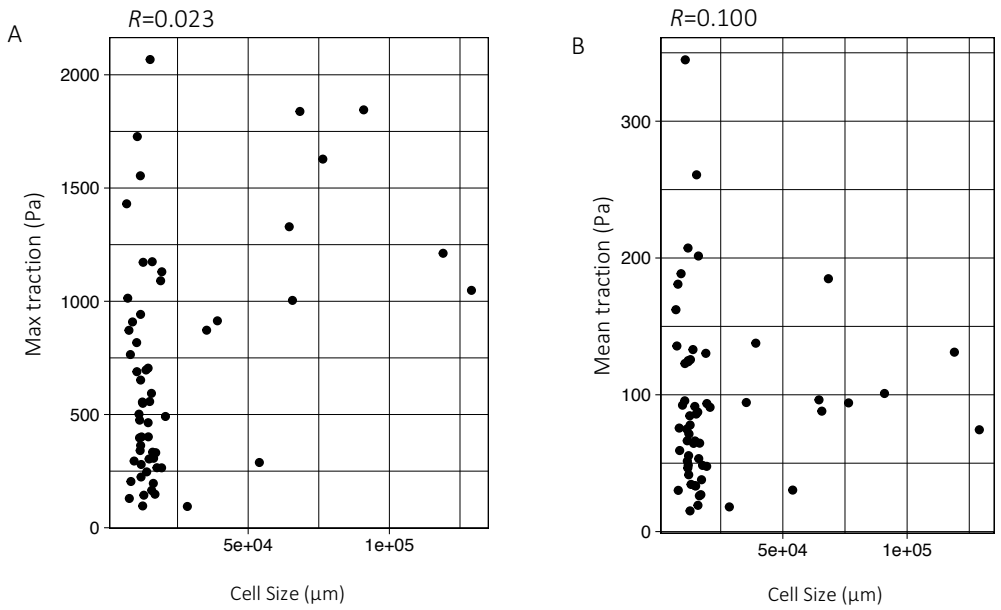


Figure 8. **No strong correlation between total cell area and peak or average traction force in human myofibroblasts**. A-B) Scatterplots showing peak (A) and average (B) traction forces exerted by cells against total cell area. Non-parametric spearman correlation coefficient (*R*) was used to calculate correlation. *n*=45 cells from at least three independent experiments.

After this, as I noted no correlation between area and max and average traction forces, I wanted to investigate whether there were any differences in force generation at distinct stages of cell spreading. For this, I calculated and compared the average and maximum forces in myofibroblasts at 15 minutes and six hours after seeding (Figure 7,9). In accordance with the correlation analysis this showed a conservation of energy between the two states and no significant difference in these parameters during early and late stages of cell spreading. Finally, immunofluorescence of F-actin

in spreading myofibroblasts confirm a circular morphology with dendritic actin networks positioned at the cell periphery (Figure 9C).

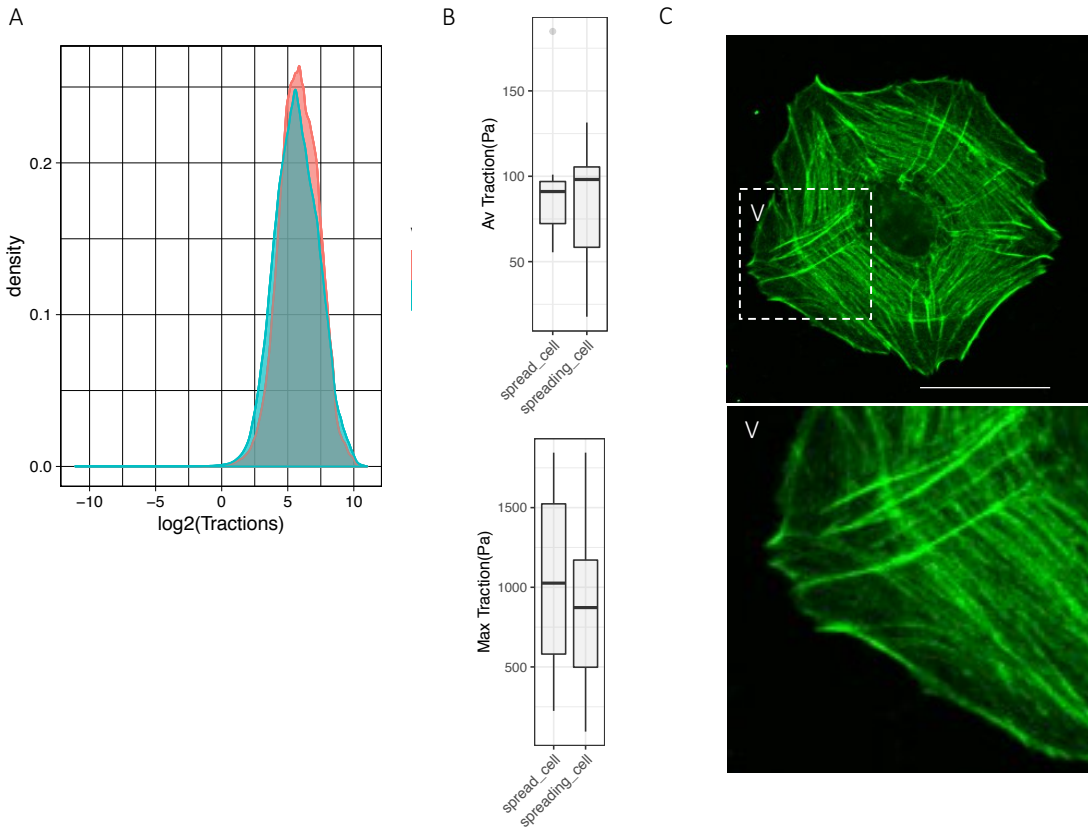


Figure 9. **Cellular traction forces are at equilibrium during cell spreading.** A) Density plot in log_2 scale showing traction forces within spreading (blue) and spread (red) myofibroblasts. Data represents all traction forces exerted by single myofibroblasts. *n*=12 myofibroblasts per condition. B) Box and whisker plots showing maximum (Max) and mean (Av.) traction forces in spreading and spread myofibroblasts. *n*=12 myofibroblasts per condition. C) Confocal image of F-actin staining (Phallodin-488) in spreading myofibroblast showing disorganized actin structures without clear stress fibres. Scale bar is 10µm. Pa = Pascal. V = zoomed area outlined by white square.

Next, I focused on examining force generation in spreading myofibroblasts (Figure 10). This confirmed the contrasting force profile as compared to fully spread myofibroblasts. Areas of high force magnitude were localized at the periphery, in a tight band around the cell circumference (figure 10A). This corresponds to the area of the cell enriched with dendritic actin networks identified using immunofluorescence (Figure 9C). When comparing force generation within these

structures (2µm from cell edge) to the rest of the cell, it was apparent that almost the entire force generation was localized to the cell periphery (Figure 10). The remaining cell area generated little force above noise levels (Figure 10C). I also noted that there were no discernible force foci in myofibroblasts undergoing spreading. Together, these results suggest the profile of myofibroblast force generation is highly dynamic during spreading but there is a conservation of energy during this process.

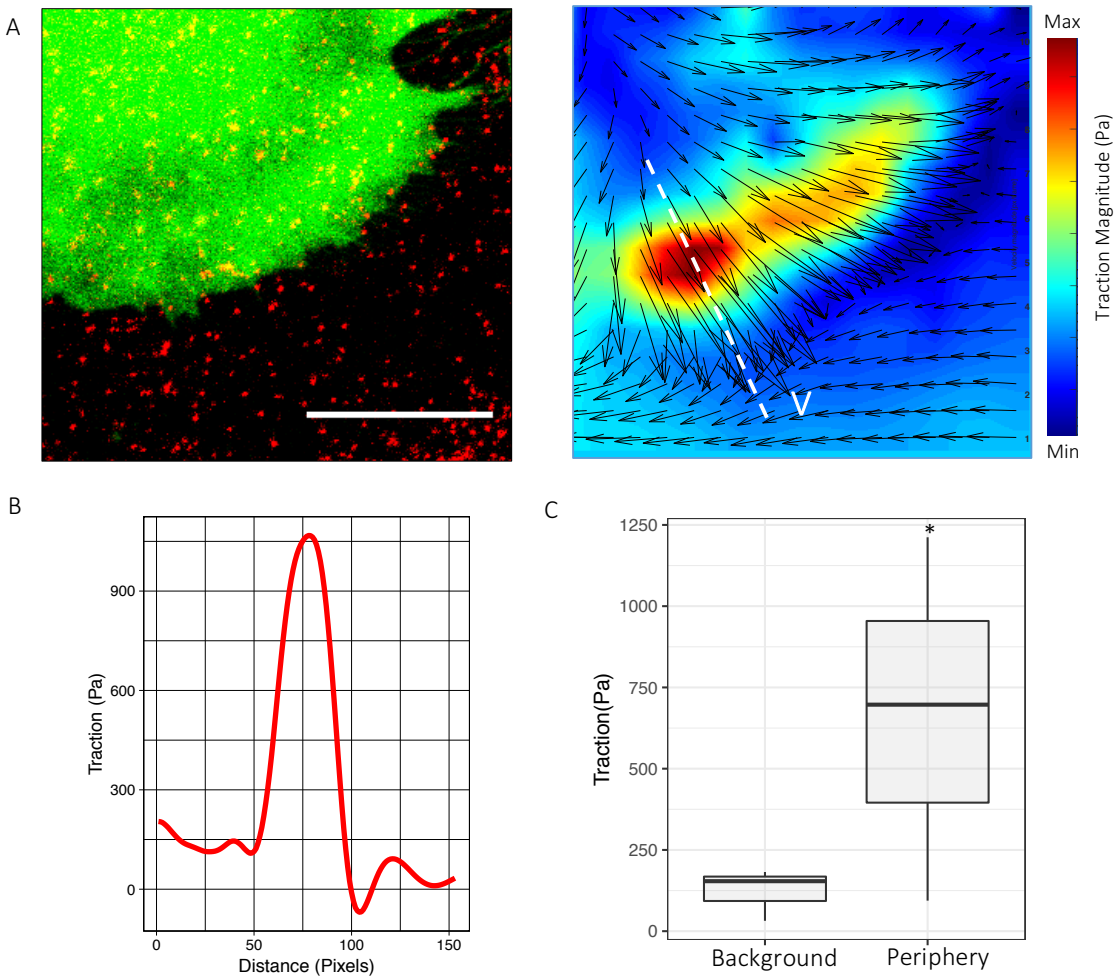


Figure 10. **Force generation in spreading myofibroblasts is localized to the cell periphery.** A) Confocal image of myofibroblast (Green) on PAA gel with marker beads (Red) and corresponding traction stress heatmap showing localized area of high traction force in red. Scale bar is 5µm. B) Line profile of Line V showing distribution of forces along one representative cell periphery. C) Box and whisker plot showing mean traction forces at cell periphery (2µm from cell edge) and the remaining cell surface (Background) during spreading. *n*=20 from three independent donors. p-values calculated by Wilcoxon Rank Sum test. * = $p < 0.05$. Pa = Pascal

VII.III.IV *Distinct biophysical communities in human myofibroblasts*

Single cell technologies have illuminated crucial biological process in diverse fields such as immunology(163), cell biology(150) and cancer(174). These techniques usually quantify a surrogate marker of cell function such mRNA molecules or protein abundance and subsequently ascribe functional heterogeneity on measured parameters. However, we are still to apply our growing knowledge in single cell biology to explore the functional behaviour of human cells. Given this, I finally sought to perform a detailed characterization of human myofibroblast force generation based on the knowledge I have gained from my single cell RNA-seq data.

I first quantified single cell force profiles of human myofibroblasts (n=59) using traction force microscopy. Hydrogel stiffness and imaging parameters were kept constant to allow for comparison between cell types. After this, I normalized force profiles to cell area, centred and scaled the data before input to principal component analysis (PCA). This was undertaken as PCA is a variance maximizing procedure and scaling aims to prevent bias in high variance single cell profiles dominating the signal. On inspection of the PCA plot, cells did not separate into obvious clusters (Figure 11). Instead, along the first two principal components there was a small nidus of myofibroblasts spreading out to more heterogeneous cells (Figure 11).

Next, I visualized single cell force profiles after partitioning the data using a clustering routine. I used the partitioning around medoids (PAM) algorithm to separate the data. PAM clustering was used as it is less sensitive to outliers and noise as compared to k-means. PAM separated the data into three distinct clusters and I explored the parameters of each (Figure 12). Interestingly, I found

that each cell cluster had distinct biophysical properties with one composed of highly contractile cells. This provides the first evidence that human myofibroblasts exists as distinct biophysical communities with unique force profiles.

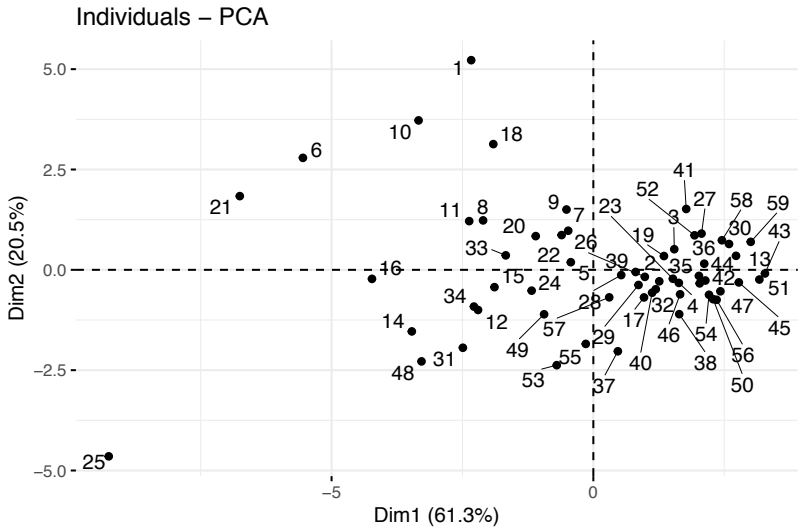


Figure 11. **Principal component analysis of human myofibroblast force signatures**. Scatterplot projecting single cell force profiles along the first two principal components (Dim1 and Dim2). Each point represents one cell, and each is labelled by the order in which it was collected (*n*=59, from >5 independent donors). % represents the proportion of variance explained by each principal component (Dim1-2).

