The role of the synovial microenvironment in Rheumatoid Arthritis

Jean-Baptiste Michel Marie Richard



University of Oxford
New College

Supervisors
Professor Kim Midwood
Professor Christopher Buckley

Thesis submitted for the degree of
Doctor of Philosophy

October 8th 2024

Table of Contents

Acknowledgments ***5***

Statement of originality ***6***

Abbreviations ***7***

Abstract ***12***

Chapter 1 | Introduction ***14***

1 Introduction ***14***

1.1. Tissue biology **15**

1.1.1. Immune-Mediated Inflammatory Diseases: a clinical unmet need 15

1.1.2. When form informs function 16

1.1.3. Principles of tissue biology: the cellular and the extracellular 16

1.1.4. Spatial omics 18

1.1.5. The dawn of tissue biology: networks, neighbourhoods, and niches 20

1.2. Rheumatoid Arthritis **23**

1.2.1. Principles of synovial joint tissue architecture 23

1.2.2. Synovial tissue cellular and molecular composition in homeostasis 23

1.2.3. Clinical presentation of Rheumatoid Arthritis 24

1.2.4. Epidemiology of Rheumatoid Arthritis 24

1.2.5. The molecular and cellular basis of Rheumatoid Arthritis 25

1.2.6. Past, current, and future therapeutic strategies for RA 30

1.3. Type VI collagen **35**

1.3.1. Principles of collagen biology 35

1.3.2. Overview of type VI collagen 39

1.3.3. Roles for collagen VI in human pathology 42

1.4. Project aims and summary **46**

Chapter 2 | Materials & Methods ***48***

2.1 Bioinformatics **49**

2.1.1 Datasets 49

2.1.2 AUCell Gene Set Enrichment Analysis 49

2.1.3 Pseudotime analysis 50

2.1.4 gsfisher Gene Set Enrichment Analysis 50

2.1.5 Alluvial label transfer 50

2.1.6 *De novo* fibroblast clustering 50

2.1.7 Matrisome-focused clustering 51

2.1.8 Differential expression analysis 51
2.1.9 Target prioritisation 52
2.1.10 Proteomic analysis 53
2.1.11 Cell subtype distribution in collagen VI rich and poor regions 53
2.2 Laboratory reagents and solutions 55
2.3 Immunofluorescence 59
2.3.1 CellDIVE multiplexed immunofluorescence – human tissue 59
2.3.2 Confocal immunofluorescence – mouse tissue 65
2.4 Protein methods 66
2.4.1 Protein purification 66
2.4.2 Endotoxin testing 67
2.4.3 Polyacrylamide Gel Electrophoresis (PAGE) 67
2.4.4 Native PAGE 67
2.4.5 Silver staining 68
2.5 In vitro cell-based methods 68
2.5.1 Human Peripheral Blood Mononuclear Cell (PBMC) and Monocyte isolation 68
2.5.2 Human macrophage differentiation and polarisation 68
2.5.3 Adhesion assay 70
2.5.4 Transwell migration assay 71
2.5.5 Ibidi channel migration assay 72
2.6 Experimental statistics 72
Chapter 3 | Bioinformatic analysis of matrisome gene expression in RA 73
3.1 Chapter introduction 74
3.2 Results 76
3.2.1 Matrisomal gene set enrichment in synovial cell subsets 76
3.2.2 Fibroblast and macrophage cluster-defining matrix markers 78
3.2.3 Pseudotime analysis of matrisome genes in RA fibroblasts 82
3.2.4 AMP1-like and matrisome-focused clustering 84
3.2.5 Disease type- and disease activity-specific matrix expression profiles 90
..... 92
3.2.6 Cross-disease matrisomal remission signature 94
3.2.7 Target prioritization of key matrisomal genes and proteins 95
3.2.8 IGF1 signalling gene expression profiles 98
3.2.9 Basement membrane gene expression profiles 99
3.3 Technical discussion 101
3.3.1 The matrisome gene list 101

3.3.2 Transcriptomics as an atlas corner stone 102
3.3.3 Dataset selection 102
3.3.4 scRNAseq analysis pipeline selection 103
3.3.5 Differential expression analysis 104

3.4 Biological discussion 105

3.4.1 Fibroblasts and macrophages share the burden of defining tissue composition 105
3.4.2 Synovial ECM is spatially heterogenous, with potential functional implications 105
3.4.3 Cross-tissue disease remission matrix signature 107
3.4.4 Target prioritisation and selection: unbiased ranking, biological interest, and pragmatism 108
3.4.5 Bioinformatic-guided hypotheses: basement-like membrane turnover and maladaptive responses to stressors 109

Chapter 4 | Spatial distribution of key matrisomal proteins in OA and RA 112

4.1 Chapter Introduction 113

4.2 Results 115

4.2.1 Multiplexed immunofluorescence staining of key matrisome proteins 115
4.2.2 Synovial cell type classification and enumeration in OA and RA 118
4.2.3 Subsynovial niche characterisation and cellular composition 121
4.2.4 Histological feature enumeration 123
4.2.5 Key matrisome protein abundances in subsynovial niches 125
4.2.6 Key protein distances to immune cell types 127
4.2.7 IGFBP5 localises to the lining layer macrophages and fibroblasts 129
4.2.8 IGFBP7 localises to the luminal side of synovial vasculature 132
4.2.9 T-cell aggregates localise to collagen VI-rich basement membrane vasculature 135
4.2.10 Perivascular collagen VI levels correlate with MerTK^{+} macrophage abundance 138
4.2.11 Lymphoid aggregates are restricted to collagen VI *dark zones* 141
4.2.12 Collagen VI deposition is associated with inflammation resolution in mouse AIA 143

4.3 Technical discussion 145

4.3.1 Rationale and limitations of immunofluorescence sample selection 145
4.3.2 Protein spatial distribution characterisation: technique and analysis selection 146

4.4 Biological discussion 147

4.4.1 COL4A1 staining is absent from the lining layer 147
4.4.2 The roles of vascular IGFBP7 148
4.4.3 IGFBP5 is a lining layer-deposited pro-fibrotic agent 149
4.4.4 COL6A1 defines two vasculature subtypes with differential T-cell associations 150
4.4.5 Perivascular COL6A1 correlates with remission-associated macrophages 151

Chapter 5 | Functional roles for type VI collagen in RA disease pathology 153

5.1 Chapter Introduction 154

5.2 Results 156

5.2.1 Collagen VI alone is unable to induce MerTK upregulation in macrophages 156

5.2.2 Collagen VI dark zones harbour most tissue immune cell subsets 159

5.2.3 Collagen VI dark zones are found cross-disease 161

5.2.4 Collagen I does not form dark zones 161

5.2.5 Collagen VI dark zones are associated with specific immune subsets 163

5.2.6 Collagen VI dark zones are outlined by a degradative front 167

5.2.7 Monocytes line dark zone degradative fronts and colocalise with CHP in tissue 168

5.2.8 Proteomics confirms the observed collagen VI-specific deposition and degradation patterns 170

5.2.9 Collagen VI is not adhesion permissive to THP1, Jurkat, and Daudi cell lines 175

5.2.10 Collagen VI acts as a molecular mimic obstructing cell adhesion 177

5.2.11 Collagen VI inhibits cell migration in preliminary in vitro experiments 180

5.2.12 Collagen VI degradation products are bioactive 183

5.3 Technical discussion 185

5.3.1 Immunofluorescent imaging sample selection and analysis 185

5.3.2 In vitro functional assay refinements and limitations 186

5.4 Biological discussion 187

5.4.1 MerTK regulation: new insights into an attractive yet elusive target 187

5.4.2 Collagen VI acts as a cellular sieve filtering immune ingress 188

5.4.3 Monocyte ingress regulates the cellular sieve and modulates immune cell establishment in tissue 190

Chapter 6 | Final discussion 193

6.1 Conclusions 194

6.1.1 A complex transcriptomic landscape with hints at dysregulation 194

6.1.2 Key ECM proteins define subsynovial niches with functional implications 195

6.1.3 Collagen VI is a remission-deposited regulator of immune ingress which is remodelled by infiltrating monocytes 196

6.2 Clinical relevance 197

6.3 Future work 198

Appendix Figures 201

Bibliography 217

Acknowledgments

First, I would like to thank my supervisory team. Kim has been an exceptional supervisor. Her expertise in matrix biology and in navigating the world of academic research have been incredibly valuable, and over the course of a sometimes personally difficult four years, she always found time to listen and support me. Chris' overarching guidance, his insights, and his consistently positive attitude were a true support throughout my PhD. Thank you also to my thesis committee, particularly Stephen Sansom and Matthias Friedrich for stimulating discussions and advice. I am also grateful to Mathilde Pohin who, while not being one of my official supervisors, grilled me when I needed it and fanned the flames of my curiosity and scientific enthusiasm. Thank you to RACE for funding this project and allowing me to make a small contribution to efforts to improve patient lives, and to the patient research partners involved in the programme.

I am grateful to past and present Midwood lab members. Thank you in particular to my colleagues Libby, Linda, Žofia, Luca, Caroline and Florrie, for putting up with me for four years and indulging my wild scientific (and less scientific) speculations over lunch or a cup of coffee. Thank you to the wider Kennedy Institute members for their generosity in laughs, reagents, and advice. Thank you to the friends I made during my five years in Oxford for making this city my first real home away from home. Thank you to my partner Eve for her unconditional support.

Finally, Papa, Maman, merci d'avoir fait de moi la personne que je suis, j'espère vous avoir rendu fiers et de continuer à le faire. Merci à vous ainsi qu'à mes frères – Paul-Auguste, Gaspard, et Basile - pour votre amour et soutien infaillible. Merci enfin au reste de ma famille, les Ceccherelli, Perdoux, et Richard.

They say it takes a village to raise a child, thank you all for making up mine.

Statement of originality

I declare that all data, words, figures, and results included herein are my own work and have not been submitted for another award, degree, or diploma at any other university or learning institution. I certify that this thesis is the result of my own work and all collaborations and use of published public data have been acknowledged at the appropriate locations within the text.

This project was completed at the Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford.

Abbreviations

ACK	Ammonium-Chloride-Potassium
ACPA	Anti-Citrullinated Protein Antibodies
ACR	American College of Rheumatology
AIA	Antigen-Induced Arthritis
AID	Activation Induced Deaminase
AMP	Accelerating Medicines Partnership
ANOVA	Analysis of Variance
APRIL	A-proliferation inducing ligand
AR	Antigen Retrieval
ARS	Antigen Retrieval Solution
ATCC	American Type Culture Collection
AUC	Area Under the Curve
BAFF	B-cell Activating Factor
BGN	Biglycan
BM	Bethlem Myopathy
BMDM	Bone Marrow Derived Macrophages
BMI	Body Mass Index
BMP	Bone Morphogenic Protein
BSA	Bovine Serum Albumin
BXT	Buffer XT
CBD	Collagen Binding Drugs or Domains
CCH	Collagenase Clostridium Histolyticum
CCLE	Cancer Cell Line Encyclopaedia
CD	Cluster of Differentiation
CHP	Collagen Hybridizing Peptide
CIA	Collagen Induced Arthritis
CILP	Cartilage Intermediate Layer Protein
CNS	Central Nervous System
CODEX	Co-Detection by Indexing
COL	Collagen
COVID	Coronavirus Disease
CRD	Collagen VI-Related Dystrophies
CREBBP	CREB-Binding Protein
CRLF	Cytokine Receptor-Like Factor
CSF	Colony Stimulating Factor
CTGF	Connective Tissue Growth Factor
CTLA-4	Cytotoxic T-Lymphocyte Associated protein 4
CTSK	Cathepsin K
CV	Column Volumes
DABCO	1,4-diazabicyclo[2.2.2]octane
DAPI	4′,6-diamidino-2-phenylindole

DB	Data Base
DC	Dendritic Cells
DCN	Decorin
DD	Dupuytren's Disease
DEA	Differential Expression Analysis
DEG	Differentially Expressed Genes
DEX	Dexamethasone
DMARD	Disease Modifying Anti-Rheumatic Drug
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSS	Dextran Sulfate Sodium
EBI	European Bioinformatics Institute
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-Linked Immunosorbent Assay
ELN	Elastin
EMBL	European Molecular Biology Laboratory
ER	Endoplasmic Reticulum
ETP	Endotrophin
EU	Endotoxin Units
EULAR	European Alliance of Associations for Rheumatology
FACS	Fluorescence-activated cell sorting
FAK	Focal Adhesion Kinase
FAP	Fibroblast Activation Protein
FBG	Fibrinogen
FBLN	Fibulin
FBN	Fibrillin
FBS	Foetal Bovine Serum
FC	Fold Change
FDR	False Discovery Rate
FFPE	Formalin Fixed Paraffin Embedded
FGFBP	Fibroblast Growth Factor Binding Protein
FGL	Fibrinogen-Like Protein
FMO	Fluorescence Minus One
FN	Fibronectin
FS	Frozen Shoulder
GBM	Glioblastoma
GM-CSF	Granulocyte Macrophage Colony Stimulating Factor
GP	Giga Pascal
GZMB	Granzyme B
HBSS	Hank's Balanced Salt Solution
HLA	Human Leukocyte Antigen

HNC	Head and Neck Cancer
IBD	Inflammatory Bowel Disease
IF	Immunofluorescence
IFN	Interferon
IGF	Insulin-like Growth Factor
IGFBP	Insulin-like Growth Factor Binding Protein
IHC	Immunohistochemistry
IL	Interleukin
IMID	Immune Mediated Inflammatory Diseases
IMS	Industrial Methylated Spirit
INHBA	Inhibin-A
IPF	Idiopathic Pulmonary Fibrosis
ITGA	Integrin α
ITGB	Integrin β
IV	Intravenous
JAK	Janus Kinase
JKT	Jurkat
LAL	Lumulus Amebocyte Lysate
LL	Lining Layer
LN	Lymph Node
LOESS	Locally Estimated Scatterplot Smoothing
LOXL	Lysyl Oxidase Like
LPS	Lipopolysaccharide
LSM	Laser Scanning Microscope
LTBP	Latent TGF-beta Binding Proteins
MAST	Model-based Analysis of Single-cell Transcriptomics
MDK	Midkine
MDM	Monocyte Derived Macrophages
MES	2-(N-morpholino)ethanesulfonic acid
MFAP	Microfibrillar-Associated Proteins
MFI	Mean Fluorescence Intensity
MGP	Matrix Gla Protein
MHC	Major Histocompatibility Complex
MMP	Matrix metalloproteinase
MRI	Magnetic Resonance Imaging
MZB1	Marginal Zone B and B1 cell specific Protein 1
NDMRB	Nuffield Department of Medicine Research Building
NET	Neutrophil Extracellular Trap
NHS	National Health Service
NHSBT	National Health Service Blood and Transplant
NID	Nidogen
NOTCH	Neurogenic locus notch homolog protein
NOV	Nephroblastoma Overexpressed

NSAID	Non Steroidal Anti-Inflammatory Drug
NTN	Netrin
OA	Osteoarthritis
OCT	Optimal Cutting Temperature
OI	Osteogenesis Imperfecta
PAGE	Polyacrylamide Gel Electrophoresis
PBMC	Peripheral Blood Mononuclear Cells
PBS	Phosphate-Buffered Saline
PBST	Phosphate-Buffered Saline Tween
PBT	Phosphate-Buffered Saline Triton
PC	Principal Component
PCA	Principal Component Analysis
PDPN	Podoplanin
PECAM	Platelet endothelial cell adhesion molecule
PFA	Paraformaldehyde
PLP	Periodate-Lysine-Paraformaldehyde
PMA	Phorbol 12-myristate 13-acetate
PPAR	Peroxisome proliferator-activated receptors
PTM	Post-Translational Modification
PTX	Pertussis Toxin
RA	Rheumatoid Arthritis
RACE	Research into Inflammatory Arthritis Centre Versus Arthritis
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
SDS	Sodium dodecyl-sulfate
SL	Sublining
SLE	Systemic Lupus Erythematosus
SLRP	Small Leucine Rich Protein
SMA	Smooth Muscle Actin
SPARC	Secreted Protein Acidic And Cysteine Rich
SPARCL	Secreted Protein Acidic And Cysteine Rich Like
SSc	Systemic Scleroderma
STIA	Serum Transfer Induced Arthritis
TAM	Tyro3, Axl, and Mer
TC	Tissue Culture
TGF	Transforming Growth Factor
TIMP	Metalloproteinase inhibitor
TK	Tyrosine Kinase
TLR	Toll Like Receptor
TLS	Tertiary Lymphoid Structure
TME	Tissue Microenvironment
TNC	Tenascin-C

TNF	Tumour Necrosis Factor
TRAFIC	Targeting synovial fibroblast proliferation in rheumatoid arthritis
UCMD	Ullrich Congenital Muscular Dystrophy
UPA	Undifferentiated Peripheral Arthritis
UV	Ultra Violet
VCAN	Versican
VWF	von Willebrand Factor
WB	Western Blot

Abstract

Rheumatoid Arthritis (RA) remains plagued by poor treatment responses and a lack of curative solutions despite a dense armamentarium. A deeper investigation into the tissue biology of the RA synovium, including its cells but also the remodelled extracellular matrix (ECM) that characterises it, represents a promising new approach in the quest for therapeutic tissue repair and inflammation resolution.

Publicly available single cell RNA sequencing datasets were analysed to define the matrisomal gene expression landscape in RA. Distinct fibroblast subsets were found to express distinct matrisomal gene profiles, thereby defining the composition of specific subsynovial microarchitectures. Differential expression analysis identified common and unique matrix gene expression cassettes between fibroblast subsets, between osteoarthritis (OA) and RA, and between active RA and RA in remission. IGF signalling and basement membrane compositional dysregulation were identified as key features of RA pathology and investigated further at the protein level.

A multiplexed immunofluorescence dataset was generated to look at the spatial distribution of IGFBP5, IGFBP7, COL4A1, and COL6A1 in human OA and RA synovium. COL6A1 was notably found to correlate positively with remission-associated MerTK^{+} macrophage abundance in tissue. Its deposition pattern was found to highlight specific areas devoid of collagen VI, termed dark zones, where immune cells localised.

Despite the correlation with MerTK^{+} macrophages in tissue, collagen VI was unable to polarise monocyte derived macrophages towards this remission associated phenotype. Looking more closely into the dark zones however, all immune cell subsets apart from monocytes were tightly locally restricted to collagen VI dark zones. Monocyte ingress into collagen VI-rich regions was found to be associated with active collagen degradation, suggesting potential degradation-dependent migration. Preliminary data point to a direct role for collagen VI in preventing immune cell adhesion and thereby their migration through tissue.

Fragments resulting from the cleavage of collagen by migrating monocytes were detectable specifically in human RA synovium.

These data point to a novel mechanism of immune migration restriction by synovial stromal cells. Further investigation of this process may hold significant promise in the reestablishment of tissue homeostasis and the curtailing of immune cell infiltration in RA.

Chapter 1 | Introduction

1.1. Tissue biology

1.1.1. Immune-Mediated Inflammatory Diseases: a clinical unmet need

Immune-Mediated Inflammatory Diseases (IMIDs) are a collection of inflammatory diseases driven by immune dysregulation[1]. Despite their heterogeneity, IMIDs broadly centre around immune cell activation and infiltration, and are often associated with cytokine secretion and signalling dysregulation. They include but are not limited to inflammatory bowel diseases (IBD), systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and Sjogren's syndrome. The incidence of IMIDs has sharply increased over the last three decades, requiring the development and implementation of novel preventative strategies to curb this trend[2]. Because of their poorly understood complexity and multifactorial nature[3], IMIDs are plagued by poor response rates to available treatments, and a complete lack of curative therapeutic avenues[4,5]. Their rise in incidence adds both urgency to a significant unmet clinical need, and pressure to healthcare resources already stretched to their limit in the wake of the COVID-19 pandemic[6].

Major IMID therapeutic advances in the last decade have centred around the blockade of individual pathways or cell types involved in the underlying pathogenesis of IMIDs. Adalimumab or etanercept are highly successful examples of the former, through the targeting of TNFα, one of the key cytokines driving inflammatory states in RA and IBD[7–9]. This success inspired the development of Janus Kinase (JAK) inhibitors like tofacitinib, which block the signalling of a plethora of inflammatory cytokines by inhibiting kinases involved in signal transduction[10]. Targeting cell type trafficking has also proved relatively successful in instances where a specific cell type has been implicated in disease pathology. This is the case in IBD for example, where vedolizumab blocks α4β7 integrins used by $CD4^+$ T-cells to traffic to the gut[11]. The overarching strategy behind these is the quenching of the inflammation characteristic of the chronic stages of IMIDs through the inhibition of a set of molecular and cellular mediators. While this strategy has proved effective in inducing remission in some

subsets of patients, it has indeed fallen short of providing curative solutions or even generating any response at all in some patients.

1.1.2. When form informs function

A conceptual shortcoming of current therapeutic strategies is the lack of structural context surrounding the targeted pathways. Biological processes do not happen in a vacuum, they require spatial context. American architect Louis Sullivan brilliantly summarised the idea of rational design under the axiom “form follows function”, implying that the starting point for the design of any structure should be its purpose or function[12]. Scientists and engineers have since embraced the axiom, finding it summarises an almost universal truth. Function defines the shape or structure that an organism, a tissue, or an object can take. Examples of this include the design requirements of an airplane wing to generate lift in aeronautical engineering, or the evolution of the opposable thumb for grasping branches and objects in evolutionary biology. This axiom also implies a form of reciprocity used in other fields, such as archaeology, relying on the observation of “form” or structure, to infer function.

The expansion of this concept to biomedical research is ushering in a new age of spatial tissue biology. It aims to address the current shortcomings of IMID research by harnessing modern technologies, particularly spatial transcriptomics and proteomics. Its promises are three-fold. Firstly, to bring much needed context and granularity to already identified pathological mechanisms, which may enable effective patient stratification, clinical trial design, and drug repurposing. Secondly, to uncover new molecular and cellular ecosystems involved in disease pathology, and yet unidentified at a systemic level. Thirdly, to build an understanding of the structural characteristics of healthy and disease tissue, to provide a path and a target for curative therapeutic strategies.

1.1.3. Principles of tissue biology: the cellular and the extracellular

A healthy adult body is the result of a fragile harmony between roughly 30 trillion cells[13]. The interplay between these is only rendered possible by the existence of multiple levels of organisation. Three major physiological scales exist within this: individual cells, the tissues in

which they are assembled, and the organs they make up. While individual cells lack context and organs are too large to study effectively, tissues provide the ideal scale to investigate homeostatic cellular functions and the perturbations that plague them in disease.

A tissue is defined as a collection of cells and the complex 3-dimensional network of extracellular matrix (ECM) in which they are embedded[14]. The cells of a tissue are its living actors, performing individual functions which, in unison, generate the functional output of a tissue. In the main parenchymal tissue of the liver for example, hepatocytes arranged hexagonally around a central vein undertake the detoxification of circulating blood[15]. The ECM is a set of proteins and associated molecules produced by the cells of the tissue[16]. In an effort to exhaustively characterise the ECM and facilitate research efforts, a classification termed the *matrisome* was generated by Alexandra Naba and colleagues[17]. The core matrisome can be broken down into glycoproteins, collagens, and proteoglycans. These form the building blocks of the matrix surrounding cells. Adding to this core matrisome are a set of matrisome-associated proteins which include proteins with domains homologous with core matrisome proteins, ECM regulators including proteases such as matrix-metalloproteinases (MMPs) but also cross-linkers like lysyl oxidases, and secreted factors including cytokines, chemokines, and growth factors. These matrisome-associated proteins, particularly the ECM regulators, work to modulate the composition and integrity of the deposited core matrisome components. The final composition of the deposited ECM gives the tissue its mechanical properties and reflects its function. The ECM of bone for example is composed of 90% collagen type I and 10% of other proteins including glycoproteins such as osteonectin and thrombospondins, and proteoglycans such as biglycan and decorin[18–22]. Together these give bones their strength and load bearing properties. Cartilage on the other hand is primarily composed of collagen type II, assembled in parallel fibres along the external surface of cartilage and perpendicular fibres deeper in the structure, and the proteoglycan aggrecan acting as a sponge to absorb impacts and restore shape after the impact[23].

The ECM of a tissue also informs cell behaviour and interactions. The sensing of mechanical forces is a key regulatory mechanism for cells, particularly during the scarring

process of injury repair[24]. Strong associations exist between higher mechanical tension on the injury site and excessive ECM deposition, leading to hypertrophic scar formation, increased tissue stiffness, and loss of organ function[25,26]. The deposited ECM can also regulate cellular distribution by modulating adhesion and migration of resident and recruited cells, as well as creating gradients of chemotactic and growth factors[27].

The tissue microenvironment (TME) can therefore be defined as the delicate balance between the activity of cellular neighbourhoods which define ECM composition, which in turn affects cellular functions and organisation.

1.1.4. Spatial omics

With a growing understanding of the importance of tissue-level dynamics in the outcomes of pathway dysregulation and modulation, a need for novel technologies capable of investigating these dynamics has emerged. Biomedical research has derived tremendous benefit in the past from single-cell technologies, both in the form of flow cytometry and single cell transcriptomics[28,29]. These technologies have provided unequalled resolution in deciphering the identity, and activation states of tissue cells. They have provided the first steps towards unbiased systems biology and immunology, yet they lack spatial context with crucial implications for tissue-level outcomes. A new era of omics technologies, termed spatial omics has since filled this niche[30]. Crudely these can be separated into spatial transcriptomics and spatial proteomics.

Modern spatial transcriptomics methods broadly rely on arrays of spatially barcoded probes distributed on commercially available chips, evolving as an improvement on fluorescence in situ hybridization methods[31] (Table 1.1). Tissue sections can then be placed onto the chips, where tissue RNA is released and captured by the probes. The captured RNA can then be sequenced, and assigned to a specific tissue localisation based on the spatial barcode it contains. This simple concept is the basis for the plethora of commercial methods to have emerged recently, including Visium by 10x Genomics, CosMx by nanoString, and Slide-seq[32–34]. These provide novel untapped insights in the spatial networks of gene

expression that are likely interacting with each other. The main current limitation of these methods is their resolution, but novel methods such as Visium HD aim to solve this problem and generate spatial transcriptomic signatures for each individual cell of a given tissue[35].

The information provided by transcriptomic methods is highly valuable, as it is able to capture the breadth of gene expressed and hint at the transcriptional programs activated in a given cell in a given tissue. Yet they remain an indirect representation of the functional output of a tissue, driven by its proteins. This is made worse by the only poor correlation that exists between mRNA levels and protein levels[36]. Spatial proteomics therefore also have a role to play in this emerging field of tissue biology and immunology (Table 1.1). These can be broadly split conceptually between Multiplexed immunohistochemistry/imaging mass spectrometry (IMS) such as Deep Visual Proteomics[37] or Hyperion[38], and cyclic immunofluorescence methods including antibody-based detection using either fluorophore-conjugated antibodies in CellDIVE or antibodies with DNA barcodes and fluorescent dNTP analogues used in CODEX[39]. While IMS methods are still limited by their throughput and the scale of the data generated, antibody-based methods are limited by the availability of validated antibodies against the investigated targets of interest. Method selection is therefore heavily reliant on project type and the nature of the questions asked of the data. Because of these limitations, spatial transcriptomics and proteomics should be used in tandem to characterise the composition of tissues of interest in health and disease.

Table 1.1 Modern spatial transcriptomics and proteomics methodologies.
Adapted and expanded on from Bressan and colleagues[30].

	Multiplexed Immunhistochemistry/Mass Spectrometry	Cyclic immunofluorescence	Multiplexed In Situ Hybridization	Spatial barcoding
Methods	IMC[1], MIBI[2], DVP3	CyCIF[1],4i[2], CODEX[3], MACSima[4], Cell DIVE[5], COMET[6]	MERFISH[1], seqFISH[2], seqFISH+[3], osmFISH[4], SCRINSHOT[6], EEL-FISH[6]	Visium[1], SlideSeqV2[2], Stereo-seq[3], HDST[4], Seq-Scope[5], PIXEL-seq[6]
Detection system	Antibodies (all) and mRNAs (1)	Antibodies	RNA (all), DNA (1 and 3)	RNA (all), DNA (2), antibodies (1)
Resolution	1 pm (1), 100s of nanometers (2 and 3)	Diffraction limited	Diffraction limited	<1 to 100 jim
Acquired area	Up to centimeters	Up to centimeters	Up to centimeters	4 mm-2cm
Tissue type	Frozen, FFPE	Frozen, FFPE	Frozen (all), FFPE (1 and 5)	Frozen (all), FFPE (1)
Commercially available as	SBT Hyperion (1), IonPath, MIBISCOPE (2)	Leica Cell DIVE (5), Akoya Phenocycler (3), Miltenyi MACSima (4), Lunaphore COMET (6)	Vizgen MERSCOPE, 10X Xenium, Nanostring CosMX, Resolve biosciences, Rebus Esper	10X Visium, Curio Seeker

1.1.5. The dawn of tissue biology: networks, neighbourhoods, and niches

In applying these modern methodologies towards the characterisation of tissue architecture and dynamics in pathological settings, research groups are beginning to generate valuable insights. Exemplars of the early applications of these methodologies can provide insights into the value generated and the direction the field is headed in.

One of the earliest applications of spatial transcriptomics platforms was performed on human leprotic lesions and the formation of granulomas[40]. The samples used in the study were collected from human leprotic lesion samples undergoing either disseminating leprosy (L-lep) or Reversal Reactions (RR), a transitory phase towards self-limiting leprosy and the formation of granulomas. ScRNAseq was first performed to define the cellular identities and phenotypes present in both types of lesions. The authors identified 12 cell types (T-cells, B-cells, plasmablasts, myeloid cells, Langerhans cells, mast cells, keratinocytes, fibroblasts, smooth muscle cells, endothelial cells, eccrine gland cells, and melanocytes), creating a blueprint for the building blocks of the tissue. At this level already, phenotypic differences were

identified in the T-cell compartment, with greater expression of cytotoxic factor *GZMB* in RR T-cells compared to L-lep T-cells. Looking further into cytokine and cytokine receptor expression and their theoretical interactions, an antimicrobial network could be hypothesised, underlying the differences between RR and L-lep samples, and gravitating around the expression of IL-1β from Th17 T-cells and IFNγ from Langerhans and Dendritic cells. Spatial sequencing then enabled the authors to put tissue coordinates on these putative networks, forming the cellular neighbourhoods responsible for deciding the fate of the leprotic niche. This additional layer of information enabled the authors to observe the architecture and spatial relationships between the identified cellular actors of the networks, and to build a model of a self-limiting leprotic granuloma. Macrophages were found to form clusters at the centre of the granuloma, surrounded by T-cells, Dendritic cells, and fibroblasts in an outward-radiating structure. Importantly, this encasing of the macrophages and anti-microbially active T-cells was associated with high levels of expression of anti-microbial signatures associated with IL-1β and IFNγ responses by dendritic cells and fibroblasts. This tightly regimented architecture is critical to the self-limiting potential of the infection and sheds a new light on the deep underlying mechanisms of leprosy and the crucial roles for the IL-1β and IFNγ in its regulation.

Early applications of spatial proteomics have provided similarly impactful insights. One of the first to do so was a study focused on prostate cancer[41]. Prostate cancer is typically detected through multiparametric MRI (mp-MRI), yet this technology fails to identify about 15% of cases. To untangle the different characteristics of visible and invisible tumours via mp-MRI, the authors of the study collected tumour samples from both conditions and subjected them to CellDIVE multiplexed immunofluorescence. The samples were stained for epithelial markers (PCK26, NaKATPase), endothelial marker CD31, immune marker CD45, T-cell markers CD4 and CD8, macrophage marker CD68, and cell cycle marker Ki67. Because of the number of cell types that could be detected within each sample, the authors could then assess the architecture of tissue niches and compare them based on the origin of the sample. Upon first inspection, mp-MRI invisible tumours were found to have much simpler structures than mp-MRI visible tumours. Mp-MRI visible tumours were also found to have greater cell

density and tumour cell neighbourhood sizes. The complexity and density of the tumour niche may therefore be playing a role in the detectability of the tumour via MRI. Interestingly, the density and neighbourhood size of the tumours correlated significantly with the likelihood of patient relapse. Despite a limited immune panel, the authors then looked into the differences in immune composition at various sites of the tissue. Importantly, malignant epithelium appeared to lack immune cells, regardless of the condition, which may be linked to the poor response rates to immune checkpoint therapies in prostate cancer. Some differences were however identified, with a greater abundance of $CD4^+$ T-cells in mp-MRI visible tumour malignant stroma. These observations suggest important differences between these two subtypes of prostate cancer, and importantly identify novel structural features associated with poor prognosis, which appear to be also directly related to mp-MRI visibility. In harnessing an early iteration of spatial proteomic technologies, the authors may therefore here have identified a novel clinically implementable indicator of prostate cancer prognosis.

These two examples of studies harnessing either spatial transcriptomics or proteomics showcase the power of these modern methodologies and of the conceptual shift towards tissue biology in elucidating the deep underlying mechanisms associated with human pathologies. While the study of communicable diseases and cancer has benefited the most from the advent of these novel methodologies, the field of IMID research is slowly catching up. In this context, the study of Rheumatoid Arthritis presents a particularly exciting opportunity, sitting at the intersection between clinically unmet need and gaps in the understanding of its tissue composition

1.2. Rheumatoid Arthritis

1.2.1. Principles of synovial joint tissue architecture

Synovial joints are the most common type of joint in the human body and are characterised by the presence of a fluid-filled cavity[42]. Broadly, synovial joints are made up of articular cartilage, synovial fluid, and surrounded by an articular capsule. The function of articular cartilage is to minimise the friction between the multiple opposing bone surfaces, as well as absorb impacts to prevent bone structural damage. As previously discussed, the composition of cartilage reflects these function, with collagen type II and aggrecan giving it its unique mechanical properties[23]. The articular capsule is composed of two distinct layers, an outermost fibrous layer and an inner synovial tissue layer. The fibrous layer is ligamentous and holds the opposing bones together, while also protecting the inner synovial layer. These ligaments are typically collagen type I-rich, giving them their tensile strength[43]. The synovial fluid within the articular is produced by the synovial tissue layer and aims to reduce friction between the moving parts of the joint. This synovial fluid is composed primarily of hyaluronic acid and lubricin[44]. Through their structural properties, the two molecules reduce adhesion and shear rates of the cartilaginous surfaces of the bones[45,46]. In addition to producing synovial fluid, the synovial layer, or synovial tissue, also plays an important role in maintaining the nutrient balance[47]. It is highly vascularised, providing a regulated exchange of blood-derived materials and oxygen to the joint capsule, and also controlling the homeostatic coming and going of cells into the fragile ecosystem of the joint.

1.2.2. Synovial tissue cellular and molecular composition in homeostasis

As with other tissues of the joint capsule, the function of the synovial membrane is reflected by its cellular and molecular composition. The synovial membrane can be divided into two main structures, at the macroscopic and microscopic level[48]. The lining layer, the innermost part of the synovium, is a 1-3 cell thick layer composed of tissue resident macrophages and fibroblasts, distributed through a basement-like membrane[49]. This layer acts as a selective sieve for the exchange of nutrients from the blood and the synovial fluid,

and maintains the integrity of the cavity and of the articular capsule. This is in part mediated by specialised differentiation of the macrophages in the lining layer, which express tight junction proteins aimed at the formation of a neat separative layer. The lining layer is also responsible for the bulk of the production of lubricin (PRG4), expressed in vast quantities by its fibroblasts. Below the lining layer resides the sublining layer of the synovium, containing fibroblasts, tissue resident macrophages, blood and lymphatic vasculature, nerves, and adipocytes[48]. The function of this layer is primarily to vascularise and enervate the synovium. Together the lining and sublining layers of the synovium work towards the protection of the joint cavity from mechanical wear and tear and maintain the homeostatic balance of the articular capsule (Figure 1.1).

1.2.3. Clinical presentation of Rheumatoid Arthritis

Rheumatoid arthritis (RA) is an autoimmune disease of poorly understood aetiology, primarily characterised by inflammation and swelling of the synovial joints, and which can lead to loss of function and disability[50]. The affected joints can include the knees, hips, and proximal interphalangeal joints of the hand, while typically sparing the distal interphalangeal joints. The condition is often associated with secondary afflictions ranging from cardiovascular complications to significant psychological burden[51,52]. These all contribute to making RA one of the worst conditions in terms of impact on health-related quality of life[53]. The clinical course of RA is thought to develop from sub-clinical undetected inflammation to undifferentiated arthritis, through to fully diagnosable RA. Patients are diagnosed with RA according to the ACR/EULAR criteria[54]. The disease course is a difficult one, in which patients suffer through cycles of active disease flares and periods of disease remission, typically leading to worsening symptoms over time[55].

1.2.4. Epidemiology of Rheumatoid Arthritis

RA is thought to affect between 0.5 and 1% of the population of Western nations[56], with the scarce data available from other parts of the world suggesting a

similar prevalence[57]. It therefore carries a significant global socio-economic burden, explaining the considerable research effort it motivates. With such a critical medical and scientific undertaking at hand, a large body of work has focused on unravelling the complexities of the primary site of inflammation in RA, the synovial tissue.

1.2.5. The molecular and cellular basis of Rheumatoid Arthritis

In RA, both the architecture and cellular landscape of the synovium are significantly altered (Figure 1.1). At the structural level, the lining and sublining layers are markedly expanded due to cellular infiltration and proliferation[58]. At the cellular level, recent studies harnessing the power of modern sequencing technologies, particularly scRNAseq, have enabled an in-depth characterisation of the players involved in the observed tissue inflammation. We now know more about the various cell subsets implicated in inflammation of the RA synovium. These can be broadly separated into cellular components of the innate and adaptive immune response.

While not a member of the haematopoietic lineage, fibroblasts play an important role in the innate immune landscape of RA (Figure 1.1). Broadly, the resident homeostatic fibroblasts of the synovium become expanded in RA, and their homeostatic functions are gradually corrupted towards pro-inflammatory and degradative states. They are notably responsible for the dysregulated expression of pro-inflammatory cytokines like IL-6[59], and tissue degradative factors such as MMPs or RANKL[60,61]. In the last decade, the understanding of their functions have been refined through the identification of at least 4 spatially and functionally distinct fibroblast subsets. $CD90^-$ lining layer and $CD90^+$ sublining layer fibroblasts were first found to play distinct roles in driving degradation or inflammation respectively when injected into the knees of mice[62]. Within the $CD90^+$ sublining fibroblasts, 3 further subsets have been identified based on gene expression signatures, the $CD34^+$ SC-F1 fibroblasts, $HLA\text{-}DRA^+$ SC-F2 fibroblasts, and the $DKK3^+$ fibroblasts[63]. The SC-F2 fibroblasts are suggested to be the most expanded RA-associated population of fibroblasts[63], and to secrete

many of the pro-inflammatory factors associated with the pathology such as IL-6, or CCL19 and CCL21 which play a crucial role in the recruitment and retention of lymphocytes. The positional identity of these fibroblast subsets appears to be regulated by a gradient of NOTCH3, expressed by endothelial cells and diffusing outward from the vascular niche[64]. Fibroblast polarisation towards inflammatory phenotypes is thought to be one of the key mechanisms behind the chronicity of RA inflammation, as repeat inflammatory insults have been shown to induce inflammatory "priming" of fibroblasts[65].

The second key players of the innate immune landscape in RA are monocytes and macrophages (Figure 1.1). As previously discussed, tissue resident macrophages are crucially important in homeostasis maintenance. In RA, the balance between homeostatic tissue resident macrophages and recruited inflammatory macrophages is shifted unfavourably[66]. At the level of the lining layer, $TREM2^{+}$ $MerTK^{+}$ macrophages are gradually polarised into and replaced by $TREM2^{-}$ $MerTK^{-}$ cells lacking the structurally crucial tight junctions and primed for expression of pro-inflammatory factors such as IL-1β and neutrophil-attracting chemokine IL-8[67]. In the sublining, recruited and polarised inflammatory macrophages lead to the establishment of an inflammatory milieu with high levels of IL-1β, IL-6, and TNFα but also chemokines such as CCL2 or CCL3[67]. The expression of these chemokines then leads to further dysregulated recruitment of monocytes, which differentiate into these inflammatory sublining macrophage populations. Different studies report contributing roles for both non-classical and intermediate monocytes to the development of inflammatory arthritis[68,69]. Broadly however, RA monocytes appear to be metabolically dysregulated and primed for excessive pro-inflammatory activation, as verified by ex vivo stimulation of circulating RA monocytes[70]. These compounding factors likely contribute to the inflammatory polarisation of nearby resident cells such as fibroblasts and endothelial cells, further worsening the established inflammatory milieu.

Neutrophils are another set of myeloid cells found to play an interesting role in RA (Figure 1.1). Circulating neutrophils from RA patients are primed to released greater amounts of reactive oxygen species than in healthy controls[71,72]. They also express higher levels of FcRγ which may make them more sensitive to activation by ACPA complexes. In addition to responding more readily to ACPAs, their presence in the synovium may promote directly the generation of these ACPA complexes through the release of Neutrophil Extracellular Traps (NETs)[73]. Finally, recent work implicates neutrophils directly in the onset of inflammatory arthritis in the mouse antigen-induced model of arthritis[74].

Dendritic cells (DCs) are also involved in RA pathology and bridge the gap between innate and adaptive immune compartments of the dysregulated chronic immune response[75]. $CD11c^+$ classical DCs can be divided into cDC1 and cDC2 based on the markers they express and their functions. cDC1 are particularly equipped for cross-presentation of antigens to $CD8^+$ T-cells, while cDC2 form the main subtype of cDCs and perform typical antigen presentation, primarily to $CD4^+$ T-cells. Circulating cDC2 are more abundant in RA patients compared to controls and the costimulatory signals they harbour are sufficient to activate $CD4^+$ T-cells and promote the release of IFNγ, IL-4, and IL-17[76]. Transient depletion of these cDCs in mice was also found to alleviate arthritis[77], showcasing the important of dendritic cells in driving inflammatory arthritis both in mice and humans.

Through their dysregulation, dendritic cells potentiate T-cells, one of the key pathological effector subsets in RA (Figure 1.1). Specific T-cells clones are indeed found to be expanded in early RA biopsies, likely specific to the typical set of citrullinated proteins associated with RA onset[78]. Evidence also points to 5 amino acid positions in the HLA-DRβ1, -B, and -DP conferring significantly increased risk of RA[79]. Amongst T-cells, Th1 and Th17 subsets were historically thought to be the primary drivers of adaptive immunity in RA, yet recent work points to a new subset of peripheral helper T-cells (Tph). Tph cells express high levels of PD-1, IL-21, and CXCL13[80]. Tph are also capable of promoting antibody production

by B-cells, and produce TNFα, IFNγ, and GM-CSF, all implicated cytokines in RA pathogenesis. These Tph were importantly detected in the RA synovium of patients in recent scRNAseq-based studies[63], and likely play a role in germinal centre formation and activity.

Potentiated by $CD4^+$ Tph cells, B-cells are some of the terminal effectors involved in RA pathology (Figure 1.1). In response to stimulation by T-cells specific for citrullinated auto-antigens, B-cells and differentiated plasma cells secrete ACPAs which lead to additional targeted tissue damage and fuel the inflammatory milieu of the joint. In addition to Tph, fibroblasts of the synovium support the establishment of B-cell niches through the expression of B-cell survival factors BAFF, APRIL, and IL-6. Together, BAFF and APRIL promote the activity of activation-induced deaminase (AID) through TLR3 stimulation[81]. The generation of tertiary lymphoid structures (TLS) in the synovium, likely brought on by the expression of CCL19 and CCL21 by fibroblasts and the recruitment of Tph, leads to somatic hypermutation and non-classical affinity maturation in the tissue and towards these citrullinated self-peptides[82].

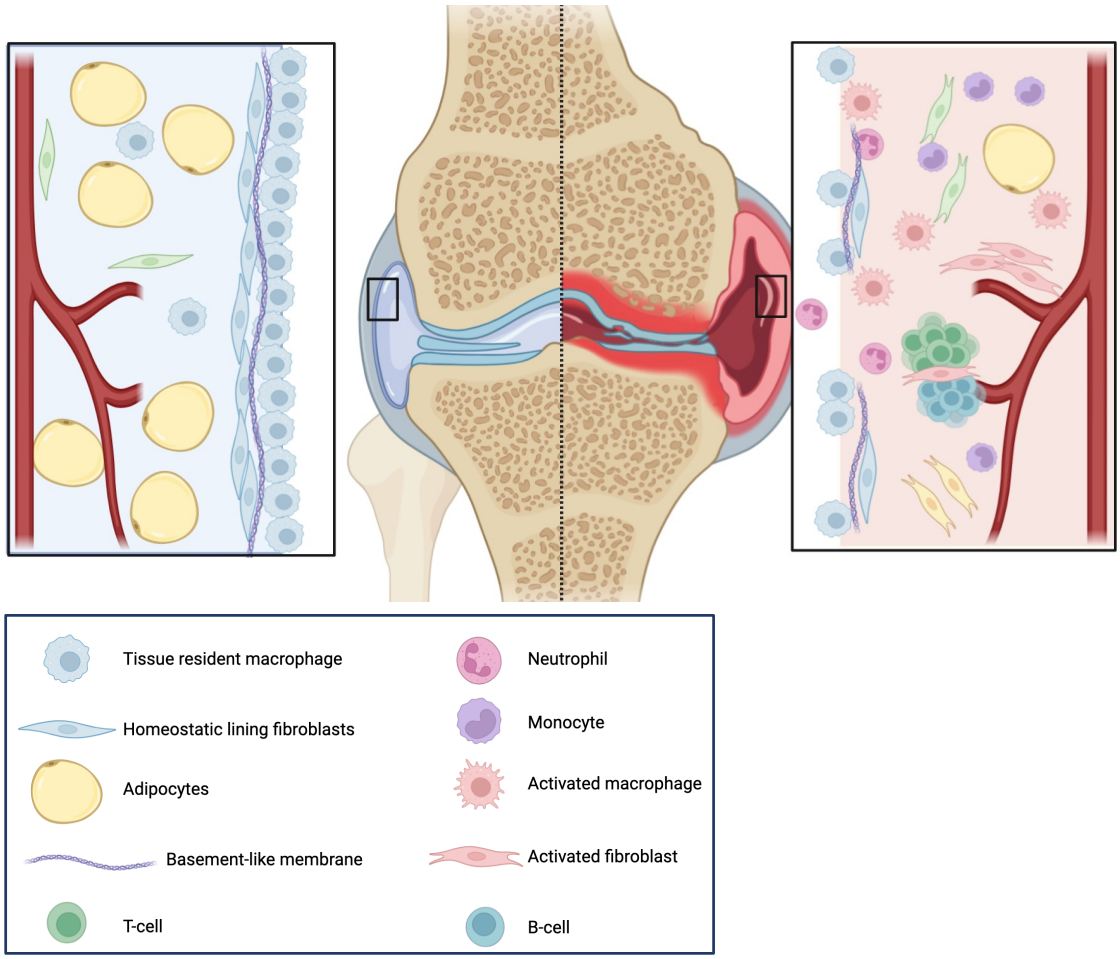


Figure 1.1 Cellular composition of the synovial membrane in health and Rheumatoid Arthritis. A knee joint is here split between schematic representations of the joint in health (left) and RA disease (right). Schematic created in Biorender.

RA is characterised by the concurrent dysregulation of a plethora of inflammatory cell subtypes. While little is known on the roles of the ECM in the RA synovium, some functions have been elucidated and hinted at. The ECM provides the synovium with a specific architecture which can be divided into three key structural niches: the sublining, which is further divided into the perivascular space, characterised by endothelium-associated basement membranes, and the loose collagenous interstitial space beyond, and the lining layer basement-like membrane[83]. The ECM also provides spatial patterning of critical cellular signals such as chemokines, cytokines, and growth factors through their capture and release by heparan sulfate proteoglycans for example[84]. Finally, some ECM proteins affect

phenotypes directly by signalling through cellular receptors. This latter effect is highlighted by one of the few ECM factors known to contribute to RA pathogenesis, tenascin-C (TNC). TNC is a large glycoprotein expressed at low levels in healthy adult tissue but transiently expressed upon tissue injury, during which the fibrinogen-like globe (FBG) domain of TNC drives inflammation through TLR4-mediated signalling[85]. Chronic TNC expression and the associated TLR4 activation have been implicated in the pathogenesis of RA, with antibody-mediated blockade of the FBG domain slowing disease progression in the rodent type II collagen-induced arthritis[86].

The example of TNC is an exciting prospect for future therapeutic development. More importantly, it highlights the role of dysregulated tissue-intrinsic factors in inflammatory diseases, including RA, and points to the lack of knowledge about synovial ECM composition as a major barrier to the understanding of the tissue in health and disease. A better understanding of synovial ECM composition may enable more effective targeting of the microenvironment, with the hope of improving treatment outcomes and patient lives. An in-depth analysis of the ECM factors present in the synovial microenvironment is therefore needed to address this gap in the literature.

1.2.6. Past, current, and future therapeutic strategies for RA

Treatment strategies for RA patients have evolved significantly since the 1920s[87]. Modern improvements to the standard of care require an understanding of where the therapeutic landscape has developed from and where it is headed. In these early days, diagnostic criteria for RA relied heavily on joint deformity and loss of function, biasing the therapeutic window towards established disease[88]. Still rudimentary laboratory methods left clinicians and scientists guessing as to underlying pathological mechanisms, and therefore had to rely on long term glucocorticoids and non-steroidal anti-inflammatory drugs (NSAIDs) for symptomatic treatment[87]. These long term symptom-driven strategies had no effect on disease remission and accumulating toxicities. Early disease-modifying anti-rheumatic drugs

(DMARDs) such as gold salts and penicillamine were also employed, with low efficacy and high toxicities[87]. The general outlook for patients during this era of RA therapeutics was poor, with only rare disease remission and high toxicity, only tolerated because of the scale of the disease burden.

In current practice, some legacy classes of drugs remain in frequent use, particularly as first line therapeutics. This is the case for glucocorticoids like prednisolone, and DMARDs. Glucocorticoids act via signalling through the glucocorticoid receptor, which translocates to the nucleus and modulates the transcription of close to 1% of the genome, and particularly inflammatory genes[89]. A recent study focused on RA showed that low-dose prednisolone in RA patients above the age of 65 may be somewhat beneficial for disease control[90]. Long-term glucocorticoid usage is however associated with increased risks of infections, osteoporosis, and diabetes[91], making it a suboptimal long term option. DMARDs on the other hand remain the first line of therapeutics given to many RA patients. The most common DMARDs used in RA include methotrexate, sulfasalazine, leflunomide, and hydroxychloroquine. None of these were originally developed to treat RA, but clinical benefit was observed with all of them, along with reasonable safety profiles. Methotrexate for example, was initially developed as a chemotherapeutic agent for cancer patients but is thought to potentiate anti-inflammatory adenosine signalling, leading to positive outcomes in RA particularly when combined with glucocorticoids[92].