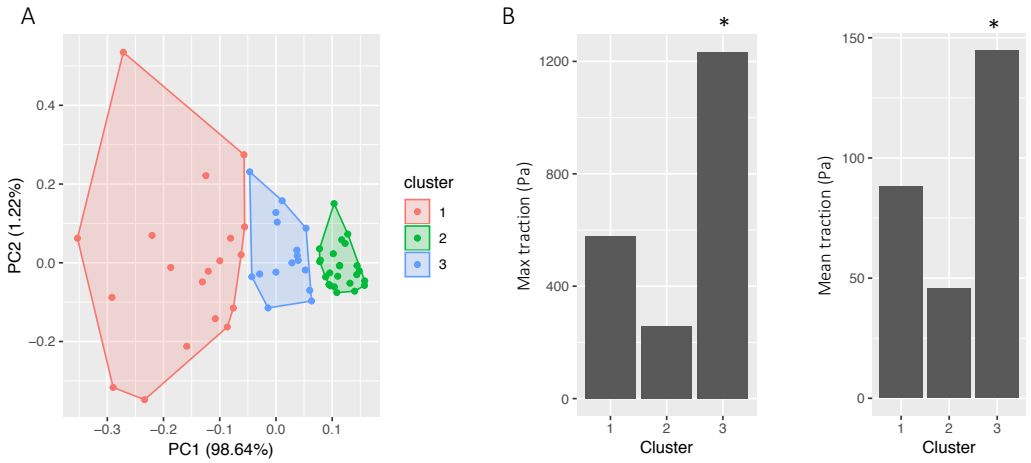


Figure 12. **Partitioning around medoids (PAM) clustering results in three biophysical communities of human myofibroblasts.** A) Scatterplot projecting single cell force profiles along the first two principal components with overlay of PAM clusters (*k*=3). PC = principal component. % represents the proportion of variance explained by each principal component (Dim1-2). B) Bar plots of max and mean traction forces per PAM cluster (1-3). p-values calculated by Wilcoxon Rank Sum test. * = p-value < 0.05, comparing cluster 2 with Cluster 3. Pa = Pascals.

After this, I applied a separate methodology to resolve potential sub-communities in human myofibroblasts. To cluster single cells by their force profiles, I used an unsupervised clustering routine, based on the Louvain algorithm, following recent approaches described for single cell RNA-seq. For this, I constructed a *k*-nearest-neighbour (*k*-NN) graph on the data using, for each pair of cells, the Euclidean distance between the scores of significant PCs to identify *k* nearest neighbours (see Chapter II.II.XVIII). The *k*-NN graph was then used as the input to the Louvain algorithm implemented using the 'cluster_louvain` function from the 'igraph' R package. This method returned three clusters, and the *k*-NN graph placed each along a linear trajectory, in accordance with the PCA projection (Figure 13).

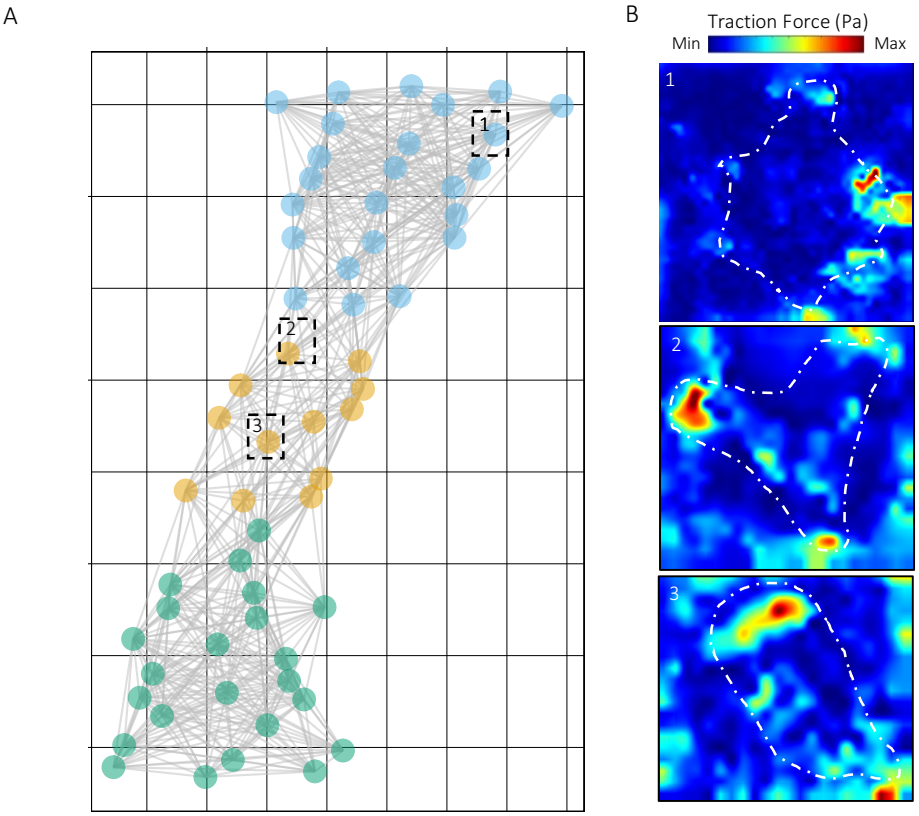


Figure 13. **Graph-based clustering reveals three myofibroblast communities along a linear trajectory.** A) *k*-NN graph of PCA meta-signatures (force profiles) coloured by Louvain cluster. Each point represents a single myofibroblast. B) Corresponding traction force heatmaps of cells identified in dashed black squares in (A) with Scale bar in Pa. Pa = Pascal.

VII.III.V *Functional heterogeneity along myofibroblast activation continuum*

Finally, I sought integrate data of biophysical functional readouts with results obtained from single cell transcriptomics to further interrogate functional heterogeneity in human myofibroblasts. In Chapter V.III.III I described a transcriptomic signature of myofibroblast activation and a novel cell surface marker, CD82, that followed this trajectory. I aimed to provide evidence for phenotypically distinct myofibroblast populations along this trajectory. To do this, I sorted freshly isolated $CD82^{high}$ and $CD82^{low}$ myofibroblasts from Dupuytren's nodules and performed traction force microscopy to measure cell contractility at single-cell resolution (Figure 14). This demonstrated that myofibroblasts from human fibrotic tissue are functionally heterogeneous, with $CD82^{high}$ marking a highly contractile population. In addition, $CD82^{+}$ myofibroblasts had distinct biophysical profiles, with discrete areas of high force magnitude enabling these cells to bind and remodel matrix proteins (Figure 15B).

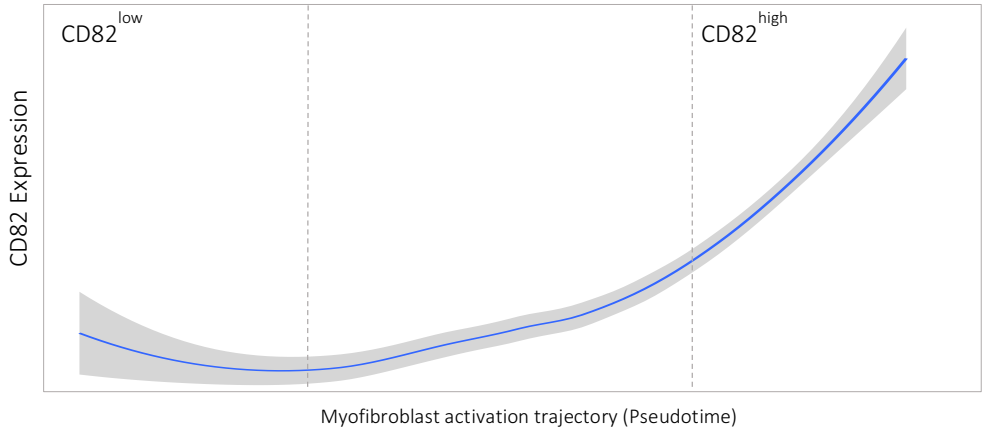


Figure 14. **CD82 expression follows pseudotime trajectory of myofibroblast activation.** Line profile showing gene expression of *CD82* (loess fit of normalized counts) along modelled pseudotime. Mean +/- SE plotted. Modelled pseudotime represents myofibroblast activation trajectory described in Chapter V.III.III. CD82 was shown to be highly correlated with this path and this plot shows a gradual increased in *CD82* expression along pseudotime.

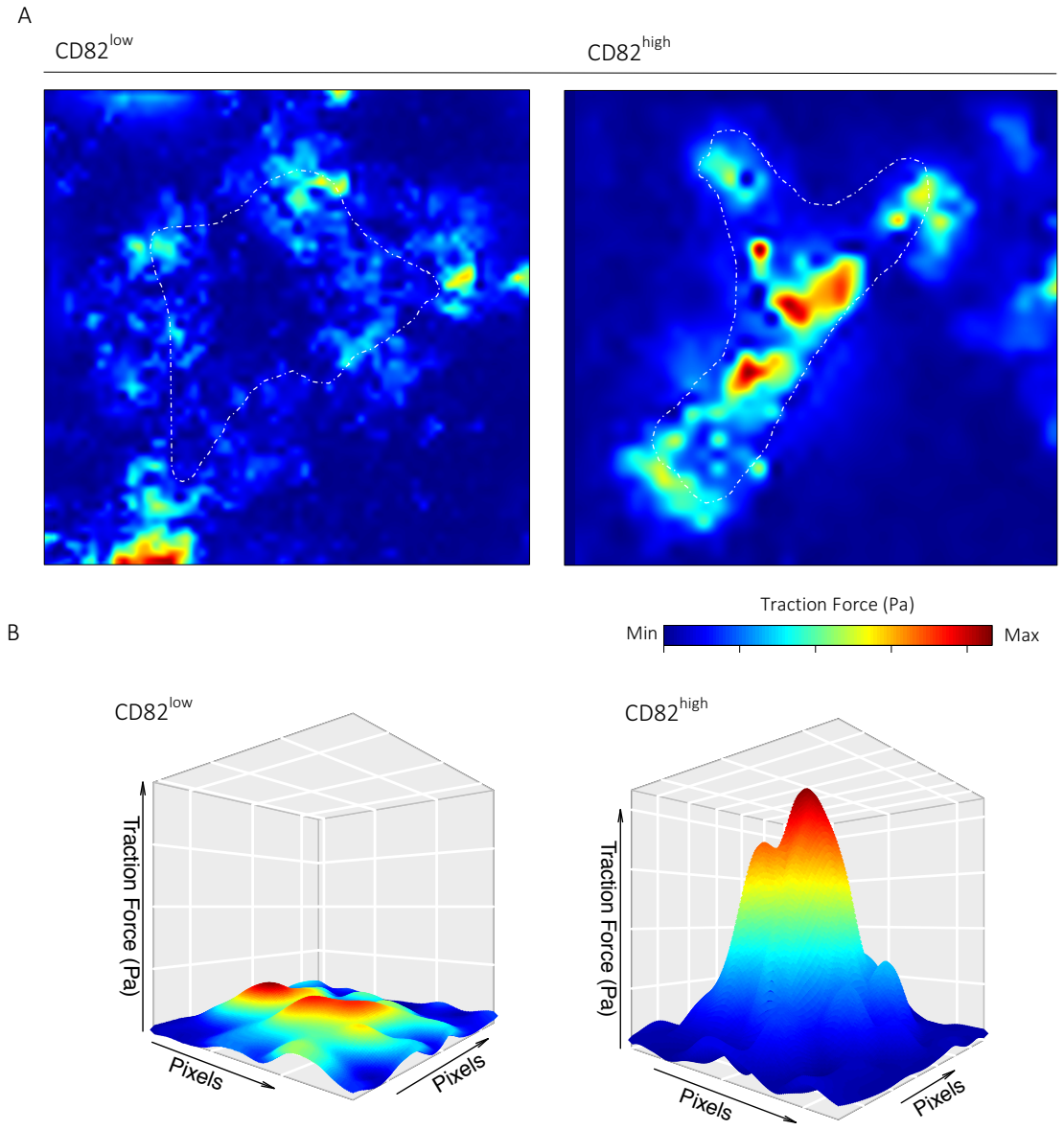


Figure 15. **CD82^{high} myofibroblasts are biophysically distinct.** A) Representative traction stress heatmaps showing force profiles of CD82^{low} and CD82^{high} myofibroblasts. Scale bar is 20µm. CD82^{high} and CD82^{low} defined by top and bottom 20% of cells based on CD82 expression (see Chapter II.II.XII). White dashed line is cell boundary. B) Representative 3D surface plots of force focus in CD82^{low} and CD82^{high} myofibroblasts. These were selected by segmenting force foci in myofibroblast populations that represented discrete areas of high force magnitude. Z axis is scaled to 800 Pa and colour represents traction force range across single cells. Pa = Pascals.

CD82^{high} myofibroblasts had an average maximum traction force of 2050 Pa (+/- 499 Pa) and mean traction force 385 Pa (+/- 78 Pa). In contrast, CD82^{low} myofibroblasts had an average maximum traction force of 533 Pa (+/- 26 Pa) and mean traction force of 72 Pa (+/- 17 Pa). Traction forces were significantly different between the two myofibroblasts (p-value < 0.01) (Figure 16). Finally, unsupervised tSNE clustering of force profiles grouped CD82^{high} and CD82^{low} myofibroblasts into separate clusters, suggesting they represent distinct communities (Figure 16C).

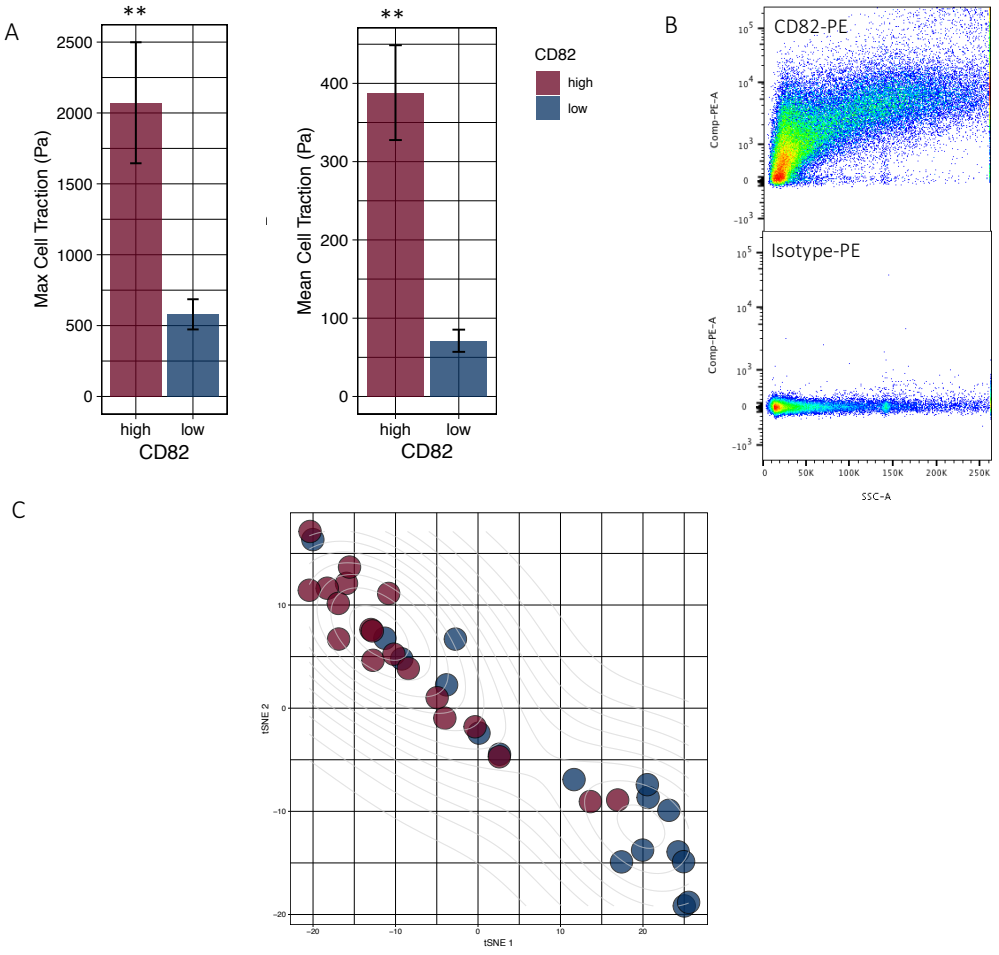


Figure 16. **Functional heterogeneity along myofibroblast activation trajectory marked by CD82.** A) Bar plots showing mean and max traction forces in sorted populations (CD82^{low} and CD82^{high}). *n*>15 cells from three independent donors. Pa = Pascal. B) Representative scatter plot of FACS analysis showing isotype control for CD82 antibody. C) tSNE projection of single cell force profiles in sorted populations illustrating clear separation of distinct myofibroblast communities. Cells coloured by sorted population. n>15 cells from three independent donors. ** = p-value < 0.01 (Wilcoxon Rank Sum test). Pa = Pascal.

VII.IV. Discussion

The results presented in this chapter aim to provide the first detailed measurements of force generation in human myofibroblasts. A major motivation for this pertains to myofibroblast contractility being key to their pathogenic potential *in vivo*(64, 178, 259). A description of this process may in future facilitate mechanistic insights to be gained from established or merging anti-fibrotic treatments. I envisage functional readouts obtained using traction force microscopy (TFM) complementing traditional readouts of myofibroblast activation, such as gene and protein expression, to provide a more complete perspective on how therapeutics may influence cell function.

Methodological developments in TFM may also allow mechanical forces to be quantified at greater depth, with greater precision, extending my observations at a micrometer scale. One exciting novel development is the integration of TFM with super-resolution microscopy, overcoming the diffraction limit of light and providing force measurements at nanometer resolution(81). Super-resolved TFM may help uncover further complexity in myofibroblast force generation. In this chapter, I have undertaken a close examination of sub-cellular biophysical structures in myofibroblasts, demonstrating forces are generated in discrete and localised foci. Moving forward, super-resolved TFM may provide further information on the precise morphology and topography of these structures.

The large ovoid regions of maximal force in myofibroblasts corresponds to structures termed focal complexes(214). Focal complexes are 5-16 μm and are a mature form of focal adhesion that functions to remodel ECM proteins(214). 3D visualization of force profiles in myofibroblasts matches in size and topography the predicted morphology of a focal complex binding fibrillar ECM proteins(54, 266), and could be speculated to represent the biophysical profile of this structure. Of note, segmenting and comparing force generated within and outside these structures confirmed that the remaining cell area exerted force comparable to baseline. Thus, as force is required for cell-ECM binding, it can be surmised that these localised areas of force are responsible for myofibroblasts binding to the ECM. As the gel surface was coated with type I collagen, it would be interesting to see how these structures change when binding other matrix proteins, such as fibronectin. It could be hypothesized that distinct matrix proteins require differential force for cell binding and thus may impose unique biophysical signatures on myofibroblasts. Extending the results presented in this chapter, we should now look to test how these structures adapt and alter when exposed to different matrix proteins.

One interesting observation, that there was no clear correlation between cellular area and force generation, conflicted with previous reports that suggested a tight correlation between the two(267, 268) (Figure 8). One reason for this disparity could relate to differences in methodology. In several previous studies, the exact cell area was tightly controlled though a process termed micropatterning, in which cells are confined to a discrete region of the hydrogel(269). Consequently, cell morphology and structure are finely regulated, but this may also impose changes to the measured force profile(269). It could be reasoned that the complex geometry of

spreading myofibroblasts leads to characteristic force generation and makes direct correlations between cell area and traction forces challenging.

Another finding from the TFM was the conservation of energy during cell spreading. Although the distribution of traction forces changed dramatically during this process, the overall maximum and mean traction force per cell remained constant. One could conceptualize a biological reason for such a phenomenon in the context of myofibroblast function. Myofibroblasts are crucial mediators of normal wound healing during which they effectively restore tissue architecture through remodelling matrix proteins(69, 109). Consequently, it could be reasoned that spread myofibroblasts exerting forces in distinct foci capable of remodelling matrix proteins should exert significantly more force than a cell undergoing spreading, where the cell would devote more energy during its primary function. However, the data presented in this Chapter contradict this hypothesis and show equal cellular force magnitude before myofibroblasts have adopted an elongated morphology facilitating matrix remodelling. One explanation for this could be the need for myofibroblasts to obtain an elongated and spread morphology, enabling ECM remodelling, as quickly as possible. In this regard, a cell would devote significant energy to hasten spreading to adopt a phenotype that enables its key functions(270).

Single cell biology has uncovered many critical cellular processes and fostered a deep appreciation for the significance of heterogeneity in biology(141, 151, 167). As traction force microscopy provides a readout of contractility at single cell resolution, I was interested in exploring diversity in human myofibroblasts based on this functional parameter. Unbiased clustering of

myofibroblasts resulted in three distinct biophysical communities. In addition, PCA and graph-based approaches placed each cluster along a visible trajectory structure. One could reason that the overall topography of these data suggests that cells exist along a continuum, rather than as discrete entities. This hypothesis is consistent with my previous results in Chapter V.III.III showing transcriptomically, myofibroblasts form an activation continuum towards a novel $CD82^{+}$ population. The concordance between gene expression and biophysical profiles supports the presence of a strong trajectory structure in myofibroblasts. Looking forward, we must now aim to integrate functional parameters with gene expression or proteomic readouts in the same single cells to enable correlation between gene markers and myofibroblast function. Furthermore, distinct biophysical communities require further validation, including whether the reported communities are conserved over time, or rather show fluctuating and dynamic force profiles.

In this chapter, I sought to integrate findings obtained from single cell RNA-seq with functional readouts at single cell resolution. The overarching aim of this was to prove phenotypic diversity along a modelled transcriptomic signature of myofibroblast activation. This would not only help validate the gene expression data but provide support that $CD82^{high}$ myofibroblasts represent a functionally distinct community. The results presented here demonstrated that $CD82^{high}$ myofibroblasts are a phenotypically distinct and mark a highly contractile population. In addition, the force profile of these cells was characterized by force foci of high force magnitude enabling these cells to bind to and remodel matrix proteins efficiently. Given the highly contractile nature of $CD82^{high}$ myofibroblasts, it is possible that *in vivo* these cells are key pathogenic drivers. Just as I have undertaken using traction force microscopy, further work should systemically investigate

the functional properties of this population. In addition to a contractile phenotype, they may also express ECM proteins, be highly motile or even promote inflammation. These are important questions but are beyond the scope of this thesis. In future, further characterization of $CD82^{+}$ myofibroblasts may help illuminate novel therapeutic targets in DD.

In summary, the results presented in this chapter explore functional heterogeneity in human myofibroblasts and integrate observations made in single cell gene expression data. Here, myofibroblasts are shown to exist in distinct biophysical communities and exhibit phenotypic diversity along a modelled differentiation trajectory.

Chapter VIII. Conclusion

VIII.I. Dissecting the multicellular ecosystem of human fibrosis with single cell RNA-seq

Investigations into the mechanisms of fibrosis depend upon accurate identification and categorization of cell types(45, 241, 271, 272). Assignment of molecular features and transcriptomic profiles to single cells will allow fundamental processes to be understood with greater precision(273, 274). Furthermore, comprehensive cellular taxonomies enable a comparison between disease states and across different technologies(274). At present, we are in the infancy of understanding the complexity of stromal cells in fibrosis, and their precise role in disease progression.

Fibrotic disorders represent some of the most devastating and poorly treated conditions in the Western World. Despite this, effective therapeutic interventions have yet to be identified for the majority of these conditions(275). A major barrier for successful clinical translation is a poor understanding of developing fibrotic microenvironments in humans given their dynamic and fluctuating nature. In this thesis I use Dupuytren's disease (DD), a localized fibrotic condition of the hand, as a model to construct a stromal cell atlas in human fibrosis at single cell resolution. Subsequently, I leveraged this atlas with functional assays to validate phenotypically distinct stromal cell subsets and states in DD. I hope these results embody the first step towards knowing how stromal cells orchestrate fibrosis, their interactions, and whether one or more represent crucial therapeutic targets.

The fibrotic microenvironment includes fibroblasts, myofibroblasts, endothelial, immune and perivascular cells(1, 165, 201, 276). A major goal of this thesis was to explore not just myofibroblasts, but the entire spectrum of cell types in DD nodules and to focus initially on immune cells and pericytes. Pericytes are rounded fusiform-shaped cells positioned adjacent to blood vessels with branching cytoplasmic processes(201, 277). In normal tissues, pericytes play a crucial role in vascular development and homeostasis through the secretion of soluble factors and direct cell-cell communication with the endothelium. Since their first description, many pericyte markers have been described, the most commonly used including platelet-derived growth factor receptor beta (PDGFRβ)(278) and desmin(279). Nevertheless, we still lacked a complete census of human pericytes or a validated universal pericyte marker. Many markers only labelled subpopulations and others showed dynamic expression under varied pathological states(279). To add to this complexity, in heterogeneous disease systems such as fibrosis, it becomes challenging to distinguish pericytes from other resident or invading mesenchymal cells. In Chapter III.II.IV, I provide the first comprehensive single cell atlas of pericytes in human fibrosis, capturing over 7,000 cells from 12 DD patients. Clustering of the entire DD nodule dataset identified a clear pericyte population, transcriptomically distinct from other stromal cells and housing four subsets (pericyte 1-4). Subsequently, *MCAM* (CD146) was validated as a specific pericyte marker, showing expression across all four subsets (pericyte 1-4) and as a cell surface protein provides the opportunity for future FACS sorting studies. Functionally, CD146 is a surface glycoprotein shown to regulate PDGFR-β-induced pericyte recruitment(280) and coordinate endothelial-pericyte communication in the blood brain barrier(281). We are still missing an exact role of CD146 in DD, but this could be an exciting avenue for future research.

Besides their role in vascular homeostasis, pericytes are of significance in fibrosis as myofibroblast precursors in various disease models. For example, in a murine model of kidney fibrosis(208), fate mapping revealed pericytes derived from metanephric mesenchyme contributed the majority of α-SMA^{+} myofibroblasts. Despite the elegance of this study, it highlights the challenge of studying pericytes in fibrosis. In this publication, a *FoxD1-Cre* driver bred into the ROSA26 reporter mouse lead to LacZ expression in stromal cells that included mesangial cells, vascular smooth muscle cells and pericytes. After this, only three markers (CD73, PDGR-β & α-SMA) were used to characterize the progression of interstitial pericytes to activated myofibroblasts. Considering this, my results in Chapter III.III.IV show that *PDGFRB* marks only a proportion of pericytes, emphasizing the risk of bias when using only a few markers. An important next step will be to repeat previous murine models of fibrosis and undertake single cell profiling. This may provide insights into which subsets are preferentially upregulated in fibrosis, and whether the subsets described in DD are conserved across models of fibrosis.

In summary, the pericyte census presented in Chapter III.III.IV will hopefully allow a more detailed understanding of this population in fibrosis. One vital next step is to test whether the pericyte subsets (pericyte 1-4) in DD carry distinct functional properties. This will be important to validate the findings in this thesis and may uncover pericytes with discrete functions. Ideally, our future goal will be to build an objective nomenclature of pericytes in health and disease and single cell RNA-seq will be invaluable for this task.

Low-grade chronic inflammation is firmly associated with the nascent stages of human fibrosis(165). In line with this, my fibrotic atlas revealed a complex immune landscape, with predominant cell types being macrophages, T cells and dendritic cells (see Chapter III.III.V). In Chapter I.I.I, I introduced a simplified macrophage nomenclature describing classically activated (M1) and alternatively activated (M2) macrophage phenotypes. Applying single cell RNA-seq to DD nodules (see Chapter III.III.V), I could identify immune cell clusters enriched for both M1 and M2 markers. Future work should aim to expand upon this observation and look to characterize distinct macrophage lineages in DD. After this, correlating transcriptomic profiles with protein readouts and clinical parameters may help classify pro-fibrotic signatures of macrophages, which remains active area of research.

With regards to the immune cells profiled there are two major limitations of the single cell dataset. Firstly, the low sequencing depth obtained using droplet-based technologies may prevent lowly expressed transcripts being captured. Secondly, the relatively low numbers of immune cell profiled (*n*=1033). These factors make it challenging to discern potentially meaningful sub-clusters in the data and major conclusions must be restricted to the identification of major immune cell types. Another potential source of bias is technical artefacts introduced during tissue processing, as immune cells may be more susceptible to injury. To obtain a more detailed view of the immune compartment in DD it would be useful to sort and purify $CD45^{+}$ cells freshly isolated from nodules and sequence this enriched population. By increasing the number of cells sequenced, it would improve the resolution to resolve intermediate cell states or rarer populations. In addition, it could help dissect the macrophage populations in greater detail.

Another interesting future direction would be to further characterize T cells in DD. Due to the limitations listed above, the results presented in Chapter III.II lack the required resolution to draw strong conclusions regarding the presence of distinct T cell lineages. One next step could be to undertake paired single cell RNA-seq with T-cell repertoire sequencing on the enriched $CD45^+$ population from DD nodules, looking for clonally expanded T cell subsets. It has been proposed that DD may represent a T-cell mediated autoimmune disorder and this project could help test this hypothesis(30).

Together, the data presented in thesis provide a perspective on the transcriptomic complexity of immune cells in DD, whilst remaining descriptive. Hopefully, this will be the first step towards uncovering the mechanism of immune cell recruitment and activation in this common fibrotic condition.