In recent years however, a concerted push has been made towards mechanism-driven drug discovery, which has led to the development of targeted biologics and small molecules (Figure 1.2). Biologics target specific components of the underlying mechanisms of RA pathology identified through biomedical and clinical research efforts. The first and most popular of these were anti-TNFα biologics, particularly with adalimumab which became the best -selling drug in the world for a decade under its commercial name Humira[93]. Etanercept is a recombinant dimer of the TNF-R2 which binds to circulating TNFα[94], while infliximab and adalimumab are antibodies raised against TNFα[95]. They respectively achieve good to moderate EULAR response rates in 81%, 79%, and 87% of patients[96]. IL-1β and IL-6 are two

other chronically dysregulated inflammatory cytokines in RA, and the target of anakinra and tocilizumab respectively. Anakinra achieves ACR20 responses in up to 46% in certain studies when combined with methotrexate[97], while tocilizumab achieves ACR20 responses of up to 50%[98]. Beyond simple cytokine blockade, biologics have also been developed and used in RA to block specific cell types. This is the case for example with abatacept, a T-cell co-stimulation blocker designed to contain the extracellular domain of CTLA4 and an Fc-domain. In the AGREE trial, abatacept was found to induce remission in 46.1% of patients after one year, and 55.2% of patients after two years[99]. One final explored strategy has been the depletion of pathology-associated cells, harnessed by rituximab. Rituximab is a biologic targeting CD20 which leads to the depletion of most B-cells, thereby eliminating pathogenic auto-reactive B-cells in RA[100]. 29% of methotrexate-refractory patients treated with rituximab in combination with methotrexate were found to achieve an ACR50 response[101]. In tandem with biologics, small molecules targeted to similar inflammatory pathways have also been developed, most successful of which have been the JAK inhibitors. Janus Kinases (JAK) sit at the start of the signal transduction cascade for numerous inflammatory cytokines. JAK inhibitors are able to block the activity of multiple of these cytokines at once by targeting the signal transduction cascade itself. This is the case for upadacitinib, baricitinib, and tofacitinib, which target JAK1, JAK1/2, and JAK1/3 respectively. In one study, tofacitinib notably achieved ACR70 at 6 months in 37.7% of patients[102]. Together, these therapeutics form part of a dense armamentarium, yet the broad immune suppression they induce leads to significant risks of infections[103].

Therapeutic development for the future and ongoing clinical work aims to address some of the outstanding issues with the available armamentarium. Novel exploratory mechanism-driven strategies are at the forefront of this drive. One such novel strategy is the harnessing of adoptive cell transfers, whether it be to restore immune tolerance or target pathogenic cells. Tolerogenic DCs are a subset of DCs with acquired tolerogenic capabilities, particularly through exposure to IL-10 or TGF-β[104]. Clinical work now aims to induce tolerogenic phenotypes in patient DCs using autologous synovial fluid, and to reinfuse them

into patients in a bid to re-establish tolerance at the inflamed joint site[105]. A phase I trial showed no clear signs of efficacy, but demonstrated the intervention to be safe, forming a strong basis for further refinements. Future therapeutic development may also centre around the depletion of fibroblasts as a clear source of inflammatory cytokines and tissue remodelling, and primed inflammatory states which may explain the chronic nature of the disease. FAP^+ fibroblast depletion using diphtheria toxin indeed led to arthritic symptom alleviation in mice undergoing serum transfer-induced arthritis[62]. Coincidentally, particularly hard to treat RA patients have been found to have more fibroblast-rich synovial tissue than treatment responders in some studies[106]. New trials are therefore currently ongoing to assess the efficacy of targeting patient fibroblasts with selicilib, a fibroblast proliferation inhibitor, after positive safety signals from a phase I trial[107]. Finally, a push towards early targeted therapeutic interventions will likely be the next great revolution in RA therapeutics. Decades of clinical experience has demonstrated that early treatment leads to significantly better outcomes for patients, including for currently approved drugs like abatacept[108]. Patient stratification based on tissue composition is also opening the door to rational drug selection from the available armamentarium[106,109]. A combination of these two strategies will likely provide the best chance of significant quality of life improvements for patients, which should be the goal.

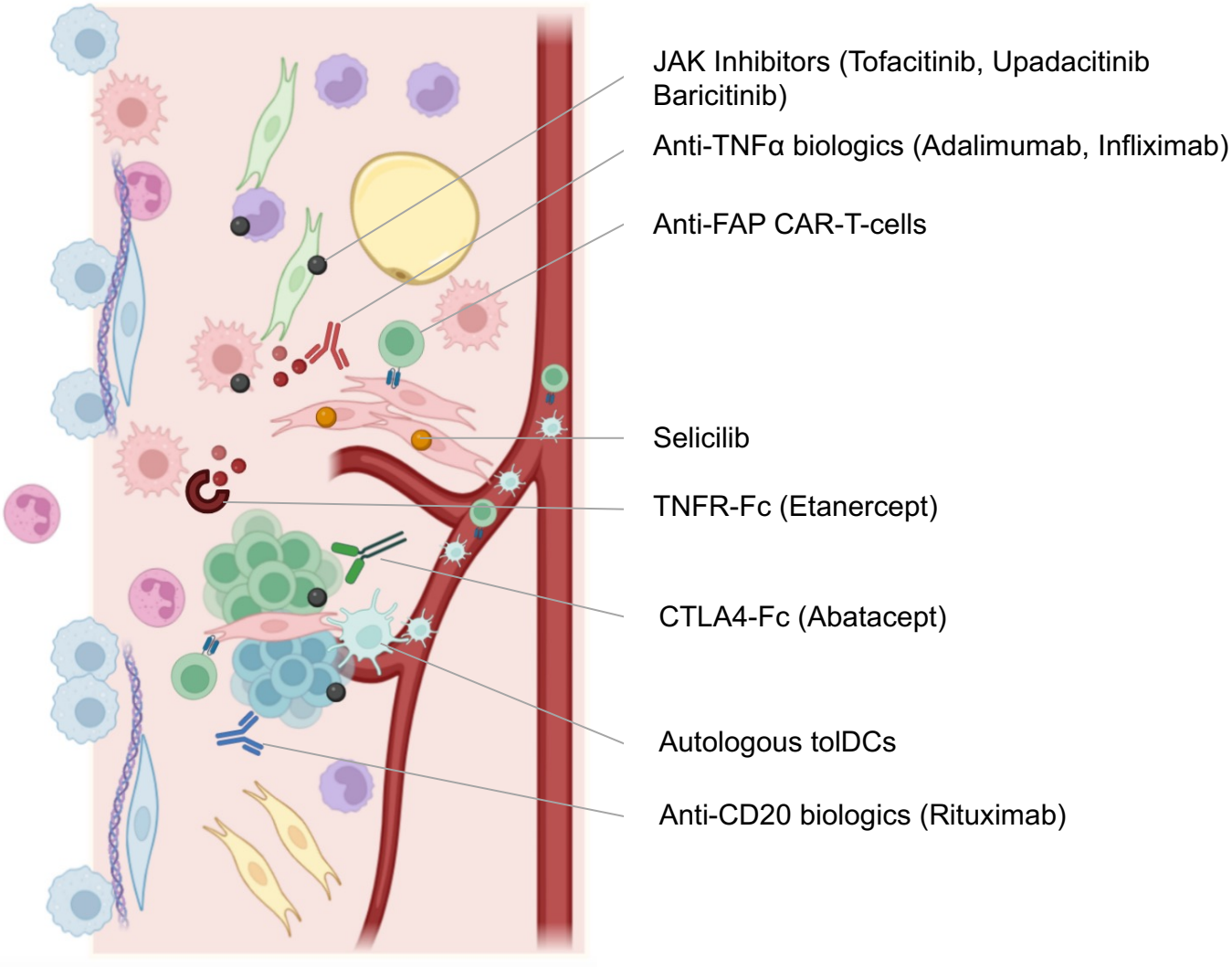


Figure 1.2 Targets of the modern and potential future mechanism-driven armamentarium for RA. Modern biologics and future cell-based therapeutics target many of the known underlying mechanisms of RA pathology. TolDCs = Tolerogenic Dendritic Cells. CAR-T = Chimeric Antigen Receptor T-cells. Schematic created in Biorender.

1.3. Type VI collagen

Through this study's exploration of ECM dysregulation in RA, type VI collagen was found to be of particular interest and therefore warrants an in depth exploration.

1.3.1. Principles of collagen biology

Collagen is one of the core matrisomal building blocks of any tissue, representing one third of total protein content of a human body[110]. While subtypes of collagen exist, they converge around a uniting structural assembly. Collagen is characterised by a right-handed triple helical structure, formed by three parallel polypeptide chains[111]. The tight structural assembly of these helices is reliant on every third amino acid to be a Glycine residue, forming X-Y-Gly sequence repeats, with X and Y being any amino acid. The proportion and location of these helices vary across the subtypes of collagen. The most common residues completing the repeating triplet are proline and hydroxyproline, with Pro-Hyp-Gly being the most common motif[112].

Higher order structures are reliant on the structural composition of the individual subtypes of collagen (Table 1.2). These subtypes can broadly be defined based on these higher order structures into fibril forming and non-fibril forming collagens. Collagen type I, II, III, V, XI, XXIV, and XXVII are fibril forming[113]. The deposition of these classical fibril forming collagens requires the assembly of individual collagen strands into a procollagen triple helix, which is trimmed by N-terminal and C-terminal proteases, into a tropocollagen triple helix, which can then self-assemble into microfibrils. These micro-assemblies, typically narrower than 5nm, are made possible by a 67nm stagger between tropocollagen helices[114]. Lysyl oxidases, based on availability and activity, may then cross-link these microfibrils to create much larger collagen fibres[115], which can reach 500nm in width and 1cm in length. The size of these assemblies explains the critical importance of collagens in delineating the architecture of tissues.

In addition to the size of collagen supramolecular assemblies, its mechanical properties also contribute significantly to the definition of the tissular ecosystem in which it is

deposited. Because of the multi-level assembly process, the mechanical properties of collagen can be assessed at different levels. Tropocollagen triple helices for example are exceptionally strong with an estimated fracture strength of 11GPa, while that of a full collagen fibril is 0.5GPa[116], and that of a tendon is roughly 100MPa[117]. The strength of a collagen strand therefore sets a theoretical fracture strength upper limit on tissular assemblies, while the crosslinking between these strands sets the practical limit. With each collagen subtype being associated with different chain structures and assembly architectures, their distinct expression patterns in distinct tissues lead to distinct properties (Table 1.2).

Proteases and crosslinking enzymes also play an important role in the regulation of collagen deposition, fibre organisation and strengthening. Collagen metabolism sits in constant balance between deposition and degradation, mediated by proteases such as MMPs and their inhibitors TIMPs[118,119]. MMPs are secreted as pro-MMPs requiring activation, which can be done by ROSs and other proteases, which then enables the degradation of specific collagens based on subtype-specific affinities[120]. The crosslinking of collagens on the other hand is mediated by lysyl oxidases, transglutaminases, and advanced glycosylation end-products[121]. Recent work also points to regulation of collagen deposition and degradation at the circadian level[121].

Because of the critical importance of collagen in tissue architecture, and therefore function, dysregulation of any components of the deposition or degradation machinery can lead to pathological complications. A number of musculoskeletal conditions are the result of mutations in the genes encoding collagen strands, including the classical subtypes of osteogenesis imperfecta (OI). OI, also known as brittle bone disease, is a heterogenous condition united around bone fragility and a high fracture susceptibility[123]. While multiple genetic mutations are implicated in the aetiology of the disease, they all revolve around a dysregulated biosynthesis of type I collagen, the primary component of bone. Over 1500 mutations in *COL1A1* and *COL1A2* have been reported to be linked to the condition[124], by preventing adequate production of collagen type I, thereby weakening bone structure. Fibrotic conditions are characterised by unbridled collagen deposition, which is further enhanced by

lysyl oxidase-mediated tissue stiffening. This disrupted balance leads to a chronic shift towards collagen anabolism, which gradually leads to loss of tissue and organ function. An example of this is Idiopathic Pulmonary Fibrosis (IPF), in which collagen types I and III are disproportionately deposited[125] and form larger and stiffer fibres due to *LOXL1* overexpression[126]. Certain types of cancer also appear to be at least partly driven by collagen imbalances. Recent work demonstrated that collagen type I homotrimers, formed of three α1 strands, are produced specifically by pancreatic cancer cells and promote cell proliferation and survival via signal transduction through α3β1 integrins[127]. Collagens also play other immunomodulatory roles in cancer, with stiff highly crosslinked collagen promoting proliferation and metastasis[128], and dense collagen matrix preventing infiltrating CD8+ T-cell migration[129]. Collagen dysregulation is therefore a common feature underlying many subsets of pathological conditions, and has implications for immune function.

Because of this critical importance of collagen subtypes, their modulation is now being harnessed in the context of these collagen-associated pathologies. Dupuytren's disease is for example characterised by fibrotic nodules and contractures. In a number of phase 3 clinical trials, patients were shown to benefit in the short term from the injection of collagenase clostridium histolyticum (CCH)[130]. This likely occurs through the tilting of the disrupted balance between collagen deposition and degradation through the injection of an exogenous collagenase. Another therapeutic avenue has been the development of collagen binding drugs or domains (CBD)[131]. These are classical drugs or recombinant proteins conjugated to collagen-binding domains with the goal of improving drug retention at sites of dysregulated collagen deposition. This is particularly interesting in the context of cancer, where excessive collagen deposition is suspected to affect drug tumour penetration and retention[132]. In selectively increasing the affinity of the drug for collagen-rich tissue regions, CBDs aim to harness this inherently pro-tumorigenic feature. This CBD conjugation was for example shown to improve the toxicity profile and efficacy of checkpoint inhibitors anti-CTLA4 and anti-PDL1 and of the cytokine IL-2 in mouse models[132]. These examples, far from being exhaustive,

highlight the tractability of targeting collagen subtypes as promising therapeutic avenues in conditions characterised by various flavours of collagen dysregulation.

Table 1.2 Tissue distribution and higher order organisation of the major collagen subtypes.
FACIT = Fibril-Associated Collagens with Interrupted Triple Helices.

Subtype	Chains	Tissue expression	Higher order structures
Collagen I	α1(I), α2(I)	Bone, dermis, tendon, ligaments, cornea	Fibrils
Collagen II	α1(II)	Cartilage, vitreous body, nucleus pulposus	Fibrils
Collagen III	α1(III)	Skin, vessel wall, reticular fibres of most tissues (lungs, liver, spleen, etc.)	Fibrils
Collagen IV	α1(IV), α2(IV), α3(IV), α4(IV),α5(IV), α6(IV)	Basement membranes	Basement membranes
Collagen V	α1(V), α2(V), α3(V), α4(V)	Lung, cornea, bone, fetal membranes; together with type I collagen	Fibrils
Collagen VI	α1(VI), α2(VI), α3(VI), α4(VI), α5(VI), α6(V)	Dermis, skeletal muscle, tendon, cartilage, placenta, lungs, vessel wall, intervertebral disc	Microfibrils
Collagen VII	α1(VII)	Skin, dermal–epidermal junctions; oral mucosa, cervix,	Anchoring fibrils
Collagen VIII	α1(VIII)	Epidermis, hair follicle, endomysium, intestine, chondrocytes, lungs, liver	Hexagonal networks
Collagen IX	α1(IX), α2(IX), α3(IX)	Cartilage, vitreous humor, cornea	FACITs
Collagen X	α1(X)	Hypertrophic cartilage	Hexagonal networks
Collagen XI	α1(XI), α2(XI), α3(XI)	Cartilage, vitreous body	Fibrils
Collagen XII	α1(XII)	Perichondrium, ligaments, tendon	FACITs
Collagen XIII	α1(XIII)	Epidermis, hair follicle, endomysium, intestine, chondrocytes, lungs, liver	Transmembrane collagen
Collagen XIV	α1(XIV)	Dermis, tendon, vessel wall, placenta, lungs, liver	FACITs
Collagen XV	α1(XV)	Fibroblasts, smooth muscle cells, kidney, pancreas,	Multiplexins
Collagen XVI	α1(XVI)	Fibroblasts, amnion, keratinocytes	Multiplexins
Collagen XVII	α1(XVII)	Dermal–epidermal junctions	Transmembrane collagen
Collagen XVIII	α1(XVIII)	Lungs, liver	Multiplexins
Collagen XIX	α1(XIX)	Human rhabdomyosarcoma	FACITs
Collagen XX	α1(XX)	Corneal epithelium, embryonic skin, sternal cartilage, tendon	FACITs
Collagen XXI	α1(XXI)	Blood vessel wall	FACITs

1.3.2. Overview of type VI collagen

Within collagens, type VI collagen is particularly interesting. Similarly to other collagens, collagen VI is reliant on the formation of a triple helix made of three distinct chains encoded across 6 genes: *COL6A1*, *COL6A2*, *COL6A3*, *COL6A4*, *COL6A5*, *COL6A6*[133]. The three main chains are α1, α2, and α3. The α3 chain bears structural similarities to 3 alternative chains α4, α5, and α6 (Figure 1.3)[134]. While α4 does not appear to be functional in humans, α5 and α6 can substitute for α3 in the triple helix[135]. Each collagen VI chain revolves around a similar core structure. They contain a short helical domain of 335-336 amino acids in length, surrounded by N- and C-terminal domains homologous to Von Willebrand Factor type A (VWFa) domains. Chains α1 and α2 are the smallest of the collagen VI chains, containing the helical domain, a single N-terminal and two C-terminal VWFa-like domains, resulting in proteins between 130-150kDa[136]. Chains α3-6 are larger, ranging from 220 to 300kDa. The α3 chain for example contains the helical domain, 10 N-terminal and 2 C-terminal VWFa-like domains, and 3 C-terminal domains[137]. These 3 additional domains are a unique domain, a fibronectin type-III domain, and Kunitz-like domain. Alternatively spliced α2-C2 and α3-N3/N5/N7/N8/N9/N10 have also been observed in mouse and human cells[138–140]. These chain-specific structural peculiarities lead to higher order assembly differences compared to other collagens. Indeed, α1/α2/α3 chains are first assembled intracellularly into a roughly 500kDa triple helix monomer (Figure 1.3). This monomer is then assembled into anti-parallel dimers bringing the size of the complex to roughly 1000kDa. Finally, these dimers are paired up further to form roughly 2000kDa tetramers via disulphide bonds[141–143]. These can then be secreted, and extracellularly assemble into beaded microfilaments. These beaded microfilament networks differ from the structures created by classical fibrillar collagens like collagens type I and II. Through their aligned fibrils, they form strength-providing structures, while collagen VI creates meshwork-like structures. Part of the assembly process of collagen VI is heavily reliant on the processing of collagen VI chains by specialised proteases. Bone Morphogenic Protein (BMP)-1 is the best characterised protease to do this. Following the intracellular assembly of the collagen VI tetramers, BMP-1 cleaves the α3 chain of collagen

VI at the C-terminal end of the C4 fibronectin type-III domain, leading to release of the C5 domain and enabling beaded microfilament assembly immediately following secretion[144,145]. This is of particular importance due to the unique functions of the released C5 domain, also known as endotrophin (ETP)[146], and is explored further in the next section. Through the same body of work, Furin-like Pro-Protein Convertase (PCSK3) was found to release a larger fragment of the α3 chain of collagen VI containing ETP, through cleavage between the C1 and C2 VWFa-like domains[145].

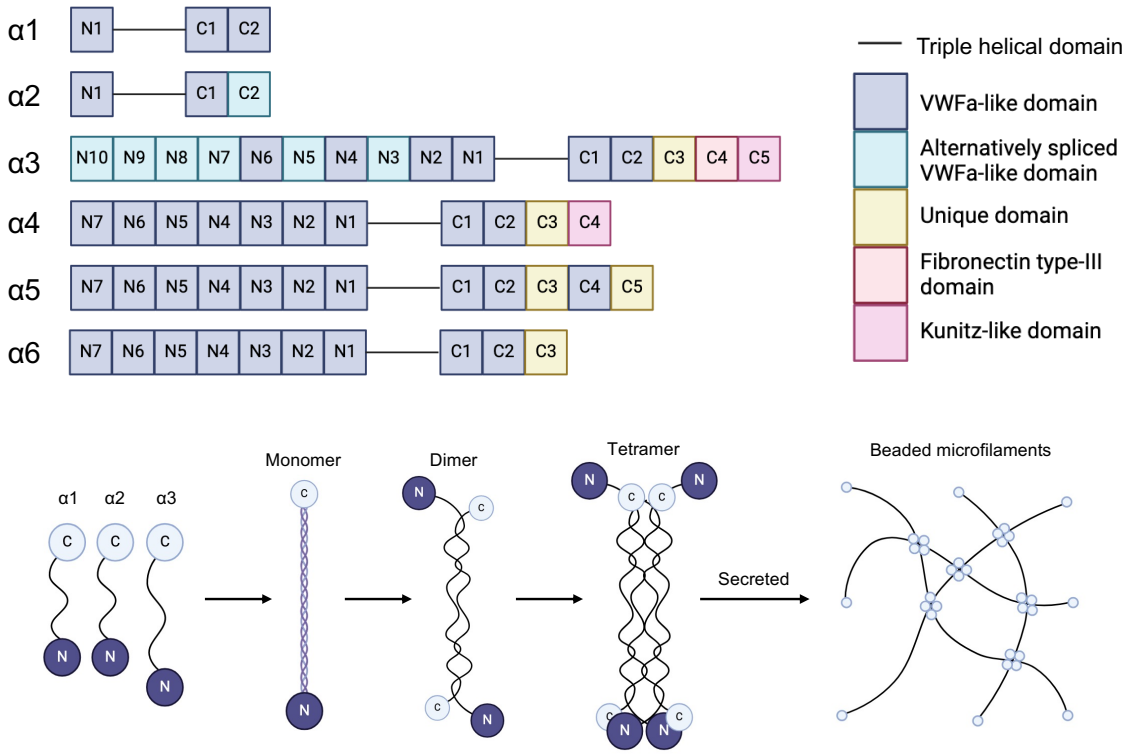


Figure 1.3 Collagen VI chain structure, assembly, secretion, and deposition. Collagen VI is made up of a heterotrimeric monomer, typically composed of chains α1, α2, and α3. This monomer is then assembled into an anti-parallel dimer, and further assembled into tetramers intracellularly. These tetramers can then be secreted and assemble extracellularly into networks of beaded microfilaments in the tissue. Schematic created in Biorender.

Beyond these two proteases known to be involved in the processing and assembly of collagen VI, three other proteases have been shown to cleave the α3 chain of collagen VI with unclear consequences. MMP-2 and MMP-9 notably have been shown to cleave collagen VI in multiple locations, including towards the start of the C1 domain of the α3 chain[147]. Finally, Furin-like protease PCSK7 has also been suggested to regulate the cleavage of the α3

chain[148]. Because of their role in neutralising and potentiating different domains of collagen VI, the dynamic temporally and spatially regulated distribution of these proteases must be evaluated to study the effective deposition of collagen VI in tissue.

Form follows and therefore informs function in the case of collagen VI. Its structural assemblies play a number of different roles in the tissues in which it is deposited. Skeletal muscle is one of such tissues. Collagen VI is one of the most highly abundant ECM components of muscle[149]. In muscle, collagen VI is primarily expressed by interstitial fibroblasts, and acts to support the basement membrane surrounding muscle fibres. This basement membrane association appears to be partly due to direct interactions with type IV collagen. Its importance in skeletal muscle is highlighted by the large spectrum of collagen VI-associated myopathies and is explored further in the next section. Collagen VI is also abundant in tendon, expressed predominantly by resident fibroblasts, and is distributed pericellularly due to the lack of basement membranes to support its focalisation[150]. Collagen VI is also an important component of the central and peripheral nervous system. Collagen VI was indeed found to be expressed and deposited by meningeal cells along with other components of the interstitial matrix and basement membrane such as collagen types I, III, and IV, nidogen, and laminin[151]. In the peripheral nervous system, collagen VI was found to be expressed by Schwann cells as part of their maturation process[152]. Adipose tissue was also found to contain large amounts of collagen VI[153]. In the same study, expression of *COL6A3* was found to strongly correlate with BMI and fat mass. *Col6* knockout mice on a high fat diet were interestingly found to have unrestricted enlarged adipocytes with improved inflammatory profiles. This finding suggests a potential role for collagen VI in the restriction of adipocyte size but also in the modulation of the inflammatory states associated with obesity[154]. Bone and cartilage, while being predominantly composed of collagen types I and II respectively, also harbours collagen VI-specific niches. The pericellular matrix surrounding chondrocytes, cells responsible for the preservation and consolidation of cartilage, is indeed rich in collagen VI. *Col6* knockout mice were interestingly found to have a higher propensity to osteoarthritic joint degeneration. These were attributed to changes in cartilage properties

due to the lack of collagen VI in the pericellular matrix of chondrocytes[155]. Finally, skin is one of the major sites of collagen VI expression and deposition. Its importance in skin is exemplified by the impact of mutations to any of the three major chains leading to skin abnormalities in humans, such as keloids or “cigarette paper” scars[156]. The wide array of tissues in which collagen VI is expressed make it an important structural and functional component of the human body. Its dysregulation, whether it be lack of deposition or overexpression, can therefore lead to a varied set of pathologies.

1.3.3. Roles for collagen VI in human pathology

Because of the important structural role that collagen VI plays in skeletal muscle, loss of function mutations in collagen VI typically lead to dystrophies. The collagen VI-related dystrophies (CRDs) can be placed along a spectrum of severity ranging from least severe Bethlem myopathy to the most severe Ullrich Congenital Muscular Dystrophy (UCMD). Bethlem myopathy is characterised by proximal muscle weakness, distal hyperlaxity, and joint contractures[157]. The myopathy progresses slowly, but can still lead to skin affections such as keloids, and to breathing restrictions in certain cases. Roughly 50% of patients living with Bethlem myopathy will need ambulatory assistance by the age of 60[158]. UCMD is much more severe. It is characterised by early life muscle weakness which culminates in the systematic loss of ambulation by age 10. This muscle weakness eventually culminates in scoliosis and ventilation supported breathing[159,160]. Both disorders are “umbrella” clinical presentations for loss of function mutations in any of the three main collagen VI chains, leading to less or no collagen VI being secreted, and a loss of integrity in collagen VI-dependent tissues. Endoplasmic Reticulum (ER) stress, induced by the inability to assemble and secrete collagen VI tetramers, has also been hypothesised as an underlying mechanism of CRDs, but currently lacks evidence to support it[161]. CRDs appear to be predominantly caused by mutations in the triple helical domain of the collagen VI chains (80% of mutations in one study[162]). This may be due to the crucial structural role of the triple helical domains in the assembly of functional collagen heterotrimers[111]. Mutations in this domain could therefore be more likely to cause

severe structural disruptions to the triple helix and affect collagen VI secretion. The severity of the clinical phenotype generated by a mutation is likely also determined by the different biochemical properties that arise from the amino acid substitution. One study exemplifies this particularly well, reporting two patients with a point mutation in Gly271 of the α2 chain[162]. Patient 2 presents with UCMD due to a Gly271Arg mutation, inserting an amino acid with a large and positively charged side chain. Patient 12 presents with Bethlem myopathy because of a Gly271Ser mutation, inserting an amino acid with a small uncharged side chain instead. The severity of CRDs therefore appears to be defined both by the location and the biochemical properties of the substituted amino acid, which likely impact the ability to secrete functional collagen VI tetramers.

Beyond the simple loss of collagen VI, dysregulated deposition of collagen VI has been heavily associated with numerous fibrotic phenomena. In Dupuytren's Disease (DD), characterised by fibrotic nodules in the palm, collagen VI was found to be overexpressed as a result of epigenetic dysregulation by CREBBP/EP300[148]. Collagen VI knockdown also decreased the profibrotic functions of tested DD fibroblasts. In the skin, increased collagen VI has been observed at the edge of keloids and as part of the tissue repair process in wound healing[163,164]. Some evidence also exists for increased collagen VI deposition in human skin fibrotic diseases such as systemic scleroderma (SSc). Indeed, higher levels of the α2 chain were detected in five SSc patients compared to healthy controls[165]. Increased collagen VI was also detected in the subcutaneous adipose tissue of SSc patients[166]. Kidney fibrosis also appears to be associated with collagen VI. Some deposition is detectable at the glomerular basement membrane in heathy adults[167], but interestingly significant increases were seen in patients with fibrotic renal failure[168]. The clearest example of an association between collagen VI and fibrosis remains however lung fibrosis. The lung is heavily reliant on basement membranes as an interface organ to external inhaled products. Collagen VI is expressed throughout the lung's basement membranes and underlying interstitial tissue[169]. This deposition was found to be exacerbated in lung fibrosis regardless of the aetiology of the fibrotic event[169]. Separate studies confirmed these findings, observing increased collagen VI

in interstitial lung disease tissue, chronic obstructive pulmonary disease, and idiopathic pulmonary fibrosis[170,171]. Mechanistically, this effect is partly dependent on ETP, the C5 domain of the α3 chain cleaved and released by BMP-1. Adipocyte-derived ETP has indeed been found to stimulate TGF-β expression and to promote fibrosis, angiogenesis and inflammation[146]. The exact mechanism through which ETP generates this effect remains however unclear. ETP does not yet have an established receptor. The intact collagen VI itself may also have a yet undescribed pro-fibrotic role. Some suggested hypotheses include the stiffening of tissue ECM from collagen VI deposition, which may lead to unrestricted collagen deposition[172].

A number of different cancer types are also characterised by excessive collagen VI deposition. Some examples include but are far from limited to lung cancer[173], pancreatic cancer[174], breast cancer[175], and colon cancer[173]. Collagen VI upregulation is notably associated with poor prognoses in a number different cancers including bladder and cervical cancer[176,177]. These poor prognoses fit with a general profile of increased cancer aggressiveness, with collagen VI suggested to play a role in metastases development in pancreatic cancer[178], as well as the formation of brain metastases in breast cancer[179]. In glioblastoma (GBM), collagen VI is also thought to contribute to the maintenance of the stemness of GBM cells, as well as promoting chemotherapy resistance through an associated cassette of co-regulated genes[180]. ETP also appears to play a pro-tumorigenic role, promoting tumour progression in a mammary mouse tumour model[146]. Its inhibition on the other hand appeared to render tumours cisplatin therapy-sensitive[181].

Emerging data points to a potential dysregulation of collagen VI deposition and processing in IMIDs. Indeed, Nordic Bioscience has begun to explore the use of peptide by-products of collagen synthesis and degradation as biomarkers. Their work points to increases in a biomarker termed C6Mα3, a C-terminal cleavage product of collagen VI by MMP9, in inflammatory bowel diseases[182]. Increased levels of C6Mα3 also appear to be associated with severity scores in the Crohn's disease and Ulcerative Colitis[183]. Interestingly, the opposite effect is seen with PRO-C6, a biomarker that corresponds to ETP. As discussed previously,

ETP release is the result of normal collagen VI assembly, and therefore represents levels of active collagen deposition. Recent data supports a similar trend in RA[184]. Indeed, the authors of the study found decreases in C6Mα3 levels to be associated with effective treatment responses to tocilizumab. Lack of reductions in C6Mα3 levels by 4 weeks appeared to predictive of lower odds of achieving any response on the drug. The implementation of such a biomarker could therefore facilitate the rationalisation of therapeutic regimen for patients, and the quicker identification of effective regimen. Together these data suggest a potential correlative, and possibly functionally causative effect of collagen VI deposition and degradation in IMIDs. The lack of mechanism underlying this effect makes it an attractive target for investigation, and is explored in this study.

1.4. Project aims and summary

As described throughout this introduction, the study of human disease at a tissue level is ushering in a new age of biomedical research. RA represents the perfect opportunity to apply this new approach. RA is complex in its aetiology and in the breadth of cellular actors involved in its onset and progression. Yet the spatial and tissular context in which these actors are at play remains poorly understood. The current therapeutic armamentarium is used to fight a losing battle in this context, with an incomplete understanding of the mechanisms underlying the pathology. One of the clear knowledge gaps surrounding RA is the ECM composition dynamics of the synovial tissue, and its interactions with the cellular compartment. Given the clear impact of ECM dysregulations in musculoskeletal conditions, fibrotic disorders, and cancer, ECM dysregulation may also play a role in RA.

Based on this, the first aim of this project was to compile an atlas of ECM expression in the synovium. Using the matrisome gene list and publicly available single-cell RNA sequencing datasets, the objective was to identify the major cellular regulators of ECM composition in the synovium. This would enable a more focused analysis into these cell subsets specifically. From there, matrix expression could then be compared between OA and RA, and between RA in remission and active RA, as per the conditions represented in the analysed datasets. Such an analysis would enable the identification of disease type and disease activity status-associated differentially expressed ECM genes. ECM genes of interest to come out of this analysis could then be prioritised to yield a list of key ECM genes associated with RA pathology. Chapter 3 of this thesis lays out these bioinformatic explorations, from an exhaustive characterisation of ECM gene expression in the RA synovium to the identification of key ECM genes to explore further at the protein level, including collagen VI. These transcriptomic findings were then complemented with the creation of a multiplexed immunofluorescence dataset, combining markers for the major immune and stromal cell subsets present in RA, as well as our key ECM proteins. Chapter 4 describes these proteomic

investigations, from the creation of a dataset to the spatial distribution of key ECM proteins in the synovium.

The second aim of this study was to apply this novel analysis pipeline to generate and test functional hypotheses surrounding our key ECM proteins. Chapters 3 and 4 provide evidence for the selection of collagen VI as our target of interest, and outline novel hypotheses surrounding collagen VI deposition and degradation in RA. Chapter 5 focuses on these functional explorations. The result of these explorations is the discovery of a new mechanism through which fibroblasts and monocytes regulate immune ingress and foci formation in the synovium. Together these data shed a new light on the extensive ECM dysregulation of the RA synovium. These findings point to new roles for the tissue microenvironment in the modulation of immune responses in tissue. This work forms the first step towards further exploratory work which we hope may culminate in novel therapeutic strategies to address a severe clinical unmet need.

In summary the questions I sought to answer through this work are the following:

1) What ECM is expressed in the RA synovium and does this change with disease activity?
2) Spatially, where is this ECM deposited and how does it associate with synovial cells?
3) What functional insights can be gained from these ECM-immune interactions?

Chapter 2 | Materials & Methods

2.1 Bioinformatics

Unless otherwise stated, all bioinformatic analyses were performed using R version 4.0.3 (2020-10-10), in RStudio version 1.3.1093 for MacOS. Plots, unless otherwise stated, were generated using *ggplot2* version 3.3.5.

2.1.1 Datasets

Table 2.1 Publicly available datasets used for bioinformatic analyses. scRNAseq = single cell RNA sequencing, OA = Osteoarthritis, RA = Rheumatoid Arthritis, STIA = Serum Transfer-Induced Arthritis, FS = Frozen Shoulder.

Species	Type	Tissue/Cells	Disease/ Model	Source	Accession no.	In-text name
Human	scRNAseq	Sorted T-cells/B-cells/Macrophages/ Fibroblasts from synovial biopsy	OA and RA patients	Zhang et al. (2019)[63]	SDY998	AMP1
Human	scRNAseq	Unsorted cells from synovial biopsy	Active RA and RA in remission patients	Alivernini et al. (2020)[67]	E-MTAB-8322	Alivernini
Mouse	scRNAseq	Unsorted synovial cells from synovial dissections	STIA	Wei et al. (2020)[64]	GSE145286	Wei
Human	scRNAseq	Unsorted cells from shoulder capsule	FS and comparator patients	Ng et al. (2024)[185]	ERP143358	FS

2.1.2 AUCell Gene Set Enrichment Analysis

Mapping of the RA synovial matrix was performed both on the AMP1 and Alivernini datasets using the *Seurat* and *AUCell* packages[186,187]. The major matrix and matrix subcategory-expressing cells were identified using the AUCell package following the documentation vignette (available through *vignette("AUCell")* in R). Notable changes made to the script laid out in the vignette were the use of the matrisome subcategories defined by the Matrisome Project as the gene sets of interest[17], and the increase of the threshold of cells used for gene set enrichment calculations from 5% to 10%. The gene set enrichment scores for each cell

and each matrisome subcategory, or AUC scores, were then used to generate the boxplots and feature plots, using the *ggplot2* package.

2.1.3 Pseudotime analysis

Pseudotime analysis was performed on the AMP1 dataset using the *slingshot* package[188]. Fibroblasts were computationally subsetted from the rest of the dataset's cells, and the lining layer SC-F4 were selected as the starting cluster for trajectory analysis based on their discrete physiological location relative to sublining SC-F1/2/3. Trajectories were plotted in PCA space. Temporally expressed genes and matrix genes were then identified by fitting a Generalised Additive Model, with a LOESS term for Pseudotime. These were plotted using the *pheatmap* package[189].

2.1.4 gsfisher Gene Set Enrichment Analysis

Gene Set Enrichment Analysis was performed using the *gsfisher* package, following the documentation vignette available in the package's GitHub repository (https://github.com/sansomlab/gsfisher).

2.1.5 Alluvial label transfer

Comparisons between the *de novo*-clustered and matrisome-clustered Alivernini fibroblasts were visualised using alluvial plots from the *ggalluvial* package[190]. Metadata columns containing the annotated clusters of each cell using each method were taken from the analysed *Seurat* object, enumerated, and wrangled into a data frame that could be taken by the package as input.

2.1.6 *De novo* fibroblast clustering

To enable direct comparisons across datasets, Alivernini dataset cells were integrated and clustered *de novo* using the current gold standard method for scRNAseq data integration, Harmony[191]. Data from remission and active RA samples from the Alivernini dataset was integrated using Harmony using the samples of origin as the covariates. A first round of clustering was then performed in *Seurat*, with a resolution of 0.3, to identify and subset stromal

cells based on the expression of PTPRC^{-} and PECAM1^{+}, FAP^{+} or PDPN^{+}. All clustering steps were performed using the *clustree* package to select the ideal resolution to optimise cluster robustness[192]. A second round of clustering was done, with a resolution of 0.3 again, to identify fibroblasts based on the expression PECAM1^{-} and PDPN^{+}, FAP^{+} or THY1^{+}. With the fibroblasts computationally sorted, a third round of clustering was performed with a resolution of 0.5, which revealed high levels of condition-specific clusters, hence unlikely to be biologically relevant. To account for these effects likely due to the large number of cells present in the initial integration, the cell identification codes for all the sorted fibroblasts were extracted and used to subset only fibroblasts from the original non-integrated dataset, and a new round of Harmony was run in these cells only. Fibroblast-specific integration was considerably improved as a result and a final round of integration with 0.4 resolution revealed AMP-like clusters, used as input for the Differential expression analysis (DEA) pipeline.

2.1.7 Matrisome-focused clustering

De novo clustered fibroblasts from the Alivernini dataset were also subjected to matrix-focused clustering. The matrisome gene list was first manually curated to create a new list excluding cytokines and chemokines from the Secreted Factors category of the matrisome (Supplementary Table 1). This curated list was then used as the *features* input for the *RunPCA()* function of the *Seurat* clustering pipeline, limiting the PCA input to curated matrisome genes and thereby limiting the downstream graph-based clustering to be defined by differences in matrisome gene expression. The basic level of matrix clustering presented here in this study was performed with a resolution of 0.3, while the deepest level presented here was performed with a resolution of 0.5.

2.1.8 Differential expression analysis

The Seurat pipeline described in the relevant vignette (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html) was used to identify cluster-defining markers between fibroblast subsets in each dataset. Specifically, the *FindAllMarkers*() function

was used, testing only for genes detected in at least 25% of cells within the cell subset and with an average log-fold change of at least 25% between two cell subsets, and using the default Wilcoxon Rank Sum test to identify differentially expressed genes between clusters. A more stringent differential expression analysis (DEA) pipeline was also set up and applied to the AMP and Alivernini datasets to compare OA and RA, and remission and active RA samples respectively. The pipeline comprises of two established DEA methods, edgeR and MAST[193,194], which both form part of the best practice in single cell analysis at the time of the analysis[195]. They are used here in parallel to provide an additional level of coverage of and confidence in potential differentially expressed genes (DEGs). The stringent bulk RNAseq method edgeR was run following the relevant documentation vignette (available through *edgeRUsersGuide*() in R) on raw counts data sum-pseudobulked by cell cluster and sample and with the p-values taken from the output of the *glmLRT*() function of the *edgeR* package. MAST was run directly on the log-normalised data using the *Seurat FindMarkers*() function, and p-values were adjusted using the *p.adjust*() function using the Benjamini-Hochberg adjustment approach[196].

2.1.9 Target prioritisation

To sieve through the large number of DEGs identified by both methods, a target prioritisation scheme was set up combining three different metrics. The first metric is the overlap between the DEA methods. The second metric is the conservation of the DEG expression pattern in the Wei dataset analysed using MAST, comparing control mice and Serum Transfer-Induced Arthritis (STIA) mice[64]. The third and final metric is a differential expression score calculated as:

$$|-log10(pval).log2(FC)|$$

The latter component avoids the common analytical downfall of defining genes as “statistically significant”, which is heavily caveated particularly when working with scRNAseq data. It relies instead on placing genes on a spectrum of differential expression.

2.1.10 Proteomic analysis

Proteomic analyses were performed by Dr. Anna Hoyle. Raw data was accessed from PRIDE, project PXD027703[197]. Data was reanalysed using Proteome Discoverer (version 3.1) using the TMTe 6plex quantification method, and searched against the Uniprot human database using SequestHT with a maximum false discovery rate (FDR) of 1%. Fixed modifications included in the search were carbamidomethyl addition to cysteine (+57.021Da) and TMT 6plex modifications at peptide N-termini and lysine residues (+229.163Da). Dynamic modifications of oxidation of methionine (+15.995Da), hydroxylation of proline (+15.995Da), deamination of asparagine/glutamine (+0.984) and phosphorylation of serine/threonine/tyrosine (+79.966) were also included. 2 missed cleavages were allowed. Both tryptic and semi-tryptic searches were run. Peptides were only included if they met the threshold 1% FDR and known contaminants were removed. Data was exported and onward analysis was conducted in Perseus (v2.0.11), R (v4.3.2) and Graphia (v4.2). Abundances were normalised using median subtraction. Semi-tryptic to tryptic ratios were calculated from average abundances of the relevant peptides for the protein of interest. Protein sequence mapping was done using normalised semi-tryptic peptide abundance. Network analysis was performed using Graphia[198]. Proteins are modelled as nodes and the edges between them are based on correlation in expression. Edges were only included if their Pearson Correlation Value was over 0.85. Within this, the data was organised based on Louvain clustering with a granulation of 0.5.

2.1.11 Cell subtype distribution in collagen VI rich and poor regions

The tissue sections stained for the validation panel of markers and imaged on the CellDIVE system were run through the SCIMAP python toolkit for in depth analysis of extracellular collagen VI fibres and cell phenotyping by Shihong Wu[199]. Following the pipeline,

cells and collagen VI fibres were segmented and cellular subtypes were assigned based on the decision matrix laid out in Supplementary Table 2. Local fibre density scores were calculated for each individual fibre through nearest neighbour analysis, and each fibre was assigned to COL6A1-rich or COL6A1-poor areas. This assignment was based on a cut off value set at the median local fibre density of each section, which was reproducibly found to be 0.01. Fibres were then clustered, and the outermost fibres were selected to label the boundaries of the COL6A1-rich and poor regions. The coordinates of these fibres were then joined up into a polygon, which was expanded further by 50μm to define the regions of interest. Cell subtype frequencies were then measured in each of the regions and the frequencies were exported into R, where a Fisher's exact test was run.

2.2 Laboratory reagents and solutions

Table 2.2 List of reagents used in the study

Reagent	Company	Catalogue #
96-well flat plates	Falcon	353072
96-well Nunc MaxiSorp (ELISA 96-well plate)	VWR	735-0083
Ammonium-Chloride-Potassium (ACK) Lysing Buffer	Gibco	A1049201
BD Cytofix™	BD	554655
Biolock	Fisher	15653608
BioLock – Biotin blocking reagent	Fisher	15653608
Bovine Serum Albumin (BSA)	Sigma – Aldrich	A7906
Buffer XT-R	Fisher	16654714
Carbo-free blocking solution 10x	2B Scientific	SP-5040-125
CellPath Optimal Cutting Temperature (OCT) Embedding Matrix	CellPath	15212776
Stock antigen unmasking solution	Vector Laboratories	H-3300
Corning™ 96-Well Clear Polypropylene Corner Notch Microplates	Corning	11313595
Corning™ Cell Lifter	Corning	10508292
Corning™ Costar™ 12-well plate Clear Tissue Culture-treated	Corning	10253041
4′,6-diamidino-2-phenylindole (DAPI) Stock Solution	ThermoFisher Scientific	D3571
Daudi	ATCC	CCL-213
Donkey Serum	Sigma – Aldrich	D9663
EasySep™ Human Monocyte Isolation Kit	Easysep	19359
Ethylenediaminetetraacetic acid (EDTA)	Alfa Aesar	33312
ELISA 3,3',5,5' tetramethylbenzidine (TMB)	Insight Biotechnology Limited	50-76-03
Ethanol	Honeywell	32221-2.5L
Fetal Bovine Serum (FBS)	Gibco	10500-064
Fc Receptor Blocking Reagent	Miltenyi	130-059-901
FluorSave Reagent	Merck	345789-20ML
Glycerol	Sigma – Aldrich	G5516
Hank's Balanced Salt Solution (HBSS) (10X), calcium, magnesium, no phenol red	Gibco	14065056
Histopaque®-1077	Merck	10771-100ML
Human M-CSF	Peprotech	300-25-1MG
Human TruStain FcX™ (Fc Receptor Blocking Solution)	Biolegend	422302
Hydrogen Peroxide	Sigma – Aldrich	216763
Ibidi µ-Slide VI 0.1	Ibidi	80666

Jurkat E6-1	ATCC	TIB-152
Kimberly-ClarkTM KimtechTM Science Precision Wipes (Kimwipes)	Fisher Scientific	11768188
Leica SurgiPath Coverglass 24 x 50 mm coverslips (Size 1)	Leica	3800145G
PyroGene endotoxin testing kit	Lonza	50-658NV
NativePAGE™ 4 to 16%, Bis-Tris, 1.0 mm, Mini Protein Gels	ThermoFisher	BN1002BOX
Non-Tissue Cutlure-treated Petri Dish	Falcon	351029
NuPAGE 2-ethanesulfonic acid (MES) buffer	ThermoFisher	NP0002
NuPAGE 4-12% gradient pre-cast gradient gels	ThermoFisher	NP0322PK2
Pap pen	ThermoFisher	8899
10% Paraformaldehyde (PFA)	ThermoFisher	47317.9L
Penicillin Streptomycin	Gibco	15140-122
Phosphate Buffered Saline (PBS) 10X	VWR	437117K
Propyl Gallate	Sigma – Aldrich	2370
PyroGene™ Recombinant Factor C Endotoxin Detection Assay	Lonza	50-658NV
RMPI 1640	Gibco	21875-034
Sodium Bicarbonate	Sigma – Aldrich	S6297
Sodium dodecyl sulphate (SDS)	MP Biomedical	102918
Sodium Hydroxide Pellets	Sigma – Aldrich	30620-1KG-M
Stock Antigen Unmasking Solution	Vector Labs	H-3300-250
Strep-Tactin XT Starter kit	Fisher	16660854
Strep-Tactin XT Superflow gravity flow column	iba	2-1209-001
Strep-TactinXT high capacity resin	Fisher	16684724
Super Frost Plus Glass Slides	Fisher Scientific	10149870
SYPRO® Orange Protein Gel Stain	Merck	S5692
SyTox Blue dye	ThermoFisher	S11348
T75 TC-treated Falcon Flasks	Fisher scientific	10508662
THP1	ATCC	TIB-202
Tris base	ThermoFisher	T1503
Triton X-100	Sigma – Aldrich	T9284
Trizma® Base	Sigma – Aldrich	T6066
TrypIE™ Express	Gibco	12604-021
Tween 20	VWR	663684B
Xylene	Sigma – Aldrich	534056
β-mercaptoethanol	Sigma Aldrich	M6250

Table 2.3 List of solutions and their composition

Solution	Composition
10x Tris Antigen Retrieval Solution	12.1g 10x Tris, 3.7g EDTA, 5.0mL Tween 20, 995mL dH2O
Annode Buffer	50 ml 20x NATIVE PAGE Running Buffer & 950 ml ultrapure H2O
Buffer BXT	100 mM Tris/HCl, 150 mM NaCl, 1 mM EDTA, 50 mM D-biotin, pH 8
Buffer W	100 mM Tris/HCl, 150 mM NaCl, 1 mM EDTA, pH 8
Cathode Buffer	12.5 ml 20x NATIVE PAGE Running Buffer, 1.25 ml 20x NATIVE PAGE Cathode Buffer Additive, 236.25 ml ultrapure H2O
CellDIVE mounting media	10mL 1xPBS, 90mL glycerol, 4.0g Propyl Gallate, 1.0g DABCO
Citrate Antigen Retrieval Solution (ARS1)	5% Stock Antigen Unmasking Solution, 95% dH2O
DAPI Stock solution	10mg DAPI dilactate, 2.0mL dH2O
ELISA coating buffer	1.59g Na2CO3, 2.93g NaHCO3 in 50mM Carbonate Buffer
ELISA stop solution	6% sulfuric acid in dH2O
FACS Buffer	2% FBS in 1xPBS
Laemelli buffer	0.0625 M Tris base, 0.07 M SDS, 50% glycerol, 25% β-mercaptoethanol, 0.5% (w/v) bromophenol blue
Macrophage differentiation media	10% FBS, 1% Pen/Strep, 100ng/mL in RPMI
muIF Blocking Solution	1x Carbo-free blocking solution, Triton-X free, 0.3 M Glycine, 0.08% NaN3
muIF Staining Solution	1x Carbo-free blocking solution, Triton-X free
PBS-Tween	0.01% Tween 20, in 1xPBS
PBST	0.05% Tween 20, in 1xPBS
PBT	0.3% Triton-X in 1xPBS
PLP Buffer	1% methanol-free PFA, 1.3% wt/vol L-lysine, and 0.2% wt/ vol sodium metaperiodate, in 1xPBS
RIA Buffer	1% BSA, 350mM NaCl, 10mM Tris/HCl pH=7.6, 1% Triton-X100, 0.5% Na-deoxycholate, 0.1% SDS
Silver Developer	40% Formaldehyde, 2% Sodium Carbonate, in dH2O
Silver Fixative	40% Ethanol, 10% Acetic Acid, in dH2O
Silver Reagent	0.1% Silver Nitrate Solution, in dH2O
Silver Sensitiser Solution	0.02% Sodium Thiosulfate, in dH2O
Silver Stopper Solution	5% Acetic Acid, in dH2O
Slide Blocking Solution	10% donkey serum, 3% BSA, 1xPBS
Standard media	10% FBS, 1% Pen/Strep, in RPMI
Starvation media	1% Pen/Strep, in RPMI
Stock 10x Tris Antigen Retrieval (AR) Solution	12.1g Tris, 3.7g EDTA, 5.0mL Tween-20, 995mL dH2O
Tris AR Solution (ARS2)	10% Tris AR stock, 90% dH2O
WB Transfer Buffer	200mL of 5x Transfer buffer (Biorad), 200mL pure ethanol, 600mL dH2O

Table 2.4 List of equipment

Equipment	Company	Link/Catalogue #
AE-6050 mini gel tank	Atto	P195
Biocare Decloaker	Biocare Medical	https://biocare.net/produ cts/instrumentation/declo aking-chamber-nxgen/
Cell DIVE multiplexed imaging platform	General Electric	https://www.ge.com/rese arch/project/multiplexed- tissue-imaging-platform
Curix 60 automatic film developer	AGFA	https://pdf.medicalexpo.com/pdf/agfa/curix-60/70894-214465.html
Fluostar Omega Plate Reader	BMG Labtech	https://www.bmglabtech.com/en/fluostar-omega/
Incucyte® SX5	Sartorius	https://www.sartorius.com/en/products/live-cell-imaging-analysis/live-cell-analysis-instruments/sx5-live-cell-analysis-instrument
Leica CM3050 cryostat	Leica	https://www.leicabiosystems.com/en-gb/histology-equipment/cryostats/leica-cm3050-s/
LSRFortessa	BD	https://www.bdbiosciences.com/en-gb/products/instruments/flow-cytometers/research-cell-analyzers/bd-lsrfortessa
Manual Rotary Microtome HistoCore BIOCUT	Leica	https://www.leicabiosyste ms.com/en-gb/histology-equipment/microtomes/hi stocore-biocut/
NanoDropTM One/OneC Microvolume UV-Vis Spectrophotom eter	ThermoFish er Scientific	ND-ONE-W
ZEISS LSM 980 Confocal	ZEISS	https://www.zeiss.com/microscopy/en/products/light-microscopes/confocal-microscopes/lsm-980-with-airyscan-2.html?utm_source=google&utm_medium=cpc&utm_campaign=C-00011305&gad_source=1

2.3 Immunofluorescence

2.3.1 CellDIVE multiplexed immunofluorescence – human tissue

Multiplexed immunofluorescence (mIF) imaging was performed on the CellDIVE (Leica microsystems). Synovial tissue samples from OA and RA patients were obtained from the University of Oxford (Prof. Sarah Snelling) and Birmingham (Mr. David Gardner) (Table 2.5). Samples were obtained with written, informed consent and approval from the West Midlands Black Country Research Ethics committee (12/WM/0258) and the Oxford Musculoskeletal Biobank (09/H0606/11 and 19/SC/0134) in compliance with the Declaration of Helsinki. After tissue collection, biopsies were fixed in 4% PFA for 6-18h and Formalin Fixed Paraffin Embedded (FFPE) block preparation was performed according to the histology unit guidelines specific to each institute. FPPE blocks were sectioned on a microtome at 5µm thickness and deposited onto SuperFrost Plus™ Glass slides. Immediately after sectioning, slides were baked in a laboratory oven at 60°C overnight to promote tissue adhesion. This step is crucial as ensuring the adhesion of tissue sections to slides enables the extensive rounds of staining. After baking, slides could be kept at room temperature for storage.