VIII.II. Fibroblasts in the fibrotic milieu

Resident fibroblasts make up a significant proportion of tissue structure and are central for diverse biological processes(282). Despite this, a detailed understanding of their function in fibrosis remains incomplete, stimulating a major aim of this thesis; to characterize fibroblasts in DD with regards to their expression profile and potential subpopulations. The results presented Chapter IV suggest that fibroblasts in DD could perform an immune regulatory function and were shown as the major source of stromal-cell derived chemokines (IL8 & IL6). Chemokine secretion could allow these cells to promote and maintain immune cell recruitment into nodules that express factors (e.g. TNF) to drive stromal-cell activation. My results parallel findings in breast cancer, where fibroblasts have been ascribed immune modulatory functions(235). In one study, Costa A et al (2018) described four subsets of carcinoma-associated fibroblasts (CAFs) in human breast cancer, with one (CAF-S1 fibroblasts) attracting $CD4^{+}CD25^{+}$ T lymphocytes and promoting their differentiation to $CD25^{High}FOXP3^{High}$ regulatory T cells, whereby inducing an immunosuppressive environment. In the context of cancer, CAF-S1 fibroblasts may be of translational significance and could be targeted to augment treatment strategies using immunotherapy. With regards to DD, dynamic $ICAM1^{+}$ fibroblasts may also represent an interesting therapeutic target.

In line with my description of $ICAM1^{+}$ fibroblasts, a recent study using murine models and a human organoid system, discovered a population of CAFs in pancreatic ductal adenocarcinoma (PDAC) that selectively secreted IL-6 and other inflammatory mediators(171). Indeed, it has long been appreciated that stromal-cell derived chemokines drive immune cell recruitment in various

diseases, but we are just beginning to localize their source to discrete cell populations(283-285). An increasing use of single cell technologies will refine our understanding pathogenic stromal cells and hopefully illuminate common expression profiles of immune regulatory fibroblasts(154). My data support this concept, as ICAM1^{+} fibroblasts in DD share features of inflammatory fibroblasts in breast and pancreatic cancer, including high expression of *IL6* and *IL8*. Moving forward, this population needs to be temporally mapped in the context of tumorigenesis and other fibrotic diseases. If, as in DD, ICAM1^{+} fibroblasts are dynamic, they could represent a target for inhibiting inflammation. Further experimental validation using immune cell and fibroblast co-culture systems will provide important insights into how ICAM1^{+} fibroblasts may directly regulate immune cell function.

Despite limited single cell RNA-seq studies of stromal cells, it appears there may be common genetic markers of pathogenic fibroblasts. In human melanoma, single cell RNA-seq was applied to 4,645 single cells isolated from 19 patients with metastatic disease(174). In this study, differential gene expression between immune cells and CAFs identified key marker genes for the latter as *SERPING1*, *C1R*, *C1S* and *CXCL12*. These markers were also highly specific for fibroblasts in DD. As *SERPING1*, *C1R* and *C1S* all encode proteins of the complement family(286), this highlights a putative pathway through which fibroblasts could promote inflammation. The complement system is a strong stimulus for multiple features of inflammation, including immune cell recruitment, vasodilation and endothelial activation and hence its activation by fibroblasts could perpetuate tissue injury and fibrosis(286, 287). With a growing number of single cell studies, it will be vital to assimilate and integrate molecular profiles to investigate conserved genetic

markers of pathogenic fibroblasts and ask whether the complement system may endow these cells with pro-inflammatory properties.

Fibroblasts have been investigated in various anatomical sites and diseases using single cell RNA-seq, including human skin(288), early murine wounds(245) and hair follicles(289). In human skin fibroblasts, Tabib et al (2018) described two major populations characterized by the co-expression of *SFRP2*$^{+}$*/DPP4*$^{+}$ and *FMO1*$^{+}$*/LSP1*$^{+}$(288). In another study, Philippeos et al (2018) identified five fibroblast populations in the skin: upper and lower dermal fibroblasts, pericytes, and two uncharacterized populations(290). In DD, the three major subsets of fibroblasts (CD34^{+}, PDPN^{+} & ICAM1^{+}) were not distinguished by markers reported in dermal fibroblasts, suggesting a phenotypic distinction between skin and fibrotic fibroblasts. In contrast, DD fibroblasts do share molecular features with recently described rheumatoid arthritis (RA) fibroblasts(154). In this disease, three major fibroblasts were also reported (CD34^{+}, CD34^{-}THY^{-} and THY^{+}) and each carried distinct functions and spatiotemporal relation to the disease. Fibroblast markers in RA conserved with DD fibroblasts included *PDPN, CDH11* and *CD34.* One clear distinction is between the chemokine-enriched population, as in RA CD34^{+} fibroblasts are the major source of chemokines as opposed to ICAM1^{+} fibroblasts in DD. This indicates that immune regulatory subsets may have contrasting molecular profiles between rheumatic and fibrotic disorders.

Another opportunity afforded by single cell transcriptomics is the ability to examine potential relationships between cell types. In addition, intermediate cellular states and phenotypic

transitions can be explored. Pseudotemporal ordering is a general term for computational methods aimed at elucidating underlying developmental paths or trajectories. In Chapter IV.III, diffusion maps were applied to place fibroblasts in a pseudotemporal order, uncovering discrete trajectories from $CD34^+$ and $ICAM1^+$ fibroblasts towards $PDPN^+$ fibroblasts. This may signify an underlying developmental path along which $PDPN^+$ fibroblasts are an intermediate state. From this analysis, one could speculate that the $PDPN^+$ fibroblast represent a proto-myofibroblast, and supporting this they also showed low expression of *ACTA2* and other myofibroblast markers. What must be remembered is that pseudotemporal orders are directionless, and thus $PDPN^+$ fibroblasts may in fact represent the precursor population. The logical follow up of this observation is functional validation of distinct fibroblast subsets. Perhaps, as in RA, each carry unique molecular functions contributing to the disease. Another important question is whether genes showing a significant association with transitional states in the pseudotemporal order may be key regulators instructing the differentiation of fibroblast subsets in DD. To test this, individual fibroblasts ($CD34^+$, $PDPN^+$, $ICAM1^+$) could be isolated and subjected to perturbation experiments to test for phenotypic transition between subsets.

VIII.III. A detailed classification of human myofibroblasts in fibrosis

Since their first description by Gabbiani(56), myofibroblasts are now known as central to forms of fibrosis(56, 60). Despite this, a precise molecular classification of this cell was missing, in part due to their dynamic nature *in vivo* and *in vitro*. Alpha-smooth muscle actin (α-SMA) remains the only consistent myofibroblast marker(60) and the results in Chapter IV.III.I show *ACTA2*, alongside others such as *SPON2* and *FABP5*, as one of the most significant differentially expressed genes between myofibroblasts and fibroblasts.

In Chapter V, one of the most exciting results was elucidating a novel cell surface marker of myofibroblasts, CD82, that showed a tight co-expression with α-SMA. This represents the first myofibroblast specific cell surface protein in DD and in future may help inform experimental strategies to isolate and study this cell. As more single cell datasets are generated, it will be important to ascertain if CD82 is a conserved myofibroblast marker in different organs, or whether its expression is specific to DD. In Chapter V, I provide an initial exploration of this question, integrating the single cell atlas of DD with a single cell RNA-seq dataset of murine lung fibrosis, showing *CD82* marks contrasting stromal cell types in lung and hand fibrosis. Future work should ask whether $CD82^{+}$ stromal cells carry discrete functional properties in fibrotic disorders across anatomical sites.

A central question in fibrosis is the origin of myofibroblasts(241, 242, 271). Although is it likely that activation of local tissue fibroblasts contributes a major proportion, numerous studies describe a

diverse range of precursor cells in multiple organs, including pericytes(277), circulating fibrocytes(291), neighbouring endothelial cells(292, 293) and epithelial cells(294, 295). Indeed, distinct precursors are associated with specific organs undergoing fibrosis, for example circulating fibrocytes in the lung(296) and endothelial cells in heart(293). Considering this, there may be varied mechanisms that instruct the development of a mature myofibroblast and mapping the ontogeny of myofibroblast differentiation *in silico* could provide valuable insights into regulatory programs governing this process. In one example, Bach et al (2017) used diffusion maps to elucidate the hierarchy of mammary epithelium during its development and concluded that *in vivo* this process forms part of a continuum(297). In Chapter V, I apply a similar methodology and provide evidence for an activation trajectory in myofibroblasts, moving towards a pathogenic $CD82^{+}$ phenotype. The strong concordance between linear dimensional reduction projections and non-linear modelling highlights this activation trajectory as a significant feature in the data. Looking forward, transcription factors showing an association with the inferred trajectory, such as *MAFB*, should be functionally tested with the aim of identifying central regulators of myofibroblast differentiation in DD. Notably, the modelled trajectory failed to ascribe a precursor population and represented only a differentiation path through the myofibroblasts. An important next step would be to expand upon the pseudotime analysis in Chapter V and ask if discrete trajectories exist between myofibroblasts, endothelial cells, pericytes and fibroblasts. This, in turn, may help determine a candidate for a myofibroblast precursor in DD. One interesting cell to start at would be pericytes, as the pseudotime analysis in Chapter III.III.IV suggest pericyte 4 may represent an intermediate state between pericyte and myofibroblast.

Discovery of $CD82^{+}$ myofibroblasts presents an exciting opportunity to investigate novel therapeutic targets in DD and may hold significance in other fibrotic disorders. This population shows high expression of several genes (e.g. *TNFSF4, IL11* and *MMP14*) strongly implicated in fibrosis across multiple organs. For example, *TNFSF4* (OX40L) has been demonstrated as crucial for inflammation driven fibrosis in the skin(298). In addition, IL-11 is now recognized as a central mediator of cardiovascular fibrosis where it acts to augment TGF-β1 signalling(299). To extend the observations made in Chapter V.III.I, $CD82^{+}$ myofibroblasts require further functional characterization across a range of fibrotic diseases. As in DD and lung fibrosis, CD82 may not only mark distinct populations, but also hold contrasting functions to promote fibrosis. In DD, it regulates cell cycle progression but in other organs it may control cell contraction, adhesion or migration.

While important insights can be drawn from computational methods, findings require appropriate biological validation. With regards to pseudotemporal analysis, most studies so far have been undertaken in well-studied models such as murine and human haematopoiesis. Collectively, I have aimed to associate a modelled transcriptomic signature of myofibroblast activation with functional heterogeneity along its path (see Chapter VII). Using CD82 as a marker of myofibroblast differentiation, I was able to demonstrate phenotypic diversity along this structure. However, due to experimental limitations, related to the relatively few cells obtained from FACS sorting and a dependence on enough cells as input for the traction force assay, I was restricted to measuring force profiles on only two populations ($CD82^{high}$ and $CD82^{low}$). Ideally, to prove this path signifies

a continuous process I would aim to sort cells at multiple points, measure cell contraction, to prove a continuum of distinct myofibroblast populations or states.

VIII.IV. The membrane tetraspanin CD82 in fibrosis

Despite several limitations, the bleomycin model is the most commonly used murine model of lung fibrosis(22). I took advantage of a recent study performing single cell RNA-seq in this model to see whether discoveries in DD were conserved(141). Remarkably, *Cd82* marked a distinct *Acta2*$^{-}$ *Col13a1*$^{+}$ fibroblast, which was the most upregulated population during fibrosis progression. In addition, pseudotemporal analysis and marker genes infer that *Acta2*$^{-}$*Col13a1*$^{+}$ fibroblasts may be a proto-myofibroblast, able to expand in fibrosis and providing a pool of α-SMA^{+} myofibroblasts. In contrast to this hypothesis, a second possibility is that *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts exist alongside myofibroblasts with a separate function. In the murine model, *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts did not exhibit high expression of genes encoding chemokines or cytokines and lacked a proliferative signature. Considering this, one potential role of *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts could be to support the survival and function of myofibroblasts, either through direct cell-cell communication or secretion of a soluble factor. A future direction would involve further characterisation of *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts with regards to their functional properties in lung fibrosis. As part of this project, it will be important to spatiotemporally map *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts further in the human lung, extending initial immunohistochemistry results reported in Chapter V.III. An understanding of the exact location and function of *Cd82*$^{+}$*Col13a1*$^{+}$ fibroblasts in human fibrotic disease will be an important avenue for future research.

Functions of CD82 have been most extensively investigated in the context of cancer and T cell biology(300-302). CD82 is a heavily glycated cell surface protein and a member of the tetraspanin

superfamily(300-302). In cancer, CD82 is a metastasis suppressor involved in cell adhesion, migration and apoptosis(303). Downregulation of CD82 expression is associated with the advanced stages of many human cancers and correlates with the acquisition of metastatic potential(303). In addition, the complex mechanisms leading to CD82 loss of function suggest a central role in controlling the metastatic spread of tumours(304, 305). Despite this, only few reports have examined CD82 expression in fibroblasts(306, 307), and even fewer have studied a molecular function in this context. The most comprehensive to date investigated CD82 expression in RA fibroblasts(306, 307). Here, CD82 was expressed on fibroblasts adjacent to blood vessels, and regulated cell adhesion and migration. In lung fibroblasts, CD82 is expressed in freshly isolated cells, but its function remains unknown(306). My results show CD82, as in cancer, acts to regulate cell cycle progression and maintains this process by promoting entrance into mitosis. Looking forward, more work is warranted to tease apart the molecular pathways underpinning the function of CD82. As in cancer, CD82 may regulate p53 activation in myofibroblasts.

VIII.IV. What is a single cell?

Multicellular organisms have developed a panoply of specialized features encoded by the genome to allow diverse functional programs to exist simultaneously. The human body beautifully compartmentalizes both structurally and functionally complementary features to optimize highly specialized homeostatic tasks. How this compartmentalization is constructed and maintained is at the heart of scientific enquiry and yet we have only recently developed tools to explore this question.

This thesis aims to build a complete single cell atlas of a human fibrotic disorder. The broad aim of this project was to deepen our understanding of the complex multicellular ecosystem that exist in the fibrotic microenvironment. Through this, we have a complete taxonomy of stromal and immune phenotypes in DD and I report novel cellular states and subtypes.

Despite the enormous opportunities afforded by single cell genomics it, like many scientific techniques, highlights our crude appreciation of established biological processes. I hope this thesis will be the first step towards a unified catalogue of fibrotic single cell atlases across different anatomical sites and diseases. Only then will we possess a robust understanding of the molecular landscape driving human fibrosis.

References

1. Wynn TA. Cellular and molecular mechanisms of fibrosis. The Journal of pathology. 2008;214(2):199-210.
2. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. New England Journal of Medicine. 2018;378(19):1811-23.
3. Gieseck Iii RL, Wilson MS, Wynn TA. Type 2 immunity in tissue repair and fibrosis. Nature Reviews Immunology. 2017;18:62.
4. Wynn TA, Ramalingam TR. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. Nature medicine. 2012;18(7):1028-40.
5. Wynn TA, Vannella KM. Macrophages in tissue repair, regeneration, and fibrosis. Immunity. 2016;44(3):450-62.
6. Richter J, Bode JG, Blondin D, Kircheis G, Kubitz R, Holtfreter MC, et al. Severe liver fibrosis caused by Schistosoma mansoni: management and treatment with a transjugular intrahepatic portosystemic shunt. The Lancet Infectious diseases. 2015;15(6):731-7.
7. Rogler G, Hausmann M. Factors Promoting Development of Fibrosis in Crohn's Disease. Frontiers in medicine. 2017;4:96-.
8. Ho YY, Lagares D, Tager AM, Kapoor M. Fibrosis—a lethal component of systemic sclerosis. Nature Reviews Rheumatology. 2014;10:390.
9. Rybinski B, Franco-Barraza J, Cukierman E. The wound healing, chronic fibrosis, and cancer progression triad. Physiological genomics. 2014;46(7):223-44.
10. Jaster R, Sparmann G, Emmrich J, Liebe S. Extracellular signal regulated kinases are key mediators of mitogenic signals in rat pancreatic stellate cells. Gut. 2002;51(4):579-84.
11. Thomas D, Radhakrishnan P. Tumor-stromal crosstalk in pancreatic cancer and tissue fibrosis. Molecular cancer. 2019;18(1):14-.
12. Cao Q, Harris DC, Wang Y. Macrophages in kidney injury, inflammation, and fibrosis. Physiology (Bethesda, Md). 2015;30(3):183-94.
13. Nikolic-Paterson DJ, Wang S, Lan HY. Macrophages promote renal fibrosis through direct and indirect mechanisms. Kidney International Supplements. 2014;4(1):34-8.
14. Braga TT, Agudelo JSH, Camara NOS. Macrophages During the Fibrotic Process: M2 as Friend and Foe. Frontiers in immunology. 2015;6:602.
15. Herrera J, Henke CA, Bitterman PB. Extracellular matrix as a driver of progressive fibrosis. The Journal of clinical investigation. 2018;128(1):45-53.
16. Haak AJ, Tan Q, Tschumperlin DJ. Matrix biomechanics and dynamics in pulmonary fibrosis. Matrix biology : journal of the International Society for Matrix Biology. 2018;73:64-76.
17. van Beuge MM, Ten Dam E-JPM, Werker PMN, Bank RA. Matrix and cell phenotype differences in Dupuytren's disease. Fibrogenesis & tissue repair. 2016;9:9-.
18. Wight TN, Potter-Perigo S. The extracellular matrix: an active or passive player in fibrosis? American Journal of Physiology-Gastrointestinal and Liver Physiology. 2011;301(6):G950-G5.
19. Jorgenson AJ, Choi KM, Sicard D, Smith KM, Hiemer SE, Varelas X, et al. TAZ activation drives fibroblast spheroid growth, expression of profibrotic paracrine signals, and context-

dependent ECM gene expression. American journal of physiology Cell physiology. 2017;312(3):C277-c85.
20. Piersma B, de Rond S, Werker PMN, Boo S, Hinz B, van Beuge MM, et al. YAP1 Is a Driver of Myofibroblast Differentiation in Normal and Diseased Fibroblasts. The American journal of pathology. 2015;185(12):3326-37.
21. Liu F, Lagares D, Choi KM, Stopfer L, Marinković A, Vrbanac V, et al. Mechanosignaling through YAP and TAZ drives fibroblast activation and fibrosis. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2014;308(4):L344-L57.
22. Liu T, De Los Santos FG, Phan SH. The Bleomycin Model of Pulmonary Fibrosis. Methods in molecular biology (Clifton, NJ). 2017;1627:27-42.
23. Peng R, Sridhar S, Tyagi G, Phillips JE, Garrido R, Harris P, et al. Bleomycin induces molecular changes directly relevant to idiopathic pulmonary fibrosis: a model for "active" disease. PloS one. 2013;8(4):e59348-e.
24. B Moore B, Lawson WE, Oury TD, Sisson TH, Raghavendran K, Hogaboam CM. Animal models of fibrotic lung disease. American journal of respiratory cell and molecular biology. 2013;49(2):167-79.
25. Borthwick LA, Wynn TA, Fisher AJ. Cytokine mediated tissue fibrosis. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2013;1832(7):1049-60.
26. Verhoekx JSN, Verjee LS, Izadi D, Chan JKK, Nicolaidou V, Davidson D, et al. Isometric Contraction of Dupuytren's Myofibroblasts Is Inhibited by Blocking Intercellular Junctions. Journal of Investigative Dermatology. 2013;133(12):2664-71.
27. Verjee LS, Verhoekx JS, Chan JK, Krausgruber T, Nicolaidou V, Izadi D, et al. Unraveling the signaling pathways promoting fibrosis in Dupuytren's disease reveals TNF as a therapeutic target. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(10):E928-37.
28. Wynn TA. Integrating mechanisms of pulmonary fibrosis. The Journal of experimental medicine. 2011;208(7):1339.
29. McHedlidze T, Waldner M, Zopf S, Walker J, Rankin AL, Schuchmann M, et al. Interleukin-33-dependent innate lymphoid cells mediate hepatic fibrosis. Immunity. 2013;39(2):357-71.
30. Baird KS, Alwan WH, Crossan JF, Wojciak B. T-cell-mediated response in Dupuytren's disease. Lancet (London, England). 1993;341(8861):1622-3.
31. Rankin AL, Mumm JB, Murphy E, Turner S, Yu N, McClanahan TK, et al. IL-33 induces IL-13-dependent cutaneous fibrosis. Journal of immunology (Baltimore, Md : 1950). 2010;184(3):1526-35.
32. Li D, Guabiraba R, Besnard AG, Komai-Koma M, Jabir MS, Zhang L, et al. IL-33 promotes ST2-dependent lung fibrosis by the induction of alternatively activated macrophages and innate lymphoid cells in mice. The Journal of allergy and clinical immunology. 2014;134(6):1422-32.e11.
33. Lodyga M, Cambridge E, Karvonen HM, Pakshir P, Wu B, Boo S, et al. Cadherin-11–mediated adhesion of macrophages to myofibroblasts establishes a profibrotic niche of active TGF-β. Science Signaling. 2019;12(564):eaao3469.

34. Gieseck RL, 3rd, Ramalingam TR, Hart KM, Vannella KM, Cantu DA, Lu WY, et al. Interleukin-13 Activates Distinct Cellular Pathways Leading to Ductular Reaction, Steatosis, and Fibrosis. Immunity. 2016;45(1):145-58.
35. Buckley CD, Barone F, Nayar S, Benezech C, Caamano J. Stromal cells in chronic inflammation and tertiary lymphoid organ formation. Annual review of immunology. 2015;33:715-45.
36. Kaviratne M, Hesse M, Leusink M, Cheever AW, Davies SJ, McKerrow JH, et al. IL-13 activates a mechanism of tissue fibrosis that is completely TGF-beta independent. Journal of immunology (Baltimore, Md : 1950). 2004;173(6):4020-9.
37. Manohar M, Kandikattu HK, Verma AK, Mishra A. IL-15 regulates fibrosis and inflammation in a mouse model of chronic pancreatitis. American Journal of Physiology-Gastrointestinal and Liver Physiology. 2018;315(6):G954-G65.
38. Weng HL, Liu Y, Chen JL, Huang T, Xu LJ, Godoy P, et al. The etiology of liver damage imparts cytokines transforming growth factor beta1 or interleukin-13 as driving forces in fibrogenesis. Hepatology (Baltimore, Md). 2009;50(1):230-43.
39. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. Nature reviews Immunology. 2015;15(11):669-82.
40. Nanchahal J, Ball C, Davidson D, Williams L, Sones W, McCann FE, et al. Anti-Tumour Necrosis Factor Therapy for Dupuytren's Disease: A Randomised Dose Response Proof of Concept Phase 2a Clinical Trial. EBioMedicine. 2018;33:282-8.
41. Deshmane SL, Kremlev S, Amini S, Sawaya BE. Monocyte chemoattractant protein-1 (MCP-1): an overview. Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research. 2009;29(6):313-26.
42. Duffield JS, Forbes SJ, Constandinou CM, Clay S, Partolina M, Vuthoori S, et al. Selective depletion of macrophages reveals distinct, opposing roles during liver injury and repair. Journal of Clinical Investigation. 2005;115(1):56-65.
43. Zhou D, Yang K, Chen L, Wang Y, Zhang W, Xu Z, et al. Macrophage polarization and function: new prospects for fibrotic disease. Immunology And Cell Biology. 2017;95:864.
44. Bai L, Liu X, Zheng Q, Kong M, Zhang X, Hu R, et al. M2-like macrophages in the fibrotic liver protect mice against lethal insults through conferring apoptosis resistance to hepatocytes. Scientific Reports. 2017;7(1):10518.
45. Shook BA, Wasko RR, Rivera-Gonzalez GC, Salazar-Gatzimas E, López-Giráldez F, Dash BC, et al. Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair. Science. 2018;362(6417):eaar2971.
46. Taylor DK, Mittereder N, Kuta E, Delaney T, Burwell T, Dacosta K, et al. T follicular helper–like cells contribute to skin fibrosis. Science translational medicine. 2018;10(431):eaaf5307.
47. Yamamoto T, Nakamura T, Noble NA, Ruoslahti E, Border WA. Expression of transforming growth factor beta is elevated in human and experimental diabetic nephropathy. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(5):1814-8.
48. Meng X-m, Nikolic-Paterson DJ, Lan HY. TGF-β: the master regulator of fibrosis. Nature Reviews Nephrology. 2016;12:325.

49. Taniguchi H, Ebina M, Kondoh Y, Ogura T, Azuma A, Suga M, et al. Pirfenidone in idiopathic pulmonary fibrosis. European Respiratory Journal. 2010;35(4):821-9.
50. Smith GP. Chapter 19 - Animal Models of Fibrosis in Human Disease. In: Conn PM, editor. Animal Models for the Study of Human Disease. Boston: Academic Press; 2013. p. 435-58.
51. Zeisberg M, Soubasakos MA, Kalluri R. Animal models of renal fibrosis. Methods in molecular medicine. 2005;117:261-72.
52. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. Nature Reviews Disease Primers. 2017;3:17074.
53. Bochaton-Piallat ML, Gabbiani G, Hinz B. The myofibroblast in wound healing and fibrosis: answered and unanswered questions. F1000Research. 2016;5.
54. Klingberg F, Hinz B, White ES. The myofibroblast matrix: implications for tissue repair and fibrosis. The Journal of pathology. 2013;229(2):298-309.
55. Verjee LS, Midwood K, Davidson D, Essex D, Sandison A, Nanchahal J. Myofibroblast distribution in Dupuytren's cords: correlation with digital contracture. The Journal of hand surgery. 2009;34(10):1785-94.
56. Gabbiani G. The myofibroblast: a key cell for wound healing and fibrocontractive diseases. Progress in clinical and biological research. 1981;54:183-94.
57. Hsia LT, Ashley N, Ouaret D, Wang LM, Wilding J, Bodmer WF. Myofibroblasts are distinguished from activated skin fibroblasts by the expression of AOC3 and other associated markers. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(15):E2162-71.
58. Rockey DC, Weymouth N, Shi Z. Smooth Muscle α Actin (Acta2) and Myofibroblast Function during Hepatic Wound Healing. PLOS ONE. 2013;8(10):e77166.
59. Shinde AV, Humeres C, Frangogiannis NG. The role of alpha-smooth muscle actin in fibroblast-mediated matrix contraction and remodeling. Biochimica et biophysica acta Molecular basis of disease. 2017;1863(1):298-309.
60. Gabbiani G. The myofibroblast in wound healing and fibrocontractive diseases. The Journal of pathology. 2003;200(4):500-3.
61. Meng XM, Nikolic-Paterson DJ, Lan HY. TGF-beta: the master regulator of fibrosis. Nature reviews Nephrology. 2016;12(6):325-38.
62. Walton KL, Johnson KE, Harrison CA. Targeting TGF-β Mediated SMAD Signaling for the Prevention of Fibrosis. Frontiers in Pharmacology. 2017;8(461).
63. Zehender A, Huang J, Györfi A-H, Matei A-E, Trinh-Minh T, Xu X, et al. The tyrosine phosphatase SHP2 controls TGFβ-induced STAT3 signaling to regulate fibroblast activation and fibrosis. Nature communications. 2018;9(1):3259-.
64. Liu F, Lagares D, Choi KM, Stopfer L, Marinkovic A, Vrbanac V, et al. Mechanosignaling through YAP and TAZ drives fibroblast activation and fibrosis. American journal of physiology Lung cellular and molecular physiology. 2015;308(4):L344-57.
65. Plotnikov SV, Pasapera AM, Sabass B, Waterman CM. Force fluctuations within focal adhesions mediate ECM-rigidity sensing to guide directed cell migration. Cell. 2012;151(7):1513-27.

66. Shinde AV, Humeres C, Frangogiannis NG. The role of α-smooth muscle actin in fibroblast-mediated matrix contraction and remodeling. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2017;1863(1):298-309.
67. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nature reviews Molecular cell biology. 2014;15(12):786-801.
68. Schwartz MA. Integrins and extracellular matrix in mechanotransduction. Cold Spring Harbor perspectives in biology. 2010;2(12):a005066-a.
69. Liu F, Haeger CM, Dieffenbach PB, Sicard D, Chrobak I, Coronata AM, et al. Distal vessel stiffening is an early and pivotal mechanobiological regulator of vascular remodeling and pulmonary hypertension. JCI insight. 2016;1(8).
70. Beningo KA, Hamao K, Dembo M, Wang Y-L, Hosoya H. Traction forces of fibroblasts are regulated by the Rho-dependent kinase but not by the myosin light chain kinase. Archives of biochemistry and biophysics. 2006;456(2):224-31.
71. Wozniak MA, Modzelewska K, Kwong L, Keely PJ. Focal adhesion regulation of cell behavior. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2004;1692(2):103-19.
72. Humphries JD, Wang P, Streuli C, Geiger B, Humphries MJ, Ballestrem C. Vinculin controls focal adhesion formation by direct interactions with talin and actin. The Journal of cell biology. 2007;179(5):1043-57.
73. Elosegui-Artola A, Oria R, Chen Y, Kosmalska A, Perez-Gonzalez C, Castro N, et al. Mechanical regulation of a molecular clutch defines force transmission and transduction in response to matrix rigidity. Nat Cell Biol. 2016;18(5):540-8.
74. Follonier L, Schaub S, Meister J-J, Hinz B. Myofibroblast communication is controlled by intercellular mechanical coupling. Journal of Cell Science. 2008;121(20):3305.
75. Harris AK, Wild P, Stopak D. Silicone rubber substrata: a new wrinkle in the study of cell locomotion. Science. 1980;208(4440):177-9.
76. Polacheck WJ, Chen CS. Measuring cell-generated forces: a guide to the available tools. Nature methods. 2016;13(5):415-23.
77. Dembo M, Oliver T, Ishihara A, Jacobson K. Imaging the traction stresses exerted by locomoting cells with the elastic substratum method. Biophysical Journal. 1996;70(4):2008-22.
78. Lee J, Leonard M, Oliver T, Ishihara A, Jacobson K. Traction forces generated by locomoting keratocytes. The Journal of cell biology. 1994;127(6 Pt 2):1957-64.
79. Colin-York H, Javanmardi Y, Skamrahl M, Kumari S, Chang VT, Khuon S, et al. Cytoskeletal Control of Antigen-Dependent T Cell Activation. Cell reports. 2019;26(12):3369-79.e5.
80. Wang J, Lin F, Wan Z, Sun X, Lu Y, Huang J, et al. Profiling the origin, dynamics, and function of traction force in B cell activation. Science Signaling. 2018;11(542):eaai9192.
81. Colin-York H, Shrestha D, Felce JH, Waithe D, Moeendarbary E, Davis SJ, et al. Super-Resolved Traction Force Microscopy (STFM). Nano Letters. 2016;16(4):2633-8.
82. Colin-York H, Eggeling C, Fritzsche M. Dissection of mechanical force in living cells by super-resolved traction force microscopy. Nature protocols. 2017;12:783.
83. Feng H, Ouyang H, Peng M, Jin Y, Zhang Y, Liu Z, et al. Assessment of extracellular matrix modulation of cell traction force by using silicon nanowire array. Nano Energy. 2018;50:504-12.
84. Han SJ, Oak Y, Groisman A, Danuser G. Traction microscopy to identify force modulation in subresolution adhesions. Nature methods. 2015;12(7):653-6.