From here, slides were baked once more overnight at 60°C. The slides were removed from the oven, and deparaffinised through plunging into xylene solutions twice for 5min, followed by decreasing concentrations of ethanol (100%, 95%, 70%, 50%) twice for 5min each. Slides were then plunged into 0.3% Triton X-100 in 1xPBS for 10min for tissue permeabilization. Slides were then washed twice for 5min in PBS.

Tissue antigen retrieval was performed in a NxGen Decloaking Chamber, set up to heat up to 110°C and maintain this temperature for 4min. When the chamber temperature reached 70°C, slides were placed in a pH=6 citrate-based solution (ARS1) for 20min, with the temperature rising to 110°C and pressure building to 6psi during that time. After 20min, pressure was released and slides were moved to a pH=9 Tris-based solution (ARS2) for 20min. After 20min, slides were allowed to cool down on the bench while remaining in the Tris-based solution (ARS2). Once cool, slides were washed 4 times for 5min in 1xPBS. Finally,

slides were left in a solution of 10% donkey serum, 3%BSA in 1xPBS overnight at 4°C to block unspecific antibody binding sites. Slides were then plunged into a DAPI solution for 15min, and washed for 5min in 1xPBS. Finally, 75µL of mounting media (50% glycerol and 4% propyl gallate in 1xPBS) was added to each slide, before being coverslipped and stored at 4°C.

Slides were then ready to be imaged. Background imaging of the slides on the CellDIVE was performed within a couple days of tissue processing. After background imaging, slides were decoverslipped by being angled face down in a container filled with 1xPBS for 15min to 1h. Slides were then carefully dried with Kimwipes and placed into a staining dish. Slides were Fragment crystallizable (Fc)-blocked using Fc Receptor Blocking Reagent from Miltenyi at 1:200 in antibody diluent (3% BSA in 1xPBS), adding 100µL of the solution to each slide for 1h at room temperature. After Fc-blocking, slides could be stained overnight with any of individual stain rounds described in Tables 2.6 and 2.7**,** diluting the antibodies in 100uL of antibody diluent per slide. Staining rounds were acquired on the CellDIVE system, before decoverslipping once more, and proceeding with fluorophore bleaching. For fluorophore bleaching, slides were plunged into dye inactivation solution twice for 15min, before being washed 3 times in 1xPBS for 5min each. Slides could then be coverslipped again, and imaged. To stain for the full panels described in Tables 2.6 and 2.7, slides were cycled through the staining and bleaching rounds described above. Isotype controls described in Table 2.8 were stained and are reported for completeness in Figure 2.1.

Analyses of CellDIVE images were performed in QuPath[200]. All the pipelines used for analysis are described and reported in a published methodological paper, included for completeness in the appendix of this thesis.

Table 2.5 Patient clinical information. Patient metadata available for each of the biopsy section stained (Tissue ID) via multiplexed immunofluorescence.

Diagnosis	Tissue ID	Age	Sex	Disease duration (years)	Treatment
RA	JRP091_3	70	F	35	Hydroxychloroquine/Plaquenil and Naproxen
RA	JRP099_3	31	F	21	Pednisolone and Naproxen
RA	JRP112_6	35	F	17	Sulfasalazine/Salazopyrine/ Prednisolone/NSAID
RA	JRP118_2	66	M	20	Methotrexate/NSAID
RA	JRP122_6	60	M	/	Methotrexate, Salazopyrine, Golimumab/Simponi and Prednisolone when needed
RA	JRP129_2	72	F	44	Rituximab/Mabthera
RA	JRP136_6	56	F	36	Prednisolone
RA	JRP139d_6	67	M	7	Prednisolone and Tocilizumab (stopped 2 weeks prior to biopsy)
RA	JRP141_3	58	F	/	/
OA	JRP144_3	74	F	9	/
OA	S00293087_3	70	F	/	/
OA	97SM_15	64	M	/	/
OA	47SM_14	67	M	/	/
OA	83SM_14	72	M	/	/
OA	S00292623_3	69	F	/	/
OA	S00292477_3	88	F	/	/

Table 2.6 CellDIVE multiplexed immunofluorescence exploratory panel. FcR = Fc-receptor, NA = Not applicable.

	Target	Clone	Catalog #	Fluorophore	Concentration
	FcR	NA	130-059-901	NA	1:200 dilution
Round 1	PDPN	NZ-1.3	14-9381-82	AF488	10ug/mL
	IGFBP5	Polyclonal	PA5-37369	AF555	10ug/mL
	IGFBP7	Polyclonal	AF1334	AF647	2ug/mL
Round 2	NaKATPase	EP1845Y	ab197713	AF488	10ug/mL
	RS6	C-8	sc-74459	AF546	5ug/mL
	Cadherin	EPR1792Y	ab195203	AF647	5ug/mL
Round 3	Vimentin	EPR3776	ab185030	AF488	5ug/mL
	CD3	SP162	ab245731	AF555	5ug/mL
	CD19	EPR5906	ab196468	AF647	10ug/mL
Round 4	MERTK	Y323	ab52968	CF450	10ug/mL
	CD68	EPR20545	ab277276	AF555	10ug/mL
	COL6A1	EPR17072	ab200430	AF647	2ug/mL
Round 5	CD34	EP373Y	ab81289	AF488	10ug/mL
	CD90	EPR3132	ab181885	AF555	5ug/mL
	COL4A1	IV-4H12	51-9871-82	AF647	2ug/mL

Table 2.7 CellDIVE multiplexed immunofluorescence validation panel. FcR = Fc-receptor, NA = Not applicable.

	Target	Clone	Catalog #	Fluorophore	Concentration
	FcR	NA	130-059-901	NA	1:200 dilution
Round 1	MMP2	Polyclonal	HPA001939	AF488	2ug/mL
	CHP	NA	5276-60UG	AF555	1:2
	CD66b	G10F5	305110	AF647	10ug/mL
Round 2	COL3A1	EPR17673	ab237238	AF488	10ug/mL
	MMP9	2C3	sc-21733	AF546	5ug/mL
	CD14	ab133335	EPR3653	AF647	10ug/mL
Round 3	COL1A1	E8F4L	28368S	AF488	0.175ug/mL
	COL12A1	A-11	sc-166020	AF546	5ug/mL
	COL6A1	EPR17072	ab200430	AF647	2.5ug/mL
Round 4	CD3	SP162	ab245731	AF488	10ug/mL
	Fibronectin	F1	ab275110	AF555	5ug/mL
	CD20	EP459Y	ab198943	AF647	5ug/mL
Round 5	MERTK	Y323	ab271851	CF450	5ug/mL
	CD4	ab280849	EPR6865	AF555	5ug/mL
	COL4A1	IV-4H12	51-9871-82	AF647	2ug/mL
Round 6	CD45	RM1007	ab282747	CF450	10ug/mL
	CD68	EPR20545	ab280860	AF555	5ug/mL
	CD8	C8/144B	372902	AF647	10ug/mL
Round 7	hIgG	EPR4421	ab226069	AF488	5ug/mL
	CD11c	EP1347Y	ab279329	AF555	5ug/mL
	MZBI	22	NBP2-90320	AF647	5ug/mL
Round 8	MMP2	EPR1184	ab237473	AF488	10ug/mL
	CD11b	EPR1344	ab133357	AF555	5ug/mL
	LYVE1	EPR21857	ab219556	AF647	10ug/mL
Round 9	COL1A1	EPR7785	ab138492	AF488	5ug/mL
	MFAP5	EPR17581	ab240367	AF555	5ug/mL
	MMP1	EP1247Y	ab196905	AF647	5ug/mL

Table 2.8 CellDIVE panel isotype controls and secondary antibodies.
NA = Not applicable.

Isotype/Secondary	Clone	Catalog #	Fluorophore	Concentration/Dilution
Rabbit IgG	Polyclonal	ab37415	NA	10ug/mL
Goat IgG	Polyclonal	ab37373	NA	10ug/mL
Mouse IgG1κ	MG1-45	ab18448	NA	10ug/mL
Mouse IgG2bκ	MPC-11	400329	AF488	10ug/mL
Donkey anti-Mouse IgG1κ	Polyclonal	A-21202	AF488	1:400 dilution
Donkey anti-Rabbit IgG	Polyclonal	A-31572	AF555	1:400 dilution
Donkey anti-Goat IgG	Polyclonal	A-21447	AF647	1:400 dilution

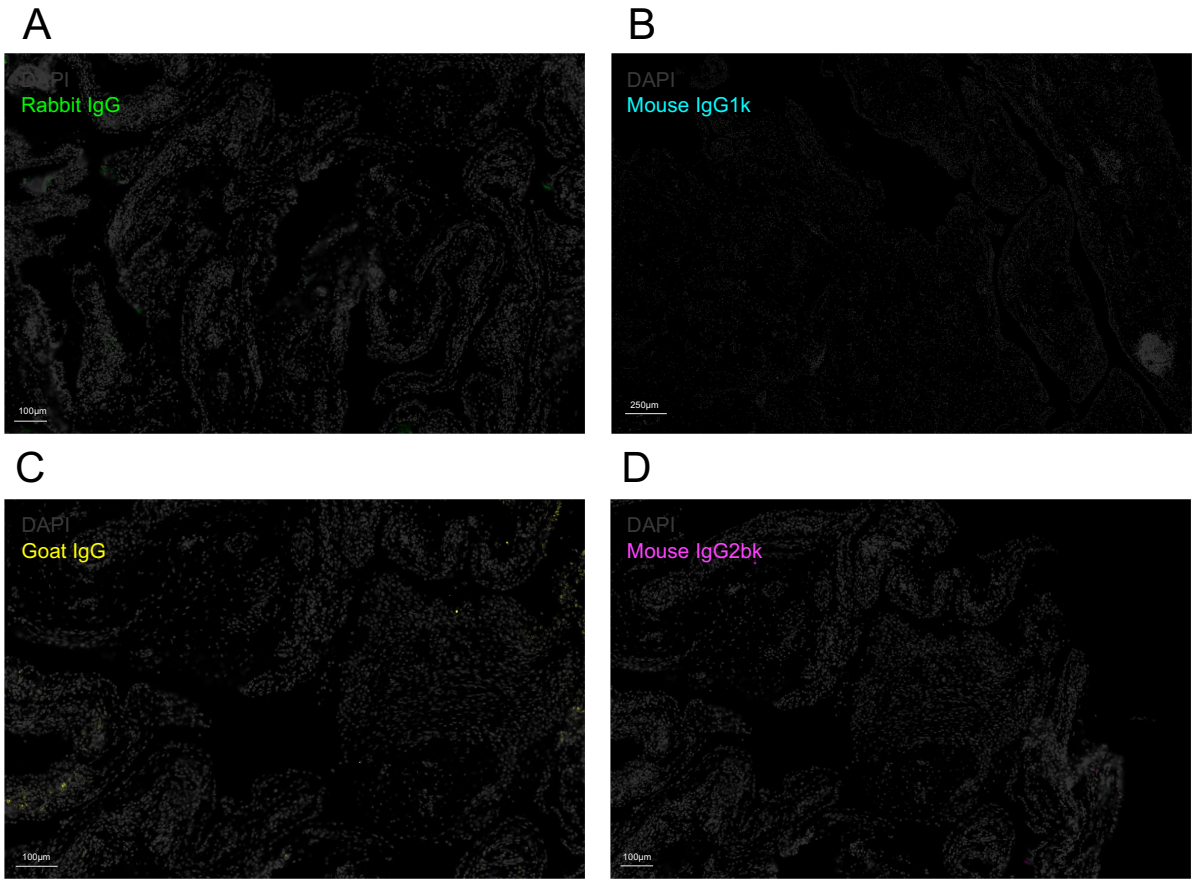


Figure 2.1 CellDIVE immunofluorescent stains for isotype controls. Isotype controls (reported in Table 2.8) for each of the isotypes for antibodies in Tables 2.6 and 2.7. **A.** Isotype control stain for Rabbit IgG (ab37415) at 10µg/mL, and Donkey anti-Rabbit IgG-AF555 (A-31572) at 1:400 dilution **B.** Isotype control stain for Mouse IgG1k (ab18448) at 10µg/mL, and Donkey anti-Mouse IgG1-AF488 (A-21202) at 1:400 dilution **C.** Isotype control stain for Goat IgG (ab37373) at 10µg/mL, and Donkey anti-Goat IgG-AF647 (A-21447) at 1:400 dilution **D.** Isotype control stain for Mouse IgG2bk-AF488 (400329) at 10µg/mL.

2.3.2 Confocal immunofluorescence – mouse tissue

Mouse knees from the Antigen-Induced Arthritis (AIA) model were stained in this study[74]. Mouse knees were excised and fixed in PLP buffer at 4°C overnight. After this, the knees were washed with 1xPBS and decalcified in 10mL of 0.5M EDTA (pH<8) for 6 days at 4°C, with the solution being exchanged twice. Knees were then placed in 10mL of 30% sucrose for 24h at 4°C, embedded in OCT, and kept in a -80°C freezer for long-term storage. 10µm-thick sections were cut on the Leica CM3050 cryostat at -20°C, and slide-bound sections were placed in a -20°C freezer. For immunofluorescent staining, slides were removed from the freezer and placed vertically in a staining dish for 1h to air dry, followed by being placed in an oven at 65°C for 20min. Once dried and cooled, the sections were circled with a PAP pen to create a hydrophobic barrier around the tissue sections. Subsequently, the slides were rehydrated with 1xPBS for 5 minutes before permeabilization with 400 µl of PBT (PBS, 0.3% Triton-X) for 5 minutes per slide. Blocking of nonspecific binding sites was achieved by adding 400µL of muIF blocking solution to each slide and incubating at room temperature for 1h in a humidified chamber. The slides were then stained by adding 400µL of antibodies, listed in Table 2.9 diluted in muIF staining solution to each slide and placed them overnight at 4°C in a humidified chamber. The slides were then washed again 3 times in 1xPBS for 5min each, before adding 400µL of 2µM SyTox Blue dye in PBT per slide and incubating them for 20min at room temperature. Slides were finally rinsed with dH_2O, dried with Kimwipes, and coverslipped by adding a drop of FluorSave mounting medium per section on the slide. Slides were lastly left to dry in the dark at room temperature overnight. Slides were then imaged using the ZEISS 980 Confocal microscope.

Table 2.9 Mouse confocal immunofluorescence panel

Target	Clone	Catalog #	Fluorophore	Dilution
CD3	17A2	100228	BV421	1:200
CD68	FA-11	137012	AF488	1:200
CD177	1171A	MAB8186	AF555	1:200
COL6A1	EPR17072	ab200430	AF647	1:200
CD31	MEC13.3	102520	AF594	1:200

2.4 Protein methods

2.4.1 Protein purification

Col6-D11-c001 (wildtype) and c002 (mutant) were expressed in Drosophila cells by Dr. Eleanor Williams, Laurent Brooke, and Korvus Wang at the NDMRB, based on the work of past Midwood group member Dr. Jesse Coker. Drosophila cell secreted proteins were harvested by centrifuging the culture at 4000g for 20mins, with the supernatant being collected. 0.1xvolumes of 10xBuffer W and 18.1mL of BioLock solution were first added per litre of supernatant and incubated for 1h at room temperature. During this incubation, the top cap of the Strep-Tactin XT column of choice was removed, before twisting the lower cap and removing the storage buffer and equilibrating it with 2x column volumes (CV) of Buffer W. After 1h, the supernatant was passed through a 0.45µM low-protein binding filter, added to the Strep-Tactin XT column of choice, and flow-through was collected throughout the passing of the supernatant. Once all supernatant had been passed through, the column was washed with 5xCV of Buffer W. 0.5xCV of Buffer BXT was then added and eluate fractions were collected. Eluate fractions were immediately nanodropped to assess protein concentration of fractions using the protein's known extinction coefficient ($12490 M^{-1} cm^{-1}$) and mass (25.323kDa), and placed in a -80°C freezer for long-term storage.

2.4.2 Endotoxin testing

All reagents from the Lonza endotoxin testing kit (50-658NV) were first allowed to equilibrate to room temperature before mixing. Meanwhile, the supplied endotoxin solution was vortexed vigorously for 15min. Endotoxin standards were then prepared according to the manufacturer's instructions, ranging from 5.0 Endotoxin Units (EU)/mL down to 0.005 EU/mL. Protein samples to be tested were prepared at a concentration of 1µM in Buffer W. 100µL of the blank (supplied LAL Reagent Water), endotoxin standards, and samples were then added to a 96 well plate, and pre-incubated at 37°C for 10min. During this incubation period, the working solution of supplied fluorogenic substrate, assay buffer, and rFC enzyme solution was prepared at a respective ratio of 5:4:1. 100µL of the working solution was then added to each well of the plate, and fluorescence (excitation at 355nm and emission at 460nm) was read at time zero with the Fluostar Omega Plate Reader. The fluorescence was then read again after 1h of incubation at 37°C in the plate reader.

2.4.3 Polyacrylamide Gel Electrophoresis (PAGE)

Recombinant proteins were loaded onto NuPAGE 4-12% gradient pre-cast gradient gels and run in an AE-6050 mini gel tank containing 1X NuPAGE 2-ethanesulfonic acid (MES) buffer. Recombinant proteins were diluted in the assay-relevant volume of 6x Laemelli buffer and incubated for 5min at 90°C. 20µL of the boiled samples were loaded into relevant wells, as well as 5µL of protein ladder. Gels were run at 20mA per gel running for approximately 2h, or until the blue running front reached the bottom of the gel without running off.

2.4.4 Native PAGE

Anode and Cathode buffers were prepared beforehand and stored for 30min at -20°C to be ice cold. Wells of a NativePAGE 4-12% gradient pre-cast gradient gel were then washed 3x with Cathode buffer and filled with Cathode buffer before loading samples and protein ladder similarly to a standard PAGE. The inner chamber of an AE-6050 mini gel tank was then filled with Cathode buffer, while the outer chamber was filled with Anode buffer. The gel was

then run at 150V for roughly 1h, or until the blue running front reached the bottom of the gel, at 4°C.

2.4.5 Silver staining

After PAGE or Native PAGE, the gels were placed in clean plastic staining trays with 100mL of Fixative for 2h with shaking. Fixative was decanted, and replaced with 100mL of fresh Fixative, left overnight. A third round of Fixative was added the next day, for 2h with shaking. Fixative was decanted the gels were wash with 30% ethanol for 10min, before decanting and adding 100mL of Sensitizing solution for 10min. The Sensitiser was decanted, and the gel was washed again with 100mL of 30% ethanol for 10min, before being washed with dH_2O for 10min. The gels were then incubated with 100mL of staining solution for 15min. After staining, the staining solution was decanted, and the gels were washed for no more than 60s with dH_2O. The gels were then incubated with developer for 4-8min or until the adequate band intensity was reached. Following this, 10mL of Stopper solution were immediately added to stop the reaction. This was decanted and the gels were subjected to one final dH_2O wash.

2.5 In vitro cell-based methods

2.5.1 Human Peripheral Blood Mononuclear Cell (PBMC) and Monocyte isolation

PBMCs were isolated from leukocyte cones provided by the NHS Blood Transfusion (NHSBT) department at the John Radcliffe Hospital in Oxford. Cones were washed and diluted with 1xHBSS, before being layered carefully on Histopaque to create a density gradient through centrifugation at 300g. From this, the layer of PBMCs was collected, washed to remove platelets, and put through erythrocyte lysis using ACK Lysis Buffer to recover PBMCs exclusively. When monocytes were required, EasySep^{TM} Human Monocyte Isolation Kits were used according to the manufacturer's enclosed instructions to isolate specifically monocytes. These monocytes could then be used fresh or stored in liquid nitrogen at $20x10^6$ cells/mL.

2.5.2 Human macrophage differentiation and polarisation

On Day 1 monocytes were resuspended in macrophage differentiation media (standard culture media with the addition of 100ng/mL of M-CSF), and were added to a non-

Tissue Culture-treated petri dish. Cells were then fed on the third and fifth day with an additional 5mL of macrophage differentiation media. On Day 7 of the differentiation, 12-well plates were coated with 400µL/well of either 1xPBS or ECM substrates at 20µg/mL, before placing the plates in the incubator for 2h. ECM substrates tested in this assay included human native collagen VI (ab7538), gel formulation human collagen VI (#354261), Matrigel (#356234), rat tail collagen I (#60-30-810). After 2h, the liquid phase of the solutions was removed from the plates, which were washed once with 1mL of 1xPBS/well, and once more with 1mL of standard culture media. Once the plate had been prepared, differentiated macrophages were removed from their petri dishes with either cold 1xPBS and gently brushing the surface with a cell lifter, or incubated with 5mL of TryplE for 5min. The lifted cells were then spun down and resuspended at $0.7x10^6$ cells/mL in standard media. From this suspension, 1mL of cells were added to each well of the coated 12-well plate, and the plate was incubated for 24h at 37°C. 10nM of LPS (ALX-581-010-L002) or 100nM of Dexamethasone (DEX) (D4902-25MG) were added with cells for negative and positive controls respectively. After 24h, the cells could be lifted with a combination of cold 1xPBS, a short 3min incubation in a 4°C fridge, and gentle lifting using a cell lifter.

Cells from each tested condition were prepared for Flow Cytometry by first adding them to a conical bottom 96-well plate. A fraction of cells from each condition was pooled in parallel to create unstained, Live/Dead (L/D), Fluorescence Minus One (FMO), and isotype controls. Cells were spun down at 500g for 5min, L/D stained at 1:1000 in 1xPBS for 15min, spun down again at 500g for 5min and washed in 1xPBS twice. Cells were then Fc-blocked for 10min at 1:200 in 1xPBS to prevent unspecific antibody binding, before being spun down at 500g for 5min, and stained with relevant antibody dilutions (Table 2.10) for 30min. Subsequently, cells were spun down at 500g for 5min and washed with 1xPBS twice, before

being fixed for 15min in BD Fix Buffer, spun at 500g for 5min and washed again, and resuspended in FACS Buffer, ready to run on a BD LSR Fortessa.

Table 2.10 Macrophage MerTK+ polarisation flow cytometry panel. NA = Not applicable

Target	Clone	Catalog #	Fluorophore	Dilution
Human TruStain FcX™	NA	422302	NA	1:200
MERTK	590H11G1E3	367612	APC	1:200
CD68	FA-11	137010	PerCP/Cy5.5	1:200

2.5.3 Adhesion assay

THP1 monocytic cells, Jurkat E6-1 T-cells, and Daudi B-lymphoblasts were cultured continuously in standard culture media (RPMI-1640 (+ L-Glutamine), 10% FBS, 1% P/S). Each cell line was split every 2 days and resuspended at a cell concentration ranging from 100k/mL to 500k/mL depending on their frequency of use, into T75 Falcon flasks, and kept at 37°C and 5% CO_2.

On day 1, cultured cell lines were spun down, resuspended at $1x10^6$ cells/mL in starvation media, and seeded into a fresh T75 flask. In parallel, a 96-well plate was coated with 100µL/well of assay-specific peptides or proteins in PBS-Tween, and left at 4°C overnight. Peptides and proteins used as coating in this assay include human native collagen VI (ab7538), human fibronectin (ab80021), purified recombinant CD11b I-domain intact or boiled at 90°C for 5min (generously donated by Dr. Anja Schwenzer), purified recombinant D11 (wildtype and mutant, in house), or purified peptide 9 of the fibrinogen domain of TNC (generously donated by Dr. Anja Schwenzer), all used at concentrations ranging from 0.05 to 40µg/mL. On day 2, the plate was washed 3 times with 1xPBS, and blocked with 200µL/well of 1% BSA in 1xPBS for 3h at 37°C. After 3h, the plate was washed with 200µL/well of 1xPBS, 50µL/well of cells resuspended at $2x10^6$ cells/mL in 1xHBSS were added to the plate, and these were incubated for 2h at 37°C. Adhesion inhibitors were added at this stage in the cell suspension and included α4β1 integrin inhibitor BIO5192 (HY-107589) and EDTA (33312) at

molarities ranging from 10 to 40nM and 0 to 10mM respectively. Non adherent cells were removed from the plate by flicking, and were gently washed out with 200µL of 1xPBS. Cells were finally resuspended in 100µL/well of 1xPBS, and well confluency was read on the Sartorius Incucyte S5.

2.5.4 Transwell migration assay

Transwell migration assays were run by Zhi Yi Wong. Corning 5µm pore transwell inserts were either left untreated or coated for 2h at 37°C with human Fibronectin (ab80021), human native collagen VI (ab7538), or fibronectin and collagen VI mixed at differing ratios. During this incubation, human PBMCs were isolated as described in section 2.5.2. The inserts were then washed twice with chemoattractant-free migration media (RPMI, 10% FBS, 1%PenStrep). 600µL of migration media with 100ng/mL of CXCL12 was added to the lower chamber of a 24-well plate. The coated and washed inserts were then added to the wells, and $1x10^6$ cells in 100µL were added in chemoattractant-free migration media to the upper chamber. As a negative control, 1µg of pertussis toxin (PTX) was added to the lower chamber. The 24-well plate was then left to incubate for 4h at 37°C, after which the cell suspensions in the bottom chambers were collected and plated into at 96-well conical bottom plate. A fraction of cells from each condition was pooled in parallel to create unstained, Live/Dead (L/D), Fluorescence Minus One (FMO), and isotype controls. Cells were spun down at 500g for 5min, L/D stained at 1:1000 in 1xPBS for 15min, spun down again at 500g for 5min and washed in 1xPBS twice. Cells were then Fc-blocked for 10min at 1:200 in 1xPBS to prevent unspecific antibody binding, before being spun down at 500g for 5min, and stained with relevant antibody dilutions (Table 2.11) for 30min. Subsequently, cells were spun down at 500g for 5min and washed with 1xPBS twice, before being fixed for 15min in BD Fix Buffer, spun at 500g for 5min and washed again, and resuspended in FACS Buffer, ready to run on a BD LSR Fortessa.

Table 2.11 Transwell migration assay flow cytometry panel. NA = Not applicable.

Target	Clone	Catalog #	Fluorophore	Dilution
Live/Dead	NA	NA	APC-Cy7	1:1000
CD45	HI30	304048	BV785	1:100
CD14	M5E2	301842	BV510	1:100
CD19	HIB19	302219	AF488	1:100
CD3	UCHT1	300434	BV421	1:100
CD4	OKT4	317436	BV650	1:100
CD8	SK1	344712	PE-Cy7	1:100

2.5.5 Ibidi channel migration assay

Similarly to adhesion assays, on day 1 cultured cell lines were spun down, resuspended at $1x10^6$ cells/mL in starvation media, and seeded into a fresh T75 flask. In parallel, Ibidi µ-Slide VI were coated with 30µL/channel of human native collagen VI (ab7538) and/or human fibronectin (ab80021) at 20µg/mL in 1xPBS and left at 37°C overnight. On Day 2, the channels were washed twice with 60µL/channel of 1xPBS. Cells were resuspended at 166.67k cells/mL in starvation media, 60µL of the suspension was added to each of the channels, and the µ-Slide was incubated at 37°C for 2h. 60µL of standard media (containing FBS) could then be added to the end well of each channel, and the µ-Slide could be incubated for 24h. Standard phase microscope images were then taken of the start and end wells of the slide at 0h and 24h.

2.6 Experimental statistics

Bioinformatic analysis and statistical testing was carried out in R version 4.0.3. Unless otherwise stated, statistical analyses surrounding experimental procedures were performed using GraphPad PRISM 9. One- or two-way ANOVA with Tukey's test for multiple comparisons or unpaired student's t test were used as indicated in figure legends. Statistical significance was set at $p < 0.05$.

Chapter 3 | Bioinformatic analysis of matrisome gene expression in RA

3.1 Chapter introduction

Tissue remodelling and degradation are key features of both OA and RA pathology. OA is a mechanical load driven joint destruction pathology. It is mainly associated with extensive cartilage degradation but also some degree of synovial tissue remodelling[201]. RA is primarily characterised by extensive synovial tissue remodelling and inflammation, with significant tissue hyperplasia and fibroblast proliferation, but also involves some degree of cartilage degradation through the formation of a fibroblast invasive front or *pannus*[83]. Individual studies have reported specific dysregulations of individual ECM genes and proteins, and studied their involvement in the overarching aetiology and pathology of both OA and RA. Examples include but are not limited to the citrullination of aggrecan[202], enolase[203], and TNC[204] in RA, and the *MMP13* overexpression-mediated excessive degradation of cartilage collagen II in OA[205].

While the cellular basis of RA synovial pathology has been extensively characterised[63–65,67,109,206,207], unbiased characterisations of synovial ECM composition currently do not exist. The accumulating evidence for a role of the ECM in mediating RA pathology, including genetic predispositions from ECM protease polymorphisms[208,209], supports the importance and urgency surrounding generating such a resource. Matrisome-focused bioinformatic analysis of published RA scRNAseq datasets was performed to address this knowledge gap and build a transcriptomic blueprint of synovial ECM organisation. The matrisome gene categories and their breakdown, along with the datasets chosen for this analysis are summarised in Tables 3.1 and 3.2 respectively. The initial aims of this analysis were firstly to characterise the subsynovial ECM niches defined by specific cell subsets, secondly to define disease type- and disease activity-specific ECM expression profiles, and thirdly to gain insights into disease-relevant dysregulated ECM pathways for further protein-level and functional investigations.

From this analysis, fibroblasts and myeloid cells were found to be the predominant regulators of ECM composition in the RA synovium. Spatially and functionally distinct fibroblast subsets were defined by specific matrix expression signatures. RA and OA were found to have highly distinct disease-specific matrix signatures. Samples from patients with

active RA or RA in remission were also associated with distinct ECM composition. Finally, IGF1 signalling modulation and basement membrane dysregulation were identified as key targets for further exploration.

Table 3.1 Gene subcategory breakdown of the matrisome. The matrisome is a list of 1027 genes split between Core Matrisome and Matrisome-associated genes curated by Naba and colleagues[17]. The list and its categorisation are based on experimental validation and bioinformatic prediction of matrix-involvement through protein domain structural analysis.

Division (*number of genes*)	Category (*number of genes*)
Core matrisome (*274*)	Glycoproteins (*195*)
	Collagens (*44*)
	Proteoglycans (*35*)
Matrisome-associated (*753*)	ECM-affiliated (*171*)
	ECM regulators (*238*)
	Secreted factors (*344*)

Table 3.2 Published scRNAseq datasets used in the study. scRNAseq = single cell RNA sequencing, OA = Osteoarthritis, RA = Rheumatoid Arthritis, STIA = Serum Transfer-Induced Arthritis, FS = Frozen Shoulder.

Species	Type	Tissue/Cells	Disease/Model	Source	Accession no.	In-text name
Human	scRNAseq	Sorted T-cells/B-cells/Macrophages/Fibroblasts from synovial biopsy	OA and RA patients	Zhang et al. (2019)[63]	SDY998	AMP1
Human	scRNAseq	Unsorted cells from synovial biopsy	Active RA and RA in remission patients	Alivernini et al. (2020)[67]	E-MTAB-8322	Alivernini
Mouse	scRNAseq	Unsorted synovial cells from synovial dissections	STIA	Wei et al. (2020)[64]	GSE145286	Wei
Human	scRNAseq	Unsorted cells from shoulder capsule	FS and comparator patients	Ng et al. (2024)[185]	ERP143358	FS

3.2 Results

3.2.1 Matrisomal gene set enrichment in synovial cell subsets

To build a transcriptomic blueprint of synovial ECM composition, I first set out to identify cell subsets with specific enrichment for matrisome gene expression. This would allow us to capture the breadth of cellular contributions to ECM composition, while also enabling downstream analyses to be focused on the cell subsets most relevant to the shaping of the tissular landscape.

Gene set enrichment analysis with the *AUCell* package was therefore performed on the AMP1 dataset[63,186]. The AMP1 dataset was one of the most thorough scRNAseq datasets at the time of this analysis. The dataset is composed of sorted synovial T-cells, B-cells, macrophages, and fibroblasts from 18 RA and 3 OA synovial biopsies (Figure 3.1 A&B). Using the matrisome gene list (Table 3.1) and the cells of the AMP1 dataset, the R package *AUCell* was used to score the expression levels of each subcategory of matrisome genes for each cell in the dataset, based on individual cell-specific gene expression rankings. The output of this analysis was an AUC score, or enrichment score for the tested gene set of interest, for each of the categories of the matrisome in each of the individual cells of the dataset (Figure 3.1 C-D, Supplementary Figure 1 A-D). Interestingly, enrichment scores were not found to differ significantly between the cells from RA and OA samples. Clear differences were however observed between cell subsets. Fibroblasts were found to be the predominant expressors of core matrisome genes, broken down into glycoproteins, proteoglycans, and collagens, with the LL (Lining Layer) SC-F4 fibroblasts being comparatively less enriched in their expression compared to the SL (Sublining) SC-F1/2/3 (Figure 3.1 C, Supplementary Figure 1 A&B). Matrisome-associated genes, broken down into ECM-affiliated, regulators, and secreted factors, appeared to be enriched in both fibroblast and macrophage subsets (Figure 3.1 D, Supplementary Figure 1 C&D). T-cell and B-cell all displayed very low enrichment scores for matrisome gene expression, with T-cells being marginally more highly enriched for proteoglycans and secreted factors.

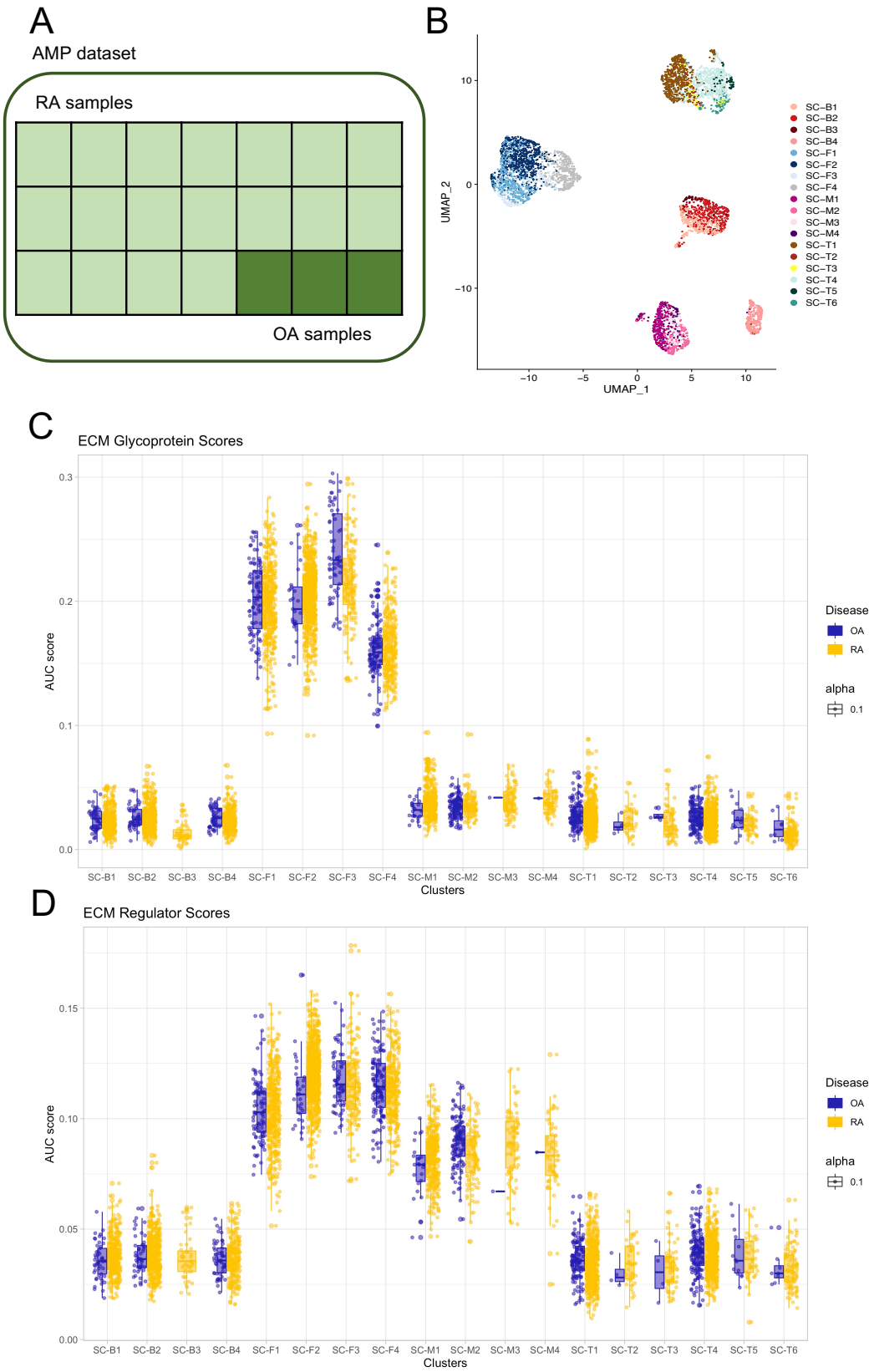


Figure 3.1 Fibroblasts and Macrophages are enriched for matrisome expression in RA and OA. Normalised single cell sequencing counts and their metadata were downloaded from ImmPort under accession SDY998. The Seurat V4 pipeline was then used to map cells to their annotations. Clustered cells and the annotated matrisome gene list were then run through the AUCell pipeline (available through vignette("AUCell") in R) to obtain enrichment scores (AUC scores) for each of the matrisome gene subcategories. **A.** Schematic breakdown of the composition of the AMP1 dataset, which each tile representing one patient biopsy sample. **B.** UMAP of the AMP1 dataset synovial cells, coloured by the clusters reported in the original study. SC-B stands for B-cells, SC-T for T-cells, SC-F for fibroblasts, and SC-M for macrophages. **C&D.** Boxplots of the AUC expression enrichment scores (Y axis) for each AMP1 cell cluster (X axis) and for matrisome subcategories: Glycoproteins (C), representative of core matrisome genes reported in Supplementary Figure 1; ECM regulators (D), representative of matrisome-associated genes. Each point on the boxplot represents an individual cell of the dataset, split between OA (blue) and RA (yellow).

Together, these data highlight the complementary roles for fibroblasts and macrophages in defining the composition of the complete synovial tissue. These also justify a specific focus on fibroblasts and myeloid cells to look at cluster-specific matrix expression profiles with increased resolution.

3.2.2 Fibroblast and macrophage cluster-defining matrix markers

The fibroblast subsets of the AMP1 dataset (Figure 3.2 A&B) are a particularly useful analytical tool in the creation of a transcriptomic blueprint, as their respective functions and distribution have been extensively characterised[62–64,83,210]. To start deciphering the specific subsynovial niches residing in the synovial tissue in RA, I therefore performed differential expression analysis (DEA) between the fibroblast subsets of the AMP1 dataset, with a specific focus on matrisome genes, using the *Seurat* V4 standard Wilcoxon Rank Sum Test[187]. Strikingly, all four fibroblast subsets in the dataset were found to have highly specific matrix-expression signatures.

The *THY1*$^{+}$ *HLA-DR*$^{+}$ SC-F2, classically thought to be immune function associated perivascular fibroblasts, were indeed characterized predominantly by inflammatory secreted factors such as *IL6, CCL2, MDK,* or *TNFSF13B* (Figure 3.2 C&D). The *THY1*$^{+}$ *CD34*$^{+}$ SC-F1 and *DKK3*$^{+}$ SC-F3 fibroblasts were found to have more similar signatures compared to the other subsets but did harbour some key defining differences. The SC-F3 were particularly enriched in *COL8A2* expression for example, while the SC-F1 had greater expression of *SFRP2, GAS6, or IGF1* (Figure 3.2 C&D), critical modulators of the Wnt, AXL, and IGF signalling pathways respectively. The *THY1*$^{-}$ *PRG4*$^{+}$ *CD55*$^{+}$ SC-F4 LL fibroblasts also displayed a unique signature, defined by restricted specific expression of *HBEGF,* and the two Semaphorins *SEMA3A* and *SEMA5A* (Figure 3.2 C&D). Interestingly *NTN4*, encoding the basement membrane protein Netrin-4, was also found to be specific to the SC-F4 lining fibroblasts and not expressed in the SC-F2 perivascular-associated fibroblasts of the sublining. This peculiar distribution of *NTN4* expression prompted us to look specifically at differential distribution of the basement membrane genes found in the lists of cluster-defining

matrix markers for the SC-F2 and SC-F4. This enabled the identification of key differences in the blueprints of the vascular basement membrane and the basement-like membrane of the synovial LL (Figure 3.2 E). Indeed, in addition to the differences in *NTN4* expression, the basement-like membrane-associated SC-F4 appeared to express much lower levels of the nidogens (*NID1* and *NID2*), key structural components of fully functional canonical basement membranes[211]. Key differences in laminin composition also appeared clearly, with *LAMB1* being predominantly vascular-associated, and *LAMB3* and *LAMA5* predominantly LL-associated.

To complement this investigation of subsynovial matrix expression niches, I repeated this analysis structure, focusing on the macrophages of the AMP1 dataset (Figure 3.3 A&B). Interestingly, highly distinct matrisome expression profiles could also be identified in macrophages. *IL1B*$^{+}$ SC-M1 macrophages were found to express large levels of secreted inflammatory and angiogenic factors such as *CCL2, CXCL2, CXCL3, IL1B,* and *VEGFA.* *NUPR1*$^{+}$ SC-M2 macrophages appeared to be more involved in ECM regulation, with the strong and specific expression of proteases *CTSK* and *HTRA1*, and protease inhibitor *TIMP2*. The expression profiles of the SC-M3 and SC-M4 were harder to dissociate from others. The SC-M3 appeared to be defined by a predominant expression of the classical complement genes *C1QA, C1QB,* and *C1QC*. The SC-M4 were characterised by the strongest expression of the *S100A8, S100A10,* and *S100A11* genes, as well as higher relative expression of key ECM structural components *FN1* and *VCAN* (Figure 3.3 C&D).

Together these data corroborate the *AUCell* matrisomal enrichment findings. At the superficial level of simple DEA between fibroblast clusters and macrophage clusters, fibroblasts appear to have clearly defined matrisomal profiles, including a number of key structural ECM genes and proteins. Macrophages on the other hand appeared to be defined more by their expression of specific subsets of secreted factors, but also some regulators of ECM turnover. In line with these observations, downstream bioinformatic characterisation of a transcriptomic tissue architecture blueprint was focused primarily on fibroblasts.

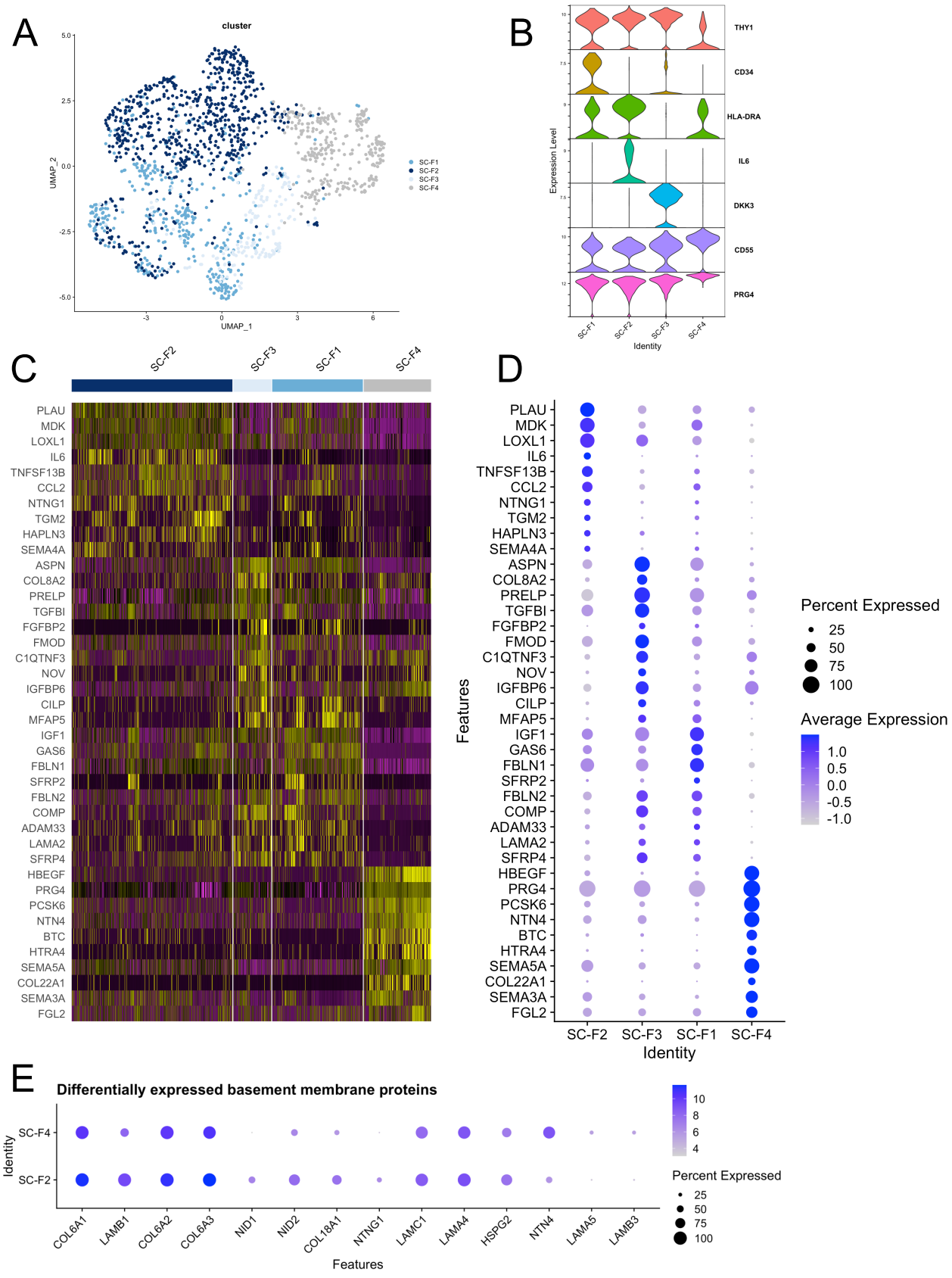


Figure 3.3.2 Distinct fibroblast subsets are characterised by distinct matrix expression profiles. Differential expression analysis was run between fibroblast subsets from the AMP1 dataset using a standard Wilcoxon Rank Sum test in Seurat V4. The output list of cluster-defining genes was then filtered for genes present in the matrisome to define fibroblast cluster-specific matrix expression profiles. **A.** UMAP of the AMP1 dataset fibroblast clusters. **B.** Violin plot of the canonical genes used to define the 4 fibroblast subsets in the original AMP1 study, namely the *THY1+ CD34+* SC-F1, *THY1+ HLA-DR+ IL6+* SC-F2, *THY1+ DKK3+* SC-F3, and *THY1- PRG4+ CD55+* SC-F4. **C.** Heatmap of the top 10 cluster-defining matrisome genes for the 4 AMP1 fibroblast subsets. Yellow corresponds to high scaled expression, and purple to low scaled expression. **D.** Dot plot of top 10 cluster-defining matrisome genes for the 4 AMP1 fibroblast subsets. **E.** Dot plot of basement membrane genes found to define clusters SC-F2 or SC-F4 specifically.