85. Sabass B, Gardel ML, Waterman CM, Schwarz US. High resolution traction force microscopy based on experimental and computational advances. Biophysical journal. 2008;94(1):207-20.
86. Brossard C, Monnier JC, BARRICAU P, Vandernoot FX, Le Sant Y, Champagnat F, et al. Principles and applications of particle image velocimetry. AerospaceLab. 2009(1):p. 1-11.
87. McCarty S, Syed F, Bayat A. Role of the HLA System in the Pathogenesis of Dupuytren's Disease. Hand (New York, NY). 2010;5(3):241-50.
88. Butler JP, Tolic-Norrelykke IM, Fabry B, Fredberg JJ. Traction fields, moments, and strain energy that cells exert on their surroundings. American journal of physiology Cell physiology. 2002;282(3):C595-605.
89. Lanting R, Broekstra DC, Werker PMN, van den Heuvel ER. A Systematic Review and Meta-Analysis on the Prevalence of Dupuytren Disease in the General Population of Western Countries. Plastic and Reconstructive Surgery. 2014;133(3):593-603.
90. Reilly RM, Stern PJ, Goldfarb CA. A Retrospective Review of the Management of Dupuytren's Nodules. The Journal of hand surgery. 2005;30(5):1014-8.
91. Wilburn J, McKenna SP, Perry-Hinsley D, Bayat A. The impact of Dupuytren disease on patient activity and quality of life. The Journal of hand surgery. 2013;38(6):1209-14.
92. Townley WA, Baker R, Sheppard N, Grobbelaar AO. Dupuytren's contracture unfolded. BMJ (Clinical research ed). 2006;332(7538):397-400.
93. BSSH. Dupuytren's disease.British Society for Surgery of the Hand. wwwbsshacuk. 2012.
94. NICE. Interventional procedures overview of needle fasciotomy for Dupuytren's contracture. National Institute for Clinical Excellence 2003;www.nice.org.uk.
95. NICE. Needle fasciotomy for Dupuytren's contracture.Interventional procedure guidanc. National Institute for Clinical Excellence. 2004;www.nice.org.uk.
96. Larsen S, Krogsgaard DG, Aagaard Larsen L, Iachina M, Skytthe A, Frederiksen H. Genetic and environmental influences in Dupuytren's disease: a study of 30,330 Danish twin pairs. The Journal of hand surgery, European volume. 2015;40(2):171-6.
97. Dolmans GH, Werker PM, Hennies HC, Furniss D, Festen EA, Franke L, et al. Wnt Signaling and Dupuytren's Disease. New England Journal of Medicine. 2011;365(4):307-17.
98. Lam AP, Gottardi CJ. β-catenin signaling: a novel mediator of fibrosis and potential therapeutic target. Current opinion in rheumatology. 2011;23(6):562-7.
99. Riester SM, Arsoy D, Camilleri ET, Dudakovic A, Paradise CR, Evans JM, et al. RNA sequencing reveals a depletion of collagen targeting microRNAs in Dupuytren's disease. BMC medical genomics. 2015;8:59-.
100. Ten Dam E-JPM, van Beuge MM, Bank RA, Werker PMN. Further evidence of the involvement of the Wnt signaling pathway in Dupuytren's disease. Journal of cell communication and signaling. 2016;10(1):33-40.
101. Ng M, Thakkar D, Southam L, Werker P, Ophoff R, Becker K, et al. A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. American journal of human genetics. 2017;101(3):417-27.
102. Surendran K, Schiavi S, Hruska KA. Wnt-dependent beta-catenin signaling is activated after unilateral ureteral obstruction, and recombinant secreted frizzled-related protein 4 alters the progression of renal fibrosis. Journal of the American Society of Nephrology : JASN. 2005;16(8):2373-84.

103. Matsushima K, Suyama T, Takenaka C, Nishishita N, Ikeda K, Ikada Y, et al. Secreted frizzled related protein 4 reduces fibrosis scar size and ameliorates cardiac function after ischemic injury. Tissue engineering Part A. 2010;16(11):3329-41.
104. Qureshi FI, Hornigold R, Spencer JD, Hall SM. Langerhans cells in Dupuytren's contracture. Journal of hand surgery (Edinburgh, Scotland). 2001;26(4):362-7.
105. Mayerl C, Del Frari B, Parson W, Boeck G, Piza-Katzer H, Wick G, et al. Characterisation of the inflammatory response in Dupuytren's disease. Journal of plastic surgery and hand surgery. 2016;50(3):171-9.
106. Satish L, LaFramboise WA, O'Gorman DB, Johnson S, Janto B, Gan BS, et al. Identification of differentially expressed genes in fibroblasts derived from patients with Dupuytren's Contracture. BMC Med Genomics. 2008;1:10.
107. Ulrich D, Ulrich F, Piatkowski A, Pallua N. Expression of matrix metalloproteinases and their inhibitors in cords and nodules of patients with Dupuytren's disease. Arch Orthop Trauma Surg. 2009;129(11):1453-9.
108. Vi L, Njarlangattil A, Wu Y, Gan BS, O'Gorman DB. Type-1 Collagen differentially alters beta-catenin accumulation in primary Dupuytren's Disease cord and adjacent palmar fascia cells. BMC musculoskeletal disorders. 2009;10:72.
109. Olaso E, Ikeda K, Eng FJ, Xu L, Wang LH, Lin HC, et al. DDR2 receptor promotes MMP-2-mediated proliferation and invasion by hepatic stellate cells. The Journal of clinical investigation. 2001;108(9):1369-78.
110. Olaso E, Arteta B, Benedicto A, Crende O, Friedman SL. Loss of discoidin domain receptor 2 promotes hepatic fibrosis after chronic carbon tetrachloride through altered paracrine interactions between hepatic stellate cells and liver-associated macrophages. The American journal of pathology. 2011;179(6):2894-904.
111. Hutchinson JW, Tierney GM, Parsons SL, Davis TR. Dupuytren's disease and frozen shoulder induced by treatment with a matrix metalloproteinase inhibitor. The Journal of bone and joint surgery British volume. 1998;80(5):907-8.
112. Wilkinson JM, Davidson RK, Swingler TE, Jones ER, Corps AN, Johnston P, et al. MMP-14 and MMP-2 are key metalloproteases in Dupuytren's disease fibroblast-mediated contraction. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2012;1822(6):897-905.
113. Johnston P, Larson D, Clark IM, Chojnowski AJ. Metalloproteinase gene expression correlates with clinical outcome in Dupuytren's disease. The Journal of hand surgery. 2008;33(7):1160-7.
114. Moeendarbary E, Harris AR. Cell mechanics: principles, practices, and prospects. Wiley interdisciplinary reviews Systems biology and medicine. 2014;6(5):371-88.
115. Wells RG. Tissue mechanics and fibrosis. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2013;1832(7):884-90.
116. Dupont S, Morsut L, Aragona M, Enzo E, Giulitti S, Cordenonsi M, et al. Role of YAP/TAZ in mechanotransduction. Nature. 2011;474(7350):179-83.
117. Ball C, Izadi D, Verjee LS, Chan J, Nanchahal J. Systematic review of non-surgical treatments for early dupuytren's disease. BMC musculoskeletal disorders. 2016;17(1):345-.
118. Smeraglia F, Del Buono A, Maffulli N. Collagenase clostridium histolyticum in Dupuytren's contracture: a systematic review. British medical bulletin. 2016;118(1):149-58.

119. Hurst LC, Badalamente MA, Hentz VR, Hotchkiss RN, Kaplan FTD, Meals RA, et al. Injectable Collagenase Clostridium Histolyticum for Dupuytren's Contracture. New England Journal of Medicine. 2009;361(10):968-79.
120. Peimer CA, Skodny P, Mackowiak JI. Collagenase clostridium histolyticum for dupuytren contracture: patterns of use and effectiveness in clinical practice. The Journal of hand surgery. 2013;38(12):2370-6.
121. Peimer CA, Wilbrand S, Gerber RA, Chapman D, Szczypa PP. Safety and tolerability of collagenase Clostridium histolyticum and fasciectomy for Dupuytren's contracture. The Journal of hand surgery, European volume. 2015;40(2):141-9.
122. Stromberg J, Ibsen Sorensen A, Friden J. Percutaneous Needle Fasciotomy Versus Collagenase Treatment for Dupuytren Contracture: A Randomized Controlled Trial with a Two-Year Follow-up. The Journal of bone and joint surgery American volume. 2018;100(13):1079-86.
123. Werlinrud JC, Hansen KL, Larsen S, Lauritsen J. Five-year results after collagenase treatment of Dupuytren disease. J Hand Surg Eur Vol. 2018;43(8):841-7.
124. Steenbeek LM, Dreise MM, Werker PMN. Durability of Collagenase Treatment for Dupuytren Disease of the Thumb and First Web After at Least 2 Years' Follow-Up. The Journal of hand surgery. 2018.
125. Gilpin D, Coleman S, Hall S, Houston A, Karrasch J, Jones N. Injectable collagenase Clostridium histolyticum: a new nonsurgical treatment for Dupuytren's disease. The Journal of hand surgery. 2010;35(12):2027-38.e1.
126. Badalamente MA, Hurst LC. Efficacy and safety of injectable mixed collagenase subtypes in the treatment of Dupuytren's contracture. The Journal of hand surgery. 2007;32(6):767-74.
127. Atroshi I, Strandberg E, Lauritzson A, Ahlgren E, Waldén M. Costs for collagenase injections compared with fasciectomy in the treatment of Dupuytren's contracture: a retrospective cohort study. BMJ open. 2014;4(1):e004166.
128. Chen NC, Shauver MJ, Chung KC. Cost-effectiveness of open partial fasciectomy, needle aponeurotomy, and collagenase injection for dupuytren contracture. The Journal of hand surgery. 2011;36(11):1826-34.e32.
129. ISRCTN1825459. Dupuytren's interventions surgery vs collagenase. 2018.
130. Au-Yong ITH, Wildin CJ, Dias JJ, Page RE. A Review of Common Practice in Dupuytren Surgery. Techniques in Hand & Upper Extremity Surgery. 2005;9(4):178-87.
131. Armstrong JR, Hurren JS, Logan AM. Dermofasciectomy in the management of Dupuytren's disease. The Journal of bone and joint surgery British volume. 2000;82(1):90-4.
132. Kan HJ, Verrijp FW, Hovius SER, van Nieuwenhoven CA, Dupuytren Delphi G, Selles RW. Recurrence of Dupuytren's contracture: A consensus-based definition. PloS one. 2017;12(5):e0164849-e.
133. Brotherston TM, Balakrishnan C, Milner RH, Brown HG. Long term follow-up of dermofasciectomy for Dupuytren's contracture. British journal of plastic surgery. 1994;47(6):440-3.
134. Citron ND, Nunez V. Recurrence after surgery for dupuytren's disease: a randomized trial of two skin incisions. The Journal of Hand Surgery: British & European Volume. 2005;30(6):563-6.

135. van Rijssen AL, ter Linden H, Werker PMN. Five-Year Results of a Randomized Clinical Trial on Treatment in Dupuytren's Disease: Percutaneous Needle Fasciotomy versus Limited Fasciectomy. Plastic and Reconstructive Surgery. 2012;129(2):469-77.
136. Selles RW, Zhou C, Kan HJ, Wouters RM, van Nieuwenhoven CA, Hovius SER. Percutaneous Aponeurotomy and Lipofilling versus Limited Fasciectomy for Dupuytren's Contracture: 5-Year Results from a Randomized Clinical Trial. Plast Reconstr Surg. 2018;142(6):1523-31.
137. Rodrigues JN, Zhang W, Scammell BE, Chakrabarti I, Russell PG, Fullilove S, et al. Functional outcome and complications following surgery for Dupuytren's disease: a multi-centre cross-sectional study. Journal of Hand Surgery (European Volume). 2016;42(1):7-17.
138. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature. 2017;541(7637):331-8.
139. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature. 2013;498(7453):236-40.
140. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature biotechnology. 2012;30(8):777-82.
141. Xie T, Wang Y, Deng N, Huang G, Taghavifar F, Geng Y, et al. Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. Cell reports. 2018;22(13):3625-40.
142. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.
143. Mills JC, Roth KA, Cagan RL, Gordon JI. DNA microarrays and beyond: completing the journey from tissue to cell. Nature Cell Biology. 2001;3(8):E175-E8.
144. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 2009;10(1):57-63.
145. West NR, Hegazy AN, Owens BMJ, Bullers SJ, Linggi B, Buonocore S, et al. Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. Nature medicine. 2017;23(5):579-89.
146. Zeissig S, Rosati E, Dowds CM, Aden K, Bethge J, Schulte B, et al. Vedolizumab is associated with changes in innate rather than adaptive immunity in patients with inflammatory bowel disease. Gut. 2019;68(1):25-39.
147. Long X, Deng Z, Li G, Wang Z. Identification of critical genes to predict recurrence and death in colon cancer: integrating gene expression and bioinformatics analysis. Cancer Cell International. 2018;18(1):139.
148. Lumb R, Buckberry S, Secker G, Lawrence D, Schwarz Q. Transcriptome profiling reveals expression signatures of cranial neural crest cells arising from different axial levels. BMC Developmental Biology. 2017;17(1):5.
149. Macosko Evan Z, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.
150. Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, et al. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell. 2016;165(4):1012-26.

151. Dahlin JS, Hamey FK, Pijuan-Sala B, Shepherd M, Lau WWY, Nestorowa S, et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. Blood. 2018;131(21):e1-e11.
152. Plasschaert LW, Žilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature. 2018;560(7718):377-81.
153. Parikh K, Antanaviciute A, Fawkner-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature. 2019;567(7746):49-55.
154. Mizoguchi F, Slowikowski K, Wei K, Marshall JL, Rao DA, Chang SK, et al. Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. Nature Communications. 2018;9(1):789.
155. Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nature Cell Biology. 2018;20(7):836-46.
156. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature. 2019;566(7745):490-5.
157. Tiklová K, Björklund ÅK, Lahti L, Fiorenzano A, Nolbrant S, Gillberg L, et al. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. Nature Communications. 2019;10(1):581.
158. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods. 2009;6(5):377-82.
159. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature Communications. 2017;8:14049.
160. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nature protocols. 2014;9:171.
161. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell reports. 2012;2(3):666-73.
162. Saliba A-E, Westermann AJ, Gorski SA, Vogel J. Single-cell RNA-seq: advances and future challenges. Nucleic Acids Research. 2014;42(14):8845-60.
163. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. Nature protocols. 2016;12:44.
164. Wynn TA. Cellular and molecular mechanisms of fibrosis. The Journal of pathology. 2008;214(2):199-210.
165. Wick G, Grundtman C, Mayerl C, Wimpissinger TF, Feichtinger J, Zelger B, et al. The immunology of fibrosis. Annual review of immunology. 2013;31:107-35.
166. Ting David T, Wittner Ben S, Ligorio M, Vincent Jordan N, Shah Ajay M, Miyamoto David T, et al. Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells. Cell reports. 2014;8(6):1905-18.
167. Chen R, Wu X, Jiang L, Zhang Y. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. Cell reports. 2017;18(13):3227-41.
168. Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature. 2015;525:251.

169. Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, et al. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell. 2016;166(5):1308-23.e30.
170. Li H, Horns F, Wu B, Xie Q, Li J, Li T, et al. Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. Cell. 2017;171(5):1206-20.e22.
171. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. The Journal of experimental medicine. 2017;214(3):579.
172. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics (Oxford, England). 2012;28(6):882-3.
173. Chung NC, Storey JD. Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics (Oxford, England). 2015;31(4):545-54.
174. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189.
175. Angerer P, Haghverdi L, Buttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics. 2016;32(8):1241-3.
176. Hastie T, Stuetzle W. Principal Curves. Journal of the American Statistical Association. 1989;84(406):502-16.
177. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology. 2012;16(5):284-7.
178. Colin-York H, Eggeling C, Fritzsche M. Dissection of mechanical force in living cells by super-resolved traction force microscopy. Nature protocols. 2017;12(4):783-96.
179. Martiel J-L, Leal A, Kurzawa L, Balland M, Wang I, Vignaud T, et al. Chapter 15 - Measurement of cell traction forces with ImageJ. In: Paluch EK, editor. Methods in cell biology. 125: Academic Press; 2015. p. 269-87.
180. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology. 2016;34:525.
181. Eyden B. The myofibroblast: phenotypic characterization as a prerequisite to understanding its functions in translational medicine. J Cell Mol Med. 2008;12.
182. Tomasek JJ, Gabbiani G, Hinz B, Chaponnier C, Brown RA. Myofibroblasts and mechano-regulation of connective tissue remodelling. Nature Reviews Molecular Cell Biology. 2002;3:349.
183. DeLeve LD. Liver sinusoidal endothelial cells in hepatic fibrosis. Hepatology (Baltimore, Md). 2015;61(5):1740-6.
184. Kato S, Inui N, Hakamata A, Suzuki Y, Enomoto N, Fujisawa T, et al. Changes in pulmonary endothelial cell properties during bleomycin-induced pulmonary fibrosis. Respiratory research. 2018;19(1):127-.
185. Sava P, Ramanathan A, Dobronyi A, Peng X, Sun H, Ledesma-Mendoza A, et al. Human pericytes adopt myofibroblast properties in the microenvironment of the IPF lung. JCI insight. 2017;2(24):e96352.

186. Hung C, Linn G, Chow Y-H, Kobayashi A, Mittelsteadt K, Altemeier WA, et al. Role of lung pericytes and resident fibroblasts in the pathogenesis of pulmonary fibrosis. American journal of respiratory and critical care medicine. 2013;188(7):820-30.
187. AlJanahi AA, Danielsen M, Dunbar CE. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Molecular Therapy - Methods & Clinical Development. 2018;10:189-96.
188. Poggio T, Mhaskar H, Rosasco L, Miranda B, Liao Q. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. International Journal of Automation and Computing. 2017;14(5):503-19.
189. Khalid S, Khalil T, Nasreen S, editors. A survey of feature selection and feature extraction techniques in machine learning. 2014 Science and Information Conference; 2014 27-29 Aug. 2014.
190. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical transactions Series A, Mathematical, physical, and engineering sciences. 2016;374(2065):20150202.
191. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir E-aD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 2015;162(1):184-97.
192. Nowotschin S, Setty M, Kuo Y-Y, Liu V, Garg V, Sharma R, et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. Nature. 2019.
193. Xu C, Su Z. Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics. 2015;31(12):1974-80.
194. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008;2008(10):P10008.
195. Cho H, Kozasa T, Bondjers C, Betsholtz C, Kehrl JH. Pericyte-specific expression of Rgs5: implications for PDGF and EDG receptor signaling during vascular maturation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2003;17(3):440-2.
196. Mitchell TS, Bradley J, Robinson GS, Shima DT, Ng YS. RGS5 expression is a quantitative measure of pericyte coverage of blood vessels. Angiogenesis. 2008;11(2):141-51.
197. Chang F-C, Chou Y-H, Chen Y-T, Lin S-L. Novel insights into pericyte–myofibroblast transition and therapeutic targets in renal fibrosis. Journal of the Formosan Medical Association. 2012;111(11):589-98.
198. Yang J, Wang M, Zhu F, Sun J, Xu H, Chong Lee Shin OL, et al. Putative endothelial progenitor cells do not promote vascular repair but attenuate pericyte-myofibroblast transition in UUO-induced renal fibrosis. Stem cell research & therapy. 2019;10(1):104.
199. Wang YC, Chen Q, Luo JM, Nie J, Meng QH, Shuai W, et al. Notch1 promotes the pericyte-myofibroblast transition in idiopathic pulmonary fibrosis through the PDGFR/ROCK1 signal pathway. Experimental & molecular medicine. 2019;51(3):35.
200. Wang N, Deng Y, Liu A, Shen N, Wang W, Du X, et al. Novel Mechanism of the Pericyte-Myofibroblast Transition in Renal Interstitial Fibrosis: Core Fucosylation Regulation. Scientific reports. 2017;7(1):16914-.

201. Guimaraes-Camboa N, Cattaneo P, Sun Y, Moore-Morris T, Gu Y, Dalton ND, et al. Pericytes of Multiple Organs Do Not Behave as Mesenchymal Stem Cells In Vivo. Cell stem cell. 2017;20(3):345-59.e5.
202. Maurya SK, Periasamy M. Sarcolipin is a novel regulator of muscle metabolism and obesity. Pharmacological research. 2015;102:270-5.
203. Layton T, Nanchahal J. Recent advances in the understanding of Dupuytren's disease. F1000Research. 2019;8:F1000 Faculty Rev-231.
204. Ribatti D, Nico B, Crivellato E. The role of pericytes in angiogenesis. The International journal of developmental biology. 2011;55(3):261-8.
205. Kramann R, Humphreys BD. Kidney pericytes: roles in regeneration and fibrosis. Seminars in nephrology. 2014;34(4):374-83.
206. Greenhalgh SN, Conroy KP, Henderson NC. Healing scars: targeting pericytes to treat fibrosis. QJM: An International Journal of Medicine. 2014;108(1):3-7.
207. Chen Q, Zhang H, Liu Y, Adams S, Eilken H, Stehling M, et al. Endothelial cells are progenitors of cardiac pericytes and vascular smooth muscle cells. Nature Communications. 2016;7:12422.
208. Humphreys BD, Lin S-L, Kobayashi A, Hudson TE, Nowlin BT, Bonventre JV, et al. Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis. The American journal of pathology. 2010;176(1):85-97.
209. Lloyd CM, Minto AW, Dorf ME, Proudfoot A, Wells TN, Salant DJ, et al. RANTES and monocyte chemoattractant protein-1 (MCP-1) play an important role in the inflammatory phase of crescentic nephritis, but only MCP-1 is involved in crescent formation and interstitial fibrosis. The Journal of experimental medicine. 1997;185(7):1371-80.
210. Belperio JA, Keane MP, Burdick MD, Lynch JP, 3rd, Xue YY, Berlin A, et al. Critical role for the chemokine MCP-1/CCR2 in the pathogenesis of bronchiolitis obliterans syndrome. The Journal of clinical investigation. 2001;108(4):547-56.
211. Blease K, Mehrad B, Standiford TJ, Lukacs NW, Kunkel SL, Chensue SW, et al. Airway remodeling is absent in CCR1-/- mice during chronic fungal allergic airway disease. Journal of immunology (Baltimore, Md : 1950). 2000;165(3):1564-72.
212. Shen JZ, Morgan J, Tesch GH, Fuller PJ, Young MJ. CCL2-Dependent Macrophage Recruitment Is Critical for Mineralocorticoid Receptor-Mediated Cardiac Fibrosis, Inflammation, and Blood Pressure Responses in Male Mice. Endocrinology. 2014;155(3):1057-66.
213. Bendall Sean C, Davis Kara L, Amir E-ad D, Tadmor Michelle D, Simonds Erin F, Chen Tiffany J, et al. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. Cell. 2014;157(3):714-25.
214. Hinz B, Dugina V, Ballestrem C, Wehrle-Haller B, Chaponnier C. Alpha-smooth muscle actin is crucial for focal adhesion maturation in myofibroblasts. Molecular biology of the cell. 2003;14(6):2508-19.
215. Lagares D, Ghassemi-Kakroodi P, Tremblay C, Santos A, Probst CK, Franklin A, et al. ADAM10-mediated ephrin-B2 shedding promotes myofibroblast activation and organ fibrosis. Nature medicine. 2017;23:1405.
216. Kendall RT, Feghali-Bostwick CA. Fibroblasts in fibrosis: novel roles and mediators. Frontiers in pharmacology. 2014;5:123-.

217. Neuzillet C, Tijeras-Raballand A, Ragulan C, Cros J, Patil Y, Martinet M, et al. Inter- and intra-tumoural heterogeneity in cancer-associated fibroblasts of human pancreatic ductal adenocarcinoma. The Journal of pathology. 2019;248(1):51-65.
218. Giang J, Seelen MAJ, van Doorn MBA, Rissmann R, Prens EP, Damman J. Complement Activation in Inflammatory Skin Diseases. Frontiers in immunology. 2018;9(639).
219. Almitairi JOM, Venkatraman Girija U, Furze CM, Simpson-Gray X, Badakshi F, Marshall JE, et al. Structure of the C1r–C1s interaction of the C1 complex of complement activation. Proceedings of the National Academy of Sciences. 2018;115(4):768.
220. Schewe M, Franken PF, Sacchetti A, Schmitt M, Joosten R, Bottcher R, et al. Secreted Phospholipases A2 Are Intestinal Stem Cell Niche Factors with Distinct Roles in Homeostasis, Inflammation, and Cancer. Cell stem cell. 2016;19(1):38-51.
221. Lu J, Chatterjee M, Schmid H, Beck S, Gawaz M. CXCL14 as an emerging immune and inflammatory modulator. J Inflamm (Lond). 2016;13:1-.
222. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.
223. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(21):7426.
224. Verhoekx JSN, Beckett KS, Bisson MA, McGrouther DA, Grobbelaar AO, Mudera V. The mechanical environment in dupuytren's contracture determines cell contractility and associated MMP-mediated matrix remodeling. Journal of Orthopaedic Research. 2013;31(2):328-34.
225. Munguia-Reyes A, Balderas-Martinez YI, Becerril C, Checa M, Ramirez R, Ortiz B, et al. R-Spondin-2 Is Upregulated in Idiopathic Pulmonary Fibrosis and Affects Fibroblast Behavior. Am J Respir Cell Mol Biol. 2018;59(1):65-76.
226. Tripoli M, Cordova A, Moschella F. Update on the role of molecular factors and fibroblasts in the pathogenesis of Dupuytren's disease. Journal of cell communication and signaling. 2016;10(4):315-30.
227. Dvorakova M, Nenutil R, Bouchal P. Transgelins, cytoskeletal proteins implicated in different aspects of cancer development. Expert review of proteomics. 2014;11(2):149-65.
228. Kinchen J, Chen HH, Parikh K, Antanaviciute A, Jagielowicz M, Fawkner-Corbett D, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. Cell. 2018;175(2):372-86.e17.
229. Vi L, Feng L, Zhu RD, Wu Y, Satish L, Gan BS, et al. Periostin differentially induces proliferation, contraction and apoptosis of primary Dupuytren's disease and adjacent palmar fascia cells. Experimental cell research. 2009;315(20):3574-86.
230. Song J, Zhang H, Wang Z, Xu W, Zhong L, Cao J, et al. The Role of FABP5 in Radiation-Induced Human Skin Fibrosis. Radiation research. 2018;189(2):177-86.
231. Lynch MD, Watt FM. Fibroblast heterogeneity: implications for human disease. The Journal of clinical investigation. 2018;128(1):26-35.