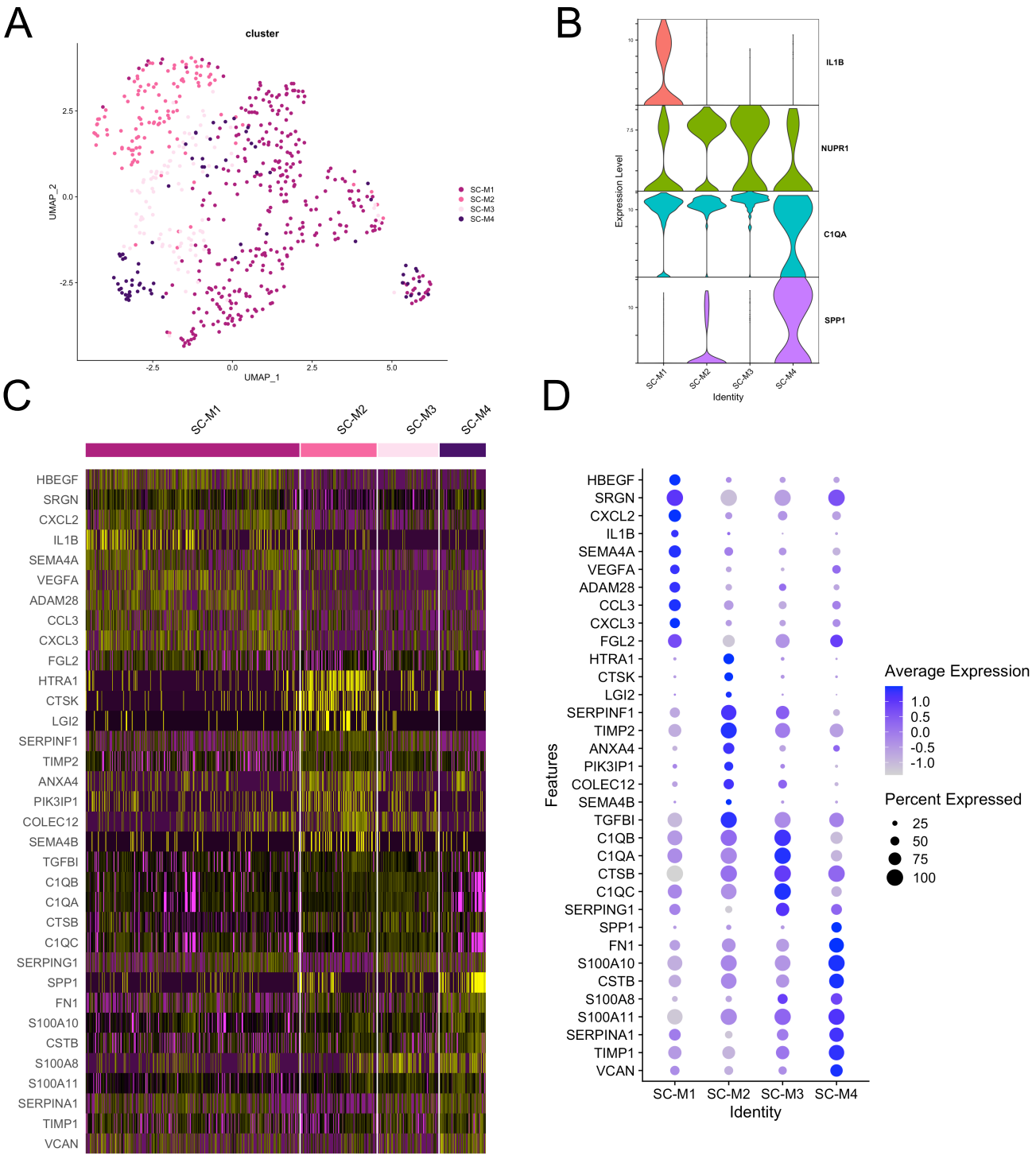


Figure 3.3 Distinct macrophages subsets have distinct matrix expression profiles. Differential expression analysis was run between macrophage subsets from the AMP1 dataset using a standard Wilcoxon Rank Sum test in Seurat V4. The output list of cluster-defining genes was then filtered for genes present in the matrisome to define macrophage cluster-specific matrix expression profiles. **A.** UMAP of the AMP1 dataset macrophage clusters. **B.** Violin plot of the canonical genes used to define the 4 fibroblast subsets in the original AMP1 study, namely the *IL1B*+ SC-M1, *NUPR1*+ SC-M2, *C1QA*+ SC-M3, and *SPP1*+ SC-M4. **C.** Heatmap of the top 10 cluster-defining matrisome genes for the 4 AMP1 macrophage subsets. Yellow corresponds to high scaled expression, and purple to low scaled expression. **D.** Dot plot of top 10 cluster-defining matrisome genes for the 4 AMP1 macrophage subsets.

3.2.3 Pseudotime analysis of matrisome genes in RA fibroblasts

The AMP1 SC-F2 and SC-F4 fibroblasts are located at tissular opposites of one another, the perivascular niche and the LL respectively[63,64,83]. This again makes the AMP1 fibroblasts a particularly potent analytical tool to assess the putative spatial distribution of ECM gene expression in the synovium. AMP1 fibroblasts were therefore analysed using the *slingshot* package in R. Trajectory analysis in principal component (PC) space using the SC-F4 as the starting cluster identified a robust trajectory (Figure 3.4 A), going sequentially through the SC-F4, SC-F3, SC-F1, and SC-F2 fibroblast clusters (Figure 3.4 B). To push this further, the dynamic nature of ECM expression across this trajectory was assessed by fitting a general additive model to model gene expression associations with Pseudotime. Interestingly, a number of matrisome genes were found to have a significant association with the trajectory, highlighting key transcriptomic niches of the synovium (Figure 3.4 C). *NTN4* for example was once again identified to be LL-restricted, along with a broader cassette of genes including *LTBP4, ADAMTS5,* or *IGFBP5*. A neat cassette of SL-restricted genes was also identified, which included but was not limited to *VCAN, COL14A1, BGN,* and *IGF1* for example. Together, these data broadly corroborate the identification of fibroblast cluster-specific matrix expression niches, and hint further at the existence of distinct sub-tissular microarchitectures in the synovium. These will need to be validated at the protein level.

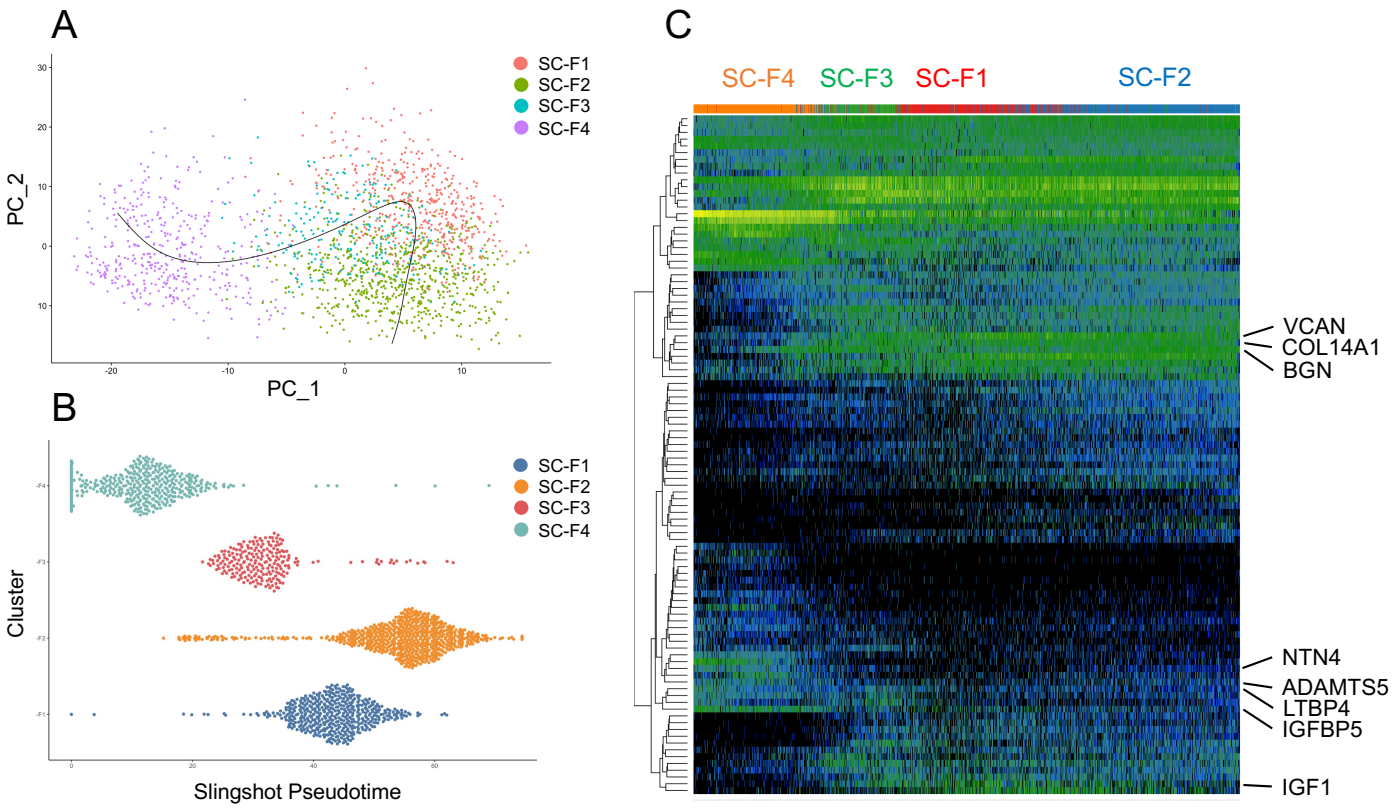


Figure 3.4 Differential expression of matrisome genes in the AMP1 fibroblasts across pseudotime. AMP1 fibroblasts were analysed using the *slingshot* package in R. Trajectory fit was performed in PC space, setting the SC-F4 LL fibroblasts as the starting cluster. Trajectory was found to run sequentially through the SC-F4, SC-F3, SC-F1, and SC-F2 fibroblasts. Temporally dynamic matrisome genes were then identfied by fitting a general additive model to test relationships between expression and pseudotime. **A.** Pseudotime trajectory overlayed on the AMP1 fibroblasts in PC space, coloured by cluster. **B.** Violin plot of the AMP1 fibroblasts, grouped and coloured by cluster, and ordered by their position along the pseudotime trajectory (X axis). **C.** Heatmap of the matrisome genes found to be differentially expressed across pseudotime (leftmost of the heatmap is minimum pseudotime, and rightmost is maximum pseudotime).

3.2.4 AMP1-like and matrisome-focused clustering

In an effort to deepen the characterisation of fibroblast cluster-specific matrix expression profiles to build the transcriptomic blueprint of synovial ECM, fibroblasts from the Alivernini dataset were clustered *de novo* to recapitulate the four fibroblast clusters defined in the AMP1 study (Figure 3.5)[63]. The dataset, comparing whole synovial cells from 4 active RA and 3 RA in remission patient samples, was selected for its greater cell number and its broader cell type coverage. As an aside, expression of all matrisome genes in each of the total synovial clusters of the dataset was analysed and is reported in Supplementary Figures 2-4. If successful this *de novo* clustering would enable us to explore disease activity specific effects on ECM expression and composition, because of the stratification of the RA samples in the dataset between samples from patients with active RA and patients with RA in remission, while retaining the analytical potency of the AMP1 clusters. Total cells from the raw Alivernini dataset were therefore integrated with *Harmony* in R *de novo*, clustered, and computationally subsetted from total synovial cells (Figure 3.5 B&C), down to stromal cells (Figure 3.5 D&E), and finally down to AMP1-like fibroblasts (Figure 3.5 F-I).

Based on the distinct fibroblast cluster-specific matrix profiles identified, I wondered whether matrisome genes alone might be sufficient in defining the AMP1 clusters. To test this, the *de novo* AMP1-like clustered fibroblasts were also clustered using exclusively matrisome genes, excluding from this cytokine and chemokine genes that are present in the secreted factors category (Figure 3.6 A). Matrisome-focused clustering was found to recapitulate the AMP1-defined clusters with striking overlap (Figure 3.6 B-C). A simplified equivalency shows cluster 0 containing 91.44% of the SC-F4 *PRG4*$^{+}$ lining fibroblasts, cluster 1 62.24% of the SC-F2 *HLA-DRA*$^{+}$ sublining fibroblasts, cluster 2 92.65% of the SC-F3 *DKK3*$^{+}$ sublining fibroblasts, and cluster 3 to be an expansion of 87.29% of the SC-F1 *CD34*$^{+}$ sublining fibroblasts fuelled by the inclusion of 22.99% of the formerly SC-F2 fibroblasts (Figure 3.6 C). The analysis also identified a fifth smaller cluster, cluster 4, characterised by the highly specific expression of *EGFLAM*, a gene associated with cell proliferation and mural cell functions[212,213].

Differential expression analysis (DEA) between the matrix-defined clusters identified distinct matrix expression profiles associated with each of these tissue architecture-defining subsets (Figure 3.6 D). Cluster-defining matrix gene cassettes, combined with pathway analysis using *gsfisher*, identified lining cluster 0 as being highly ECM regulatory, with specific expression of the proteases *MMP1/MMP3* and protease inhibitors *TIMP1/TIMP3*, while also expressing tissue repair-associated *IGFBP5* and the major structural genes *FN1* (Figure 3.6 D&E). Cluster 2 was characterised by the most unique of the sublining transcriptomic matrix signatures, with a primary pathway pointing to collagen fibril organisation, notably through *MFAP5*, *COL1A1*, *FBN1*, and *FBLN2* (Figure 3.6 D&E). Clusters 1 and 3, while being most similar in their matrix expression profiles, still harboured key differences, with cluster 3 being strongly associated with negative regulation of Wnt signalling through *SFRP1/SFRP2* and the expression of major RA-associated growth factor *IGF1*[214,215] and its regulator *IGFBP7*[216]. Cluster 1 on the other hand, canonically defined as the SC-F2 *HLA-DRA*$^{+}$ immune-associated fibroblasts[63]**,** was associated with the expression of tolerogenic dendritic cell inhibitor *MDK*, and CCN family genes *CYR61* and *CTGF.*

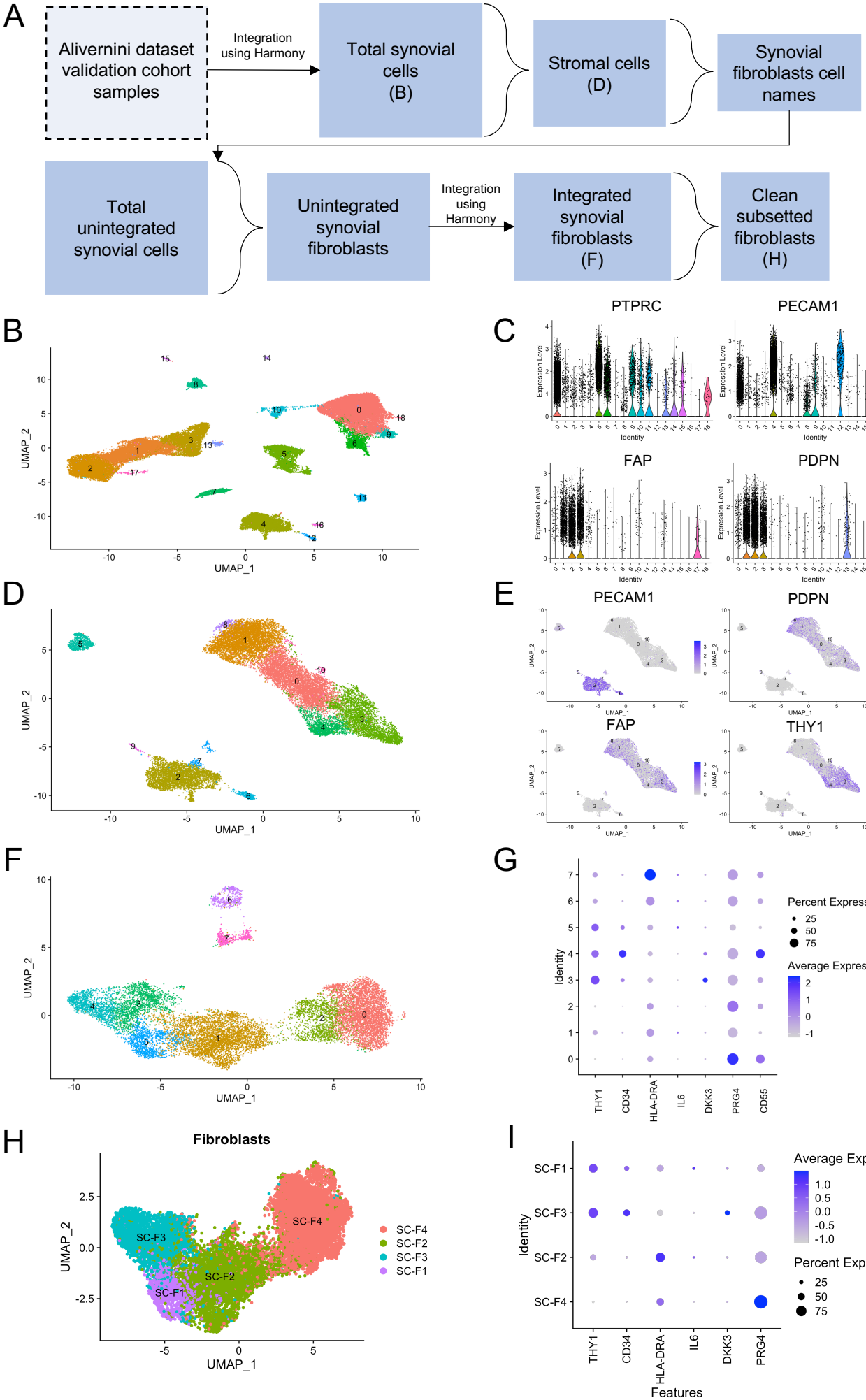


Figure 3.5 De novo clustering of Alivernini dataset fibroblasts identifies AMP1-like clusters. Full legend on next page.

Figure 3.5 *De novo* clustering of Alivernini dataset fibroblasts identifies AMP1-like clusters. Total synovial cells from the raw Alivernini dataset were downloaded from the EMBL-EBI under accession number *E-MTAB-8322*. The cells were first integrated using *Harmony*, and computationally subsetted into stromal cells by selecting *PTPRC*- and *PECAM1/FAP/PDPN*+ cells. These were then computationally subsetted further into fibroblasts by selecting *PECAM1*- and *PDPN/FAP*+ cells. Fibroblast cell identifiers were then collected and used to select only fibroblasts from the raw dataset. These were then integrated from scratch using *Harmony*, and clustered. Clusters were then merged based on the expression of canonical AMP1 fibroblast markers *THY1, CD34, HLA-DRA, IL6, DKK3, PRG4,* and *CD55* to define AMP1-lie clusters. **A.** Schematic workflow of the *de novo* clustering strategy. **B.** UMAP of the total synovial cells from the Alivernini dataset, coloured by cluster. **C.** Violin plots of the expression of markers used to computationally subset stromal cells from total synovial cells in each cell cluster. **D.** UMAP of computationally subsetted stromal cells, coloured by cluster. **E.** Feature plot in UMAP space of the expression of markers used to computationally subset fibroblasts from stromal cells. Marker expression is represented by purple colour intensity. **F.** UMAP of computationally subsetted fibroblasts, coloured by cluster. **G.** Dot plot of the expression of markers used to define the AMP1 fibroblast clusters. **H.** UMAP of computationally subsetted, reintegrated, and clustered fibroblasts, coloured by AMP1-like cluster. **I.** Dot plot of the expression of markers used to define the AMP1 fibroblast clusters.

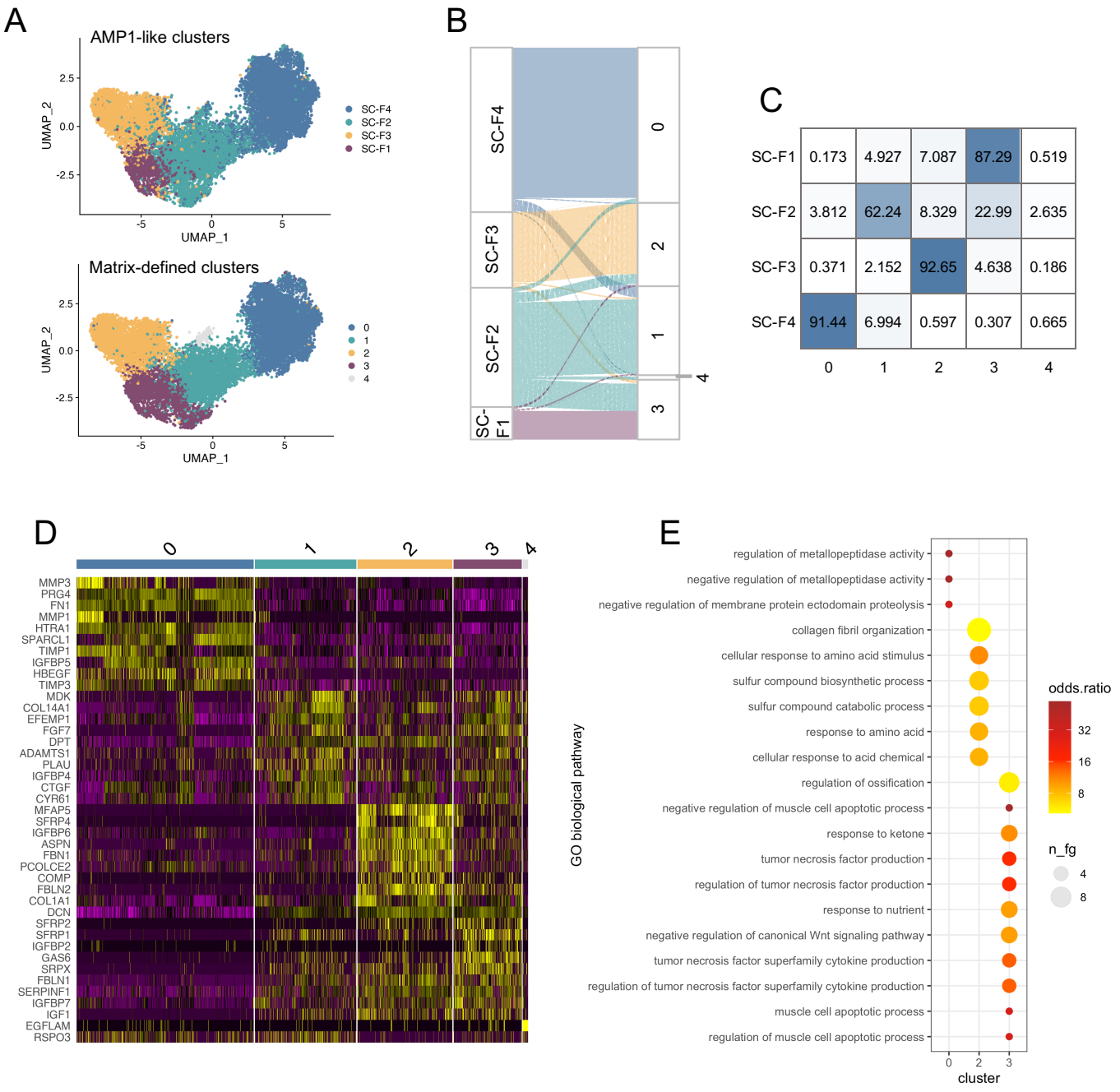


Figure 3.6 Canonical fibroblast subsets are defined by distinct matrix expression profiles. In parallel to *de novo* AMP1-like clustering of the Alivernini fibroblasts, fibroblasts were clustered using matrisome genes only (excluding cytokines and chemokines), by limiting the PCA input to this list of genes and for any downstream graph-based clustering, with a resolution of 0.3. Seurat V4 DEA using a Wilcoxon Rank Sum test was then performed as previously described to define cluster-defining matrix markers. **A.** UMAPs of the AMP1-like fibroblasts of the Alivernini dataset, coloured by their cluster attribution as a result of either AMP1-like clustering (top), or matrisome-focused clustering (bottom). **B.** Alluvial plot of the label-transfer and contributions of synovial fibroblasts in the AMP1-like clusters to the matrisome-defined clusters. **C.** Table of the percentage of cells from each AMP1-like cluster that make up each of the matrisome-defined clusters. **D.** Heatmap of the top 10 cluster-defining matrix genes for each of the matrix-defined clusters. **E.** Dot plot of the output pathway from *gsfisher*-based pathway analysis run on the matrix cluster-defining genes of each cluster. Dot size represents the number of genes driving the pathway association, while the colour scale represents the odds ratio of the pathway-cluster association.

Finally, Alivernini fibroblasts were also put through matrisome-focused clustering at a deeper resolution (resolution factor or 0.5) to look for hints at matrix micro-niches or rare fibroblast populations with interesting expression profiles. Such deeper clustering identified 8 clusters (clusters 0 through 7) (Figure 3.7 A). Though the overarching separations identified at the basic clustering depth remained broadly intact, interesting subdivisions of the original clusters could be found. The LL cluster 0 identified at the basic level was split into clusters 0 and 6, which retained broadly similar expression profiles with the exception of *WISP3* being specific to the latter (Figure 3.7 A,B&I). SL clusters 1 and 4 remained the same, becoming clusters 1 and 7 at this deeper level (Figure 3.7 A&B). SL cluster 3 became split into clusters 2 and 5, with *WISP3* again being specific to the latter. (Figure 3.7 A,B&I) Finally, cluster 2 became split into clusters 3 and 4 (Figure 3.7 A&B). These appeared to have broadly similar matrix expression profiles, with the notable difference of stronger *COL1A1* and *COL3A1* expression in the former, and stronger expression of *FBN1* and *FBLN2* in the latter (Figure 3.7 A&B). Interestingly *IGFBP5*, noted previously to be LL-specific, was found to be expressed in cluster 4 as its exclusive SL cluster of expression (Figure 3.7 E).

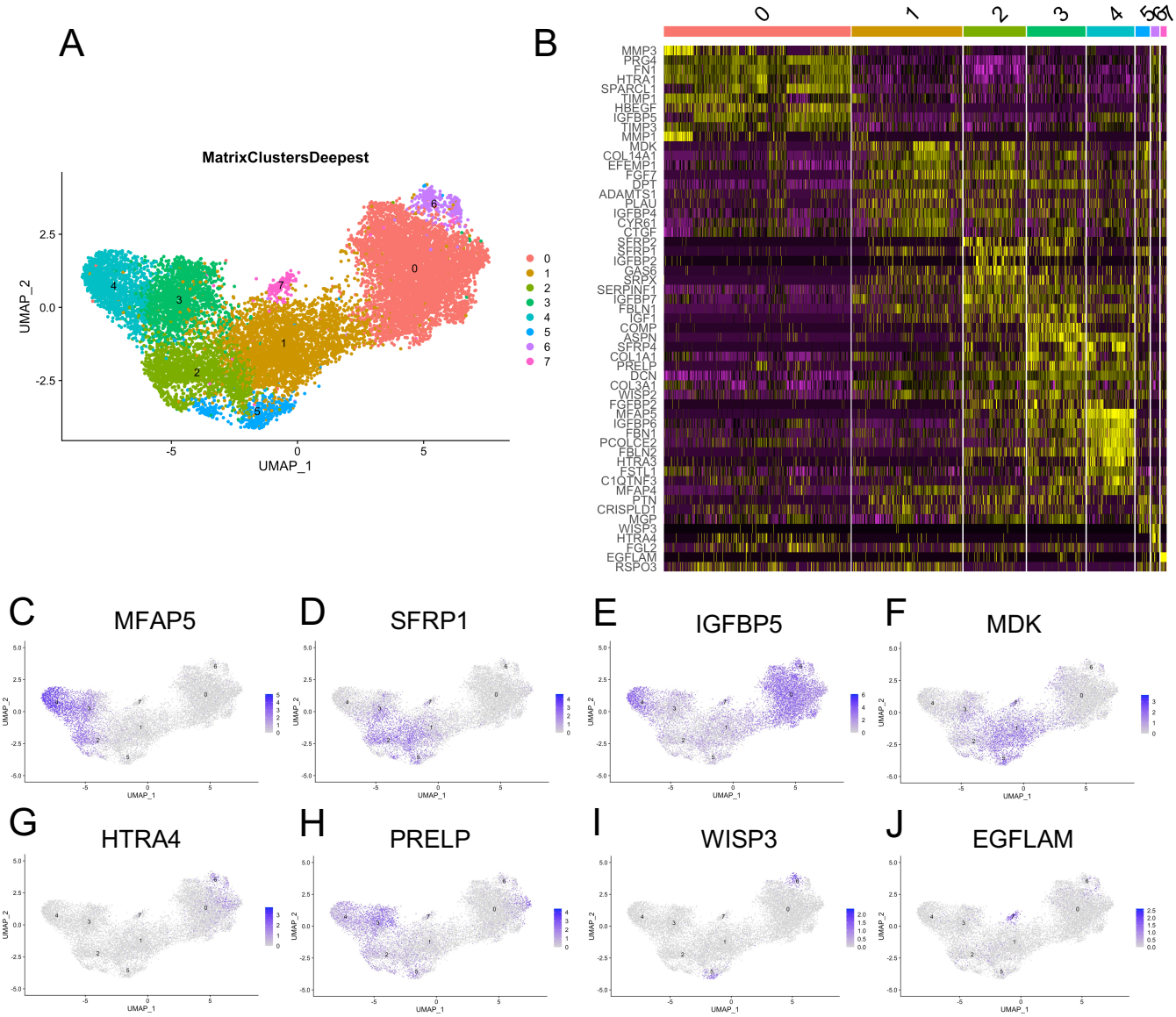


Figure 3.7 Deeper matrix-focused clustering reveals more specific matrix expression niches. To push the matrix clustering presented in Figure 3.6 further, clusters were also defined using a higher resolution of 0.5. Seurat V4 DEA using a Wilcoxon Rank Sum test was then performed as previously described to define cluster-defining matrix markers. **A.** UMAP of the fibroblasts of the Alivernini dataset, coloured by their cluster attribution as a result of either 0.5 resolution matrisome-focused clustering. **B.** Heatmap of the top 10 cluster-defining matrix genes for each of the matrix-defined clusters. **C-J.** Feature plots of top cluster defining matrix markers, with gene expression levels represented by purple colour intensity, projected onto UMAP space.

3.2.5 Disease type- and disease activity-specific matrix expression profiles

Having characterised niche-specific matrix expression landscapes in the RA synovium, I then set out to define its unique features compared to other types of synovitis, and to search for disease activity status-specific ECM expression profiles. Using the analysis pipeline described in Figure 3.8 A, DEA focused on matrisome genes was run for each fibroblast subset in the AMP1 dataset. Differentially expressed genes (DEGs) between OA and RA

samples were identified using two strategies: either sum-pseudobulking the data and running *edgeR*, or running *MAST* to improve the coverage and confidence in the identified DEGs based on the current best practices in single cell analysis[195]. An example of the output of this analysis is presented in Figure 3.8 C-F, focusing here on the SC-F4 lining layer fibroblasts (Figure 3.8 B), and the analysis outputs of SC-F1, F2 and F3 are shown in the appendix (Supplemental Figure 5). SC-F4 analysis highlighted 4 genes overexpressed in RA and 10 genes overexpressed in OA that were conserved between both DEA methods (Figure 3.8 E&F). For example, *COL6A1* was overexpressed in RA compared to OA, while *CRLF1* was overexpressed in OA. Downstream analyses do not limit themselves to these genes due to the limitations of statistical significance in single cell analysis, but these genes do provide a first confirmation of the existence of matrix compositional differences between arthritides.

Mirroring the analysis performed on the OA and RA samples of the AMP1 dataset, a DEA pipeline including the *de novo* fibroblast AMP1-like clustering was performed on the Alivernini dataset to look for disease activity-associated matrisomal DEGs (Figure 3.9 A, Supplemental Figure 6). Using the SC-F4 lining layer fibroblasts as an exemplar again, the analysis identified a similar number of DEGs, 5 overexpressed in active RA samples and 10 overexpressed in remission samples (Figure 3.9 C-F). Examples of these differences in expression include *IGFBP7* and the chemokines *CXCL2*, *CXCL3*, and *CXCL8* in the lining layer of patients with active RA, and *LTBP4* and *NOV* in the lining of patients with RA in remission. Similarly, DEGs were identified in SC-F1, 2 and 3 (Supplemental Figure 6). Building upon the AMP1 dataset analysis, these results imply that critical differences in synovial niche matrix expression and therefore composition are also present in different phases of RA itself. These differences may hold the key to a deeper understanding of the mechanisms underlying the pathogenesis and progression of RA.

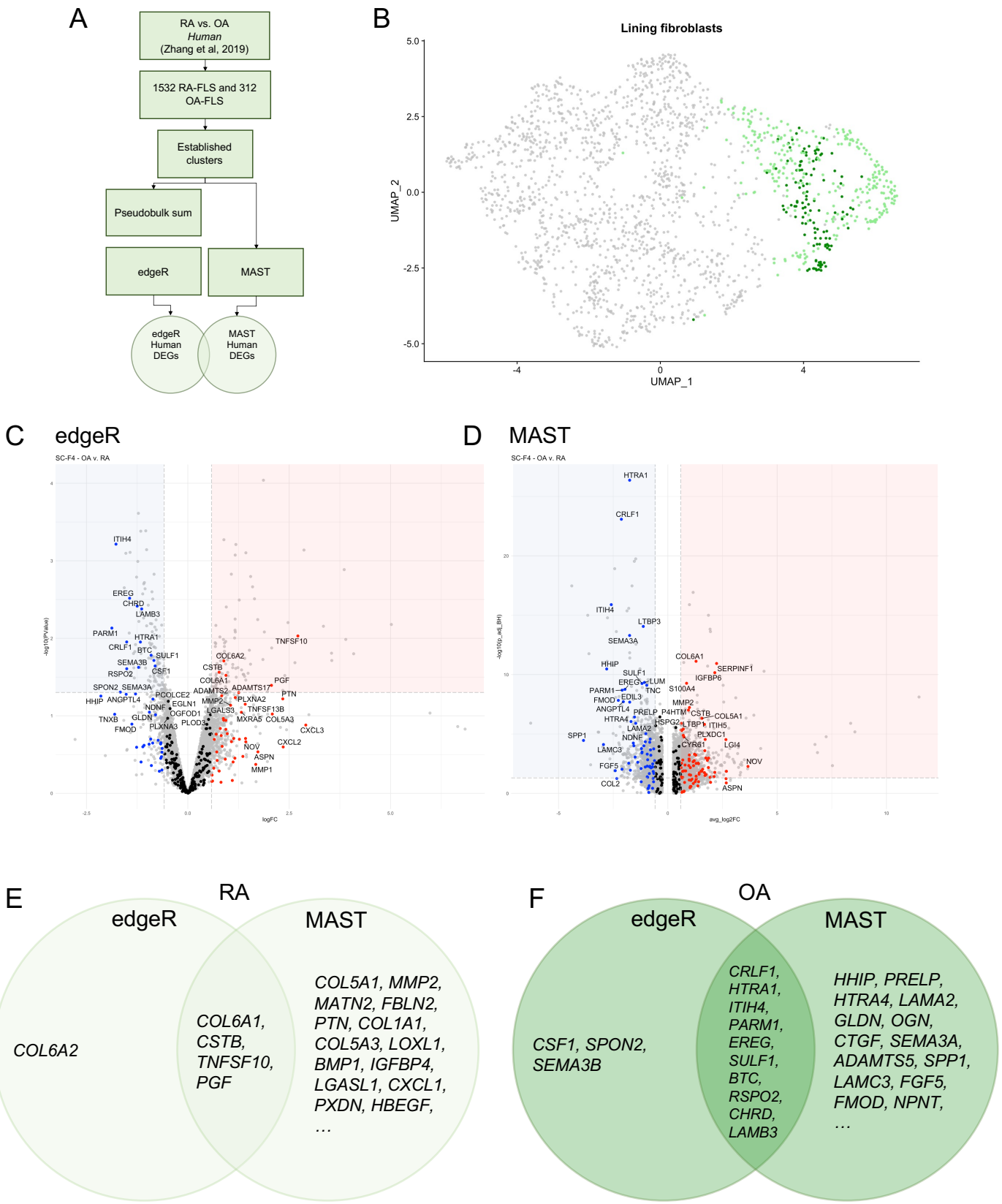


Figure 3.8 Differential expression analysis of matrisome genes in the SC-F4 of the AMP1 dataset between OA and RA samples. A. DEA strategy combining pseudobulk-summed edgeR and MAST. Fibroblasts from the AMP1 dataset were both sum-pseudobulked to be run through the *edgeR* bulk RNAseq DEA pipeline and run through *MAST* in Seurat V4. Overlap in the DEGs to come out of both methods were then used as a list of DEGs with strong associated confidence. **B.** UMAP of AMP1 fibroblasts, with lining fibroblasts from OA samples in dark green and lining fibroblasts from RA samples in light green. **C.** Volcano plot showing the output DEGs from the pseudobulk edgeR analysis. Differential expression is here defined by a fold change >1.5 and a p-value <0.05. DEGs in the blue-shaded area are differentially expressed in OA and DEGs in red- shaded area are differentially expressed in RA. Non-matrisome genes are greyed out. **D.** Volcano plot showing the output DEGs from the MAST analysis. **E.** Venn diagram showing the RA DEGs identified by edgeR and/or MAST. **F.** Venn diagram showing the OA DEGs identified by edgeR and/or MAST. Resolution better but I can't read the volcano plot labels

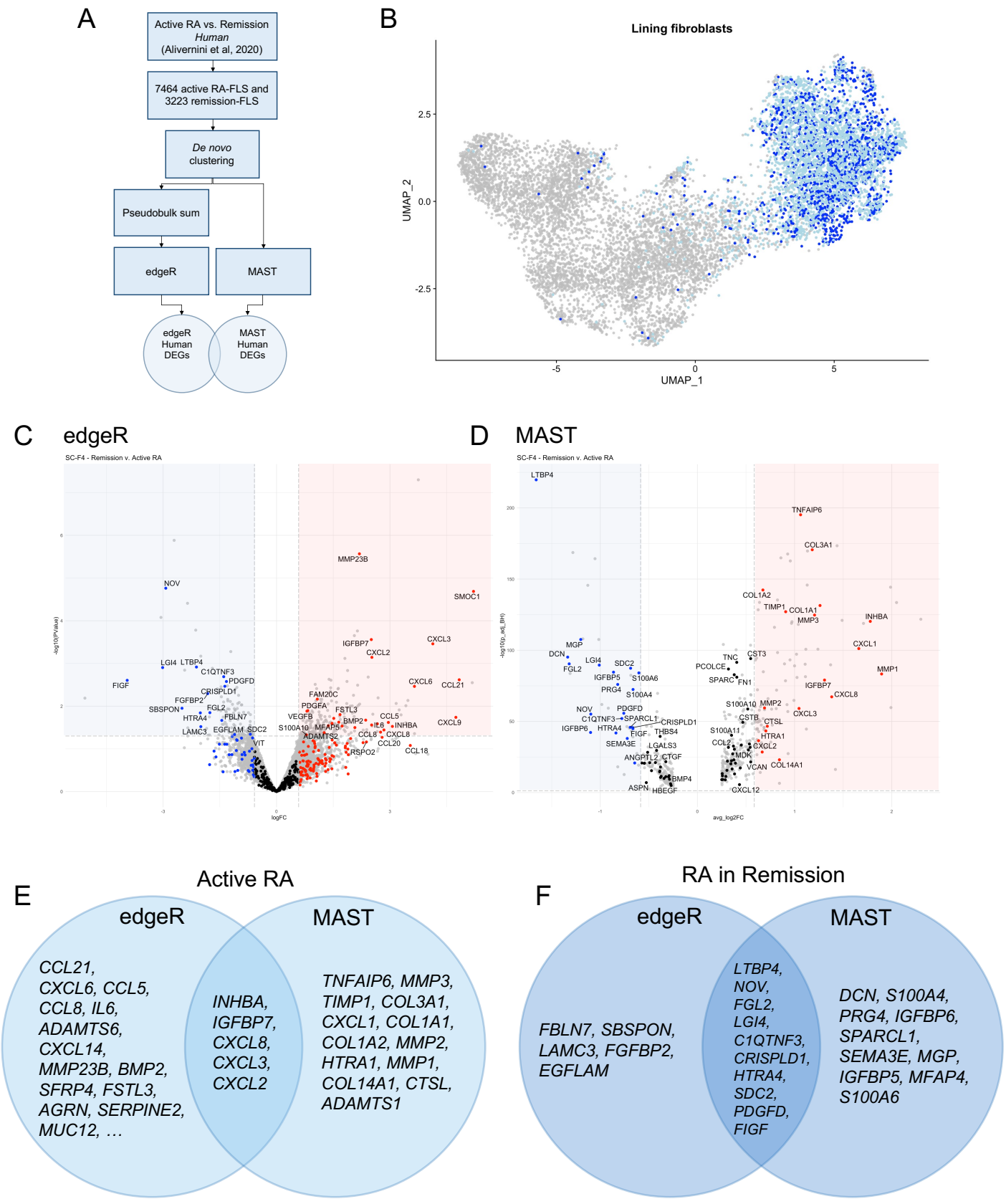


Figure 3.9 Differential expression analysis of matrisome genes in the SC-F4 of the Alivernini dataset between remission and active RA samples. A. DEA strategy combining pseudobulk-summed edgeR and MAST on de novo clustered cells. *De novo* clustered fibroblasts from the Alivernini dataset were both sum-pseudobulked to be run through the *edgeR* bulk RNAseq DEA pipeline and run through *MAST* in Seurat V4. Overlap in the DEGs to come out of both methods were then used as a list of DEGs with strong associated confidence. **B.** UMAP of Alivernini fibroblasts, with lining fibroblasts from remission samples in dark blue and lining fibroblasts from active RA samples in light blue. **C.** Volcano plot showing the output DEGs from the pseudobulk edgeR analysis. Differential expression is here defined by a fold change >1.5 and a p-value <0.05. DEGs in the blue-shaded area are differentially expressed in remission and DEGs in red-shaded area are differentially expressed in active RA. Non-matrisome genes are greyed out. **D.** Volcano plot showing the output DEGs from the MAST analysis. **E.** Venn diagram showing the active RA DEGs identified by edgeR and/or MAST. **F.** Venn diagram showing the remission DEGs identified by edgeR and/or MAST.

3.2.6 Cross-disease matrisomal remission signature

Having characterised disease activity-specific changes in ECM expression, I wondered whether some of these changes might be conserved across conditions. I therefore attempted to define common and unique active inflammation and inflammation resolution signatures across diseases by defining matrisomal DEGs in a dataset of human frozen shoulder (FS)[185]. While different tissues, the synovium and the shoulder capsule have similar structures, notably characterised by a joint cavity-facing lining layer and an underlying sublining layer[185]. Disease progression is however distinct between RA and FS, with FS being self-limiting and self-resolving. Because of this self-resolving nature of FS, I looked for potential overlap in DEGs between fibroblasts from the FS capsule and from the synovium of RA patients in sustained remission. In the lining layer, I identified cassettes of 10 remission-upregulated genes (*FGL2, HTRA4, IGFBP5, IGFBP6, LTBP4, PDGFD, PRG4, S100A6, SEMA3E, SPARCL1*) and 13 remission-downregulated genes (*ADAMTS2, ADAMTS6, COL14A1, COL1A1, COL1A2, COL3A1, IGFBP7, MMP2, MMP3, SERPINE2, SMOC1, TIMP1, TNFAIP6*) (Figure 3.10 A). In the sublining, I identified cassettes of 6 remission-upregulated genes (*CILP, FGFBP2, FGL2, LTBP4, MGP, TNXB*) and 7 remission-downregulated genes (*COL1A1, COL3A1, COL5A2, ELN, IGF1, SERPINE2, SPARC*) (Figure 3.10 B).

These cassettes point to opposing tissular landscapes based on inflammatory activity. Key features of the active inflammation cassette are the ECM remodelling and wound-healing responses mediated through the upregulation of ECM regulators such as *ADAMTS2*, *ADAMTS6, MMP2, MMP3, SERPINE2, TIMP1*, and fibril-forming type I and III collagens, which point to the activation of TGF-β signalling pathways[217–222]. IGF signalling is also a significant part of this signature through *IGF1* and *IGFBP7*, while antigen presentation promotion has been associated with the latter[223]. Tissue-level remission on the other hand is characterised by the opposite signature, notably TGF-β and IGF inhibition through *LTBP4, HTRA4, CILP, IGFBP5, and IGFBP6*[224–227]. Inhibition of antigen presentation is also a feature of the remission cassette, through the upregulation of *FGL2*[228].

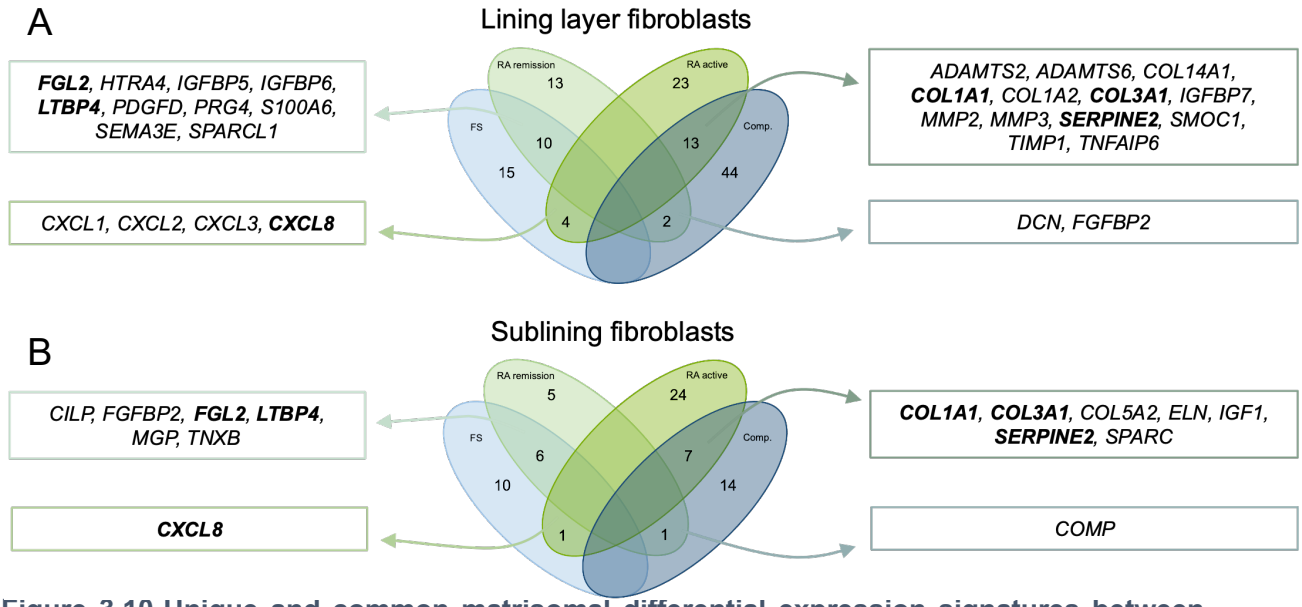


Figure 3.10 Unique and common matrisomal differential expression signatures between Rheumatoid Arthritis and Frozen Shoulder. DEA, filtered for matrisome genes, was run with *MAST* in SL and LL fibroblasts of the shoulder capsule (FS dataset) or synovium (Alivernini dataset) and between the conditions of the respective datasets. Overlap between the output DEGs for each condition was then analysed. **A.** Venn diagram of overlapping DEGs between Active RA/Remission and Frozen Shoulder/Comparator fibroblasts of the lining layer of both tissues. **B.** Venn diagram of overlapping DEGs between Active RA/Remission and Frozen Shoulder/Comparator fibroblasts of the sublining of both tissues. Genes conserved between lining and sublining fibroblasts are highlighted in bold.

3.2.7 Target prioritization of key matrisomal genes and proteins

In addition to looking for cross-disease ECM signatures, I sought to investigate further RA-specific matrisomal targets output from the DEA analysis. The underlying hypothesis behind this was that disease-associated changes in expression may translate to protein-level differences and prove to be either correlative, and serve as effective biomarkers of disease dynamics, or causative and provide novel insights into the underlying mechanisms of RA pathology. To derive the most value from this analysis, I sought to create a discrete list of priority ECM targets to be investigated at the protein level, and potentially be the focus of further functional explorations. To achieve this, individual genes that were defined as DEGs by either of the 2 DEA methods used were scored according to 3 metrics. The first of these metrics was the overlap between the 2 DEA methods to maximise confidence in the DEGs, the second was the overlap with DEGs identified in the mouse STIA model of the Wei dataset to look for species conservation and target tractability[64] (Supplementary Figure 6), and the

third was a differential expression score described further in Methods section 2.1.9. This prioritisation scheme allowed for a significant rationalisation in the number of hits of "interest" (Table 3.3).

Table 3.3 Enumerations of DEGs to come out of the initial DEA run on the AMP1 and Alivernini dataset fibroblasts, and their conservation through the prioritisation categories. "Triple overlaps" refer to genes identified as DEGs in the human datasets by MAST and edgeR, and in the Wei dataset.

Condition-Cluster	Total number of DEGs	Number of method-overlapping DEGs	Number of DEGs conserved between species	Number of triple overlaps
Active RA-F1	59	21	11	6
Active RA-F2	46	10	16	4
Active RA-F3	65	19	14	6
Active RA-F4	40	5	13	2
Remission-F1	14	4	5	2
Remission-F2	27	6	14	3
Remission-F3	28	7	11	4
Remission-F4	25	10	10	3
RA-F1	30	3	1	0
RA-F2	26	4	7	2
RA-F3	16	4	6	1
RA-F4	61	4	16	1
OA-F1	50	3	4	0
OA-F2	23	4	2	0
OA-F3	45	6	12	1
OA-F4	58	10	8	0

Importantly, this strategy allowed for a first unbiased stratification of hits (Table 3.4). Several recurrent hits across clusters and conditions could be identified at this stage. This included genes recurrently associated with remission such as *DCN* and *LTBP4*; but also genes recurrently associated with active RA such as *INHBA, SPARC,* or *COL1A1* and *COL3A1*. Interestingly, some hits were associated with inflamed conditions, namely RA compared to OA and active RA compared to remission (*IGFBP7, COL6A1,* and *COL6A3* for example), while others like *IGFBP5* were differentially expressed in less inflamed conditions (OA and RA in remission). This latter finding potentially suggests a downregulation in the active stages of RA, rather than a concurrent overexpression in remission and OA. Based on both this unbiased stratification and biological interest detailed further in the discussion of this

chapter, IGF signalling modulation and disruption of basement membrane structure were chosen as targets for further bioinformatic characterisation.

Table 3.4 Ranked top 10 prioritized matrisome DEGs for each fibroblast cluster and each condition of the AMP1 and Alivernini datasets.

Rank	Active RA-F1	Active RA-F2	Active RA-F3	Active RA-F4	Remission-F1	Remission-F2	Remission-F3	Remission-F4	RA-F1	RA-F2	RA-F3	RA-F4	OA-F1	OA-F2	OA-F3	OA-F4
1	COL1A1	MDK	COL1A1	INHBA	TIMP3	DCN	DCN	LTBP4	IGFBP7	POSTN	WISP1	COL5A1	BMPER	IGFBP5	PCOLCE2	CRLF1
2	TIMP1	COL1A1	SFRP4	TNFAIP6	S100A4	MGP	TIMP3	DCN	S100A13	CXCL9	MMP3	COL6A1	IGFBP5	OGN	ANGPTL4	HTRA1
3	CXCL14	COL14A1	BGN	MMP3	DCN	WISP2	SERPING1	S100A4	ECM2	CCL2	PAPLN	MMP2	TIMP2	ANGPTL4	TIMP3	ITIH4
4	SFRP4	INHBA	COL5A1	TIMP1	MFAP4	S100A4	CILP	FGL2	MFAP5	FNDC1	C1QTNF6	CSTB	TIMP3	INHBA	SCUBE2	HHIP
5	COL14A1	MMP3	CXCL14	COL3A1	GPC3	TNXB	LGALS3	S100A4	ANGPTL7	ASPN	TIMP4	MATN2	ADAMTS5	C1QTNF3	LTBP4	PRELP
6	BGN	LOXL1	SPARC	CXCL1	PCOLCE2	LTBP4	WISP2	LGI4	CXCL14	ADAMTS2	CTHRC1	FBLN2	COMP	COL14A1	FGL2	PARM1
7	CST3	COL3A1	COL1A2	COL1A1	VIT	IGFBP6	LTBP4	HTRA4	LAMA5	PXDN	PLXNC1	PTN	OGN	MGP	CXCL14	EREG
8	COL1A2	TIMP1	ELN	IGFBP7	LTBP4	ECM2	FIGF	PRG4	FGFBP2	COL1A1	SRGN	COL1A1	LAMB2	KAZALD1	CHRD	SULF1
9	SPARC	SPARC	COL14A1	COL1A2	IGFBP6	GAS6	CXCL12	SDC2	VIT	PCSK5	COL5A3	COL5A3	SERPINF1	WIF1	NTN1	HTRA4
10	SFRP2	COL6A3	MMP3	CXCL8	WISP2	FGL2	ANGPTL5	IGFBP6	PCSK5	RSPO3	ADAM12	TNFSF10	WIF1	SEMA3D	LAMB2	LAMA2

3.2.8 IGF1 signalling gene expression profiles

Because of its recurrent differential expression across the fibroblast clusters and conditions, and the biological interest surrounding its implication in a variety of pathologies, the IGF1 signalling pathway was chosen for focused expression analysis (Figure 3.11). Briefly, excessive IGF1 signalling through IGF1R, in turn driving cell survival and proliferation, has been reported as a feature of inflammatory arthritis[214,229]. Based on their presence in the top 10 prioritised hits (Table 3.4), I chose to examine the expression of *IGF1, IGF1R, IGFBP4, IGFBP5, IGFBP6,* and *IGFBP7*. Spatially, a clear dichotomy existed between lining layer and sublining-localised members of the IGF pathway. *IGF1, IGF1R, IGFBP4, IGFBP6, and IGFBP7* were predominantly sublining-restricted, while *IGFBP5* was limited almost exclusively to the lining layer (Figure 3.11 A). *IGF1, IGF1R, IGFBP4, IGFBP5, and IGFBP6* interestingly were overexpressed in OA compared to RA, while IGFBP7 expression was strongly associated with RA. The Alivernini dataset comparison between active RA and RA in remission corroborated the spatial distribution observations (Figure 3.11 B), but additionally shed interesting light on disease activity associations. Seemingly fitting their respective downregulation and upregulation in RA, *IGFBP5* and *IGFBP7* were strongly associated with remission and active disease respectively. *IGFBP6* was also found to be associated with remission, while *IGF1 and IGFBP4* expression were strongly overexpressed in active RA samples. Together, this expression dysregulation points to interesting potential interplays between the IGF-binding proteins and an overstimulated IGF signalling pathway in active disease.

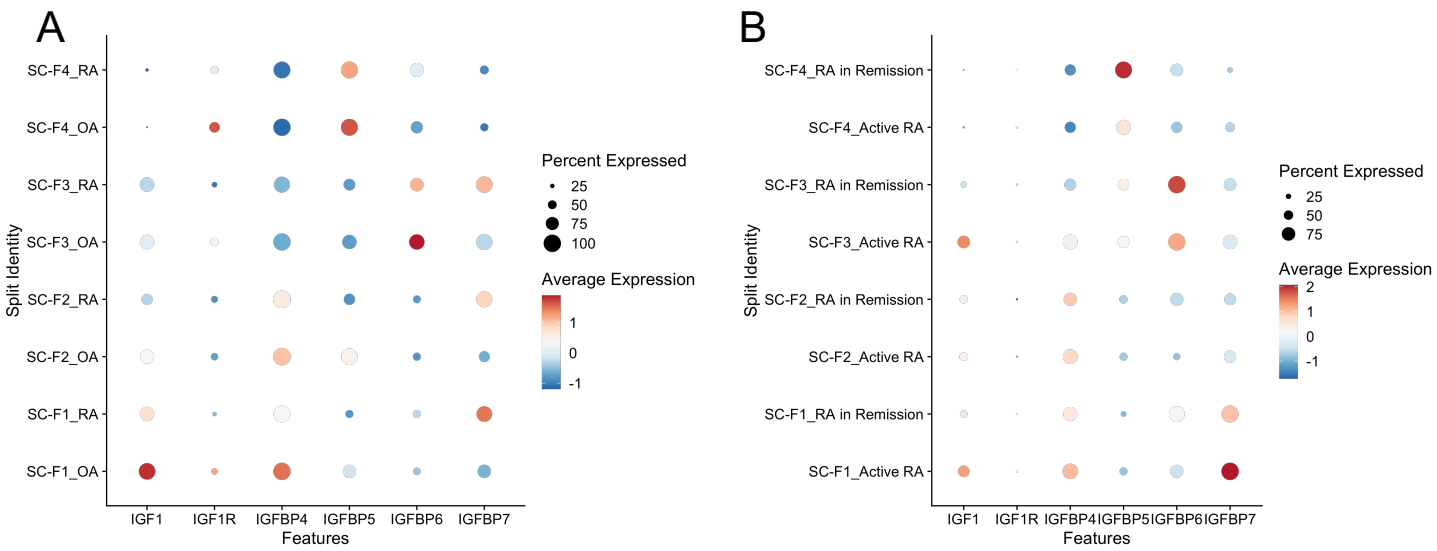


Figure 3.11 Disease-type and disease-activity expression of genes in the IGF and IGF regulation pathway. Prioritisation of output DEGs from the matrisome-focused DEA revealed a recurrent pattern of differential expression of genes in the IGF pathway and its regulators, suggesting a potential involvement of this pathway in the pathology and disease activity of RA. The genes investigated because of their recurrent in the top 10 prioritised gene list and their role in the pathway are *IGF1, IGF1R, IGFBP4, IGFBP5, IGFBP6,* and *IGFBP7.* **A.** Dot plot of the expression of the selected genes in each fibroblast subset of the AMP1 dataset, and split by condition between OA and RA. **B.** Dot plot of the expression of the selected genes in each fibroblast subset of the Alivernini dataset, and split by condition between remission and active RA.

3.2.9 Basement membrane gene expression profiles

Basement membrane-associated genes also stood out as potentially interesting from the target prioritisation efforts. Basement membranes are thin ECM structural assemblies which form architectural boundaries for a number of tissular niches, lining endothelial cells, epithelial cells, or adipocytes for example[230]. These properties are derived from their specialised protein composition, consisting notably of type IV collagen and laminin, which is altered in inflammation[14]. Based on their presence in the top 10 prioritised hits (Table 3.4), I chose to examine more closely the expression of *MMP2, TIMP3, NOV, NTN4, COL4A1,* and *COL6A1*. Spatially *MMP2, NOV,* and *COL6A1* were predominantly expressed in the sublining, while *TIMP3, NTN4,* and crucially *COL4A1* were overrepresented in the lining layer (Figure 3.12). *COL4A1* displayed an interesting spatial distribution, being restricted in expression to the lining layer of both datasets, and the SC-F2 and SC-F1 of the AMP1 and Alivernini datasets respectively. This likely reflects its contribution to basement membrane and basement-like membrane structures in the sublining and lining layer separately. Disease type

and disease activity associations appeared to be coherent, with *MMP2, NOV, COL4A1*, and *COL6A1* being associated with RA and active RA, and *TIMP3* and *NTN4* being associated with OA and remission. Interestingly again, the overexpression of *COL4A1* in active RA appeared to be restricted to its expression in the sublining, and not its expression in the lining layer.

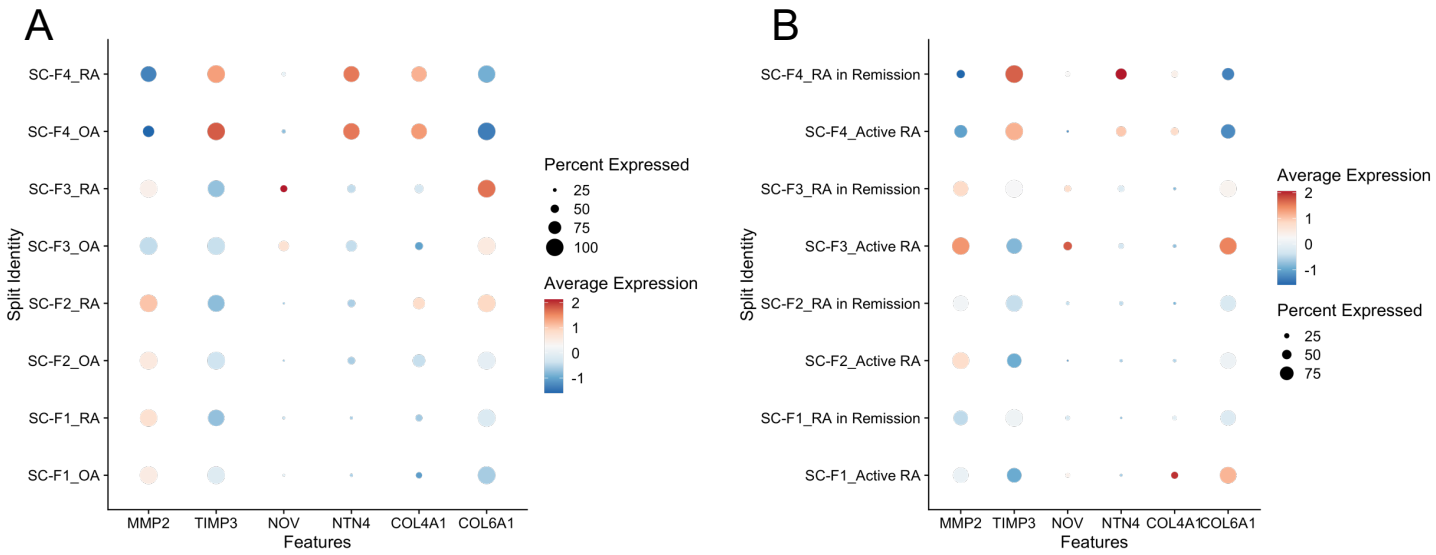


Figure 3.12 Disease-type and disease-activity expression of basement membrane-associated genes. Prioritisation of output DEGs from the matrisome-focused DEA revealed a recurrent pattern of differential expression of genes in basement membrane-associated, suggesting a potential involvement of this structure in the pathology and disease activity of RA. The genes investigated because of their recurrent in the top 10 prioritised gene list and their role in the pathway are *MMP2*, *TIMP3*, *NOV*, *NTN4*, *COL4A1*, and *COL6A1*. **A.** Dot plot of the expression of the selected genes in each fibroblast subset of the AMP1 dataset, and split by condition between OA and RA. **B.** Dot plot of the expression of the selected genes in each fibroblast subset of the Alivernini dataset, and split by condition between remission and active RA.

3.3 Technical discussion

In this chapter, matrisome gene expression was characterised across synovial cell subsets, across disease types and disease activity statuses. Fibroblasts and macrophages were found to be the predominant enriched cell subsets for matrisome expression. Distinct fibroblast subsets were found to have distinct matrisome gene expression profiles, which were sufficient in defining these subsets. Differential expression analysis focused on matrisome genes identified key distinct matrisome signatures for each of these fibroblast subsets between OA and RA, but also between RA in remission and active RA. RA in remission matrix DEGs were found to overlap significantly with matrix DEGs specific to Frozen Shoulder, a self-resolving chronic inflammatory condition of the shoulder capsule, thereby defining a cross-disease matrix remission signature. Finally, DEA and target prioritisation identified dysregulation of the IGF signalling pathway and of basement membrane deposition and degradation to be key features of the dichotomies between RA and OA, and between active RA and RA in remission. Together these data contribute to the creation of a transcriptomic blueprint of synovial ECM in RA and suggest interesting new targets to investigate further at the protein and functional level.

3.3.1 The matrisome gene list

To capture an exhaustive snapshot of the matrix composition of the synovium, the matrisome gene lists were used[17]. The primary strength of the matrisome is its breadth, combining genes shown through experimental validation to be key structural components of the matrix, with genes inferred to be associated with matrix based on their protein structural domain composition[16,231]. Its breadth can interestingly also be seen as its drawback. Indeed, in an attempt to englobe the full tissue microenvironment, the matrisome also blurs the lines of what is or is not “matrix”. Choosing to only focus on genes of the core matrisome category (glycoproteins, collagens, and proteoglycans) could be one solution. This would however fall short of the intended blueprint by ignoring the regulators and growth factors which are critical actors in the defining of tissue composition. In an effort to mediate this, I chose to run the

matrisome-focused clustering with a manually curated version of the matrisome, excluding cytokines and chemokines, to hone in on the cluster-specific tissue architecture-defining functions of fibroblasts.

3.3.2 Transcriptomics as an atlas corner stone

The choice to start with transcriptomics as the cornerstone of the atlas of synovial ECM was both logistical and biological. While RA proteomic datasets do exist, most are focused on serum protein levels, synovial fluid protein levels, or proteins from individual sorted cell subsets[232–238]. Proteomics carries greater analytical strength due to its direct functional implications, yet it lacks the breadth of samples and detected targets that comes with transcriptomics. In comparison, the transcriptomic datasets presented in Table 3.2 offer significantly more breadth of synovial cell types sampled[63,64,67,185], enabling a synovial tissue-focused yet exhaustive characterisation of the ECM gene expression of the inflamed tissue itself. Transcriptomic data also provides an interesting snapshot of tissue composition directionality, by measuring mRNA levels. Without anthropomorphising the tissue, transcriptomics does provide an idea of what tissue building blocks are 'intended' to be laid down, which has conceptual and analytical value.

3.3.3 Dataset selection

The generation of novel bioinformatic datasets, particularly scRNAseq, carries tremendous benefits. They can be a targeted approach to answering a specific question by fine tuning their experimental design to maximise insights, but also add value to a given field by enabling novel internal and collaborative explorations. In the case of this study, because of the public availability of relevant RA scRNAseq datasets, I chose to analyse published datasets rather than generate my own.

The first exploratory dataset selected for this study was the AMP1 dataset (Table 3.2), a collection of sorted synovial cells, which formed the state of the art in transcriptomic exploration of synovial tissue cells between 2018 and 2022[63]. The comparison studied in the dataset between OA and RA samples, along with the fibroblast subcluster stratification defined

in the study, build on the work of other groups by defining the premisses of functional stratification amongst fibroblasts[62,210]. The limitations of this dataset however are its small number of cells compared to novel capabilities (it is comprised of 5,265 cells in total), and its lower statistical power due to a low number of OA samples (3 OA samples vs 15 RA samples). Recent technological advances in data acquisition and sample collection have enabled the development of datasets that address these caveats effectively, including the AMP2 dataset which contains 314,000 cells and a significantly upscaled set of adequately powered condition comparisons (28 treatment naïve, 27 methotrexate-"inadequate" responders, 15 anti-TNF-"inadequate" responders, and 9 OA samples)[206]. Because this expanded dataset was released late in timeline of this study, the AMP1 dataset had to be the focus on the initial analysis. I believe that future work should focus on the AMP2 dataset for further matrisomal insights.

The second major exploratory dataset used in this study was the Alivernini dataset (Table 3.2). This dataset was a technical improvement on the AMP1 study, combining 37,072 whole synovial cells from 2 Undifferentiated Peripheral Arthritis (UPA), 4 Active RA, and 3 RA in remission samples into its "Validation cohort"[67]. The disease activity comparisons of the dataset are its main strength, providing an entirely novel view into differential tissue composition at different points of the cycle of inflammation of RA. The main limitations of this dataset include its low (but comparable across conditions) sample number, and its focus on macrophage subsets. Because of the scientific question asked by the original investigators of the study, the focus was placed on macrophage annotations, with comparatively less effort put into defining relevant fibroblast subsets. This however allowed us to produce two novel clustering strategies with interesting biological implications discussed in a later section.

3.3.4 scRNAseq analysis pipeline selection

Two major pipelines exist today for the analysis of scRNAseq datasets, *Seurat* from the Satija lab and *Scanpy* from the Theis lab, built in R and Python respectively[187,239]. At the time of the bioinformatic analysis performed in this study, *Seurat* was selected as the integrated environment of choice because of the extensive expertise of collaborators with the

pipeline and personal competency in R, as well as being the general method of choice in the literature. Since this, *Scanpy* has somewhat overtaken *Seurat* in usage, partly because of its compatibility with other Python pipelines, a language with generally greater rates of development and computational scalability than R. Future work could potentially focus on implementing equivalencies for the matrisome-focused pipeline into *Scanpy*. Additionally to selecting Seurat, *Seurat* was intentionally restricted to V4 even when V5 was released to maintain backwards compatibility of past analyses as recommended by the literature[240].

3.3.5 Differential expression analysis

The recent acceleration in the generation of bigger bulk RNAseq and scRNAseq datasets has mandated a parallel increase in the number of differential expression analysis (DEA) methods. Methods for bulk and single-cell have to contend with different technical challenges, with bulk RNAseq DEA methods being built to handle small sample numbers, and scRNAseq DEA methods having to deal with technical dropouts and high cell-to-cell variability[241,242]. Recent comparative benchmarking studies between DEA methods have identified *MAST* as one of the most effective and computationally efficient single cell-specific methods[194,195,243], while other studies have found bulk-specific methods applied to “pseudo-bulked” scRNAseq to be most stringent and effective[193,243,244]. To make the most of the breadth of developed methods, as well as the best practices at the time of the analysis, both *MAST* and *edgeR* run on the pseudo-bulked datasets were used. By looking primarily at matching DEGs between both methods, but not restricting the list of priority genes to them, an additional layer of confidence was introduced to target selection.

3.4 Biological discussion

3.4.1 Fibroblasts and macrophages share the burden of defining tissue composition

The first observation of this study was the enrichment in expression of specific matrisome gene categories in fibroblast and macrophage subsets of the AMP1 dataset (Figure 3.1). This observation is a conceptually important one in the broader context of ECM biology. The focus of the field remains greatly biased towards either the biochemistry of individual proteins of the ECM or the study of stromal cells as the primary cell subset defining tissular composition. Even now, published reviews consistently fail to address immune contributors[245]. Macrophages, and more broadly myeloid cells, are known to play critical roles in synovial homeostasis and pathological complications. Tissue resident macrophages in the lining layer are specifically equipped with tight junctions and act as an epithelial-like barrier between the sublining of the synovium and the synovial fluid in homeostasis[207]. In RA, the full spectrum of myeloid cells is at play, with specific macrophage subsets modulating an unstable balance between pro-inflammatory and pro-remission factors[66], and neutrophils likely being one of the initiator subsets of the early inflammatory transformations that the synovium undergoes[74]. This study points to a new critical function for myeloid cells in synovial inflammation, namely the direct modulation of deposited ECM via the expression of proteases, as highlighted by their stark enrichment in ECM regulator expression (Figures 3.1 & 3.3).

3.4.2 Synovial ECM is spatially heterogenous, with potential functional implications

Fibroblasts are important contributors to synovial ECM composition and turnover (Figure 3.1), justifying a specific focus on this cell type. More importantly however, this observation, coupled with the inferred spatial distribution of the fibroblast subsets of the AMP1 dataset, enabled a first unbiased assessment of the ECM heterogeneity and zonation of RA synovial tissue.

The three major sub-synovial niches – perivascular sublining, interstitial sublining, and lining layer – are all characterised by specific and distinct matrix components (Figure 3.2).

Compositional disparities were found even within similar tissue microstructures, notably between the SC-F2 fibroblast-associated perivascular basement membrane and the SC-F4 fibroblast-associated lining layer basement-like membrane. Both structures are composed of type IV collagen and laminin (Figure 3.2 E), canonical building blocks of basement membranes[246], corroborating early work suggesting the presence of a loose layer of collagen IV in the lining layer of the synovium in synovitis[247–249]. Specific differences however exist when looking into laminin subtypes, nidogens, and associated netrins. The case of Netrin-4 is particularly interesting. In cancer, Netrin-4 has been shown to modulate basement membrane stiffness by destabilising dense laminin-sheets, with functional implications for the metastatic and invasive potential of the cells within the tissue[250]. Its highly specific expression by the SC-F4 lining fibroblasts in RA suggests different functional requirements for the lining basement-like membrane compared to the perivascular basement membrane (Figure 3.2 E). In homeostasis, Netrin-4 could for example be acting to increase the porousness of the basement-like membrane to allow for the diffusion of important factors, like lubricin, which are secreted into the synovial cavity[251]. The selective absence of nidogens in the basement-like membrane, previously reported but interestingly unsupported[249], was confirmed here via transcriptomics (Figure 3.2 E). This may play a similar role in giving the basement-like membrane specific diffusion properties. Understanding the detailed mosaic of synovial ECM components may therefore tell us more about the underlying biology of the RA synovium and the dysregulation that occurs under inflammatory conditions.

The strength of these fibroblast cluster-specific matrix profiles and their zonation was further corroborated by both the Pseudotime analysis and the matrix-focused clustering. Trajectory analysis in PC space with *slingshot* placed the different fibroblast subsets consecutively, rather than mixed, along the pseudo-temporal trajectory, confirming discrete profiles with few potentially overlapping cells between clusters and many dynamically expressed genes along said trajectory (Figure 3.5). Additionally, clustering using only matrisome genes, as well as excluding cytokines and chemokines, was shown to be sufficient

to defining fibroblast subsets akin to the AMP1 clusters in the Alivernini dataset (Figure 3.6). This clustering strategy also provided additional insights thanks in part to the highly distinct cluster 1-defining cassette (Figure 3.6). Cluster 1 from the basic matrisome clustering represents the perivascular-associated or immune function-associated SC-F2 fibroblasts from the AMP1-like clustering (Figure 3.6 C)[63]. Beyond their immune-like function, the matrisome clustering demonstrated that a large portion of the fibroblasts in cluster 1 (62.24% of the SC-F2) could be defined by their matrix expression signature alone, including *COL14A1, IGFBP4, CTGF,* and *CYR61.* These fibroblasts, along with contributing to the establishment of the characteristic RA inflammatory milieu, appear to play a role in defining their own neat microarchitecture.

In the deepest level of the matrix-focused clustering we also found the *PI16*$^{+}$ "universal" fibroblasts, identified in numerous datasets by several groups across species[252,253], to be particularly characterised by the expression of matrix glycoprotein *IGFBP5*. The primary function of IGFBP5 is as an inhibitor of IGF1 signalling, a key pathway in inflammatory hyperplasic conditions such as RA, as both described in the literature and this study[215], through direct binding to IGF1[254]. Interestingly, *IGFBP5* is reportedly associated with both tissue repair responses[255], and their unbridled chronic activation in fibrosis[256]. The presence of *IGFBP5* as one of the cluster's key defining genes may suggest a yet unexplored role for this "universal" subset in injury repair responses and ensuring tissue-level re-establishment of homeostasis. As exemplified by this, I believe an interesting knowledge gap to be the conservation of these newly identified fibroblast matrix signatures across tissues with distinct function-driven designs and geographies.

3.4.3 Cross-tissue disease remission matrix signature

The cross-disease tissue remission signature identified in this study stems from such an approach. Frozen Shoulder (FS) is characterised by range of motion restriction due to chronic inflammation and fibrosis-like tissue remodelling of the shoulder capsule, features bearing a striking resemblance to the inflammatory tissue remodelling of the synovium in RA.

FS, in stark contrast to RA, is self-contained, and resolves in the majority of patients of the natural course of the condition[257]. Because of the similarities in homeostatic tissue composition and inflammatory nature of the conditions, I looked for overlapping resolution signatures of DEGs between RA in remission and FS, to define the matrisomal features of a tissue in remission. The key pathways to come out of this analysis were those of IGF-1 and TGF-β signalling modulation, particularly through expression of *LTBP4*, *HTRA4*, *CILP*, *IGFBP5*, and *IGFBP6.* Extensive literature exists surrounding the role of both signalling pathways in RA. IGF-1 signalling is heavily implicated in RA pathology, with early in vivo studies suggesting mild benefits from its inhibition on decreasing IL-6 levels and reducing fibroblast proliferation[258,259]. The availability of drugs such as Metformin and Teprotumumab, with described indirect and direct effects on IGF-1 inhibition respectively, should enable the effective exploration of this pathway as a therapeutic target[260,261]. TGF-β signalling is also widely implicated in the pathology of RA, yet attempts to inhibit it have yielded mixed successes in murine models of arthritis[262,263]. The cross-disease remission analysis points to active TGF-β signalling being involved in both chronic inflammatory active RA and acute injury responses of the comparator samples. The key to the success of anti-TGF-β signalling therapeutics may therefore lie in the chronology of the intervention. Early inflammatory phases may require TGF-β signalling as part of injury repair responses, with chronic signalling being at fault in active ongoing diseases, and a need for signalling-modulating efforts to bring a return to homeostasis.

3.4.4 Target prioritisation and selection: unbiased ranking, biological interest, and pragmatism

To efficiently narrow down the search for dysregulated matrisome genes of interest to investigate functionally, output DEGs from the DEA analysis were run through an unbiased prioritisation scheme (Table 3.3). The result of this was a list of the top 10 prioritised matrisome genes for each fibroblast cluster in the AMP1 and Alivernini datasets, and each condition in the two datasets (Table 3.4). Conscious of the high degree of attrition between scRNAseq-level hits and actual functional relevance, I sought to select 4 target genes to investigate

further at the protein level. I also chose to incorporate biological novelty and interest, as well as tractability of the prioritised hits into the selection scheme to rationalise the investigations and mitigate risks based on suggested best practices in the field[264].While this inherently introduces bias into target selection, it also ensures the selection of novel, promising, and most importantly explorable targets. Importantly, this bias is also mitigated by the restriction of target selection to my list of unbiased prioritised ranked gene hits.

DCN is a good example of a gene that was not selected to be taken forward based on this scheme. *DCN*, encoding the protein Decorin, was found to be in the top 3 ranked remission DEGs across all fibroblast clusters of the Alivernini dataset (Table 3.4). DCN is a Small Leucine Rich Proteoglycan (SLRP) involved in collagen fibril formation, which is also reported to bind to and inhibit TGF-β[265] and to act as an endogenous TLR2 and TLR4 agonist[266]. DCN has also been indirectly associated with RA through an observed overabundance of anti-DCN antibodies in RA and seronegative spondyloarthropathy patients[267]. Despite this clear biological interest and potential for novelty, *DCN* was not chosen as a target of interest for tractability purposes. The heterogeneity and variability of posttranslational modifications (PTMs) of the core proteins at the heart of proteoglycans, along with the diverse functional implications of these PTMs, make proteoglycans notoriously difficult targets to study[268]. With limited proteoglycan-research expertise in our group at the time of target selection, I elected to not pursue DCN as one of the targets of interest. Its relevance to RA pathology remains promising however, and I recommend its investigation for further research.

3.4.5 Bioinformatic-guided hypotheses: basement-like membrane turnover and maladaptive responses to stressors

Through the application of the two DEA methods and the target prioritisation methodology, two trends and ensuing hypotheses were selected to be explored in depth. One focusing on basement-like membrane turnover, and the other of maladapted tissue response to stresses.

Many basement membrane or basement membrane-associated genes were identified in the top DEGs from both datasets (Figure 3.12). The patterns of expression suggest a relative decrease in lining *COL4A1* expression in OA, combined with an increase in expression of the type IV collagenase *MMP2* and decrease in expression of its inhibitor *TIMP3* (Figure 3.12). This may suggest a loss of the type IV collagen layer hinted at by the literature[269], and is found here to be coordinated with an increase in *COL6A1* expression in the lining and sublining, a known basement membrane-associated collagen. Together these factors may be contributing to a loss of homeostatic lining layer structural integrity. This loss of integrity has been extensively described in the literature. Multiple strands of evidence include the dysregulated hyperplasia of the normally-thin lining[270], but also fibroblasts from chronically inflamed mice losing their lining-forming ability in tissue organoids and mice losing their synovial lining after repeat zymosan exposure[65]. The mechanism underlying the loss of this critical niche has however not been characterised and could hypothetically be due to a degenerative turnover of the structural ECM components supporting the lining. The loss of this compartmentalisation may also play a role in explaining the loss of the $TREM2^{+}$ $MERTK^{+}$ macrophages associated with remission in human patients[67]. The change in ECM properties of the subset's niche may promote its loss. With these questions in mind, the establishment of a multiplexed immunofluorescence panel containing markers including COL4A1 and COL6A1 (Figure 3.12), MERTK, PDPN, and CD90 could form the first step towards functional investigation of the microstructure's remodelling.

Another highlight within the top DEGs was the somewhat dichotomous expression of *IGFBP5* and *IGFBP7*, respectively IGF1 and IGF1R inhibitors, with the former being associated preferentially with OA and the lining fibroblasts and the latter with RA and the sublining fibroblasts (Figure 3.11). IGF1 signalling dysregulation is a known but poorly characterised feature of RA pathology both in humans and in mouse models of arthritis[214,229]. Little is known about the two binding proteins' roles in RA pathology, but the partial confirmation of the transcriptomic expression patterns at the protein level through multiplexed

immunofluorescence should enable their functional investigation and characterisation. Based on the literature, standout hypotheses emerge for both proteins. IGFBP7 expression appears to be an anti-tumoural response, blocking IGF1R signalling from promoting survival and proliferation, driving antigen-presentation, and promoting cell cycle arrest in an IGF1-independent manner[216,223,271]. While these effects are highly adapted and beneficial in a cancerous setting, the antigen presentation and induction of senescence with its senescence-associated secretory profiles may be contributing to the RA pathogenesis. The known matrix-binding properties of IGFBP7 also make these secondary functions particularly interesting in the context of the matrix turnover underlining active RA[272]. IGFBP5 on the other hand may act as a maladapted tissue repair response, with links to excessive ECM deposition and fibrosis reported in the literature[273,274]. Its lining-specific expression signature may contribute to the hypothesised deleterious remodelling of a lining layer shown to be critical in inducing remission[67]. IGFBP5 and IGFBP7 may therefore both act as tissue-associated double-edged swords in the synovial response to repeated inflammatory insults.

These two pathways, found to be dysregulated in RA at the transcriptomic level, carry putative pathological relevance based on their expression patterns and the published literature surrounding them. To investigate these further, I sought to evaluate their spatial distribution throughout the synovium at the protein level. Transcript to protein-level disconnects are frequent in target identification pipelines. Such investigations would serve to validate or refine the observed expression patterns. By combining matrix target protein labelling with synovial cell phenotyping, associations between protein and cell localisation could also be assessed to generate novel functional hypotheses as to their pathological involvement. To address this, chapter 4 therefore focuses on the generation and analysis of a novel multiplexed immunofluorescence dataset focused on IGFBP5, IGFBP7, COL4A1, COL6A1, and relevant synovial cell subsets.

Chapter 4 | Spatial distribution of key matrisomal proteins in OA and RA

4.1 Chapter Introduction

Through the bioinformatic analysis described in Chapter 3, IGF-1 signalling and basement membrane compositional changes were identified as two key dysregulated matrisomal pathways of interest in RA. Within the dysregulated IGF-1 signalling pathway, *IGFBP5* and *IGFBP7* were of particular interest, with lining layer remission-expressed *IGFBP5* being associated with tissue repair responses and fibrosis in the literature[273,274], and sublining active RA-expressed *IGFBP7* being associated with proliferation inhibition and antigen presentation[216,223,271]. The interest in the differential composition of the basement membrane stemmed from specific dysregulation of *COL6A1, COL4A1,* and their regulators *TIMP3* and *MMP2*, leading to potential imbalances in the ratios of both collagens. These hints towards pathological relevance at the transcriptomic level call for further characterisation at the protein level. Beyond the simple disconnect between transcript levels and protein levels, ECM proteins such as IGFBP5, IGFBP7, collagen IV, and collagen VI are secreted proteins. This renders the inference of their spatial distribution from transcriptomic information incomplete. Their cellular expressors may indeed not be located where the proteins eventually become deposited. Protein spatial characterisation should also reveal contributions to tissue architecture and associations with important immune cell subsets of the RA synovium. The aims of this chapter were therefore threefold. Firstly, to create and characterise a novel multiplexed immunofluorescence dataset of the matrisome proteins of interest and of tissue cells, secondly to define the spatial localisation of the key matrisomal proteins, and thirdly to identify putative functional involvements for at least one of the key proteins.

This chapter will therefore first describe the creation and characterisation of a multiplexed immunofluorescence dataset through the CellDIVE system. Based on the panel of markers used and automated pipelines, the defining structural features of each tissue, their subsynovial niches, and their broad cell subsets were identified. This dataset was then used to characterise the tissue distribution of the key matrisomal proteins, as well as their positions with regards to specific immune cell subsets. IGFBP5 protein distribution was found to corroborate the

transcriptomics, with predominant restriction to the lining layer, but also uncovered an interesting colocalisation with lining layer macrophages. IGFBP7 was found to localise specifically to the luminal side of the vascular endothelium, despite broad sublining fibroblast expression in the transcriptomic datasets. Collagen IV could not be detected in the lining layer, potentially due to technical reasons, but proved to be an excellent marker of vasculature through its labelling of basement membranes. Finally, collagen VI was found to have numerous interesting associations. Collagen VI levels and organisation in vascular basement membrane and the interstitial sublining were notably found to be associated with lymphocyte aggregate localisation. Collagen VI levels in the perivascular space were also found to correlate with $MerTK^{+}$ macrophages abundance in the tissue, hinting at an association with disease activity statuses. This observation was then strengthened by a concurrent observation in the mouse AIA model, with collagen VI deposition being highest in the inflammation resolution phases of the model.

Overall, the work presented in this chapter characterises the spatial distribution of IGFBP5, IGFBP7, collagen IV and collagen VI in the OA and RA synovium. Specifically, this work points to collagen VI as a key protein correlating with both remission-associated macrophage phenotypes and spatial distribution of lymphocytes in the synovium.

4.2 Results

4.2.1 Multiplexed immunofluorescence staining of key matrisome proteins

To investigate the spatial distribution of the key matrisomal transcriptomic hits, I first set out to create a multiplexed immunofluorescence dataset combining my targets with markers of the main immune cell subsets relevant to my pathways of interest. To do this, 9 RA and 7 OA samples from the University of Oxford (Prof. Sarah Snelling) and Birmingham (Mr. David Gardner) were stained (Table 4.1). The original plans were for 8 RA and 8 OA samples, but a late modification in the diagnosis of one patient from OA to RA skewed the cohort towards RA samples.

These sections were then stained for the antibodies forming part of an exploratory panel (Table 4.2). This panel included the 4 key matrix targets (IGFBP5, IGFBP7, COL4A1, COL6A1), pan-cellular markers (NaKATPase, Ribosomal protein S6), a set of stromal markers (PDPN, Cadherin, Vimentin, CD90, CD34), and immune markers (CD3, CD19, CD68, MerTK). The individual marker signals for each of these markers are reported in supplementary figure 7. Despite testing the antibodies' efficacy individually prior to their use on the experimental set of sections, on the day of the experiment the anti-CD34 antibody was found to be predominantly background signal and was therefore excluded from any further analyses (Supplementary Figure 7 F).

All the tissue sections in Table 4.1 were put through each cyclical staining round of the exploratory panel described in Table 4.2. Unfortunately, due to a higher degree of tissue fragility and subsequent poor tissue adherence to the slide, sections 47SM_14 and 83SM_14 could not be analysed, bringing the number of analysable sections down to 9 RA and 5 OA.

Together this multiplexed immunofluorescent staining laid the foundation for a new targeted protein-level dataset to analyse, made of 9 RA and 5 OA analysable sections with spatial information on 14 analysable proteins. An exemplar of this staining is presented in Figure 4.1, demonstrating the parallel information gathered on immune cell localisation, stromal cell distribution, and key matrix target expression.

Table 4.3 Patient clinical information. Patient metadata available for each of the biopsy section stained (Tissue ID) via multiplexed immunofluorescence.

Diagnosis	Tissue ID	Age	Sex	Disease duration (years)	Treatment
RA	JRP091_3	70	F	35	Hydroxychloroquine/Plaquenil and Naproxen
RA	JRP099_3	31	F	21	Pednisolone and Naproxen
RA	JRP112_6	35	F	17	Sulfasalazine/Salazopyrine/ Prednisolone/NSAID
RA	JRP118_2	66	M	20	Methotrexate/NSAID
RA	JRP122_6	60	M	/	Methotrexate, Salazopyrine, Golimumab/Simponi and Prednisolone when needed
RA	JRP129_2	72	F	44	Rituximab/Mabthera
RA	JRP136_6	56	F	36	Prednisolone
RA	JRP139d_6	67	M	7	Prednisolone and Tocilizumab (stopped 2 weeks prior to biopsy)
RA	JRP141_3	58	F	/	/
OA	JRP144_3	74	F	9	/
OA	S00293087_3	70	F	/	/
OA	97SM_15	64	M	/	/
OA	47SM_14	67	M	/	/
OA	83SM_14	72	M	/	/
OA	S00292623_3	69	F	/	/
OA	S00292477_3	88	F	/	/

Table 4.2 CellDIVE multiplexed immunofluorescence exploratory panel

	Target	Clone	Catalog #	Fluorophore	Concentration
	FcR	NA	130-059-901	NA	1:200 dilution
Round 1	PDPN	NZ-1.3	14-9381-82	AF488	10ug/mL
	IGFBP5	Polyclonal	PA5-37369	AF555	10ug/mL
	IGFBP7	Polyclonal	AF1334	AF647	2ug/mL
Round 2	NaKATPase	EP1845Y	ab197713	AF488	10ug/mL
	RS6	C-8	sc-74459	AF546	5ug/mL
	Cadherin	EPR1792Y	ab195203	AF647	5ug/mL
Round 3	Vimentin	EPR3776	ab185030	AF488	5ug/mL
	CD3	SP162	ab245731	AF555	5ug/mL
	CD19	EPR5906	ab196468	AF647	10ug/mL
Round 4	MERTK	Y323	ab52968	CF450	10ug/mL
	CD68	EPR20545	ab277276	AF555	10ug/mL
	COL6A1	EPR17072	ab200430	AF647	2ug/mL
Round 5	CD34	EP373Y	ab81289	AF488	10ug/mL
	CD90	EPR3132	ab181885	AF555	5ug/mL
	COL4A1	IV-4H12	51-9871-82	AF647	2ug/mL

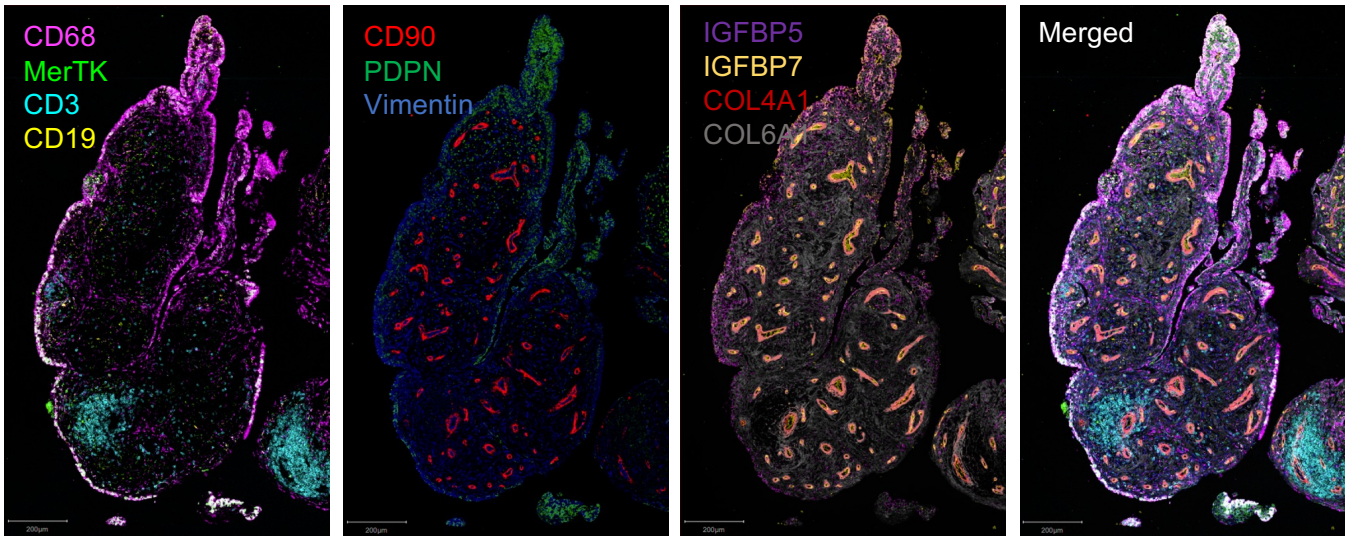


Figure 4.1 Representative multiplexed immunofluorescent imaging from the exploratory matrix panel. 5µm thick sections were cut from synovial biopsies of 5 OA and 9 RA patients and antigen retrieved (Table 4.1). The sections were then cyclically stained for rounds of 3 antibodies at a time, described in Table 4.2, and imaged on the CellDIVE system. Visualisation was then performed in QuPath to remove background fluorescence and export the reported images.

4.2.2 Synovial cell type classification and enumeration in OA and RA

To ensure the characterisation of pathologically-relevant spatial patterns of the target proteins, I set out to identify the immune cells of the tissue. To do so, cell segmentation was first performed in QuPath using $DAPI^+$ nuclei to define cellular centroids (Figure 4.2). In total, 804,845 cells were detected across the 14 analysable sections. These detected cells were then classified according to their signal intensity for CD3, CD68, MerTK, CD90, and CD19. Thresholds for each of these markers were assigned on each individual section using the mean marker intensity of the section plus one standard deviation, to exclude background signal. All $CD3^+$ cells were classified as T-cells, because of the sharp signal and low background noise of the antibody. $MerTK^+$ cells were classified as $MerTK^+$ macrophages, again because of the sharp specificity of the antibody signal. $CD68^+$ $MerTK^-$ cells were classified as $MerTK^-$ macrophages. Within the remaining $CD3^-$ $MerTK^-$ $CD68^-$ cells, $CD90^+$ cells were classified as $CD90^+$ fibroblasts, $CD19^+$ cells were classified as B-cells, and $CD90^-$ $CD19^-$ cells could not be classified. These unclassified cells likely represent a mixture of stromal cells and other immune cells not covered by the exploratory panel, such as monocytes, dendritic cells, and neutrophils.

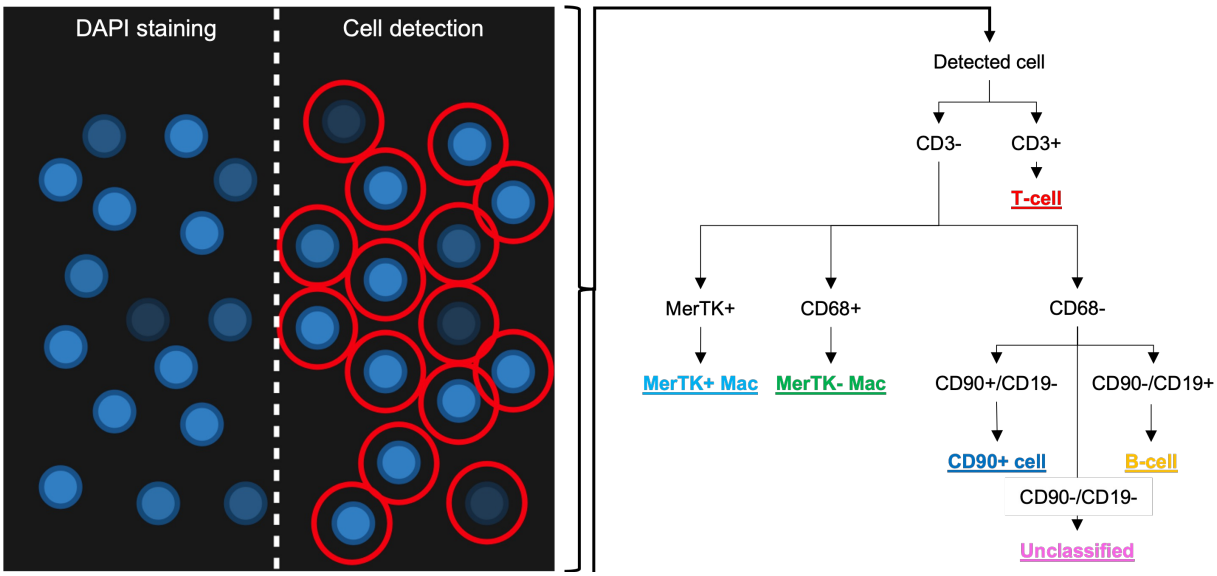


Figure 4.2 Immune cell classification scheme and cell population decision tree. Merged CellDIVE output *.tif* files for each of the stained and imaged sections were imported into QuPath and segmented using DAPI as a nuclei marker. Each individual cell that then classified as a specific cell type depending on its signal for CD3, MerTK, CD68, CD90, and CD19. Marker thresholds were set for each analysed section individually at the mean section marker intensity plus one standard deviation to avoid background signal. All CD3+ cells were classified as T-cells. Within CD3- cells, MerTK+ cells were classified as MerTK+ macrophages, MerTK- CD68+ cells were classified as MerTK- macrophages. Within CD68- cells, CD90+ cells were classified as CD90+ fibroblasts, CD19+ cells were classified as B-cells, and CD90-CD19- cells were unclassified.

With each detected cell in each section assigned to a specific cell type, the cell type composition of each section could be compared (Figure 4.3 A). Broadly, specific cell type abundances were similar between the OA and RA sections, with a high degree of inter-patient variability. Minor differences did exist with marginally higher T-cell proportions (8.28% of total detected RA cells versus 5.18% of total detected OA cells) and B-cell proportions (3.94% of total detected RA cells versus 2.94% of total detected OA cells) in RA compared to OA cells (Figure 4.3 B&C). Looking across samples, the differences in T-cell and B-cell abundances were not found to be statistically significant when a two-way ANOVA was performed (Figure 4.3 D). The most striking difference was in the proportion of unclassified cells (64.21% of total detected RA cells versus 68.54% of total detected OA cells), which may point to a higher abundance of immune cells in the RA samples. This difference was significant when looking across samples (two-way ANOVA with multiple comparisons, p-value = 0.0095). Overall these data showcase the broadly similar and immune-rich tissue composition of the synovium of the OA and RA patients in the exploratory cohort.

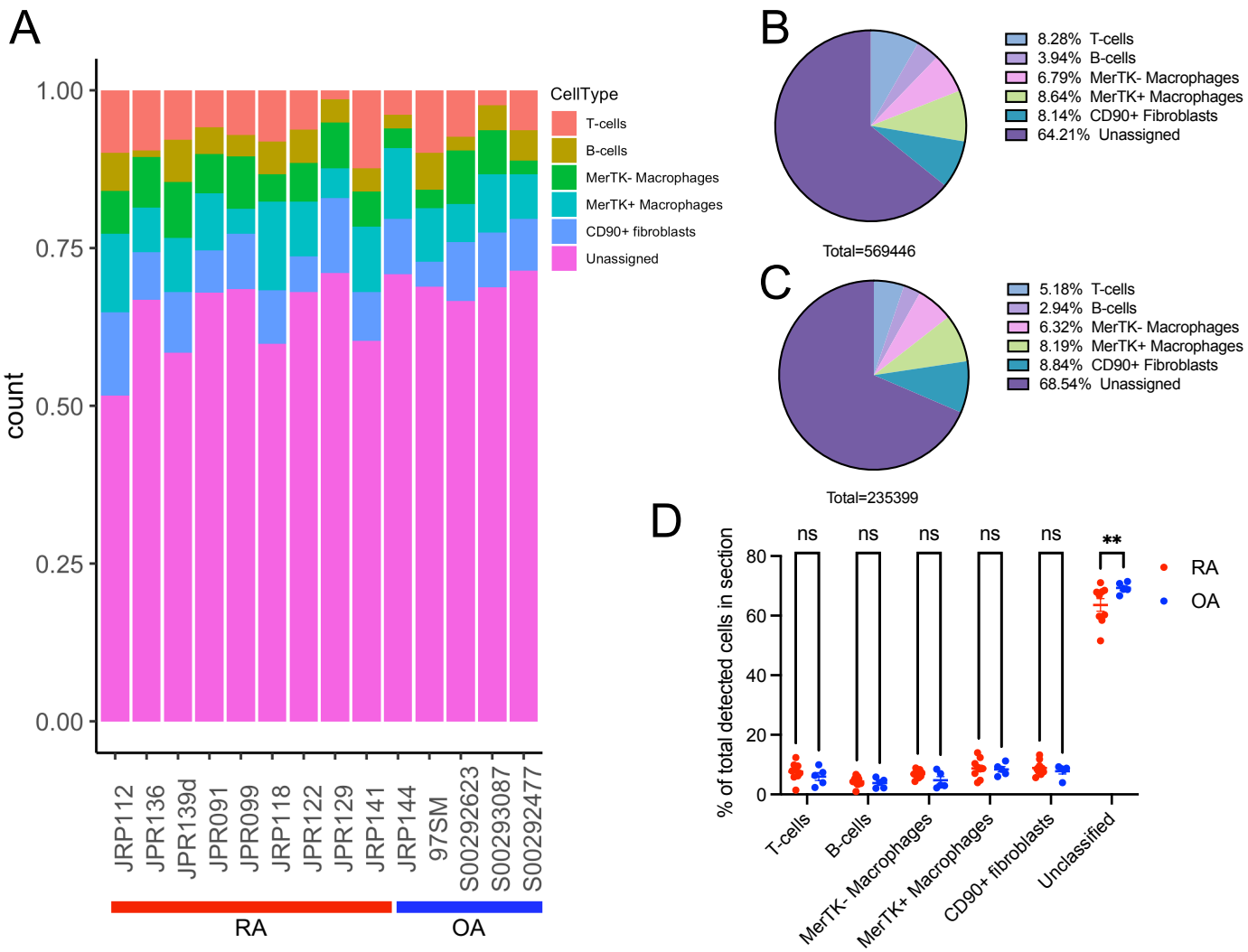


Figure 4.3 Cell type enumeration in the exploratory multiplexed immunofluorescence dataset. 804,845 cells were detected across the 14 stained and analysed sections and classified into T-cells, B-cells, MerTK- macrophages, MerTK+ macrophages, CD90+ fibroblasts, and unclassified cells. **A.** Bar plot of cell type proportions for each of the sections. This plot was generated with *ggplot2.* **B.** Pie chart of the cell type abundances in the pooled RA sections. **C.** Pie chart of the cell type abundances in the pooled OA sections. **D.** Grouped dot plot of cell type abundances in each section as a percentage of total section detected cells, split between RA (red) and OA (blue). Each point represents one section. Error bars are standard error of the mean. The statistical test is a two-way ANOVA with multiple comparisons. ns = not significant, ** = 0.0095.