232. Sorrell JM, Caplan AI. Fibroblast heterogeneity: more than skin deep. Journal of Cell Science. 2004;117(5):667.
233. Marsh E, Gonzalez DG, Lathrop EA, Boucher J, Greco V. Positional Stability and Membrane Occupancy Define Skin Fibroblast Homeostasis In Vivo. Cell. 2018;175(6):1620-33.e13.
234. Sempowski GD, Borrello MA, Blieden TM, Barth RK, Phipps RP. Fibroblast heterogeneity in the healing wound. Wound Repair and Regeneration. 1995;3(2):120-31.
235. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer cell. 2018;33(3):463-79.e10.
236. Labernadie A, Kato T, Brugues A, Serra-Picamal X, Derzsi S, Arwert E, et al. A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion. Nat Cell Biol. 2017;19(3):224-37.
237. Ng M, Thakkar D, Southam L, Werker P, Ophoff R, Becker K, et al. A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. American journal of human genetics. 2017;101(3):417-27.
238. Lee LC, Zhang AY, Chong AK, Pham H, Longaker MT, Chang J. Expression of a novel gene, MafB, in Dupuytren's disease. The Journal of hand surgery. 2006;31(2):211-8.
239. Loperfido M, Pellin D, Baricordi C, Wolock SL, Klein AM, Biasco L. The Single-Cell Transcriptional Landscape of Human Hematopoietic Progenitors. Blood. 2018;132(Suppl 1):1266.
240. Lescroart F, Wang X, Lin X, Swedlund B, Gargouri S, Sànchez-Dànes A, et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science. 2018;359(6380):1177.
241. Mack M, Yanagita M. Origin of myofibroblasts and cellular events triggering fibrosis. Kidney International. 2015;87(2):297-307.
242. Brenner DA, Kisseleva T, Scholten D, Paik YH, Iwaisako K, Inokuchi S, et al. Origin of myofibroblasts in liver fibrosis. Fibrogenesis & Tissue Repair. 2012;5(1):S17.
243. Mora AL, Rojas M, Pardo A, Selman M. Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. Nature Reviews Drug Discovery. 2017;16:755.
244. Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics. 2015;32(8):1241-3.
245. Guerrero-Juarez CF, Dedhia PH, Jin S, Ruiz-Vega R, Ma D, Liu Y, et al. Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. Nature Communications. 2019;10(1):650.
246. Rock JR, Barkauskas CE, Cronce MJ, Xue Y, Harris JR, Liang J, et al. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. Proceedings of the National Academy of Sciences. 2011;108(52):E1475.
247. Todd NW, Luzina IG, Atamas SP. Molecular and cellular mechanisms of pulmonary fibrosis. Fibrogenesis & Tissue Repair. 2012;5(1):11.
248. Zhang K, Rekhter MD, Gordon D, Phan SH. Myofibroblasts and their role in lung collagen gene expression during pulmonary fibrosis. A combined immunohistochemical and in situ hybridization study. The American journal of pathology. 1994;145(1):114-25.

249. Lagares D, Santos A, Grasberger PE, Liu F, Probst CK, Rahimi RA, et al. Targeted apoptosis of myofibroblasts with the BH3 mimetic ABT-263 reverses established fibrosis. Science translational medicine. 2017;9(420):eaal3765.
250. Uhal BD. The role of apoptosis in pulmonary fibrosis. European Respiratory Review. 2008;17(109):138.
251. Kissin E, Korn JH. Apoptosis and myofibroblasts in the pathogenesis of systemic sclerosis. Current Rheumatology Reports. 2002;4(2):129-35.
252. Thannickal VJ. Mechanisms of pulmonary fibrosis: role of activated myofibroblasts and NADPH oxidase. Fibrogenesis & Tissue Repair. 2012;5(1):S23.
253. Forsman M, Kallioinen L, Kallioinen M, Ryhanen J. Dupuytren's contracture; increased cellularity--proliferation, is there equality? Scandinavian journal of surgery : SJS : official organ for the Finnish Surgical Society and the Scandinavian Surgical Society. 2005;94(1):71-5.
254. Hung CF, Wilson CL, Chow YH, Schnapp LM. Role of integrin alpha8 in murine model of lung fibrosis. PLoS One. 2018;13(5):e0197937.
255. Plotnikov Sergey V, Pasapera Ana M, Sabass B, Waterman Clare M. Force Fluctuations within Focal Adhesions Mediate ECM-Rigidity Sensing to Guide Directed Cell Migration. Cell. 2012;151(7):1513-27.
256. Gardel ML, Sabass B, Ji L, Danuser G, Schwarz US, Waterman CM. Traction stress in focal adhesions correlates biphasically with actin retrograde flow speed. The Journal of cell biology. 2008;183(6):999-1005.
257. Achterberg VF, Buscemi L, Diekmann H, Smith-Clerc J, Schwengler H, Meister J-J, et al. The Nano-Scale Mechanical Properties of the Extracellular Matrix Regulate Dermal Fibroblast Function. Journal of Investigative Dermatology. 2014;134(7):1862-72.
258. Uroz M, Wistorf S, Serra-Picamal X, Conte V, Sales-Pardo M, Roca-Cusachs P, et al. Regulation of cell cycle progression by cell-cell and cell-matrix forces. Nat Cell Biol. 2018;20(6):646-54.
259. Marinkovic A, Mih JD, Park JA, Liu F, Tschumperlin DJ. Improved throughput traction microscopy reveals pivotal role for matrix stiffness in fibroblast contractility and TGF-beta responsiveness. American journal of physiology Lung cellular and molecular physiology. 2012;303(3):L169-80.
260. Park CY, Zhou EH, Tambe D, Chen B, Lavoie T, Dowell M, et al. High-throughput screening for modulators of cellular contractile force. Integrative biology : quantitative biosciences from nano to macro. 2015;7(10):1318-24.
261. Tomasek JJ, Gabbiani G, Hinz B, Chaponnier C, Brown RA. Myofibroblasts and mechano-regulation of connective tissue remodelling. Nature reviews Molecular cell biology. 2002;3(5):349-63.
262. Rayan GM, Tomasek JJ. Generation of contractile force by cultured Dupuytren's disease and normal palmar fibroblasts. Tissue & cell. 1994;26(5):747-56.
263. Betz T, Koch D, Lu Y-B, Franze K, Käs JA. Growth cones as soft and weak force generators. Proceedings of the National Academy of Sciences. 2011;108(33):13420.
264. Janmey PA, Miller RT. Mechanisms of mechanical signaling in development and disease. J Cell Sci. 2011;124(Pt 1):9-18.
265. Munevar S, Wang Y-l, Dembo M. Traction Force Microscopy of Migrating Normal and H-ras Transformed 3T3 Fibroblasts. Biophysical Journal. 2001;80(4):1744-57.

266. Bancelin S, Aimé C, Gusachenko I, Kowalczuk L, Latour G, Coradin T, et al. Determination of collagen fibril size via absolute measurements of second-harmonic generation signals. Nature Communications. 2014;5:4920.
267. Califano JP, Reinhart-King CA. Substrate Stiffness and Cell Area Predict Cellular Traction Stresses in Single Cells and Cells in Contact. Cellular and molecular bioengineering. 2010;3(1):68-75.
268. Rape AD, Guo W-H, Wang Y-L. The regulation of traction force in relation to cell shape and focal adhesions. Biomaterials. 2011;32(8):2043-51.
269. Vignaud T, Ennomani H, Thery M. Polyacrylamide hydrogel micropatterning. Methods in cell biology. 2014;120:93-116.
270. Kurzawa L, Vianay B, Senger F, Vignaud T, Blanchoin L, Théry M. Dissipation of contractile forces: the missing piece in cell mechanics. Molecular Biology of the Cell. 2017;28(14):1825-32.
271. Iwaisako K, Jiang C, Zhang M, Cong M, Moore-Morris TJ, Park TJ, et al. Origin of myofibroblasts in the fibrotic liver in mice. Proceedings of the National Academy of Sciences. 2014;111(32):E3297.
272. Boor P, Ostendorf T, Floege J. Renal fibrosis: novel insights into mechanisms and therapeutic targets. Nature reviews Nephrology. 2010;6(11):643-56.
273. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, et al. Molecular Architecture of the Mouse Nervous System. Cell. 2018;174(4):999-1014.e22.
274. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. eLife. 2017;6:e27041.
275. Nanchahal J, Hinz B. Strategies to overcome the hurdles to treat fibrosis, a major unmet clinical need. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(27):7291-3.
276. Habiel DM, Espindola MS, Jones IC, Coelho AL, Stripp B, Hogaboam CM. CCR10+ epithelial cells from idiopathic pulmonary fibrosis lungs drive remodeling. JCI insight. 2018;3(16).
277. Leaf IA, Nakagawa S, Johnson BG, Cha JJ, Mittelsteadt K, Guckian KM, et al. Pericyte MyD88 and IRAK4 control inflammatory and fibrotic responses to tissue injury. The Journal of clinical investigation. 2017;127(1):321-34.
278. Smyth LCD, Rustenhoven J, Scotter EL, Schweder P, Faull RLM, Park TIH, et al. Markers for human brain pericytes and smooth muscle cells. Journal of chemical neuroanatomy. 2018;92:48-60.
279. Kumar A, D'Souza SS, Moskvin OV, Toh H, Wang B, Zhang J, et al. Specification and Diversification of Pericytes and Smooth Muscle Cells from Mesenchymoangioblasts. Cell reports. 2017;19(9):1902-16.
280. Chen J, Luo Y, Huang H, Wu S, Feng J, Zhang J, et al. CD146 is essential for PDGFRβ-induced pericyte recruitment. Protein & Cell. 2018;9(8):743-7.
281. Chen J, Luo Y, Hui H, Cai T, Huang H, Yang F, et al. CD146 coordinates brain endothelial cell-pericyte communication for blood-brain barrier development. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(36):E7622-e31.
282. McAnulty RJ. Fibroblasts and myofibroblasts: their source, function and role in disease. Int J Biochem Cell Biol. 2007;39(4):666-71.

283. Burger JA, Kipps TJ. Chemokine receptors and stromal cells in the homing and homeostasis of chronic lymphocytic leukemia B cells. Leukemia & lymphoma. 2002;43(3):461-6.
284. Filer A, Raza K, Salmon M, Buckley CD. The role of chemokines in leucocyte-stromal interactions in rheumatoid arthritis. Frontiers in bioscience : a journal and virtual library. 2008;13:2674-85.
285. Döring Y, Pawig L, Weber C, Noels H. The CXCL12/CXCR4 chemokine ligand/receptor axis in cardiovascular disease. Frontiers in Physiology. 2014;5(212).
286. Sarma JV, Ward PA. The complement system. Cell Tissue Res. 2011;343(1):227-35.
287. Arend WP, Mehta G, Antonioli AH, Takahashi M, Takahashi K, Stahl GL, et al. Roles of adipocytes and fibroblasts in activation of the alternative pathway of complement in inflammatory arthritis in mice. Journal of immunology (Baltimore, Md : 1950). 2013;190(12):6423-33.
288. Tabib T, Morse C, Wang T, Chen W, Lafyatis R. SFRP2/DPP4 and FMO1/LSP1 Define Major Fibroblast Populations in Human Skin. The Journal of investigative dermatology. 2018;138(4):802-10.
289. Joost S, Jacob T, Sun X, Annusver K, La Manno G, Sur I, et al. Single-Cell Transcriptomics of Traced Epidermal and Hair Follicle Stem Cells Reveals Rapid Adaptations during Wound Healing. Cell reports. 2018;25(3):585-97.e7.
290. Philippeos C, Telerman SB, Oules B, Pisco AO, Shaw TJ, Elgueta R, et al. Spatial and Single-Cell Transcriptional Profiling Identifies Functionally Distinct Human Dermal Fibroblast Subpopulations. The Journal of investigative dermatology. 2018;138(4):811-25.
291. Garcia de Alba C, Buendia-Roldan I, Salgado A, Becerril C, Ramirez R, Gonzalez Y, et al. Fibrocytes contribute to inflammation and fibrosis in chronic hypersensitivity pneumonitis through paracrine effects. American journal of respiratory and critical care medicine. 2015;191(4):427-36.
292. Kovacic JC, Dimmeler S, Harvey RP, Finkel T, Aikawa E, Krenning G, et al. Endothelial to Mesenchymal Transition in Cardiovascular Disease. JACC State-of-the-Art Review. 2019;73(2):190-209.
293. Li Y, Lui KO, Zhou B. Reassessing endothelial-to-mesenchymal transition in cardiovascular diseases. Nature Reviews Cardiology. 2018;15(8):445-56.
294. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. The Journal of clinical investigation. 2003;112(12):1776-84.
295. Stone RC, Pastar I, Ojeh N, Chen V, Liu S, Garzon KI, et al. Epithelial-mesenchymal transition in tissue repair and fibrosis. Cell Tissue Res. 2016;365(3):495-506.
296. Heukels P, van Hulst JAC, van Nimwegen M, Boorsma CE, Melgert BN, van den Toorn LM, et al. Fibrocytes are increased in lung and peripheral blood of patients with idiopathic pulmonary fibrosis. Respiratory research. 2018;19(1):90.
297. Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, et al. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat Commun. 2017;8(1):2128.
298. Elhai M, Avouac J, Hoffmann-Vold AM, Ruzehaji N, Amiar O, Ruiz B, et al. OX40L blockade protects against inflammation-driven fibrosis. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(27):E3901-10.

299. Schafer S, Viswanathan S, Widjaja AA, Lim W-W, Moreno-Moral A, DeLaughter DM, et al. IL-11 is a crucial determinant of cardiovascular fibrosis. Nature. 2017;552:110.
300. White A, Lamb PW, Barrett JC. Frequent downregulation of the KAI1(CD82) metastasis suppressor protein in human cancer cell lines. Oncogene. 1998;16(24):3143-9.
301. Iwata S, Kobayashi H, Miyake-Nishijima R, Sasaki T, Souta-Kuribara A, Nori M, et al. Distinctive signaling pathways through CD82 and beta1 integrins in human T cells. European journal of immunology. 2002;32(5):1328-37.
302. Jones EL, Wee JL, Demaria MC, Blakeley J, Ho PK, Vega-Ramos J, et al. Dendritic Cell Migration and Antigen Presentation Are Coordinated by the Opposing Functions of the Tetraspanins CD82 and CD37. The Journal of Immunology. 2016;196(3):978.
303. Liu WM, Zhang XA. KAI1/CD82, a tumor metastasis suppressor. Cancer letters. 2006;240(2):183-94.
304. Tonoli H, Barrett JC. CD82 metastasis suppressor gene: a potential target for new therapeutics? Trends in molecular medicine. 2005;11(12):563-70.
305. Wu J, Liang S, Bergholz J, He H, Walsh EM, Zhang Y, et al. ΔNp63α activates CD82 metastasis suppressor to inhibit cancer cell invasion. Cell Death &Amp; Disease. 2014;5:e1280.
306. Heinzelmann K, Lehmann M, Gerckens M, Noskovicova N, Frankenberger M, Lindner M, et al. Cell-surface phenotyping identifies CD36 and CD97 as novel markers of fibroblast quiescence in lung fibrosis. American journal of physiology Lung cellular and molecular physiology. 2018;315(5):L682-l96.
307. Neumann E, Schwarz MC, Hasseli R, Hülser M-L, Classen S, Sauerbier M, et al. Tetraspanin CD82 affects migration, attachment and invasion of rheumatoid arthritis synovial fibroblasts. Annals of the Rheumatic Diseases. 2018;77(11):1619.