4.2.3 Subsynovial niche characterisation and cellular composition

In addition to cell type identification, the detected cells were put through semi-automated subsynovial niche classification. 3 subsynovial niches were defined at this stage: lining layer cells, lymphoid aggregate cells, and sublining cells. A small number of manual annotations were first created, assigning cells to one of these 3 subsynovial niches. These manual annotations were then fed into the QuPath Random Trees automated classifier as training. The classifier was then able to automatically classify the rest of the cells of the tissue sections based on the combined expression of all markers in the panel (Figure 4.4 A&B). A total of 202,034 lining layer cells, 53,958 lymphoid aggregate cells, and 548,853 sublining cells were detected throughout the 14 analysed sections (Figure 4.4).

Comparing the ratios of niche cells of individual sections, a high degree of variability was observed, while depicting well the expected composition of synovial tissue. Sublining and lining layer cells could indeed be detected in all sections, with some sections having a higher degree of lining layer expansion, notably JRP112, JRP118, JRP141, and S00293087 (Figure 4.4 C). Lymphoid aggregate abundance was highly variable, with JRP118, JRP141, and S00292623 having the highest content (Figure 4.4 C). No statistically significant differences could be found when comparing OA and RA niche proportions through a two-way ANOVA, despite slight trends towards expanded lining layer/lymphoid aggregate compartments and a reduced sublining compartment in RA compared to OA sections (Figure 4.4 D).

Looking deeper into these niches, stark differences in cellular composition were found. The lining layer was found to contain a large proportion of unclassified cells (47.15%), but to be otherwise dominated by macrophages (38.41%), specifically $MerTK^+$ macrophages (22.81%) (Figure 4.4 E). Some T-cells and B-cells were also found to localise to the lining layer (6.32% and 5.38% respectively). This broadly corroborates the literature up to now, with the lining layer being described as a macrophage and fibroblast-dominated niche[66]. Lymphoid aggregates were predominantly composed of T-cells (42.19%), with a small number of B-cells associated (4.94%) (Figure 4.4 F). $MerTK^+$ macrophages vastly outweighed $MerTK^-$ macrophages in this niche (11.93% to 2.03% respectively). The sublining niche finally was

found to harbour the smallest proportion of identified immune cells (13.93%, as opposed to 61.09% of lymphoid aggregates, and 50.11% of lining layer cells) (Figure 4.4 G). This niche was predominantly characterised by the largest proportion of CD90+ fibroblasts (10.77% versus 4.64% in lymphoid aggregates and 2.75% in the lining layer), which is consistent with the literature[62], and unclassified cells (75.29% versus 34.28% in lymphoid aggregates and 47.15% in the lining layer).

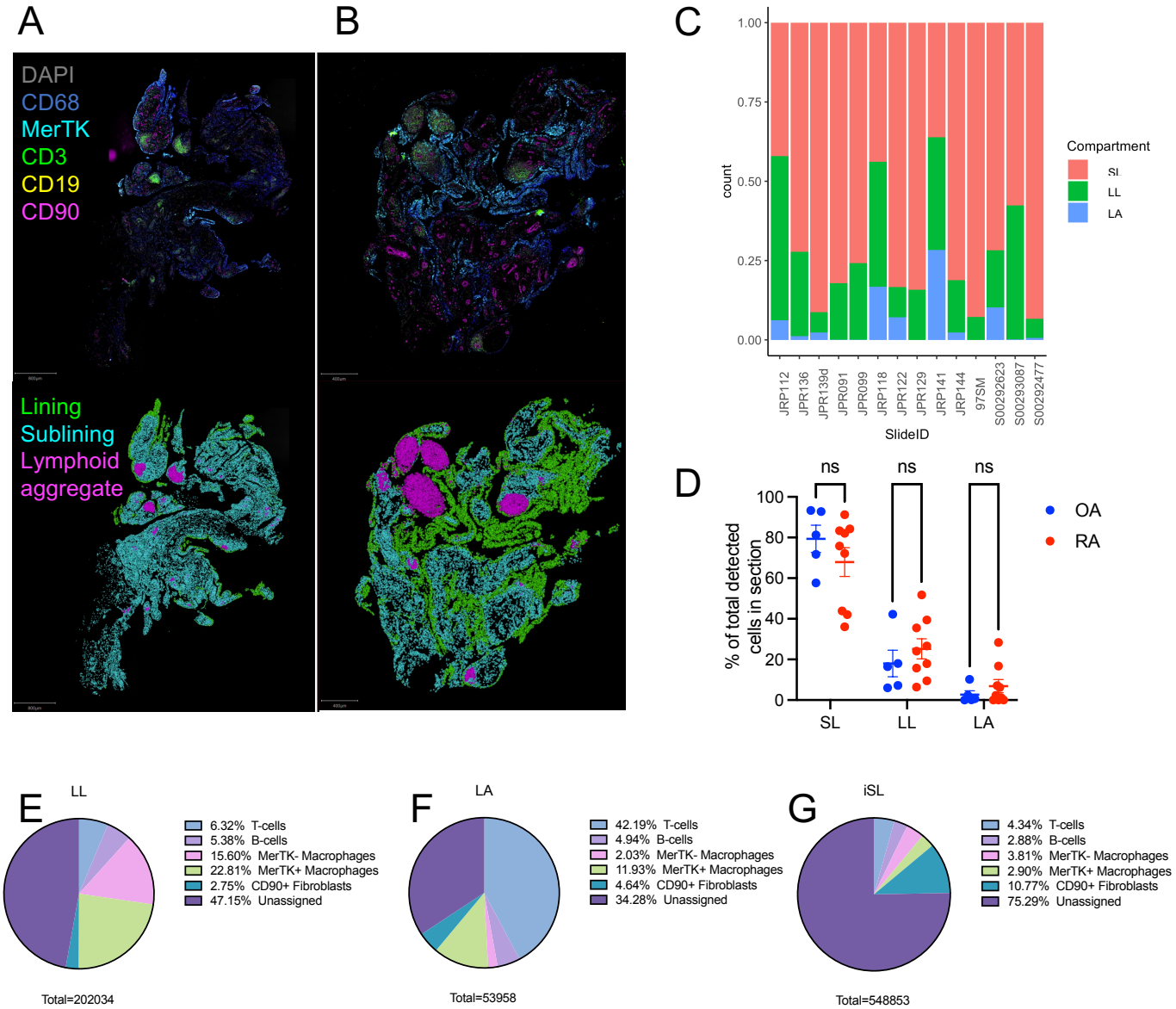


Figure 4.4 Subsynovial niche characterisation across OA and RA sections. 804,845 detected cells across the 14 stained and analysed sections were classified by manually training an automated cell classifier in QuPath, using the Random Trees algorithm. Detected cells were then classified into Lining Layer (LL), Lymphoid Aggregates (LA), and Sublining cells (SL), based on the expression of all markers in the panel. **A.** Visualisation in QuPath of section JRP122 immunofluorescent signal for DAPI, MerTK, CD3, CD19, and CD90 (top), and detected cell subsynovial niche classifications (bottom). **B.** Visualisation in QuPath of section JRP118 immunofluorescent signal for DAPI, MerTK, CD3, CD19, and CD90 (top), and detected cell subsynovial niche classifications (bottom). **C.** Bar chart of cell subsynovial niche proportions for each of the sections. **D.** Grouped dot plot of cell subsynovial niche abundances in each section as a percentage of total section detected cells, split between RA (red) and OA (blue). Each point represents one section. Error bars are standard error of the mean. The statistical test is a two-way ANOVA with multiple comparisons. ns = not significant. **E-G.** Pie chart of the cell type abundances in the pooled LL cells (**E**), LA cells (**F**), and SL cells (**G**).

4.2.4 Histological feature enumeration

To complement the characterisation of the sections in the multiplexed immunofluorescence dataset, I set out to define the key histological features of each analysed section. Because of clear visual differences between sections, I first measured their cellular density. Tissue area was first defined by setting section-specific average marker intensity thresholds. This area value in μm^2 could then be divided by the total detected cell counts in each section to get a cell density. As with other analysed metrics, high inter-sample variability was observed (Figure 4.5 A). A clear difference in tissue cellular density could however be observed between OA and RA, with mean densities of $2.206x10^{-3}$ cells/μm^2 and $4.647x10^{-3}$ cells/μm^2 respectively (p-value=0.0098, unpaired two-tailed t-test). This is consistent with the work of Mehta and colleagues, who found 3400 cells/mm^2, equivalent in our units to $3.4x10^{-3}$ cells/μm^2, to be a high sensitivity and high specificity threshold to differentiate histologically between OA and RA samples[275].

Building on top of this finally, the percentage of total tissue area occupied by vasculature was also quantified. This was done by setting section-specific COL4A1 signal intensity thresholders. Combined with the previously defined lymphoid aggregate cell percentage (Figure 4.4 C), the two metrics were used to place sections along a spectrum (Figure 4.5 C). In doing so, sections could be classified as either infiltrate heavy and relatively avascular (JRP118 and JRP122 in red), infiltrate low and highly vascular (JRP099, JRP136, S00293087, and JRP139d, in blue), infiltrate heavy and highly vascular (S00292623, JRP141, and JRP112, in purple), or infiltrate low and relatively avascular (JRP144, JRP091, 97SM, JRP129, and S00292477, in white). These classifications were made based on arbitrary cut-offs, purely for section classification purposes. Exemplar images for three of these categories are presented here, with S00293087 in Figure 4.5 D, JRP112 in Figure 4.5 E, and JRP122 in Figure 4.5 F.

Together with the cell type classifications and subsynovial niche detection, these analyses and observations culminate in the creation of an extensive multiplexed

immunofluorescence dataset. This dataset should then enable the characterisation of tissue distribution patterns of the key matrix proteins.

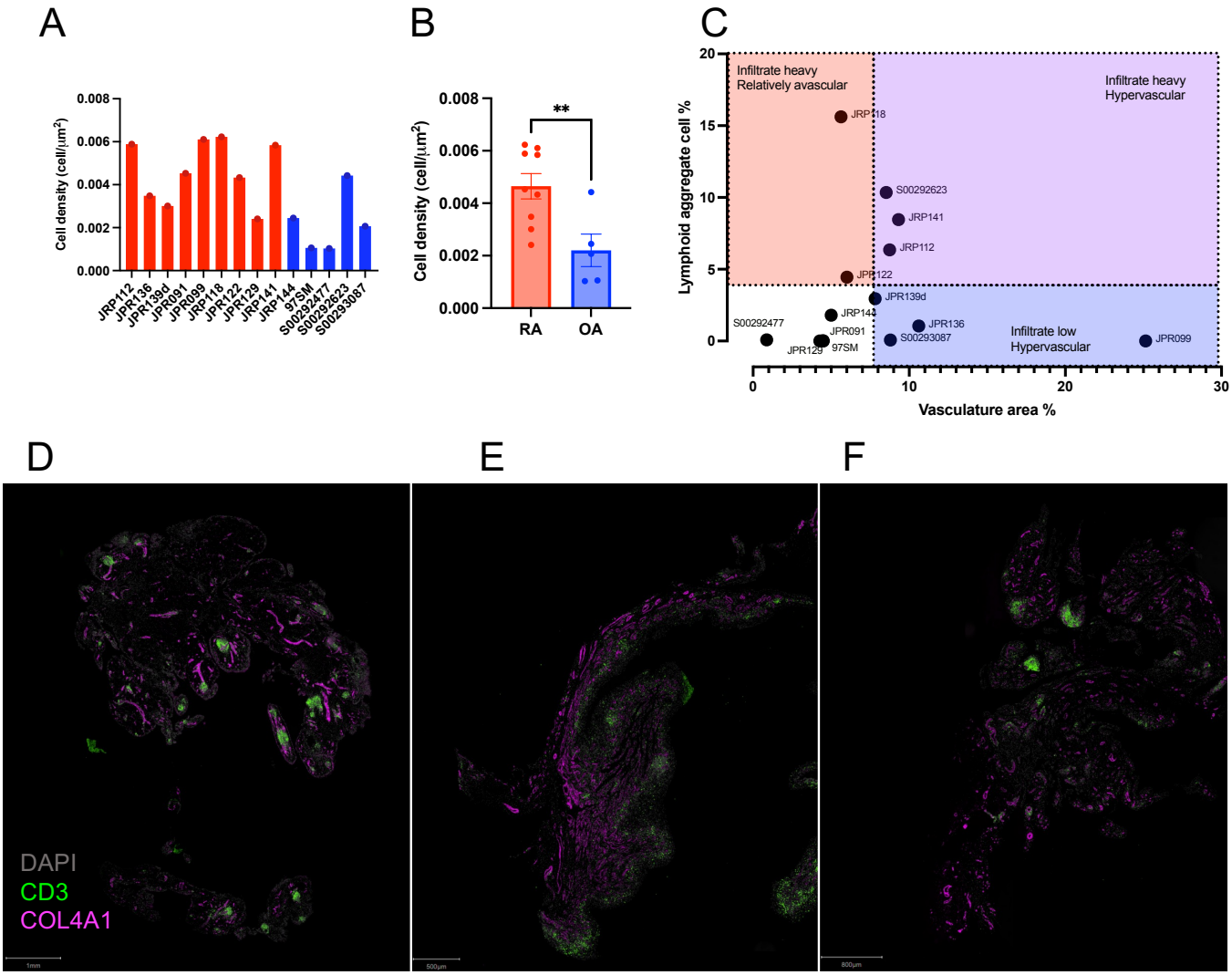


Figure 4.5 Histological feature enumeration across OA and RA sections. Following cell type classification, cell density and vascular area measurements were also performed on the 14 analysed sections. **A.** Bar plot of cellular density of each section in µm^2. RA sections in red and OA sections in blue. **B.** Comparative bar plot of cellular density of RA (red) and OA (blue) sections, with each point representing one section. Error bars are standard error of the mean. P-value=0.0098, unpaired two-tailed t-test. **C.** Plot of percentage lymphoid aggregate cells and vascular area for each of the 14 analysed sections. Sections are classified into infiltrate heavy and relatively avascular (JRP118 and JRP122 in red), infiltrate low and highly vascular (JRP099, JRP136, S00293087, and JRP139d, in blue), infiltrate heavy and highly vascular (S00292623, JRP141, and JRP112, in purple), or infiltrate low and relatively avascular (JRP144, JRP091, 97SM, JRP129, and S00292477, in white). **D-F.** QuPath visualisations of immunofluorescent imaging of DAPI, CD3 (T-cells), and COL4A1 (vasculature) in sections S00293087 (representative of n=4 infiltrate low/hyper-vascular) (**D**), JRP112 (representative of n=3 infiltrate heavy/hyper-vascular) (**E**), and JRP122 (representative of n=2 infiltrate heavy/relatively avascular) (**F**).

4.2.5 Key matrisome protein abundances in subsynovial niches

With a novel tissue protein-level dataset characterised, I first set out to investigate the spatial distribution of the key matrix proteins. Having defined the subsynovial niches of each of the cells in each section of the dataset, an automated thresholder was then applied for each of the key proteins, with the cut-off value set at the mean marker intensity of the section plus one standard deviation. This was done to exclude background signal and avoid the inherent bias of manual thresholds.

In line with the transcriptomic analysis, IGFBP5 deposition was found to be highest in the lining layer (mean detection in 35.69% of each section's cells) and lowest in the sublining (mean 7.894%) (Figure 4.6 A). In contrast to the transcriptomic analysis, IGFBP5 was also found to localise to lymphoid aggregates, seemingly with bimodal distribution, with four sections (JRP099, S00292477, JRP139d, JRP144) having ≥40% IGFBP5^{+} lymphoid aggregate cells and the rest having <20%. These four sections however fall under the previously defined "infiltrate low" tissue class, with total percentages of lymphoid aggregate cells well below 5% of tissue cells (Figure 4.4 C & Figure 4.5 C). This may explain this bimodal distribution and suggests this IGFBP5 positivity in lymphoid aggregates may be artefactual.

IGFBP7 protein deposition on the other hand appeared to be at odds with its transcriptomic distribution (Figure 4.6 B). Detection levels were comparable across the subsynovial compartments, with no statistically significant effect or difference between any of the niches (two-way ANOVA). Trends in expression were even contrary to the previously defined sublining-restricted expression of *IGFBP7*, with the highest detected deposition in the lining layer and lymphoid aggregates (mean 21.80% and 18.92% respectively) and the lowest in the sublining (mean 10.53%).

Another protein at odds with its transcriptomic distribution was COL4A1 (Figure 4.6 C). Contrary to the stark bimodal expression of *COL4A1* in the lining layer and perivascular niche fibroblasts, COL4A1 signal could only be detected in the perivascular space at the protein level, which makes up part of the sublining niche in this analysis (mean 14.59%), and not the lining layer (mean 1.874%). This discrepancy could be due to a number of interesting technical

and biological factors and is explored further in the discussion part of this chapter. Because of this discrepancy, COL4A1 is used exclusively from this point onwards as a marker of vascular basement membranes.

COL6A1 protein distribution did broadly corroborate the transcriptomics. COL6A1 was found to be marginally higher in the sublining (mean 17.18%) than in the lining layer (15.36%), despite no statistically significant difference (Figure 4.6 D). COL6A1 positivity was however significantly lower in lymphoid aggregates compared to the other two niches (mean 7.234%).

To assess potential associations with disease type, the distributions were then split and compared between OA and RA (Figure 4.6 E-H). IGFBP5 positivity was found to be marginally decreased in RA irrespective of niche, in accordance with its transcriptomic profile (Figure 4.6 E). IGFBP7 positivity was increased marginally in the lining layer and significantly in lymphoid aggregates in OA sections, which again contrasts with its transcriptomic signature (Figure 4.6 F). No differences were found between OA and RA COL4A1 positivity (Figure 4.6 G). COL6A1 lining layer positivity was decreased in RA, while its sublining positivity was increased in RA (Figure 4.6 H).

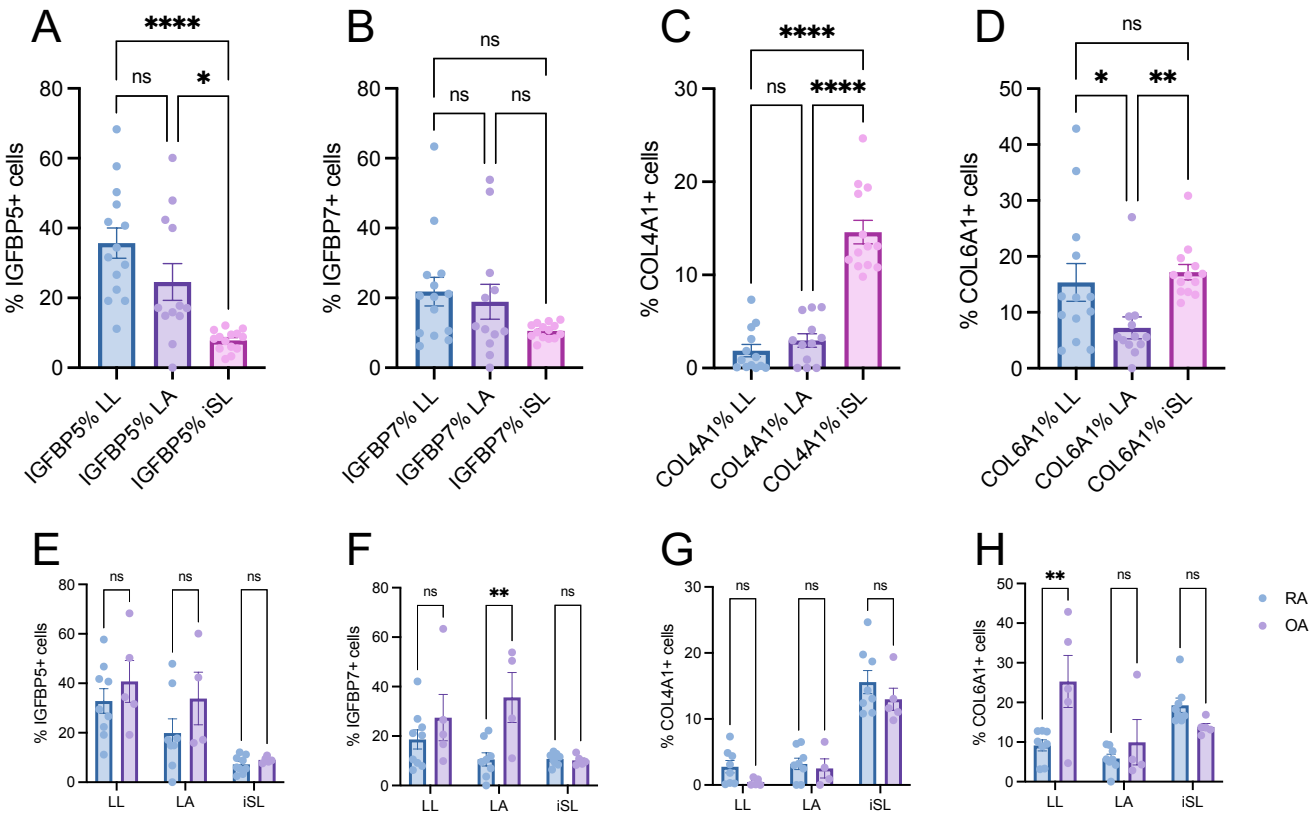


Figure 4.6 Key matrix protein subsynovial niche distribution and disease type differences. Following subsynovial niche classification of the cells in all 14 analysed sections, cells were also classified based on key protein positivity using an automated thresholder set at the mean plus one standard deviation of the signal intensity of the section analysed. **A-D.** Bar plot of the percentage positivity of each section's cells in each of its three subsynovial compartments: Lining Layer (LL), Lymphoid Aggregates (LA), or Sublining (iSL). Each point represents one section. Error bars are standard error of the mean. The statistical test is a two-way ANOVA with multiple comparisons. ns = not significant, * = p-value < 0.05, ** = p-value < 0.01, **** = p-value < 0.0001. Analysis was run for IGFBP5 (**A**), IGFBP7 (**B**), COL4A1 (**C**), COL6A1 (**D**). **E-H.** Grouped bar plot of the percentage positivity of each section's cells in each of its three subsynovial compartments, Lining Layer (LL), Lymphoid Aggregates (LA), or Sublining (iSL), split between RA (blue) and OA sections (purple). Each point represents one section. Error bars are standard error of the mean. The statistical test is a two-way ANOVA with multiple comparisons. ns = not significant, ** = p-value < 0.01. Analysis was run for IGFBP5 (**E**), IGFBP7 (**F**), COL4A1 (**G**), COL6A1 (**H**).

4.2.6 Key protein distances to immune cell types

Another relevant feature of protein spatial distribution is their position in relation to immune cells, which may inform their function. Distance measurements to the nearest key protein-positive cell was therefore performed in QuPath on every cell in the dataset. For this analysis specifically, all the cells of the dataset were pooled together to look for overarching cellular associations, rather than looking at differences between sections. This analysis was conceptually reliant on the assumption of similar cellular composition between the OA and RA

sections, which was proven in Figure 4.3. Cells could then be stratified by cell type, to look for cell type-protein associations.

Distances between IGFBP5 and synovial cell types reflected its lining layer and lymphoid aggregate-predominant distribution identified in Figure 4.6 A. Indeed, $CD90^+$ sublining-restricted fibroblasts were found to be one of the most distance cell types (Figure 4.7 A). $MerTK^+$ lining macrophages were present in much closer proximity to IGFBP5 positive cells than $MerTK^-$ macrophages. Finally, T-cells and B-cells were located in relative proximity to IGFBP5 positive cells, with B-cells were significantly closer, positioned on average within 5 cells (assuming at 10μm cell radius).

Detected IGFBP7 positive cells were found in close proximity to all cell types, within 5 cells on average of all major cell types (Figure 4.7 B). Within this close proximity, the closest cell types remained perivascular-dominant cell types, whether it be vascular-associated $CD90^+$ fibroblasts or lymphocytes which localise to the vasculature-adjacent sublining. IGFBP7 was located furthest from macrophages regardless of phenotype.

COL6A1 was again found in relatively close proximity to most cell types (Figure 4.7 C). Amongst this, $MerTK^-$ macrophages and $CD90^+$ fibroblasts were the closest on average. Lymphocytes were the furthest cell type from COL6A1, corroborating the evidence from the subsynovial niche analysis (Figure 4.6 D).

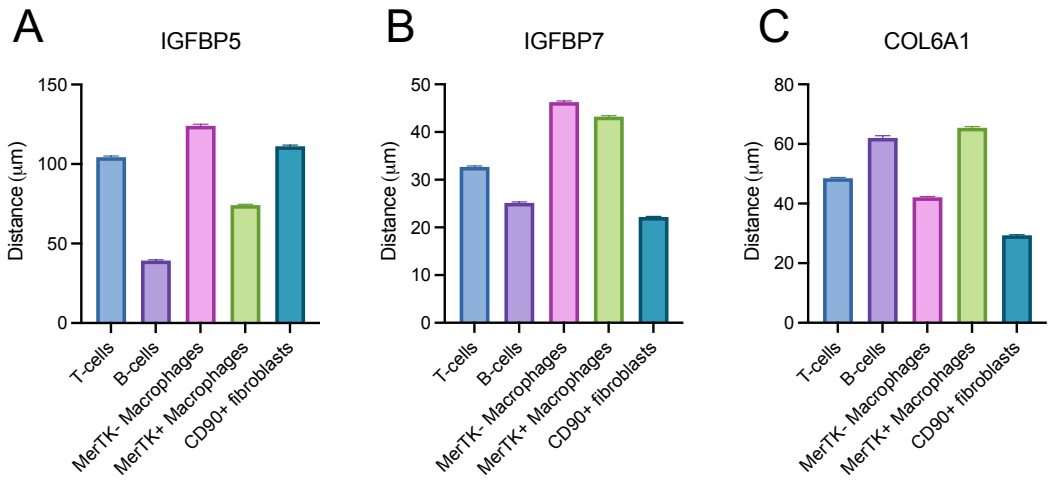


Figure 4.7 Synovial cell type distances to key matrix proteins. Distance between each detected cell in the 14 synovial sections of the dataset to the nearest cell positive for each key matrix protein was measured in QuPath. **A-C.** Bar plots of the distance of each pooled synovial cell in each of the identified cell types to the nearest cell positive for each key matrix protein. This analysis was performed and is reported for IGFBP5 (**A**), IGFBP7 (**B**), and COL6A1 (**C**). All comparisons were found to be statistically significant under two-way ANOVA with multiple comparisons testing.

4.2.7 IGFBP5 localises to the lining layer macrophages and fibroblasts

Having characterised the broad tissular distribution of each key protein, closer scrutiny was placed on each protein to look for potential functional insights. IGFBP5 was found and is reported in chapter 3 to have interesting differential expression dynamics between OA and RA, but also between disease activity statuses. Because of tissue availability, which informed dataset cohort composition, conservation of disease activity dynamics could not be assessed, while the differences between OA and RA could.

At the transcriptomic level, *IGFBP5* is a lining layer-predominant gene with higher expression in OA compared to RA (Figure 4.8 A,B&C). Automated unbiased analysis confirmed this lining layer-centric distribution (Figure 4.6 A), and OA samples were indeed found to have marginally higher levels of lining IGFBP5 than RA samples (Figure 4.6 E). Looking more closely at the sections, IGFBP5 does delineate the superficial outer layer of the synovium across sections (Figure 4.8 D-G). This distribution was found to be robust, both in RA (Figure 4.8 D&E) and OA sections (Figure 4.8 F&G). Protein level distribution and quantification therefore appear to corroborate the transcriptomic findings, solidifying IGFBP5 as a lining layer-expressed and deposited ECM glycoprotein. Its reported tissue repair and fibrosis associations prompt a deeper look into its fine positional relation to immune and tissue cells of the lining layer[273,274].

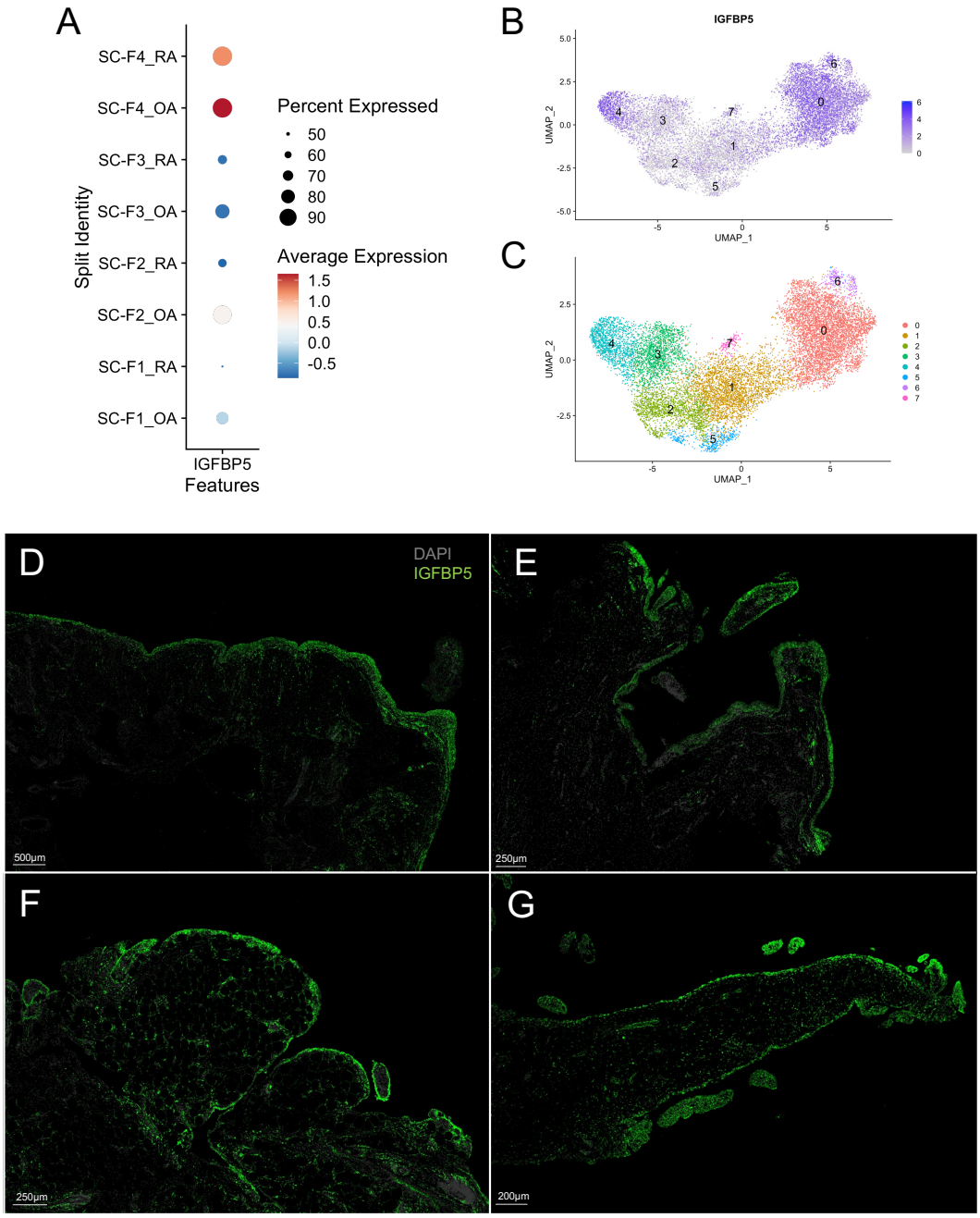


Figure 4.8 IGFBP5 is a lining layer expressed and deposited ECM glycoprotein. Immunofluorescent staining was used to validate the transcriptomic patterns of expression of IGFBP5 in OA and RA samples. **A.** Dot plot of the expression of IGFBP5 in each fibroblast subset of the AMP1 dataset, and split by condition between OA and RA. **B-C.** UMAP of the fibroblasts in the Alivernini dataset, coloured by *IGFBP5* expression (**B**) or by cluster (**C**). **D-G.** QuPath visualisations of immunofluorescent imaging of DAPI (grey) and IGFBP5 (green) in RA sections JRP099 (**D**) and JRP136 (**E**) (representative of n=9 RA sections), and S00293087 (**F**) and JRP144 (**G**) (representative of n=5 OA sections).

Unbiased analysis of distances between IGFBP5 and the annotated synovial cell types had identified MerTK^{+} macrophages as close proximal cells, notably when compared to MerTK^{-} macrophages. A closer investigation verified this, with IGFBP5 localising predominantly, yet not exclusively, to the neat layer of MerTK^{+} lining macrophages in synovial sections (Figure 4.9 A-C). This increased resolution adds value to and contrasts with the transcriptomic analysis, which described IGFBP5 as being restricted to the fibroblasts of the

lining layer. Here, lining fibroblasts in pink in Figure 4.9 A&E form an inner layer underneath the tightly organised lining MerTK^{+} macrophages (MerTK in cyan and CD68 in dark blue, Figure 4.9 A,B&D), in line with the reported structure of the synovial lining layer[207]. Within this, IGFBP5 appears to localise to the macrophages in the structure, and possibly to the interface between lining fibroblasts and macrophages.

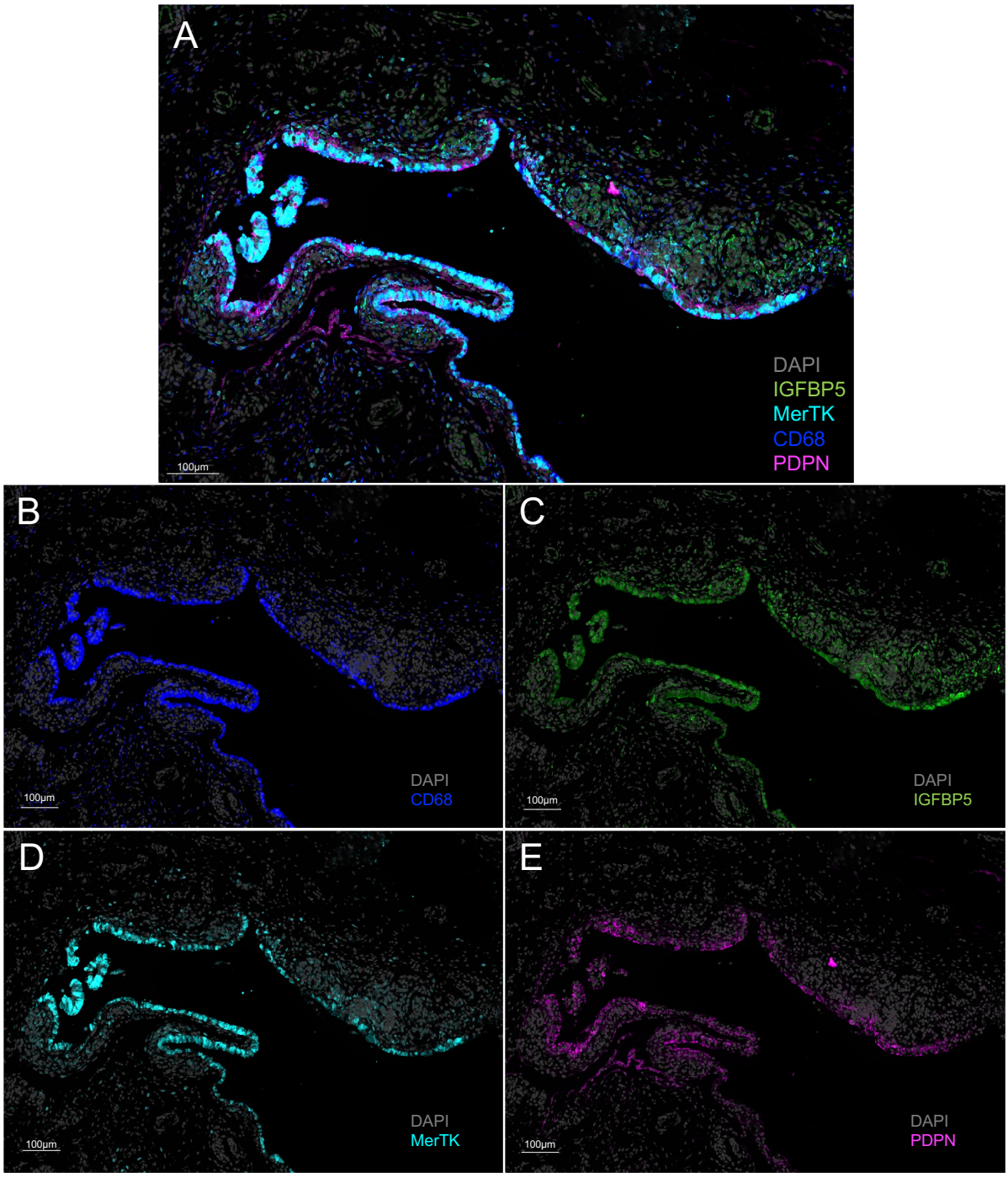


Figure 4.9 IGFBP5 distribution within the lining layer. Immunofluorescent staining was used to characterise the spatial distribution of IGFBP5 within the lining layer. **A.** QuPath visualisation of immunofluorescent imaging of DAPI (grey), IGFBP5 (green), MerTK (cyan), CD68 (blue), and PDPN (magenta) in RA section JRP122 (representative of n=14 sections). **B-E.** Single channel visualisations for key lining layer markers CD68 (**B**), IGFBP5 (**C**), MerTK (**D**), and PDPN (**E**).

4.2.8 IGFBP7 localises to the luminal side of synovial vasculature

At the transcriptomic level, *IGFBP7* appears to paint the exact opposite picture to *IGFBP5*. Indeed *IGFBP7* is highly sublining-restricted, with higher expression in RA than in OA samples (Figure 4.10 A). The previously described automated IGFBP7 protein detection was unable to confirm this, with high detected levels of IGFBP7 throughout the identified subsynovial niches (Figure 4.6 B). Disease type comparison even suggested a higher IGFBP7 positivity in OA lymphoid aggregate and lining layer cells compared to RA cells (Figure 4.6 F). Distance analysis on the other hand suggested that the sublining-restricted CD90+ fibroblasts were the most proximal cell type to IGFBP7$^+$ cells (Figure 4.7 B).

To tease these discrepancies apart, manual inspection of the stained sections was performed, as with IGFBP5. Protein deposition and detection of IGFBP7 was found to be neatly restricted to the vasculature across the 14 analysed sections (Figure 4.10 D-G). This observation was conserved between OA (Figure 4.10 D&E) and RA sections (Figure 4.10 F&G), with no clear differences based on disease type. This manual observation appears to corroborate the distance analysis (Figure 4.7 B), with the proximal CD90$^+$ fibroblasts being predominantly a perivascular cell type (Supplementary Figure 7 C).

Based on this relationship, I set out to examine more closely the position of IGFBP7 with regards to the detected vascular and perivascular cells and structures (Figure 4.11 A). I was able to neatly identify the basement membranes with COL4A1 (Figure 4.11 C) and basement-associated protein COL6A1 (Figure 4.11 B), but also perivascular CD90$^+$ fibroblasts (Figure 4.11 D). IGFBP7 localised to the inside of the structures defined by these markers, seemingly staining endothelial cells, and even lining the luminal surface of these cells (Figure 4.11 A&E).

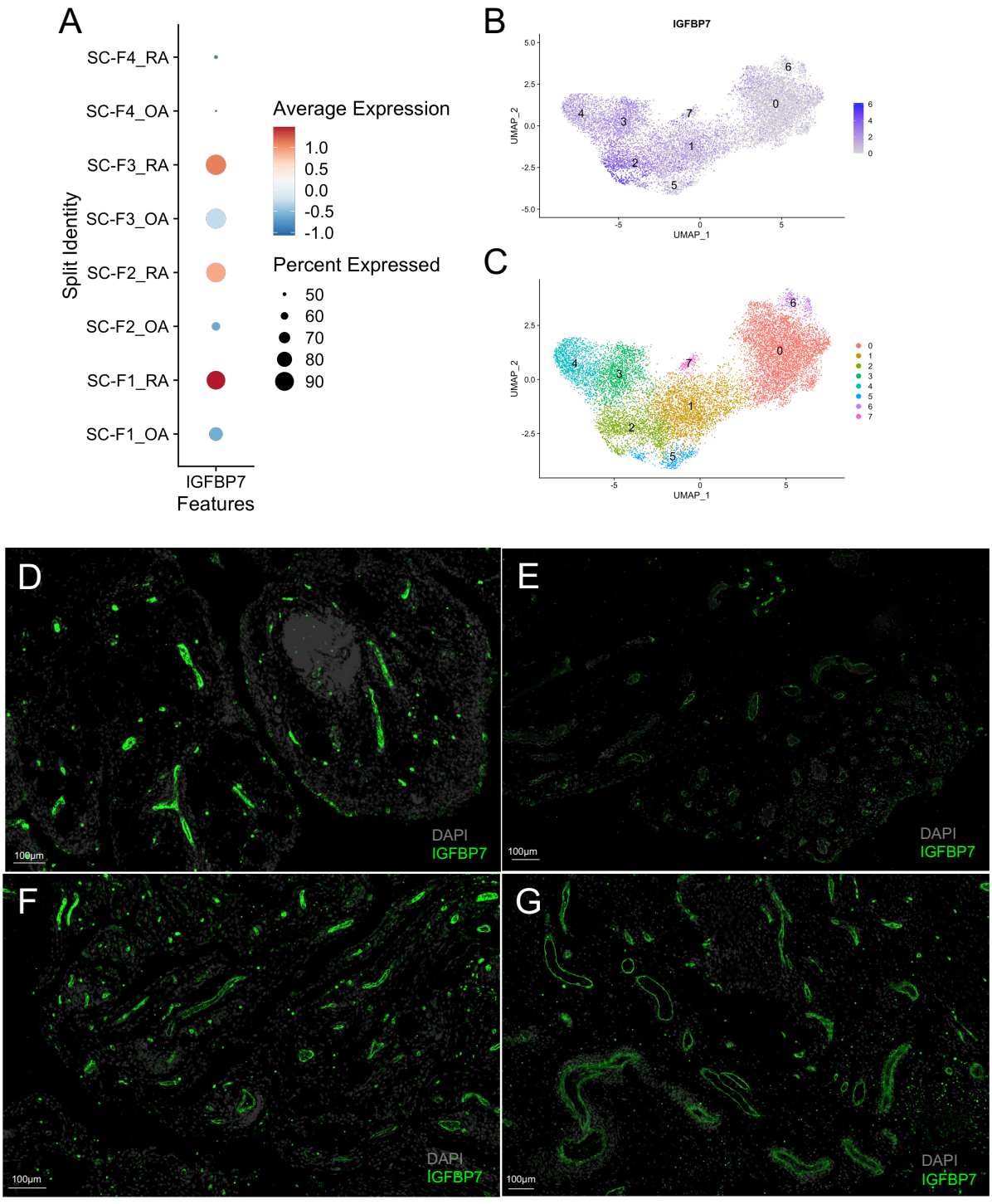


Figure 4.10 IGFBP7 is a sublining expressed but perivascularly deposited ECM glycoprotein. Immunofluorescent staining was used to validate the transcriptomic patterns of expression of IGFBP7 in OA and RA samples. **A.** Dot plot of the expression of IGFBP7 in each fibroblast subset of the AMP1 dataset, and split by condition between OA and RA. **B-C.** UMAP of the fibroblasts in the Alivernini dataset, coloured by *IGFBP7* expression (**B**) or by cluster (**C**). **D-G.** QuPath visualisations of immunofluorescent imaging of DAPI (grey) and IGFBP5 (green) in OA sections S00292623 (**D**) and 97SM (**E**) (representative of n=5 OA sections), and JRP141 (**F**) and JRP122 (**G**) (representative of n=9 RA sections).

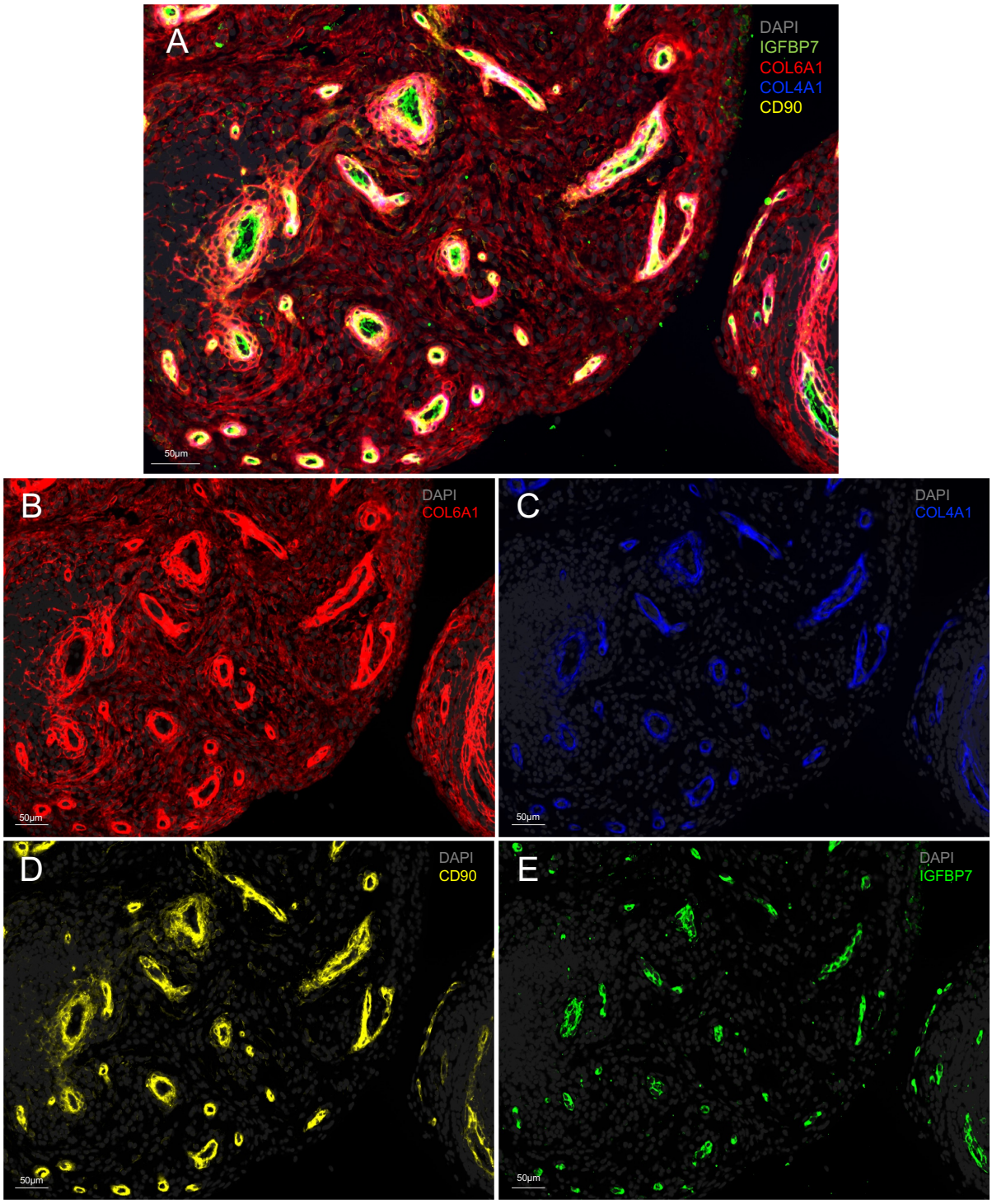


Figure 4.11 IGFBP7 distribution within the perivascular niche. Immunofluorescent staining was used to characterise the spatial distribution of IGFBP7 within the perivascular niche. **A.** QuPath visualisation of immunofluorescent imaging of DAPI (grey), IGFBP7 (green), CD90 (yellow), COL4A1 (blue), and COL6A1 (red) in RA section JRP122 (representative of n=14 sections). **B-E.** Single channel visualisations for key lining layer markers COL6A1 (**B**), COL4A1 (**C**), CD90 (**D**), and IGFBP7 (**E**).

4.2.9 T-cell aggregates localise to collagen VI-rich basement membrane vasculature

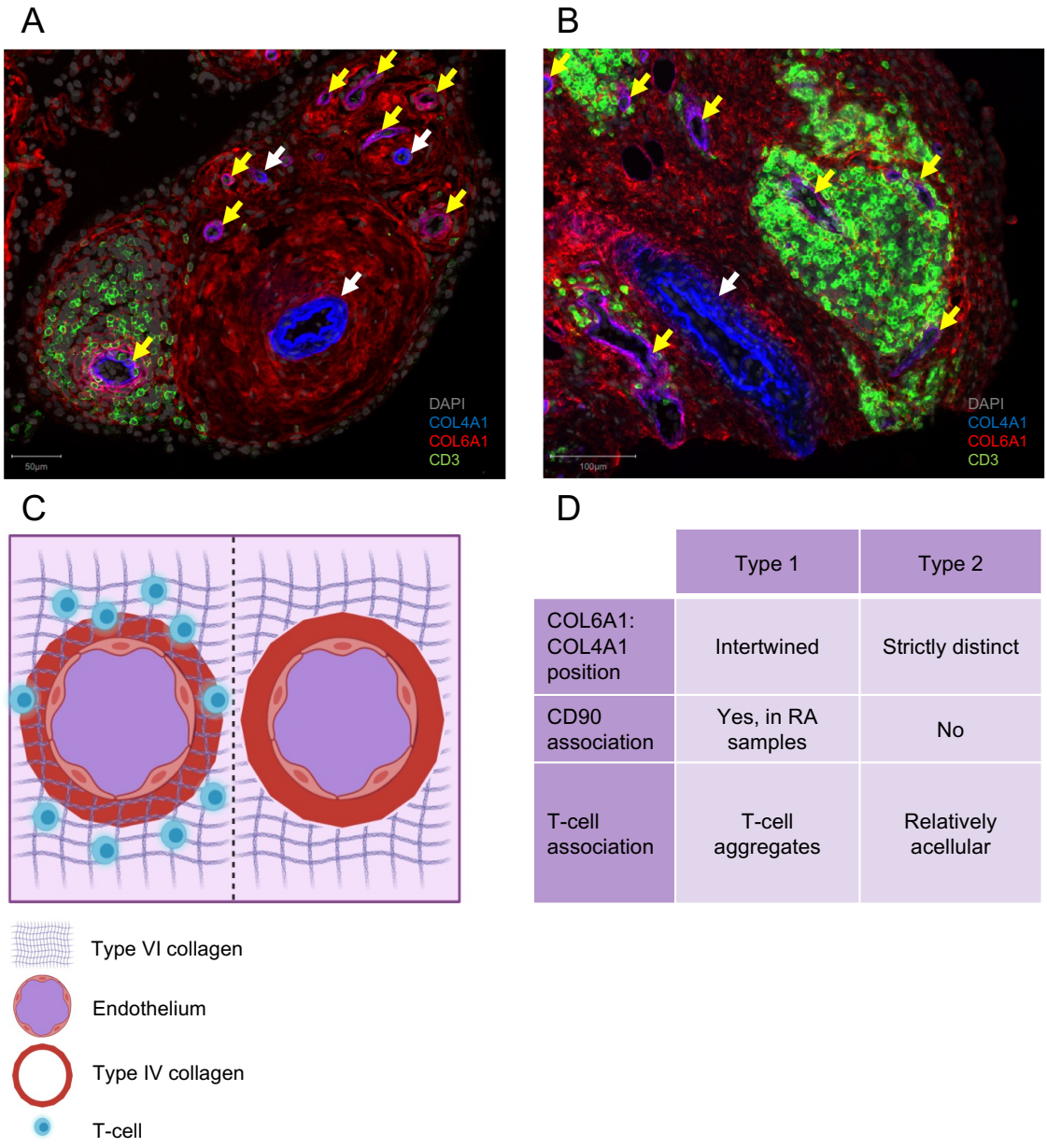
Transcriptomically, *COL6A1* was found to be potentially involved in basement membrane dysregulation in chapter 3. Automated protein level analysis highlighted high COL6A1 in the lining layer and sublining, and lower COL6A1 positivity in lymphoid aggregate cells (Figure 4.6 D). Distance analysis showed COL6A1 was close to most cell types, in particular to vascular $CD90^+$ fibroblasts (Figure 4.7 C).

While looking to investigate potential differences in basement membrane composition in the sections, a striking observation was made. Two types of vasculature could be observed based on the conjunction of COL4A1 and COL6A1 staining patterns (Figure 4.12). The first type, labelled type 1, appeared pink from the staining overlap between the red COL6A1 and blue COL4A1, while type 2 appeared bright blue from exclusive COL4A1 staining (Figure 4.12 A&B). When the $CD3^+$ T-cell marker was added to the visualisations, it appeared that the majority of $CD3^+$ T-cells would form aggregates around type 1 vasculature instead of type 2.

To verify that these were truly two subclasses of blood vasculature, and not a comparison between blood and lymphatic vasculature, a serial section was also stained for LYVE1, a marker of lymphatic endothelial cells (Supplementary Figure 8 A). In this serial section the two types of vasculature could still be observed, along with separate vasculature structures positive for exclusively LYVE1, confirming the blood vascular nature of both subtypes.

With this confirmed, I then set out to quantify this observation. Every single observable vasculature from every analysed section was manually annotated based on COL4A1:COL6A1 positivity, while remaining blind to surrounding CD3 positivity (Figure 4.13). A grading scheme was then created for T-cell aggregates. Grade 0 describing a vascular niche with no T-cells, Grade 1 a vascular niche with 5 or fewer T-cells, Grade 2 a vascular niche with an aggregate of more than 5 T-cells but smaller than 100µm in diameter, and Grade 3 a vascular niche with a T-cell aggregate larger than 100µm in diameter (Figure 4.13 A). The mean T-cell aggregate grade associated with each vasculature type for each section was then quantified. As

suspected, within a given section, type 1 vasculature mean grades were systematically higher than type 2 vasculature grades (Figure 4.13 B).



	Type 1	Type 2
COL6A1: COL4A1 position	Intertwined	Strictly distinct
CD90 association	Yes, in RA samples	No
T-cell association	T-cell aggregates	Relatively acellular

Figure 4.12 COL6A1 and COL4A1 define two vascular subtypes. Immunofluorescent staining was used to characterise the spatial distribution of COL6A1 and COL4A1 within vascular basement membranes. **A-B.** QuPath visualisation of immunofluorescent imaging of DAPI (grey), COL4A1 (blue), COL6A1 (red), and CD3 (green) in section JRP118 (**A**) and JRP122 (**B**) (representative of n=14 sections). **C.** Schematic model of the two vasculature subtypes. This model was produced in Biorender. **D.** Table of the differentiating parameters between the two vasculature subtypes.

When broken down and looking across sections, type 1 vasculature was found to contain a majority of grades 1, 2 and 3, while type 2 vasculature was primarily grade 0 (Figure 4.13 C). Indeed, most grade 0s were found to be type 2-associated, while grades 1, 2 and 3

showed a preference for type 1. Within individual sections, the differences were even more striking (Figure 4.13 D). Grade 0s were systematically associated with a higher percentage of a section's type 2 vasculature than type 1. Grade 2s and 3s were systematically associated with a higher percentage of a section's type 1 vasculature than type 2. Grade 1s appeared to be at the transition between these two presentations, with associations with a higher percentage of a section's type 1 vasculature than type 2 in all but 2 sections. Together these data contribute to the characterisation of a novel dichotomy in blood vasculature subsets, defined by differential collagen composition of their basement membranes, and associated in tissue with differential T-cell aggregation and potentially infiltration.

A

Grade	Description
0	No T-cells in PVN
1	5 or fewer T-cells in PVN
2	Aggregate of more than 5 T-cells and smaller than 100μm
3	T-cell aggregate larger than 100μm

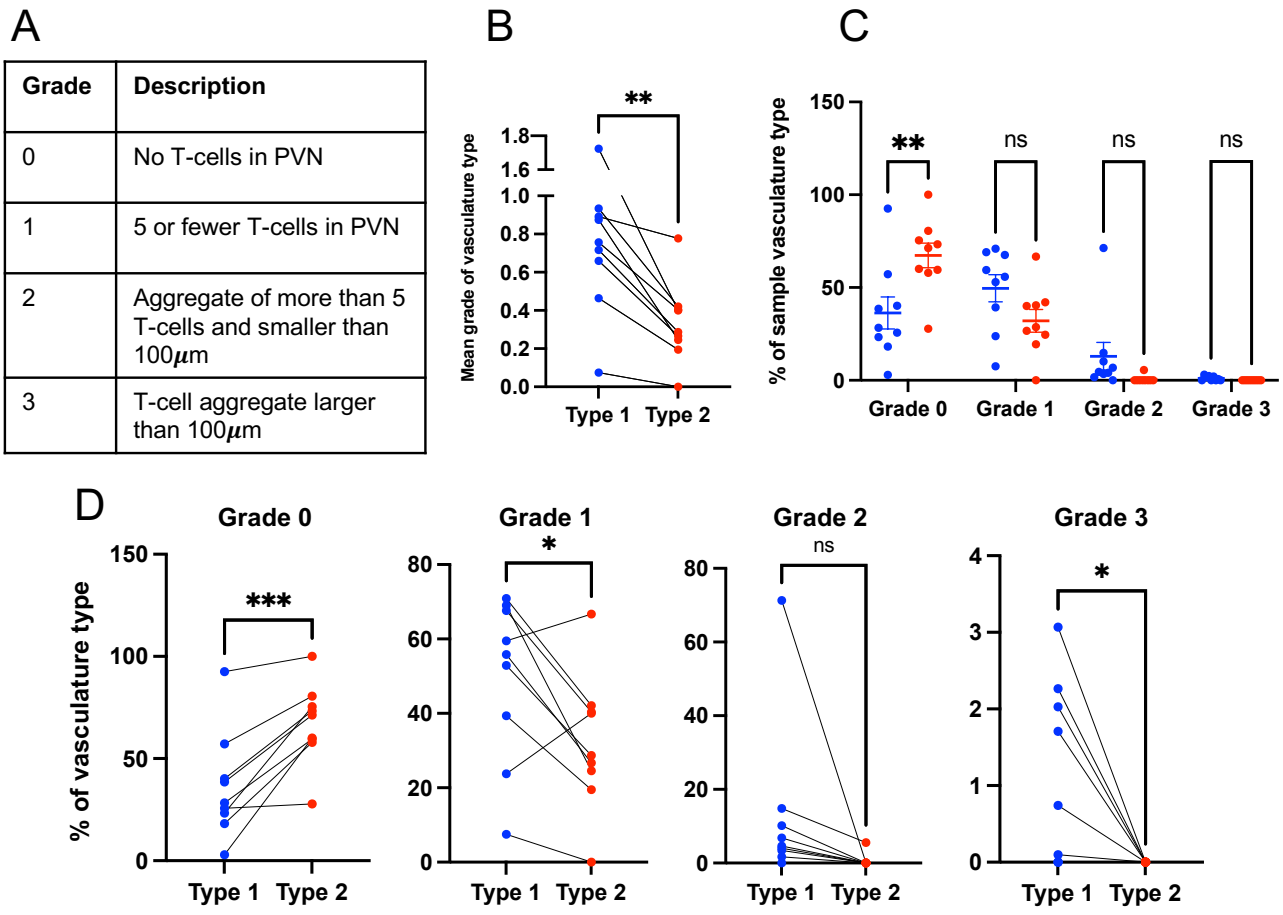


Figure 4.13 Vasculature subtypes have distinct associations with T-cell aggregate size. Each individual vasculature in each analysable section was classified manually into type 1 or type 2, while blinded to surrounding CD3 staining. Grading scheme for T-cell aggregates in the perivascular niche was then developed to score each individual vasculature. **A.** Grading scheme for T-cell aggregation around vasculature. **B.** Dot plot of the mean T-cell aggregation grade for each vascular subtype in each section. Each point in each condition represents one analysed section. Paired two-tailed t-test. **C.** Dot plot of the percentage of each vasculature subtype assigned to each grade, in each analysed section. Each point in each condition and each grade represents one analysed section. Two-way ANOVA with multiple comparisons. **D.** Dot plot breakdown of panel **C** by grade, with connecting lines between points representing the same section. Paired two-tailed t-tests. ns = not significant, * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001.

4.2.10 Perivascular collagen VI levels correlate with MerTK$^+$ macrophage abundance

In addition to its association with basement membrane dysregulation, the transcriptomics and protein-level data described so far also points to dysregulation of protein deposition in the RA sublining. scRNAseq data point to *COL6A1* being overexpressed in the sublining fibroblasts in RA compared to OA, and active RA compared to remission (Figure 3.12 E&F). Automated quantification of immunofluorescence showed a small increase in sublining COL6A1 in RA (Figure 4.6 D&H). I therefore wondered whether any associations between protein levels in the perivascular niche, the interface between vasculature and wider sublining, and key cellular players of the RA synovium existed.

To test this, I first annotated vasculature automatically using COL4A1 as a marker for it. These annotations were subsequently expanded in QuPath by a radius of 50µm, to define a new annotation representing the direct periphery of the vasculature, the perivascular niche. I then set up a pixel thresholder in QuPath, as opposed to a cellular one, to measure the percentage COL6A1 positivity of the perivascular niche area, but also of the lining layer and total tissue. I then ran Pearson correlations (with two-tailed p-values) in PRISM between these new COL6A1 area positivity metrics and the cell type frequencies in each section (Figure 4.14 A).

The first observation was one of correlative dynamics between the percentage COL6A1 positivity of the different subsynovial niches. Perivascular COL6A1 was found to correlate strongly with total tissue COL6A1 (Figure 4.14 B; Pearson's r = 0.789, p-value = 0.001), but negatively correlated with lining layer COL6A1 (Figure 4.14 C; r = -0.671, p-value = 0.012). Focusing on cell type frequency correlations, MerTK$^+$ macrophage frequency was found to be the most highly correlated individual cell type to perivascular COL6A1 levels (Figure 4.14 E; r = 0.69, p-value = 0.009). T-cells were also found to be somewhat correlated to perivascular COL6A1 (Figure 4.14 G; r = 0.40, p-value = 0.177), and strengthened the MerTK$^+$ macrophage association when combined together (Figure 4.14 F; r = 0.74, p-value = 0.004). This T-cell correlation could be observed between both the perivascular and total tissue levels of COL6A1 (Figure 4.14 A; perivascular r = 0.40, total tissue r = 0.40), while the

strength of the MerTK^{+} macrophage association was specific to perivascular COL6A1 (Figure 4.14 A; perivascular r = 0.69, total tissue r = 0.44).

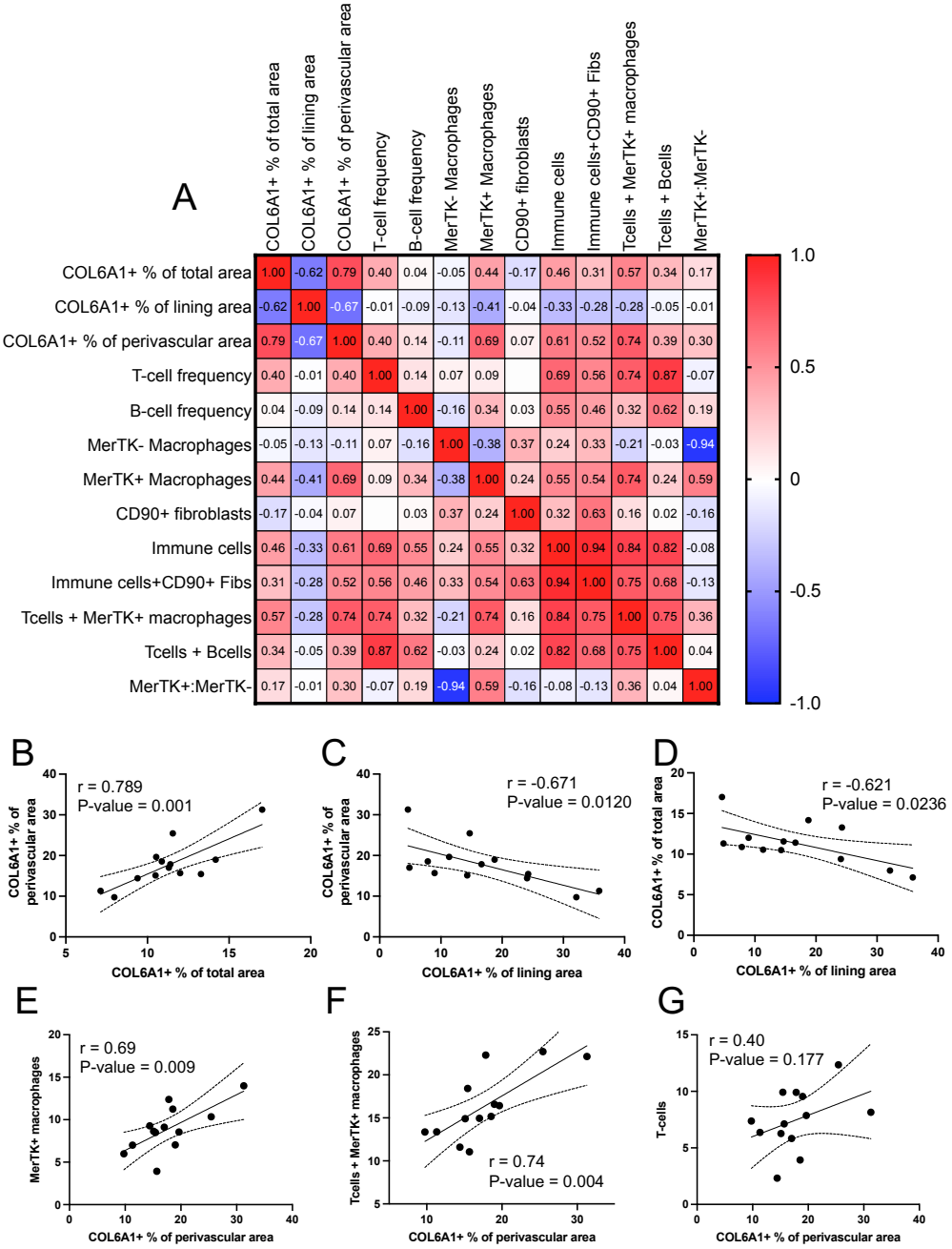


Figure 4.14 Correlations between COL6A1 niche positivity and cellular abundances. New synovial perivascular niches were defined by first automatically annotating vasculature based on COL4A1 signal in QuPath. Individual vasculature annotations were then expanded by a 50µm radius to define a new perivascular annotation. COL6A1 level were then measured using a pixel thresholder in QuPath, and were exported as the percentage of each defined area positive for COL6A1. Percentages from each section were then used to perform Pearson's correlation in PRISM, with two-tailed p-values reported. **A.** Pearson's correlation coefficient matrix for all compared variables. **B.** Correlation between COL6A1-positive perivascular area and COL6A1-positive total tissue area. Each point represents one section. Line represents simple linear regression and dotted lines represent the 95% confidence interval. **C.** Correlation between COL6A1-positive perivascular area and COL6A1-positive lining layer area. **D.** Correlation between COL6A1-positive total tissue area and COL6A1-positive lining layer area. **E.** Correlation between total section MerTK+ macrophage frequency and COL6A1-positive perivascular area. **F.** Correlation between total section T-cell and MerTK+ macrophage frequency and COL6A1-positive perivascular area. **F.** Correlation between total section T-cell frequency and COL6A1-positive perivascular area.

Manual examination of perivascular COL6A1 levels in the stained sections, with regards to MerTK positivity, validated this correlation. Tissue sections were present at both ends of a putative spectrum, with high perivascular COL6A1 and abundant MerTK^{+} macrophages on one end (Figure 4.15 A&B), and low perivascular COL6A1 and low MerTK^{+} macrophages on the opposite end (Figure 4.15 C&D). These data outline a correlation between perivascular collagen VI levels and MerTK^{+} macrophage frequency in synovial sections. The published roles for MerTK^{+} macrophages as anti-inflammatory and remission-associated may suggest, by extension, a link between collagen VI deposition and remission-associated tissular and cellular phenotypes, and are explored further in the discussion.

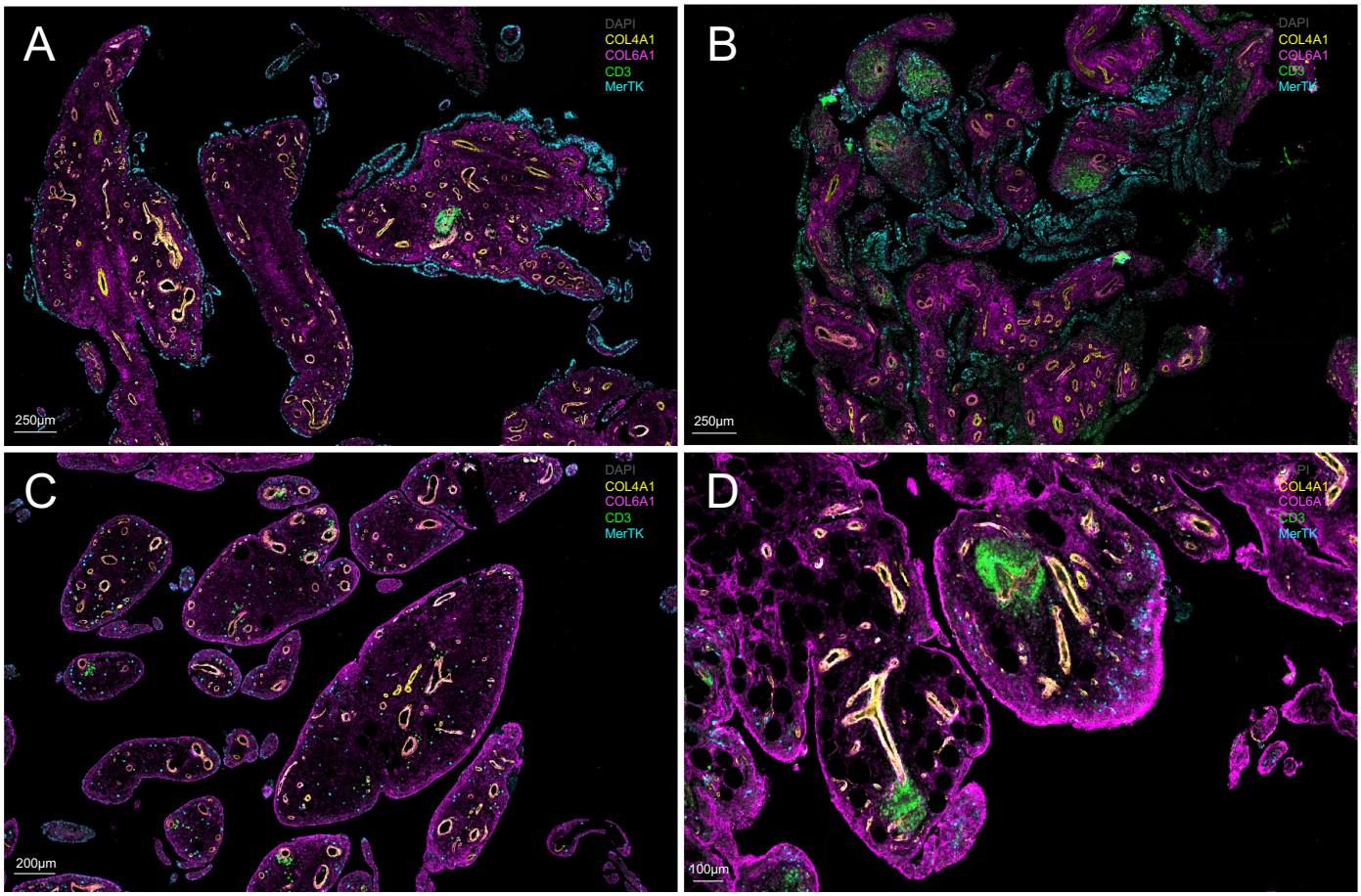


Figure 4.15 MerTK+ macrophage frequency and COL6A1 perivascular area correlate at the tissue level. A-D. QuPath visualisation of immunofluorescent imaging of DAPI (grey), COL4A1 (yellow), COL6A1 (magenta), CD3 (green), and MerTK (cyan) in sections JRP144 (**A**), JRP118 (**B**) with high perivascular COL6A1 and MerTK+ macrophages, and sections S00293087 (**C**) and S00293087 (**D**) with low perivascular COL6A1 and MerTK+ macrophages (representative of n=14 sections).

4.2.11 Lymphoid aggregates are restricted to collagen VI *dark zones*

Having identified an association between MerTK^{+} macrophages and collagen VI deposition, I set out to investigate the spatial patterns of collagen VI deposition further, particularly in relation to lymphoid aggregates. Lymphoid aggregates are key pathological elements of RA, and automated niche abundance analysis showed the lymphoid aggregate niche to be the poorest for COL6A1 (Figure 4.6 D&H).

Looking at tissue lymphoid aggregates specifically with the markers available in in the exploratory panel, COL6A1 was confirmed to be abundant throughout most of the sublining, also labelling some vascular basement membranes, but to be virtually absent in lymphoid aggregates (Figure 4.16). The absent COL6A1 appeared to form pockets, termed here *dark zones* (Figure 4.16 B, D&F), in which CD3^{+} T-cells and CD19^{+} B-cells were specifically constrained (Figure 4.16 A, C&E). This was observed across all the lymphoid aggregates, of all sizes, in all the 14 sections analysed.

This observation, together with the association between COL6A1 deposition and remission-associated macrophage phenotypes, may point to a functional role for collagen VI in directing the spatial patterning of infiltrating immune cells in the synovium and thereby regulating tissue-level inflammation.

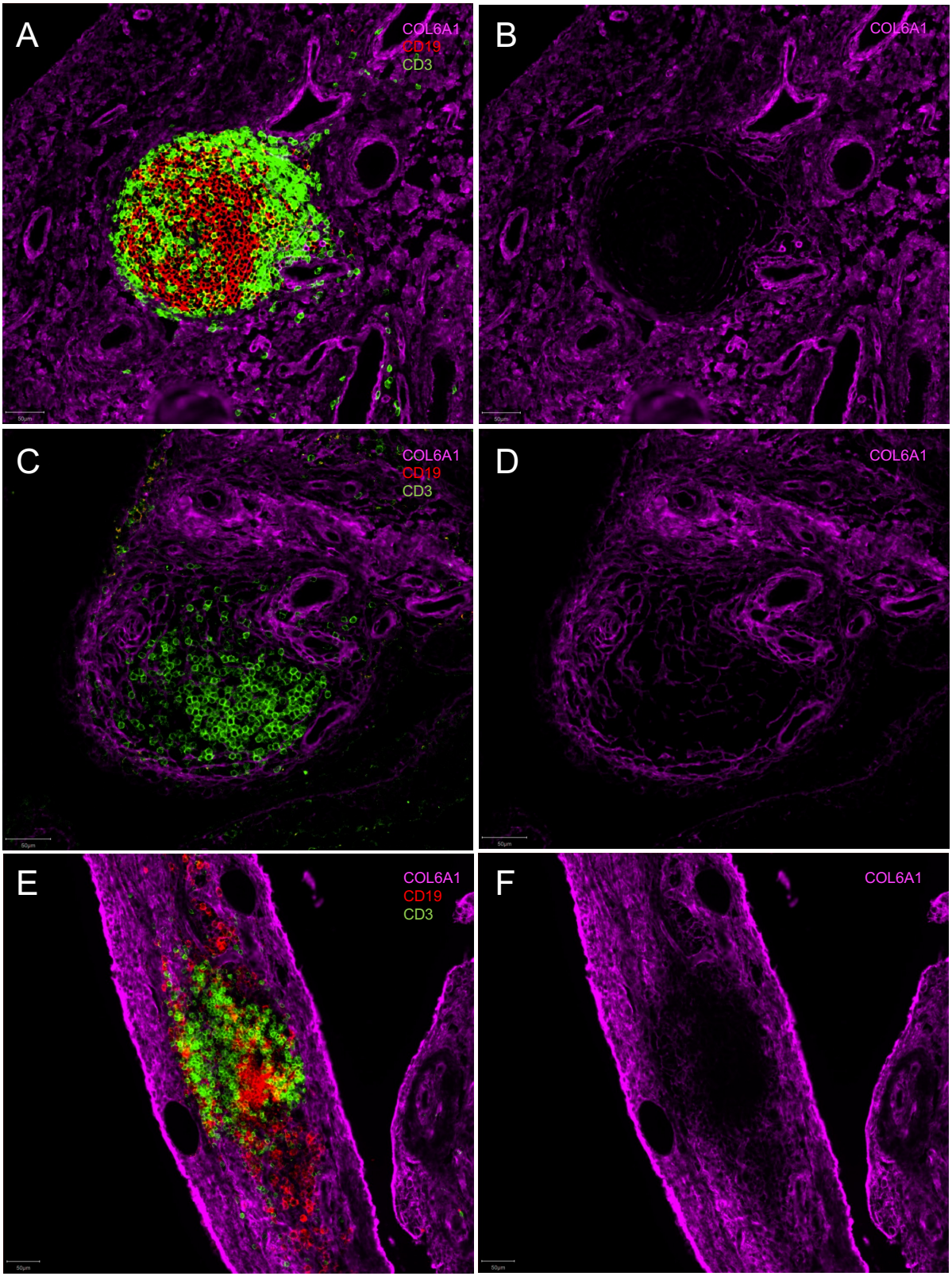


Figure 4.16 Lymphoid aggregates localise to COL6A1 *dark zones*. A-F. QuPath visualisation of immunofluorescent imaging of COL6A1 (magenta), CD3 (green), and CD19 (red) in sections JRP144 (**A&B**), JRP118 (**C&D**), and JRP122 (**E&F**) (representative of n=14 sections).

4.2.12 Collagen VI deposition is associated with inflammation resolution in mouse AIA

To further investigate the association between collagen VI deposition and remission-associated phenotypes, I turned to mouse models. These could provide insights into the chronology of COL6A1 deposition in the inflammatory and tissue repair processes.

The Antigen Induced Arthritis (AIA) model of inflammatory arthritis was induced by Dr. Barbora Schonfeldova in 6 mice, with an additional 3 control naïve mice, according to a reported method[74]. 3 mice were culled at peak inflammation on day 2, 3 mice were culled after inflammation resolution on day 14, and 3 naïve mice were used as controls prior to the induction of inflammation. Mouse knees were then processed, sectioned, and stained with a specific panel of antibodies to detect nuclei with SyTox Blue, neutrophils with anti-CD177, COL6A1, endothelial cells with anti-CD31, and macrophages with anti-CD68 (Table 2.9). Immunofluorescence was finally acquired on a Zeiss 980 Airyscan confocal microscope.

Naïve sections were found to recapitulate the canonical synovial tissue structure (Figure 4.17 A). Sparse yet regular vascularisation was observed in yellow amongst an abundance of adipocytes, the meniscus lodged between the heads of the tibia and femur, and an outer thin layer of muscle. No COL6A1 or neutrophils were detected in the synovium itself in naïve sections. COL6A1 staining could however be seen in the muscle layer, the meniscus, and peri-cellularly around the chondrocytes of the cartilage. Plenty of neutrophils could be seen in patches within the bones, in structures likely to be bone marrow. Day 2 peak inflammation sections were in stark contrast to the naïve sections (Figure 4.17 B). Tissue structure appeared severely disrupted, possibly by the abundant infiltration of neutrophils. Adipocytes looked to disappear in these sections, making place for the swarm of neutrophils, which was also associated with an apparent expansion of the synovial tissue size. A couple of early deposited COL6A1 flares could be observed in these tissue sections. As expected, Day 14 sections bared resemblance to both naïve and Day 2 sections, highlighting the kinetics of an attempt to return to homeostasis (Figure 4.17 C). The synovium in these sections appeared to remain increased in size, despite the absence of neutrophils having likely exited or been cleared from the tissue. Amongst all of this, extensive extracellular COL6A1

deposition remained, to an even greater extent than in Day 2 sections. Quantification through cell segmentation in QuPath confirmed these observations, with a peak in the percentage of synovial cells that were neutrophils on Day 2, an increased COL6A1 positivity on Day 2, and a peak in COL6A1 positivity on Day 14 (Figure 4.17 D). These data further strengthen the observation that collagen VI appears to be deposited during inflammation resolution, in both mouse and human synovium.

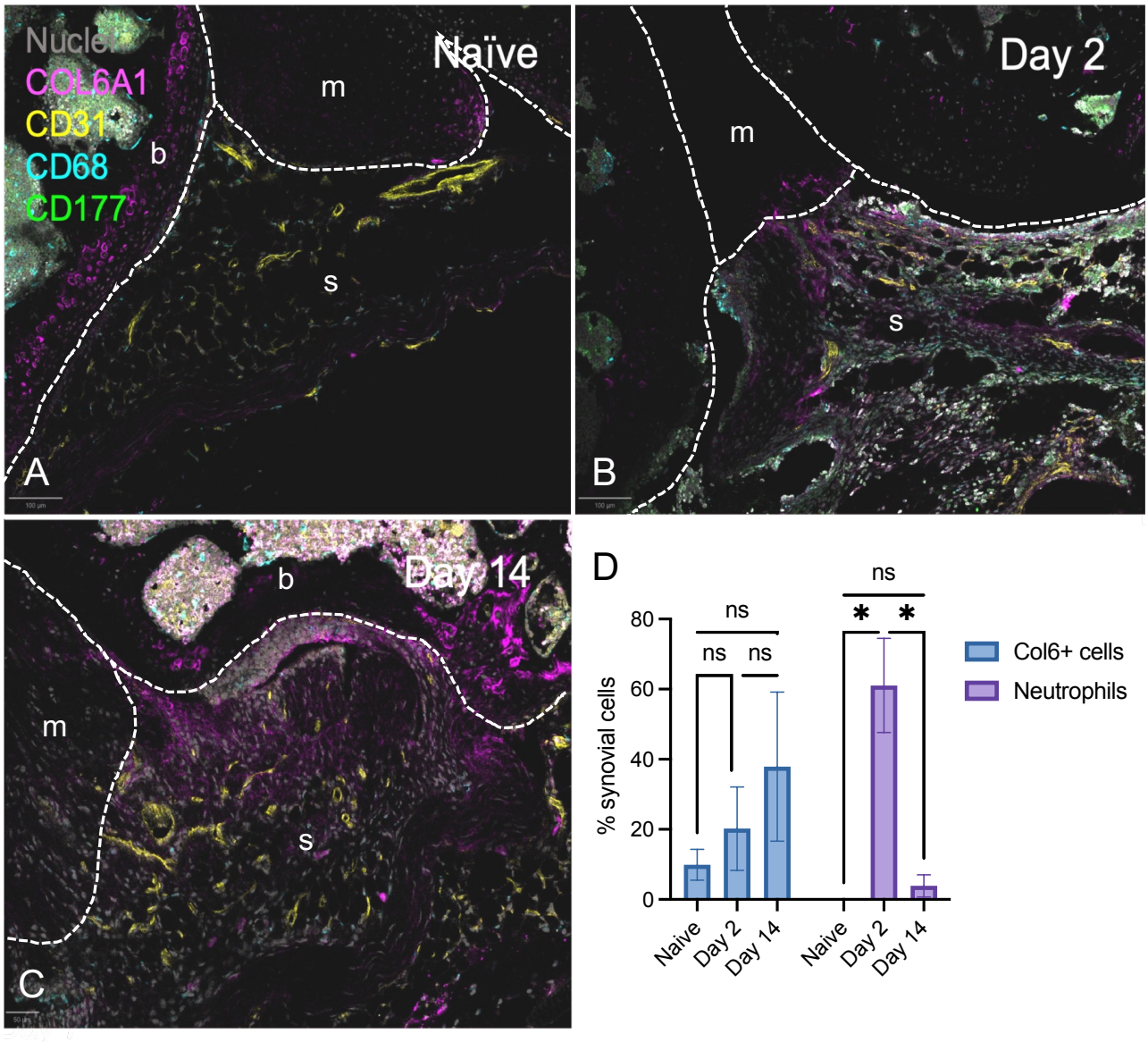


Figure 4.17 COL6A1 is highest in AIA resolution. The AIA model of inflammatory arthritis was induced in mice by Dr. Barbora Schonfeldova. Knees were processed, sectioned, stained for antibodies in Table 2.9, and imaged on a Zeiss 980 confocal microscope. **A-C.** QuPath visualisation of immunofluorescent imaging of SyTox Blue (grey), CD68 (cyan), COL6A1 (magenta), CD177 (green), and CD31 (yellow) in naïve (**A**), day 2 (**B**), and day 14 sections (**C**) (representative of n=3 sections per condition). **D.** QuPath quantification of the percentage of synovial cells defined as CD177+ neutrophils, or COL6A1+ cells in sections. Error bars are the standard error of the mean, statistical test is a two-way ANOVA with multiple comparisons. ns = non-significant, * = p-value < 0.05.

4.3 Technical discussion

4.3.1 Rationale and limitations of immunofluorescence sample selection

To complement the characterisation of the transcriptomic expression patterns of key matrisomal genes, sections from human patient biopsies were stained. Because the public datasets forming the basis of the analysis contained both OA and RA samples, I set out to stain sections from both OA and RA patients in an attempt to reinforce or disprove the identified transcriptomic associations. A strong part of the bioinformatic basis of this pipeline was also however the comparison between samples from patients with either active RA or RA in remission. Logistical and administrative obstacles prevented us from staining samples from such patients. A clear target for future studies and an improvement on this target discovery pipeline should be the incorporation of disease activity status-stratified patient samples.

To investigate the chronology of collagen VI deposition in tissue and to establish a model for future work, a mouse model of inflammatory arthritis was selected. The main three models of inflammatory arthritis used in modern research are Collagen-Induced Arthritis (CIA)[276], Serum Transfer Induced Arthritis (STIA)[277], and Antigen-Induced Arthritis (AIA)[74,278] because of their ability to model specific aspect of human inflammatory arthritis and their applicability. While all three models rely on similar immunisation and challenge stages, the speed of induction, peak inflammation, and resolution, the monoarticular nature of the inflammation, and the extensive characterisation of the immune infiltration linked to the challenge made AIA the ideal model in this case[74]. Section numbers and conditions were here selected based on availability but could be expanded in the future. Future work could for example focus on staining matched injected and non-injected knees. Additional time points could also be added in day 7 and day 21 depending on project license allowance to add a transition datapoint between the current day 2 and 14, and to investigate the stability of the increased collagen VI deposition past inflammation resolution. An interesting follow up study could also be the investigation of repeat challenges on the previously deposited collagen VI matrix, similarly to past studies using repeat injection of monosodium urate crystals[65].

4.3.2 Protein spatial distribution characterisation: technique and analysis selection

Multiplexed immunohistochemistry is amongst the fastest developing technological field in biomedical research, and therefore offers a plethora of potential methods to choose from[279]. Three of the main subcategories of methods are cyclic immunofluorescence in which rounds of IF antibodies are added and quenched, DNA barcoding in which DNA oligonucleotide-conjugated antibodies and their complimentary fluorescent oligos are added in rounds and quenched[39], and imaging mass cytometry in which metal ion-conjugated antibodies are used[38]. Because of the similar practical properties between the methods (time required for labelling and imaging and resolution), and importantly the availability on site of a Leica CellDIVE system, cyclic immunofluorescence was elected to take forward.

OA and RA sections were subsequently processed, stained, and imaged the OA and RA sections on the CellDIVE system. As one of the first projects to use this system, analysis pipelines had to be established. Analyses performed on sections stained with the exploratory panel of antibodies were performed using QuPath[200]. Because of its nature as an open source free software, QuPath benefits from a large breadth of in built and externally developed and shared analyses, which used in this study to create an extensive multiplexed immunofluorescence pipeline. One of the major limitations of QuPath however is its cell segmentation. QuPath cell segmentation is essentially only based on cell nuclei detection and expansion by a set metric to define cell “outlines”. This is appropriate in dense tissue areas with predominantly small and round cells such as lymphoid aggregates, but is less adequate in relatively acellular tissues or in defining less round cells such as fibroblasts and macrophages. In this study, QuPath cell detection was used based on the assumption that inflamed synovium is dense, but an obvious point of optimisation could be the incorporation of more advanced cell segmentation methods such as DeepCell’s Mesmer.

4.4 Biological discussion

4.4.1 COL4A1 staining is absent from the lining layer

COL4A1 was initially chosen as a key protein of interest because of its interesting expression in the SC-F2 and SC-F4 fibroblasts, putatively forming the basis of both the vascular basement membrane[246], and the lining basement-like membrane[247]. Strikingly however, COL4A1 staining was only observed in the sublining of the 14 synovial sections stained, and not in the lining layer (Figure 4.6 C). This appears to be a clear discrepancy compared to some of the published literature, which provides convincing evidence for deposited collagen IV in the lining layer[247,248]. The transcriptomic data presented in chapter 3 also appear to confirm the published works.

A few potential reasons can be found for this observation. The antibody used in this study is a monoclonal antibody (clone IV-4H12) raised specifically against the N-terminal 7S domain of COL4A1. One possible explanation could be different unreported conformations or fibre structures between the basement-like membrane and the basement membrane, leading to differential exposure of the antibody epitope. The different transcriptomic blueprints for these two structures, described in chapter 3, potentially hint at this. The absence of nidogens and presence of a greater abundance of netrins may lead to critical structural differences, which may lead to a different affinity of the antibody for the basement-like membrane compared to canonical basement membranes. The second potential explanation could be post translational modifications (PTMs). The 7S domain of COL4A1 is known to be glycosylated for example at Asn126[280]. Glycosylation status of a protein is known to affect antibody binding, and even completely abolish it in certain cases[281]. Differential PTM status of sublining and lining layer COL4A1 could therefore possibly explain this discrepancy.

Because of the role of the lining layer in synovial tissue homeostasis and its dysregulation in active disease, the basement-like membrane and its compositional dynamics are likely to be a critical importance. Further work should therefore be done to fully characterise the protein levels of the genes identified in this transcriptomic blueprint and their

spatial and temporal patterns. A first step towards this could be the use of different collagen VI antibodies, monoclonals targeting different epitopes and polyclonal antibodies, to resolve the dynamics of this elusive structure.

4.4.2 The roles of vascular IGFBP7

Despite their common role in inhibiting IGF1 signalling, the dichotomous expression of *IGFBP5* and *IGFBP7* between lining and sublining, OA and RA, and remission and active RA all respectively, hints at critical secondary functions. Transcriptomically, *IGFBP7* expression was found to be restricted to the sublining fibroblasts, and to be overexpressed in RA compared to OA and active RA compared to remission. Surprisingly however, automated protein quantification found high levels of IGFBP7 positivity across the lining, sublining, and lymphoid aggregate niches of the analysed sections (Figure 4.6 B). Distance analysis provided better resolution, identifying $CD90^+$ fibroblasts as the classified cell type in closest vicinity to $IGFBP7^+$ cells in the tissue (Figure 4.7 B). This was then confirmed manually, with IGFBP7 interestingly being predominantly found in vasculature, lining the luminal side of endothelial cells (Figure 4.11).

This spatial distribution in tissue neatly relates to the published literature surrounding roles for IGFBP7 in the modulation of angiogenesis. IGFBP7, also known as angiomodulin, appears to play a synergistic role with pro-angiogenic factor VEGF-A[282]. IGFBP7 also promotes vascular leakiness through loosening of endothelial cell tight junctions, and cellular infiltration[283]. These processes are mediated through αVβ3 integrin binding[283]. αVβ3 itself is essential to the angiogenic process[284], as well as having clear roles in activating latent TGF-β1 to promote cardiac myofibroblast differentiation[285], and activating macrophages through NFκB[286]. Together, these point to IGFBP7 being a critical regulator of chronic inflammation and angiogenic activation in the vascular and perivascular niche. The observation of its spatial localisation to RA and OA vasculature, coupled with overexpression at the transcriptomic level in RA and active RA make it an interesting target for further future investigation. These data

may potentially point to IGFBP7 as a tractable target to reduce the extensive neovascularisation and immune infiltration that is characteristic of RA.

4.4.3 IGFBP5 is a lining layer-deposited pro-fibrotic agent

Transcriptomically, *IGFBP5* was found to be predominantly restricted to the lining layer fibroblasts of both the AMP1 and Alivernini datasets and to be higher in OA than RA, and higher in remission than active RA. Protein level analyses corroborated these patterns, with the IGFBP5 deposition being lining layer-dominant and marginally increased in OA sections compared to RA sections (Figure 4.6 A&E). Looking more closely at its spatial distribution within the lining layer, IGFBP5 was found to colocalise with both fibroblasts and macrophages, with a slight preference for the latter (Figures 4.8 & 4.9).

These observations are of particular interest considering the reported associations between IGFBP5 and fibrosis. Indeed, IGFBP5 is found to be overexpressed in systemic sclerosis primary dermal fibroblasts[287], in idiopathic pulmonary fibrosis[274,288], and in atherosclerotic fibrotic cores[289]. Beyond mere associations, in vitro studies were also able to recapitulate this pro-fibrotic effect, with recombinant IGFBP5 shown to promote fibronectin, collagen, and αSMA production in mouse primary dermal[287] and lung fibroblasts[290,291]. Finally this effect was additionally confirmed in mouse models, where adenoviral-mediated overexpression of IGFBP5 in skin and lung lead to the deposition of a dense collagenous matrix[290,291]. Because of all these strands of evidence, the conceptual leap from “pro-fibrotic” to “pro-repair” has also been made in some studies, including in the context of the healing mouse gut in response to Dextran Sulfate Sodium-induced colitis[273]. IGFBP5 may well be playing a similar role in the synovial lining layer, possibly promoting the deposition ECM as part of the injury response, and thereby pushing the mechanical properties of the lining layer towards those of pseudo-fibrotic tissue.

While the extensive work on the effect of exogenous IGFBP5 on fibroblasts might hint at a synovial function, the lack of literature surrounding effects on macrophages make this an interesting additional avenue to pursue.

4.4.4 COL6A1 defines two vasculature subtypes with differential T-cell associations

Type VI collagen was selected as one of the key matrix proteins from the transcriptomic analysis based on its differential expression in the sublining fibroblasts of RA and active RA samples, as well as its interesting reported basement membrane association. Indeed, collagen VI is extensively reported to be associated to basement membrane structures in tissue[133]. This is notably due to its direct interactions with collagen IV, the primary basement membrane collagen subtype[292]. This association has also been described in mouse lung airway basement membranes, where its knockout leads to thickened and denser epithelium, leading to disrupted airway morphology[293]. Collagen VI therefore plays an important role in the composition of basement membranes and informing their morphology and effective functionality. Within this basement membrane association, collagen VI is currently thought of as promoting cell-ECM interactions[133].

Through multiplexed immunofluorescence, I identified two vasculature subtypes based on the composition of their basement membrane. Type 1 vasculature basement membranes were found to be rich in COL4A1 and COL6A1 (Figure 4.12), and to have a significantly increased frequency of infiltrated T-cells in their near vicinity (Figure 4.13). Type 2 were found to exclusively contain COL4A1 (Figure 4.12), and to have a reduced frequency of surrounding infiltrated T-cells. In light of the published roles of collagen VI in basement membranes, these observations could be suggesting a permissive role for collagen VI in the tissue infiltration of T-cells. This could potentially occur through morphological perturbations to basement membrane permeability, or through the potentiation of migration-promoting cell-ECM interactions. Collagen VI is for example known to bind to fibronectin[294], and fibronectin to promote T-cell migration[295].

Without further information on the chronology of collagen VI deposition in vascular basement membranes, it remains unclear whether type 1 vasculatures are the cause of the observed T-cell infiltration or a result of it. Collagen VI may indeed be deposited in basement membranes to prevent further cellular infiltration. Finally, these different vasculature basement membranes could also be the result of differences between arterial and venous endothelium.

Without good markers to segregate the two, this could not be explored here. Further work could however explore this by incorporating an antibody targeting membrane receptor Eph-B4, a venous endothelium-specific marker[296].

4.4.5 Perivascular COL6A1 correlates with remission-associated macrophages

In addition to basement membrane associations, automated COL6A1 detection in the three identified subsynovial niches found high levels of COL6A1 throughout the tissue, particularly in the lining and sublining, with lymphoid aggregates being least collagen VI-rich (Figure 4.6 D). This broadly corroborates the published literature on collagen VI deposition in human tissue. Through immunofluorescent labelling, collagen VI has indeed been shown to be virtually absent from normal synovium, localising exclusively to sparse vascular basement membranes, but to be extensively deposited in the lining and sublining of rheumatoid synovium[297]. This is in line with my human findings, but also my mouse data, where collagen VI was absent from naïve tissue and only deposited after inflammation induction (Figure 4.17). Few mouse synovium collagen VI data exist, with the exception of *Col6-cre* reporter mice, in which lining layer fibroblasts are labelled[207]. In the same model as the reporter studies my data, both at the transcriptomic and protein-level (Supp. Figure 6 and Figure 4.17), suggests increased expression in all fibroblasts and deposition in the lining and sublining. Further work could therefore be done in mouse tissue to unravel these discrepancies.

Looking specifically at the perivascular spaces of the sublining, a strong positive correlation was found between perivascular levels of COL6A1, and the frequency of MerTK$^+$ macrophages in a given section (Figure 4.14). In recent years, a significant amount of interest has developed around MerTK as a clinically relevant functional macrophage marker. MerTK was found to be a specific marker for a subset of synovial macrophages, predominantly of the lining layer, equipped with anti-inflammatory properties[66]. Indeed, MerTK$^+$ macrophages were found to secrete higher levels of IL-10 and lower levels of IL-1β, IL-6, TNF-α, and CCL2 in response to in vitro stimulation with LPS, compared to MerTK- macrophages[67]. MerTK$^+$ macrophages were found to be significantly more abundant in synovial tissue from patients in

remission, compared to patients with active RA. Higher MerTK$^+$ macrophage abundance was also found to correlate with sustained remission, while a lower abundance was linked to increased risk of flares[67]. MerTK$^+$ macrophage abundance may therefore play an important role in highlighting subclinical tissular dynamics amongst the cycles of active disease flares and remission that patients undergo. The strong correlation between perivascular COL6A1 positivity and this critically important macrophage subset therefore raises the question of factors driving this link. Beyond simple correlation, a few hypotheses for causative links could be explored from here.

Firstly, this association could be directly causative. In a model of peripheral nerve injury, Col6a1 knockout mice were shown to have impaired injury repair responses due to a lack of M2-polarised macrophages[298]. A putative model could therefore be that the deposited layer of collagen VI in the perivascular niche drives the polarisation of infiltrating monocytes towards anti-inflammatory MerTK$^+$ phenotypes.

Secondly, the association could be indirectly causative through different mechanisms. One such mechanism could be a broadly, yet unexplained, tolerogenic role. Indeed, COL6A1 has been shown to be a marker of poor prognosis in numerous cancer types, including cervical cancer, pancreatic cancer, triple negative breast cancer, and bladder cancer[176–178,299]. Additionally, certain subsets of T-regs have been shown to express COL6A1, with yet unknown functional importance[300]. Finally, another mechanism could be the spatial restriction of immune cells to collagen VI dark zones, identified in Figure 4.16. Collagen VI could be acting as an immune infiltration blockade, deposited in remission phases of the inflammatory course to put a stop to immune ingress and lowering overall tissue inflammation.

These two sets of hypotheses, the direct causative and indirect causative links were chosen as functional hypotheses to explore further because of their potential pathological and therapeutic implications.

Chapter 5 | Functional roles for type VI collagen in RA disease pathology

5.1 Chapter Introduction

Through the bioinformatic analysis and protein-level investigations described in Chapters 3 and 4, collagen VI was identified as a specific matrix target of interest. While transcriptomic upregulation of *COL6A1/2/3* was associated with RA compared to OA, and active RA compared to RA in remission, deposition of collagen VI protein positively correlated with a higher abundance of remission-associated MerTK$^+$ macrophages. Further staining in the AIA mouse model of arthritis corroborated these observations, with Col6a1 deposition occurring only after inflammation resolution. The distribution of COL6A1 throughout the synovial tissue was also found to be heterogeneous. Through the exploratory panel employed in Chapter 4, CD3$^+$ T-cells and CD19$^+$ B-cells were observed to be to COL6A1-poor regions, termed *dark zones*. The observational data laid out in Chapters 3 and 4 were therefore used to formulate two main hypotheses surrounding a potential role for type VI collagen in RA. The first was one of direct causation, with collagen VI directly polarising macrophages towards MerTK$^+$ phenotypes. The second, based on the spatial restriction of lymphocytes, was one of indirect causation, with collagen VI possibly restricting the tissue ingress of immune cells and thereby curtailing tissue inflammation. The aims for this chapter are therefore threefold. Firstly, to investigate a putative direct functional role of collagen VI in macrophage polarisation. Secondly, to refine and validate the spatial restriction hypothesis. Thirdly, to investigate the mechanisms behind the formation of collagen VI dark zones.

This chapter will therefore describe the investigations around a functional role for collagen VI deposition in the underlying pathology of RA. To test for a direct causative involvement of collagen VI in driving the remission-associated phenotypes of macrophages in tissue, human monocyte-derived macrophages (hMDMs) were exposed to different ECM substrates. Although a slight increase in MerTK positivity could be seen with 3D gel-like formulations of collagen VI, this effect was also observed with control Matrigel and could not be recapitulated through exposure to simple non-gel collagen VI protein. This direct causative hypothesis was therefore set aside for potential future reappraisal.

Through additional staining and the use of the validation CellDIVE panel, the restriction to dark zones of specific immune cell subsets, particularly lymphocytes, was confirmed across RA and OA samples. This spatial restriction was also observed in IBD and head and neck cancer (HNC) samples. In vitro adhesion assays were then used to demonstrate that immune cell lines do not adhere to collagen VI itself, and that collagen VI blocks the adhesion of monocytes to fibronectin. Early data showed this adhesion blocking effect could be recapitulated by the recombinant I-domain of CD11b, a structurally homologous protein to the collagen VI VWF-like domains. Preliminary migration assays suggest collagen VI may block the migration of immune cell lines, which may explain the observed spatial restriction of immune cells to the dark zones.

Finally, I sought to shed light on the formation of the collagen VI dark zones. The data hint at degradation of collagen VI being associated with this process, with collagen hybridizing peptide (CHP)-labelled active degradation rings lining the outer border of the dark zones. Amongst all stained cells, monocytes were found to line these degradation borders**.** Monocytes specifically were found to venture outside of the dark zones, and colocalized with CHP, suggesting a degradative activity which could be linked to the formation of the dark zones. Mass spectrometry-based proteomics was then used to confirm that collagen VI degradation products were more abundant in RA compared to OA, a pattern found to be specific to collagen VI compared to other collagens. D11, one of such collagen VI fragments, cleaved by monocyte-expressed MMP9, was found to be detectable in RA synovial tissue and to be permissive to monocyte adhesion.

Overall, the work presented in this chapter points to a role for collagen VI deposited by fibroblasts in response to inflammatory insults in limiting tissue ingress of immune cells. The mechanisms underlying this migration blockade remain unclear and are explored further in the discussion section of this chapter. Monocytes were found to be the primary culprits in the expansion of dark zones through their degradative potential. The fragments released by monocytes hold distinct bioactive properties compared to full length collagen VI, which should be explored further.

5.2 Results

5.2.1 Collagen VI alone is unable to induce MerTK upregulation in macrophages

The first aim of my functional investigations into collagen VI in RA was to explore a direct causative link between collagen VI and the polarisation of macrophages towards MerTK$^+$ phenotypes. To do so, human PBMCs were isolated from blood cones, from which CD14$^+$ monocytes could then be purified. Monocytes were plated and differentiated into human monocyte-derived macrophages (hMDMs) using M-CSF (Figure 5.1 A). These hMDMs were then plated onto different ECM substrates or with different soluble factors and incubated for up to 24h depending on the experimental set up. Finally, these polarised hMDMs were harvested and stained for MerTK to assess their positivity for the remission-associated marker via flow cytometry.

Conscious of the potential impact of different technical factors on the expression of MerTK, the media type and cell harvesting methods were optimised prior to the testing of a larger set of substrates and soluble factors (Figure 5.1 B-E). DMEM-based media was found to yield greater levels of MerTK expression (reported here as Mean Fluorescence Intensity or MFI) compared to RPMI (Figure 5.1 B). Both types of media were found to preserve cell viability to similar extents (Figure 5.1 C). Cell lifting methods showed greater disparities, with TryplE-based lifting proving to maintain the highest levels of both MerTK positivity and cell viability compared to cold PBS and cell scrapers, with the latter performing the worst in both metrics (Figure 5.1 D&E). The DMEM-based media and TryplE-based cell lifting were therefore chosen to take forward. Based on their published effects on MerTK positivity of macrophages, LPS and dexamethasone were chosen as negative and positive controls respectively[301,302].

The temporal dynamics of MerTK expression on macrophages were then assessed. hMDMs were plated on either plastic or on a 3D gel formulation of type VI collagen and harvested at either 6h, 12h, 18h, or 24h (Figure 5.1 E). While unstimulated cells plated on plastic exhibited stable MerTK levels throughout the time points, with a slight decrease at 24h,

cells plated on collagen VI had increasing MerTK levels reaching a peak between 18h and 24h. Coinciding with a decrease in unstimulated levels of MerTK, the 24h timepoint showcased the largest delta between the two conditions and was therefore selected as the experimental timepoint to provide the best resolution between tested conditions.

Finally, hMDMs from up to 8 donors were tested under a number of conditions for effects on MerTK positivity and based on the optimised experimental parameters (Figure 5.1 G). Of note, not all conditions could be tested for every donor due to low cell numbers and a subsequent requirement for condition rationalisation. As throughout previous optimisation experiments, LPS treatment for 24h was found to reduce cell-surface MerTK on hMDMs compared to unstimulated controls (MFI Fold Change of 0.412, adjusted p-value = <0.0001), as reported in the literature[303]. The 3D gel-like formulation of collagen VI induced a modest yet non-significant increase, with a MerTK MFI-FC of 1.291 (adjusted p-value = 0.0586). Matrigel and a gel-like formulation of rat tail collagen I were found to induce similarly slight increases on MerTK positivity compared to unstimulated controls, with respective MFI-FCs of 1.191 and 1.145 (respective adjusted p-values = 0.6487 and 0.9108). Dexamethasone, a known inducer of MerTK upregulation, generated a greater increase in MerTK (MFI-FC = 1.439, adjusted p-value = 0.0258), but was highly variable across donors. Finally, a non-gel formulation of collagen VI did not have any effect on MerTK expression (MFI-FC = 0.978, adjusted p-value = >0.9999).

These findings point to a potential role of 3D ECM scaffolds in driving MerTK upregulation in hMDMs. Non-gel collagen VI did not induce MerTK upregulation, while gel collagen VI induced a marginally, yet non-significant, greater increase than Matrigel and collagen I.

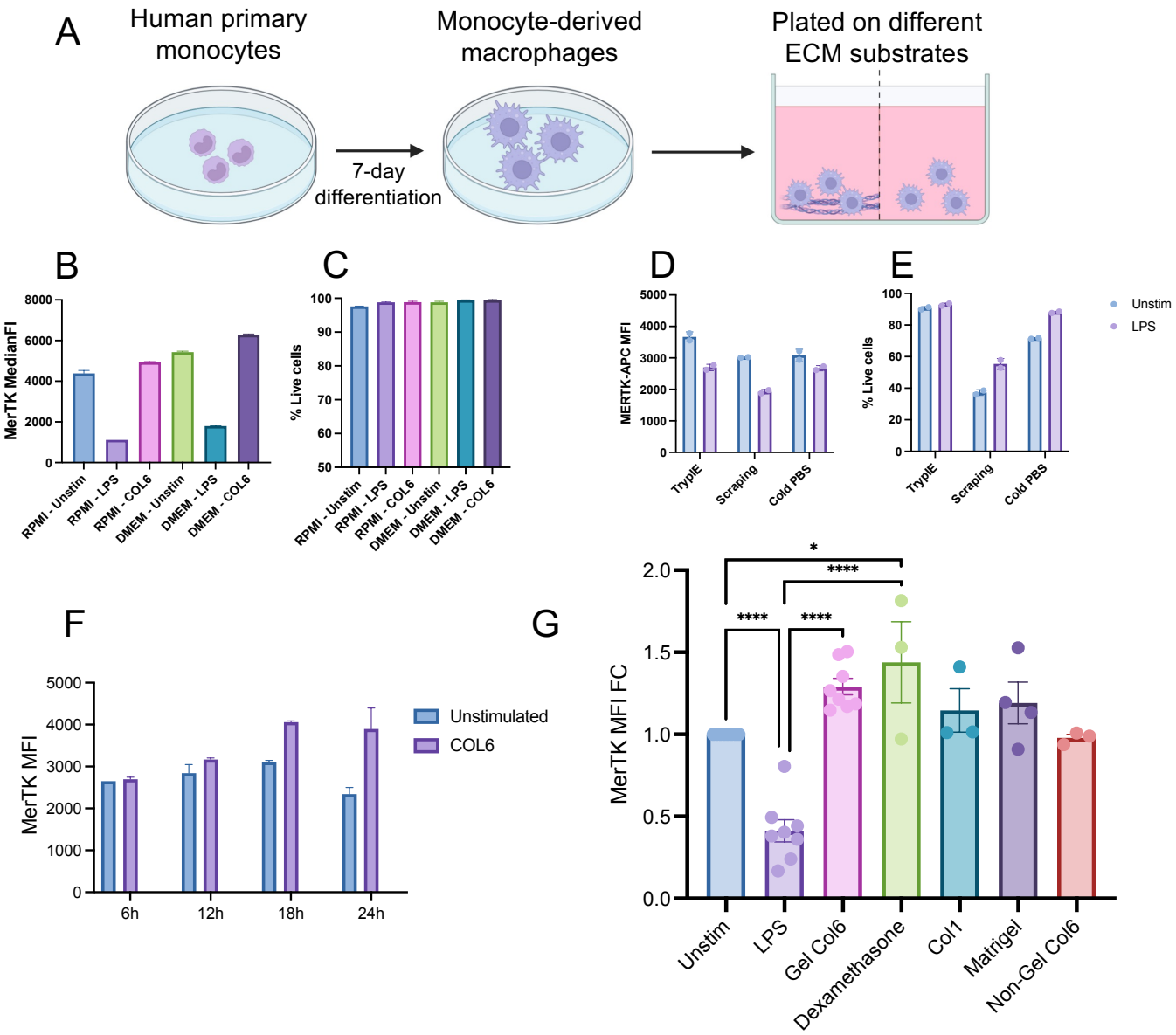


Figure 5.1 Human monocyte-derived macrophage culture and MerTK polarisation optimisation and experimental conditions. Monocytes were isolated from blood cones and differentiated into hMDMs using M-CSF for 7 days. After 7 days, the macrophages were harvested and plated onto different substrates or with different soluble factors including plastic, LPS, gel collagen VI, dexamethasone, gel collagen I, Matrigel, or non-gel collagen VI. Cells were then harvested after 24h of stimulation analysed via flow cytometry to assess MerTK positivity as mean fluorescence intensity (MFI). **A.** Schematic created in BioRender of the experimental workflow. **B.** Bar plot of MerTK MFI in media optimisation experiment, testing unstimulated, LPS-stimulated, and gel collagen VI-stimulated hMDMs in RPMI or DMEM-based media after 24h. Assay run with one donor in duplicate. **C.** Bar plot of live cell percentages in media optimisation experiment, testing unstimulated, LPS-stimulated, and gel collagen VI-stimulated hMDMs in RPMI or DMEM-based media after 24h. Assay run with one donor in duplicate. **D.** Bar plot of MerTK MFI in cell lifting method optimisation experiment, testing unstimulated and LPS-stimulated hMDMs lifted with TrypIE, scraping, or cold PBS after 24h. Assay run with one donor in duplicate. **E.** Bar plot of live cell percentages in cell lifting method optimisation experiment, testing unstimulated and LPS-stimulated hMDMs lifted with TrypIE, scraping, or cold PBS after 24h. Assay run with one donor in duplicate. **F.** Bar plot of MerTK MFI in time-course experiment, testing unstimulated and gel collagen VI-stimulated hMDMs lifted after 6h, 12h, 18h, or 24h. Assay run with one donor in duplicate. **G.** Bar plot of MerTK MFI Fold Changes (MFI-FC) induced by tested conditions compared to unstimulated cells. MFI-FCs are calculated per donor and therefore normalised by donor. Total donors n = 8, with n = 3 for dexamethasone, collagen I, and non-gel collagen VI, and n = 4 for Matrigel because of lower cell numbers. Statistical test is a one-way ANOVA with multiple comparisons via Tukey tests. All error bars are SEM.

5.2.2 Collagen VI dark zones harbour most tissue immune cell subsets

Unable to prove a direct effect for collagen VI in polarising macrophages towards MerTK$^+$ phenotypes, I turned to my second hypothesis, a role for collagen VI in the spatial restriction of immune cells. To explore this, I expanded my exploratory CellDIVE panel into a validation panel (Table 2.3.2), which included more immune subset markers including CD4 and CD8 to differentiate T-cell subsets, CD20 and MZB1 to differentiate B-cells from plasmablasts, CD14 to identify monocytes, CD66b to identify neutrophils, and CD11c to pick out dendritic cells. 3 OA and 3 RA sections were stained with this immune panel. Similarly to the implementation of the exploratory panel, collagen VI dark zones could be identified in all stained sections, both in RA and OA (Figure 5.2 A&D). Looking specifically at the lymphocyte markers in the panel, CD4$^+$ T-cells and CD20$^+$ B-cells were strictly restrained to the dark zones (Figure 5.2 B&E). CD8$^+$ T-cells and MZB1$^+$ plasmablasts localised primarily to these dark zones, but were less restricted in their localisation. DCs were fully restricted to the dark zones, colocalising with CD4$^+$ T-cells and CD20$^+$ B-cell aggregates (Figure 5.2 C&F). CD66b$^+$ neutrophils were less restricted to the dark zones. Macrophages were found outside of the dark zones, yet also outside of the deep interstitial tissue where most of the collagen VI deposition was visible, sitting on the periphery of the tissue. Monocytes were unrestricted by the collagen VI dark zones, with cells both in collagen VI-rich and collagen VI-poor regions of the tissue (Figure 5.2 C&F). Together, these observations made in all 6 of the stained RA and OA sections suggest a correlation between collagen VI deposition and immune spatial patterning. The different predilections of different immune cell subsets may suggest a role for collagen VI as a selective sieve for immune cells.

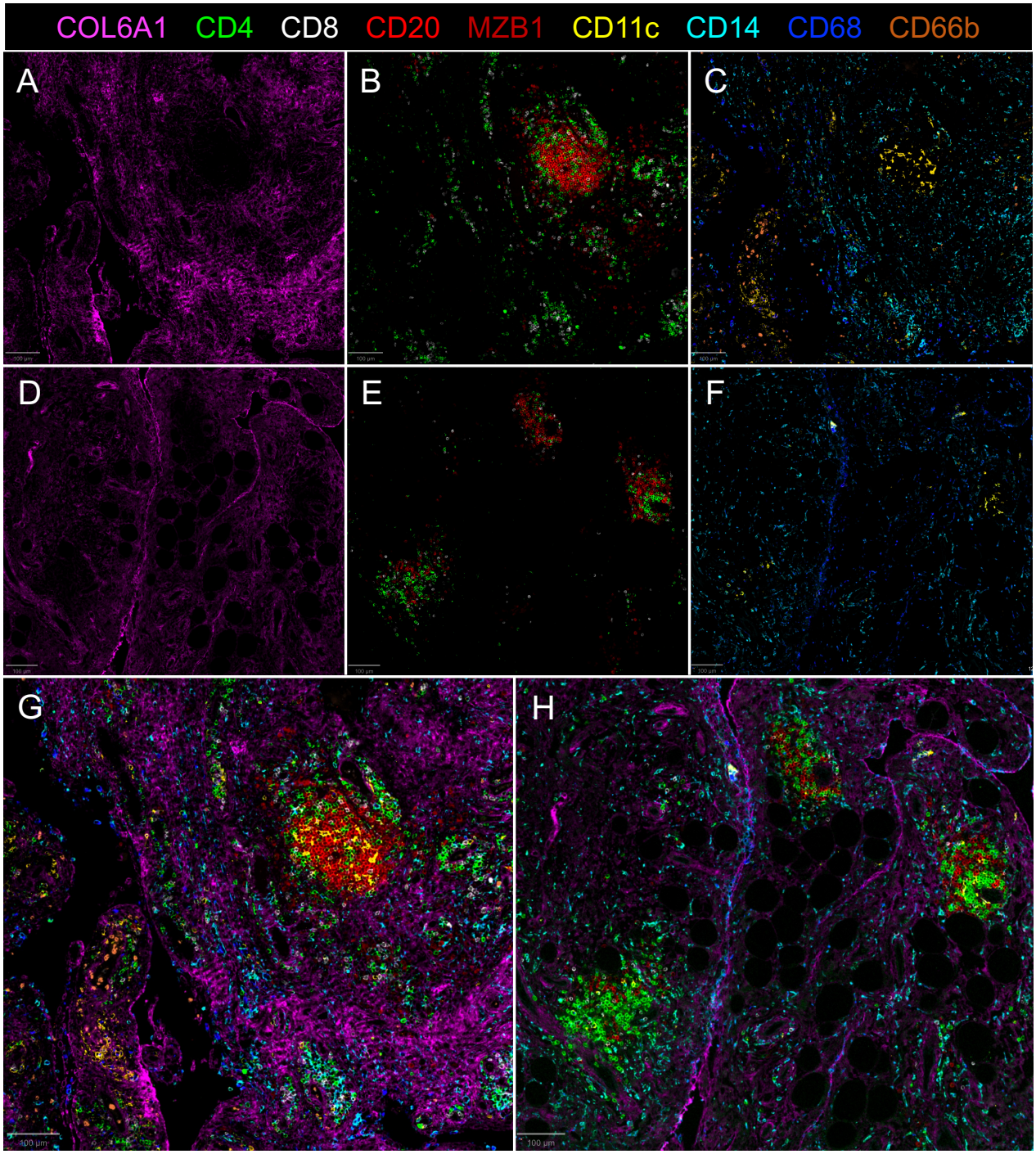


Figure 5.2 Immune subset spatial patterning in relation to collagen VI deposition in OA and RA. Immunofluorescent staining was used to characterise the spatial distribution of immune cell subsets in relation to COL6A1.QuPath was used to visualise immunofluorescent imaging of COL6A1 (magenta), CD4 (green), CD8 (white), CD20 (red), MZB1 (maroon), CD11c (yellow), CD14 (cyan), CD68 (blue), and CD66b (orange) in RA section JRP141 (representative of n=3 sections), and OA section S00292623 (representative of n=3 sections). **A.** Visualisation of COL6A1 and its dark zones in RA. **B.** Visualisation of lymphocyte markers CD4, CD8, CD20, and MZB1 in RA. **C.** Visualisation of myeloid and granulocyte markers CD11c, CD14, CD68, and CD66b in RA. **D.** Visualisation of COL6A1 and its dark zones in OA. **E.** Visualisation of lymphocyte markers CD4, CD8, CD20, and MZB1 in OA. **F.** Visualisation of myeloid and granulocyte markers CD11c, CD14, CD68, and CD66b in OA. **G-H.** Merged visualisations of COL6A1, lymphocyte, myeloid and granulocyte markers in RA (G) and OA (H).

5.2.3 Collagen VI dark zones are found cross-disease

Given the observation of dark zones in both RA and OA, I set out to investigate their presence in other disease settings. This would notably clarify the nature of this process as a synovial-specific mechanism or a cross-disease and cross-tissue one. Another IMID associated with extensive tissue and ECM remodelling is Inflammatory Bowel Disease (IBD)[304]. Sections from patients with Crohn's disease were processed and stained by Dr. Mathilde Pohin for COL6A1 and CD4, as previously described, and visualised in QuPath (Figure 5.3 A&B). Collagen VI dark zones could be observed, with immune cells labelled by CD45 displaying the same predilection for these areas. Tumour resections from patients with head and neck cancer (HNC) were also put through the same processing, staining, and analysis regimen by Dr. Caroline Morrell (Figure 5.3 C&D). Here too, collagen VI dark zones could be observed to concentrate most of the immune infiltrate found in the tumours. These data provide evidence for a cross-disease and tissue potential role for collagen VI in spatial immune patterning.

5.2.4 Collagen I does not form dark zones

To evaluate the impact of collagen VI as this immune spatial patterning factor, specificity had to be assessed. In addition to COL6A1, COL1A1 was also included in the validation panel. Collagen I is the most abundant and widely expressed collagen subtype, and is overexpressed in RA and active RA similarly to collagen VI (Chapter 3). Its protein-level tissue deposition pattern however was found to be different to that of collagen VI. Collagen I was predominantly deposited in the perivascular niche itself, surrounding collagen IV basement membranes, and importantly was found to be present in the collagen VI dark zones (Figure 5.4). This observation was reproducible across the 3 RA and 3 OA tissue sections stained. The distinct niches defined by collagen I and VI reinforce the specificity of the collagen VI dark zones, and their effect on immune restriction.

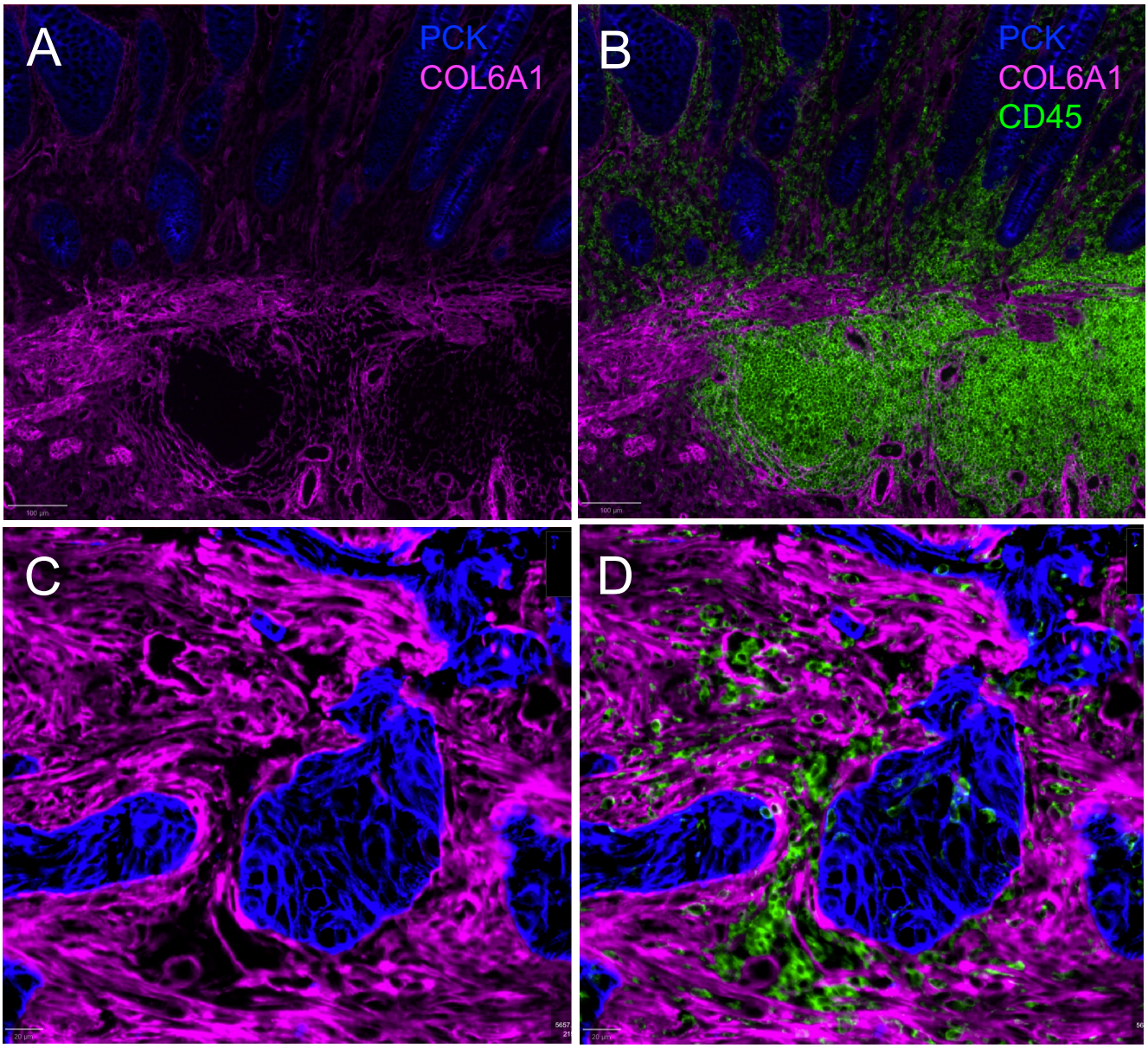


Figure 5.3 Collagen VI dark zones are present in IBD and head and neck cancer patients. Immunofluorescent staining was used to characterise the spatial distribution of immune cells in relation to COL6A1.QuPath was used to visualise immunofluorescent imaging of COL6A1 (magenta), PCK (blue), and CD45 (green) in colon section from Crohn's Disease patient (representative of n=5 sections), and head and neck cancer section (representative of n=6 sections). **A-B.** Visualisations of COL6A1, PCK+ epithelium, and CD45+ immune cells in Crohn's Disease colon section (representative of n=5 sections). **C-D.** Visualisations of COL6A1, PCK+ epithelium, and CD45+ immune cells in head and neck section (representative of n=6 sections).

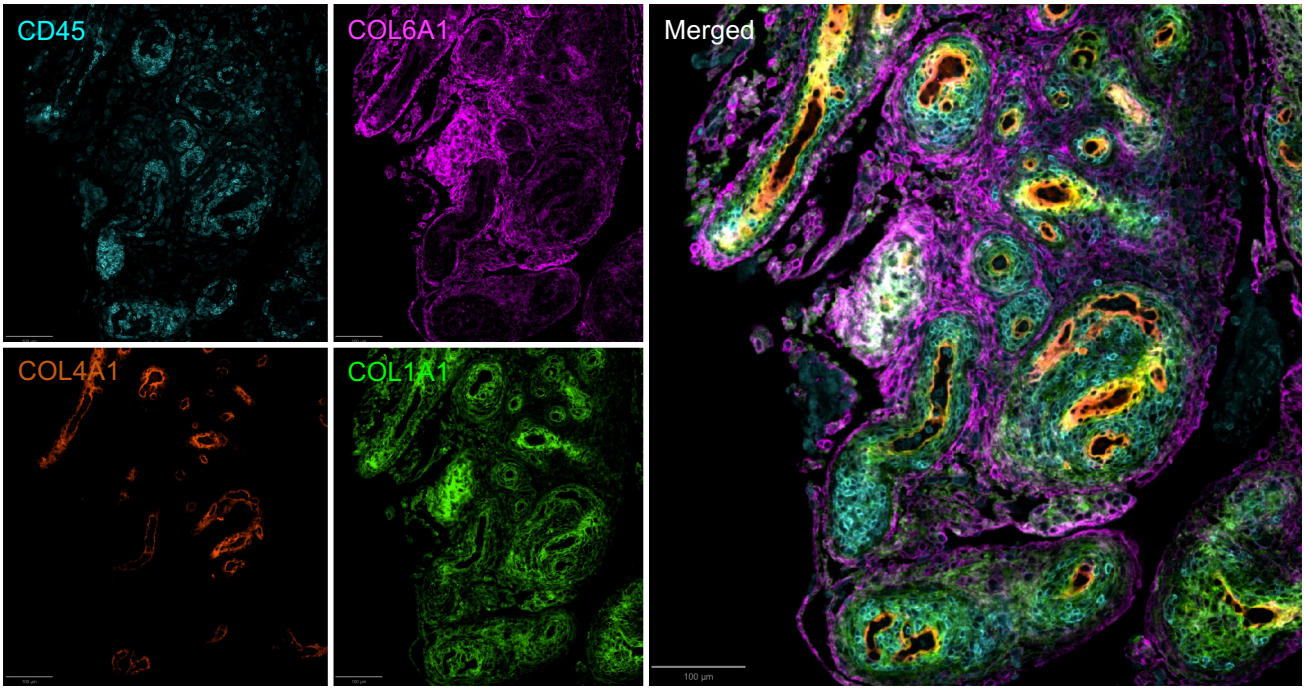


Figure 5.4 COL1A1 and COL6A1 display distinct deposition patterns in RA and OA. Immunofluorescent staining was used to characterise the spatial distribution of COL1A1 and COL6A1.QuPath was used to visualise immunofluorescent imaging of COL6A1 (magenta), CD45 (cyan), COL4A1 (orange), and COL1A1 (green) in section JRP141 (representative of 3 OA and 3 RA sections).

5.2.5 Collagen VI dark zones are associated with specific immune subsets

Following a visual appraisal of differential restriction of immune subsets in collagen VI dark zones, I sought to quantify this process. The tissue sections stained for the validation panel of markers and imaged on the CellDIVE system were run through the SCIMAP python toolkit for in depth analysis of extracellular collagen VI fibres and cell phenotyping[199]. The python-dependent part of this analysis was performed by Shihong Wu. Briefly, cell and COL6A1 fibre segmentation were implemented (Figure 5.5 A). Cell phenotyping was then performed in a semi-automated fashion, following a decision matrix reported in Supplemental Table 2. From here, local fibre density scores were calculated for each fibre through a nearest neighbour analysis. Individual fibres were then assigned to COL6A1-rich or COL6A1-poor areas based on a local density cut-off set at the median local fibre density of the sections analysed (Figure 5.5 B). This median value was found to be constant throughout the sections analysed and was set for all sections at 0.01. This local density cut-off enabled the classification of individual fibres into COL6A1-rich and COL6A1-poor (Figure 5.5 C). The boundary fibres of the

COL6A1-rich areas could then be identified based on their coordinates and used to create a polygon annotation of the COL6A1-rich areas of the tissue. This polygon was then expanded outward by 50μm to mitigate the gaps missed by the linear boundary cell-joining polygon. Cell frequencies were measured in the COL6A1-rich regions, the 50μm-expanded COL6A1-rich regions, and the background representing the full tissue including the COL6A1-poor regions.

Cell counts in each region were then exported from python into R for statistical tests. To investigate the suspected differential spatial distribution of cell subsets, Fisher's exact test was run. Odds ratios and 95% confidence intervals were obtained for each cell subset in each section, with positive log odds ratio denoting a cell frequency enrichment in COL6A1-rich regions and negative ratios denoting an enrichment towards COL6A1-poor regions (Figure 5.5 E). B-cells and $CD4^+$ T-cells were found to be associated with COL6A1-poor areas in 4 and 5 out of the 6 sections respectively. This was true across OA and RA tissue. $CD8^+$ T-cells were more nuanced, being significantly enriched in COL6A1-poor and COL6A1-rich regions in 2 sections each. This reflects the observations made previously in Figure 5.2, with less restrictive patterning than $CD4^+$ T-cells. DCs mirrored the distributions seen in both $CD4^+$ T-cells and B-cells, with a clear enrichment in COL6A1-poor regions in 4 out of 6 sections. To complete the appraisal of lymphocytes, plasmablast distribution was highly variable, with an slight predilection for COL6A1-rich regions. Neutrophils were characterised by high inter-patient variability, with significant COL6A1-poor enrichment in 3 sections and COL6A1-rich enrichment in 2 sections. Macrophages, in line with the visual observations, were predominantly found in COL6A1-poor areas. This effect was however stronger in OA sections than RA sections, which could be due to more sublining macrophage expansion in RA. Monocytes on the other hand were found to be the only cell type significantly enriched in COL6A1-rich regions across all sections. The localisation of monocytes differentiating into macrophages (labelled *Mono_Macro* in Figure 5.5) was found to be between the observed monocyte and macrophage distributions, similarly distributed across COL6A1-poor and rich regions.

These data demonstrate the preferential localisation of immune cells towards COL6A1-poor regions, with particularly restriction of $CD4^+$ T-cells, B-cells, DCs, and macrophages. Monocytes on the other hand were capable of entering collagen VI-rich regions. These findings pushed us to explore further the binding dynamics of T-cells, B-cells, and monocytes in vitro to shed a mechanistic light on these observations.

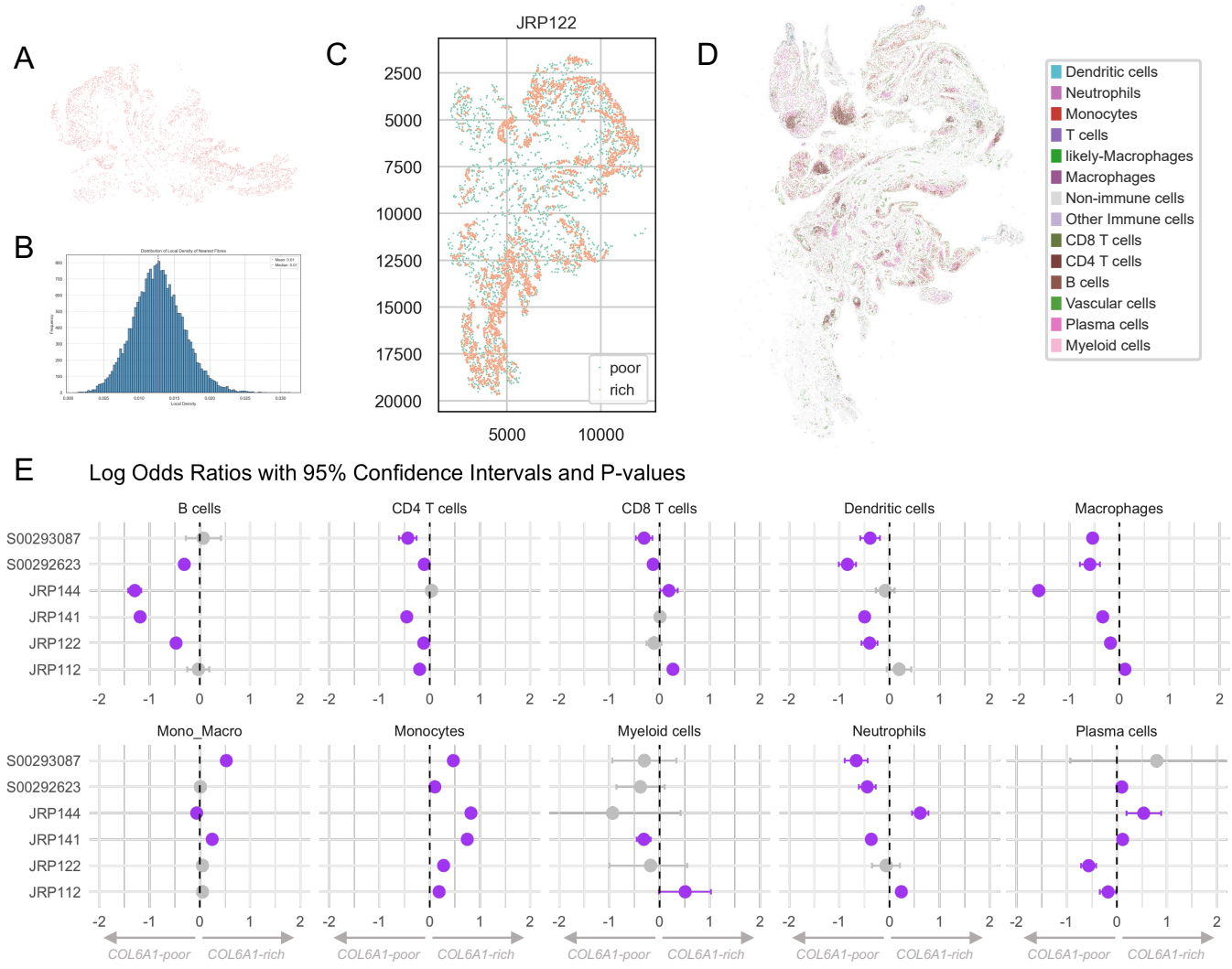


Figure 5.5 Cell type frequency in collagen VI-rich and poor zones. Immunofluorescent imaging of 6 synovial tissue sections (3 OA: S00293087, S00292623, JRP144; and 3 RA: JRP141, JRP122, JRP112) was performed for markers in the validation panel. Cell segmentation, phenotyping, and ECM fibre segmentation were performed through the SCIMAP python toolkit. Collagen VI-rich regions were identified through fibre segmentation and local density analysis. Cut-offs used to classify a collagen VI fibre as within a rich or poor region were set at the median fibre density (0.01 for each section's fibres). COL6A1-rich regions were then defined by the polygon created through the joining of boundary "rich" fibres, and an expansion constant of 50µm. Cell frequencies were then measured and extracted from each region for each analysed section, and exported into R where a Fisher's exact test was run. Exemplar image S00263087 is here presented as representative of the QC process for all 6 analysed sections (A-D). **A.** COL6A1 segmented fibres in section JRP122. **B.** Histogram of COL6A1 fibre local densities in section JRP122. **C.** Segmented fibre classification based on their belonging to COL6A1-poor (green) or rich (orange) areas in section JRP122. **D.** Cell identities phenotyped and mapped to section JRP122. **E.** Dot plots of log odds ratios, as a result of the Fisher's exact test, of cell frequencies in COL6A1-rich regions compared to background tissue. Log odds ratios are presented for each cell type identified and each analysed section. Greyed out points represents tests that yielded a p-value > 0.05, while purple points represent statistically significant data points. Error bars are the 95% confidence intervals of the test. Negative log odds ratios represent enrichment in COL6A1-poor regions, while positive ratios represent enrichment in COL6A1-rich regions.

5.2.6 Collagen VI dark zones are outlined by a degradative front

Having shown a preferential spatial restriction of most immune cells to COL6A1-poor regions, I then sought to explain the mechanisms behind the formation of these dark zones. In Chapter 3, inverse expression dynamics of MMPs and their inhibitor TIMPs pointed to increased collagen VI turnover in RA compared to OA and in active RA compared to remission, and increased deposition in OA and remission. In Chapter 4, this was corroborated by human and mouse collagen VI deposition, correlating with $MerTK^{+}$ remission-associated macrophages and with inflammation resolution respectively. Based on these data and the broad expression of collagen VI by all RA fibroblast subsets, I hypothesised that dark zone formation may be degradation-dependent, and that these zones may be active degradation sites.

To test this, I incorporated a Cy3-conjugated collagen hybridizing peptide (CHP) into the CellDIVE validation panel. CHP is a molecular mimic of the collagen helical domain's Gly-X-Y repeats. When added onto tissue sections, CHP binds exclusively to newly exposed unfolded collagen triple helical domains. This triple helical domain unfolding is a feature of collagen degradation, and enables the fluorophore-conjugated CHP to label active collagen degradation sites (Figure 5.6 A&B). In the sublining outside of dark zones, CHP staining labelled both patches and trajectory-like lines of active degradation. The clearest patterns however were observed surrounding the collagen VI dark zones, with CHP forming rings lining the circumference of the dark zones (Figure 5.6 A&B). This observation demonstrates the existence of a degradative front at the boundary between COL6A1-rich and poor regions, which may suggest an association between collagen VI degradation and the formation of the dark zones.

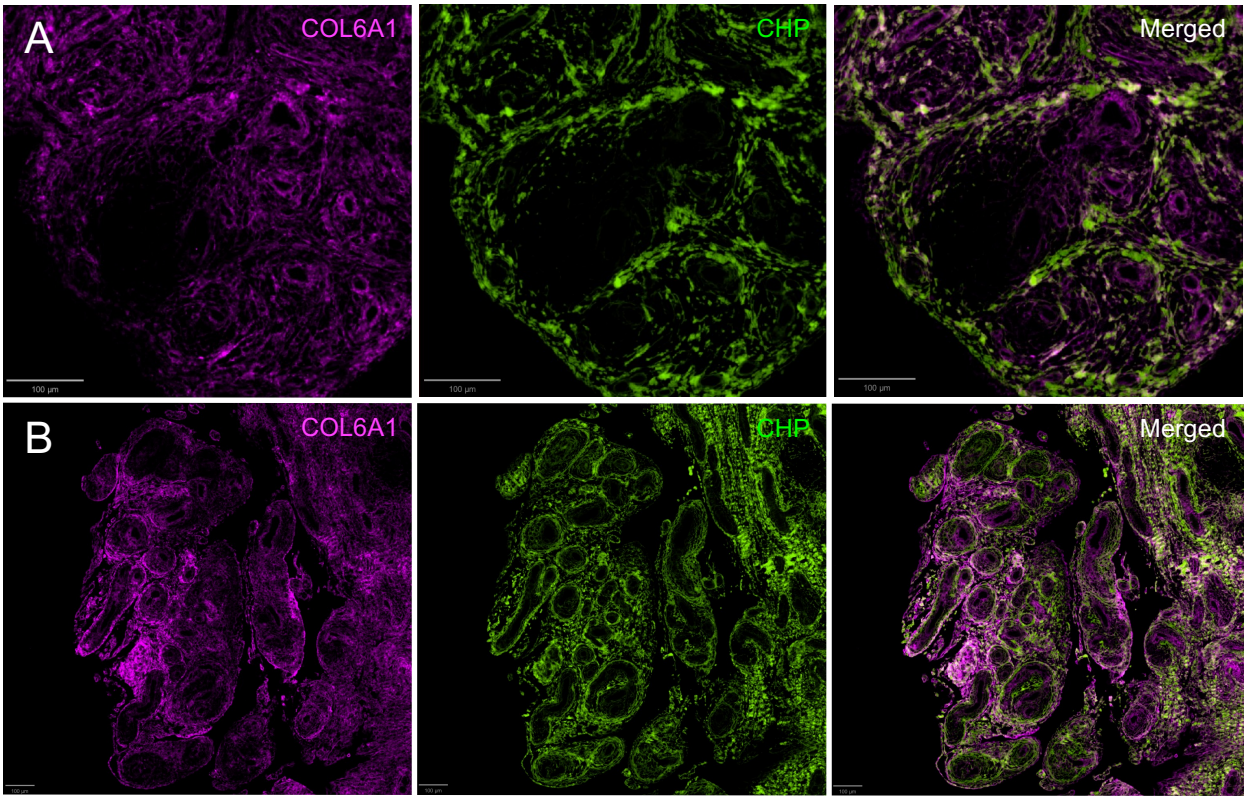


Figure 5.6 Collagen VI dark zones are surrounded by active degradation rings. Immunofluorescent staining was used to characterise the spatial distribution of Collagen Hybridizing Peptide (CHP) in relation to COL6A1.QuPath was used to visualise immunofluorescent imaging of COL6A1 (magenta), CHP (green) in RA sections JRP122 and JRP141 (representative of 3 OA and 3 RA sections). **A.** Visualisation of COL6A1 (first panel), CHP (second panel), and both markers merged (third panel) in JRP122. **B.** Visualisation of COL6A1 (first panel), CHP (second panel), and both markers merged (third panel) in JRP141.

5.2.7 Monocytes line dark zone degradative fronts and colocalise with CHP in tissue

Having identified the outer border of dark zones as active degradation fronts, I sought to identify the cellular culprit behind this process. Through the validation panel, I looked at the localisation of synovial cell subsets with regards to these CHP-labelled degradation rings surrounding the dark zones (Figure 5.7 A,B,D&E). $CD14^{+}$ monocytes were observed lining both the inner and outer surface of the degradative front, with other $CD45^{+}$ immune cells localising deeper in the dark zones (Figure 5.7 C). This was observed across collagen VI dark zones, regardless of their size, suggesting monocytes may play a role in the creation of the dark zones.

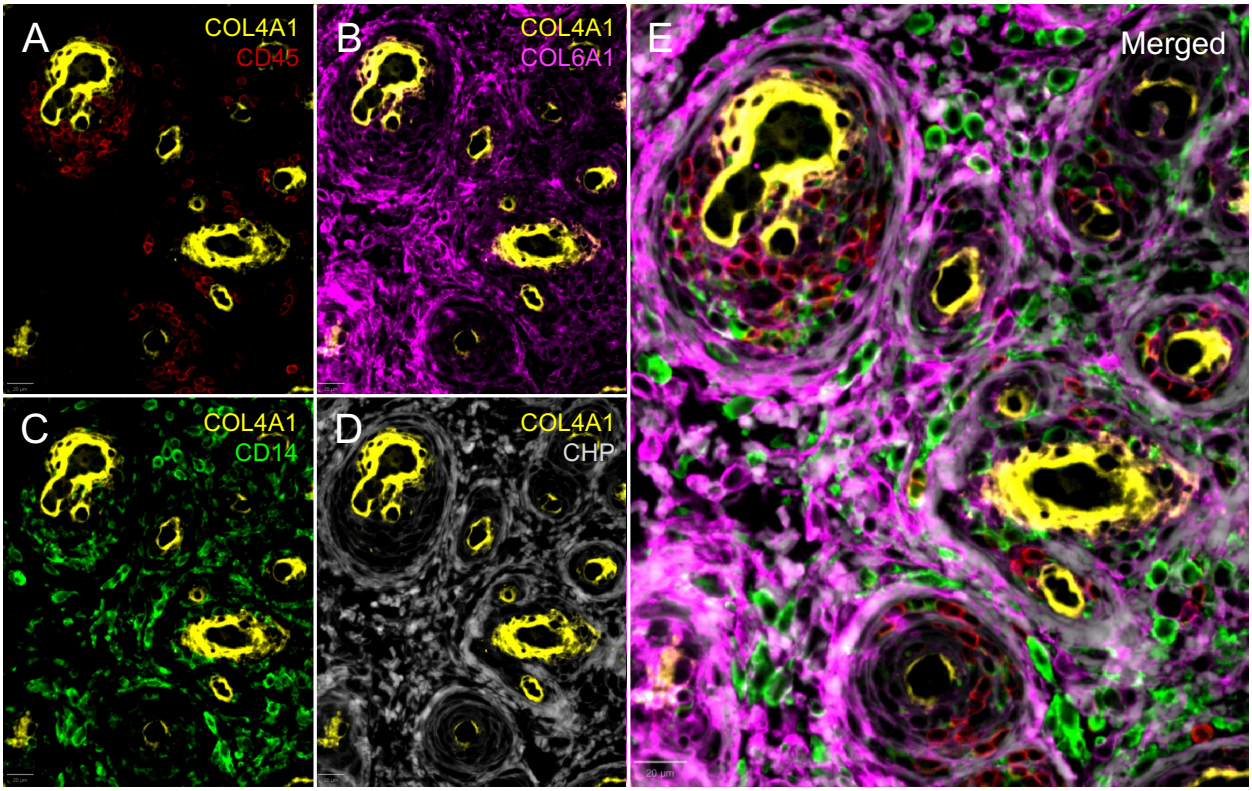


Figure 5.7 Monocytes localise to the outer border of the dark zones and line the degradation rings. Immunofluorescent staining was used to characterise the spatial distribution of COL4A1, COL6A1, CHP, CD45, and CD14. QuPath was used to visualise immunofluorescent imaging of COL4A1 (yellow, A-E), CD45 (red, A), COL6A1 (magenta, B), CD14 (green, C), CHP (white, D) in section JRP141 (representative of n=6 sections).

As previously established in Figure 5.5, monocytes are the one cell type identifiable with this validation panel to be reproducibly found in collagen VI-rich areas, and to not be restricted to the dark zones. In addition to their localisation near the degradative front, I therefore also looked at their tissue distribution with regards to CHP (Figure 5.8). $CD14^{+}$ monocytes outside of the dark zones were also found to colocalise with CHP (Figure 5.8 B,C&E). Together with their outlining of dark zones, these data point to monocyte tissue ingress being associated with collagen degradation. Monocyte infiltration itself may therefore act as a collagen VI remodelling factor.

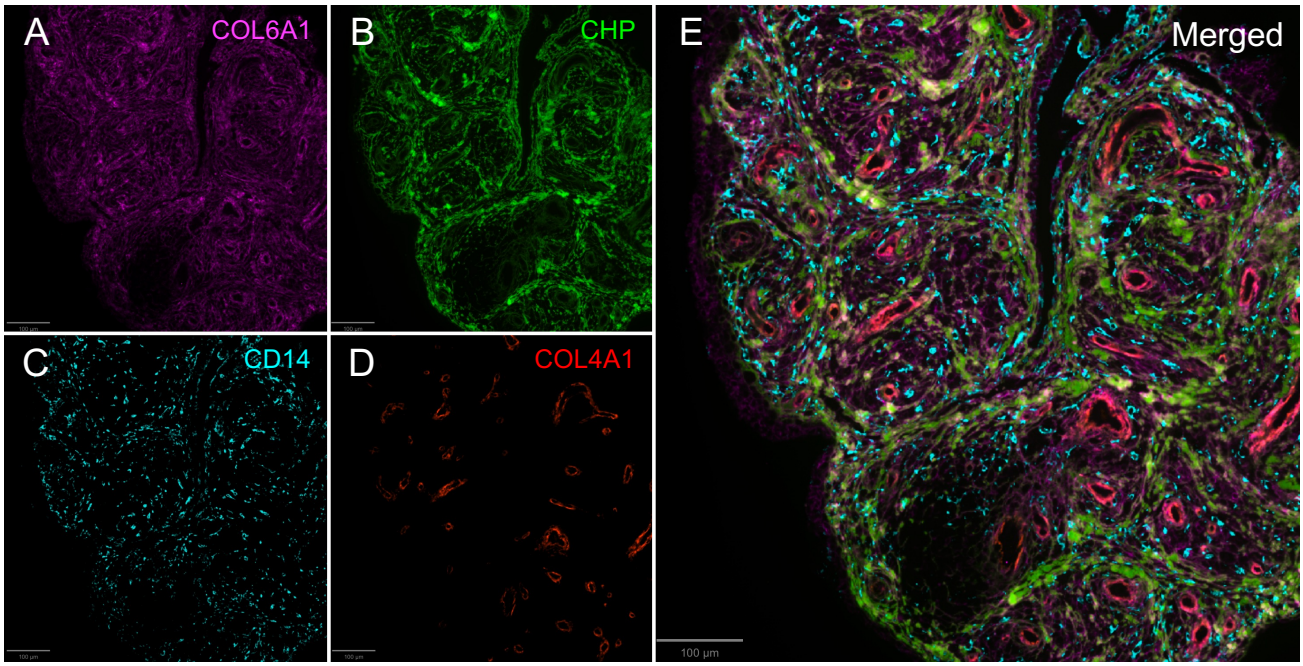


Figure 5.8 Deep tissue monocytes localise to CHP paths. Immunofluorescent staining was used to characterise the spatial distribution of COL4A1, COL6A1, CHP, and CD14. QuPath was used to visualise immunofluorescent imaging of COL6A1 (magenta, **A**), CHP (green, **B**), CD14 (cyan, **C**), COL4A1 (red, **D**) in section JRP122 (representative of 3 OA and 3 RA sections).

5.2.8 Proteomics confirms the observed collagen VI-specific deposition and degradation patterns

Based on the increased expression of MMPs and decreased expression of TIMPs in RA, combined with the observation of effective collagen VI deposition in tissue remission-associated conditions, the data pointed to increased collagen VI degradation in RA. After showing the specificity of the collagen VI deposition pattern in Figure 5.4, and highlighting the degradation associated with monocytic ingress, I sought to measure whether collagen VI was indeed specifically degraded. A published mass spectrometry dataset of synovial tissue from RA and OA patients was analysed by Midwood lab member Dr. Anna Hoyle[197]. Individual detected peptides were mapped to their protein of origin, and analyses were run to look for semi-tryptic and tryptic cleavage products. This enabled the quantification of proteins of interest, but also the quantification of their degradation products through the calculation of a semi-tryptic to tryptic peptide ratio.

After quality control and data pre-processing, protein abundances could be calculated and used for network analysis (Figure 5.9 A) using the Graphia platform[198]. Proteins clusters were formed based on their upregulation in OA or RA and their correlations. Amongst these, a number of collagens were identified. COL1A1, COL5A1/2, COL3A1, and COL12A1 were found to be more abundant in RA compared to OA (Figure 5.9 B). COL2A1, COL6A1/2/3/6, and COL15A1 were found to be more abundant in OA compared to RA. Based on their similar transcriptomic profiles yet different protein deposition patterns in tissue (Figure 5.4), we performed more focused analyses on collagen types I and VI. Looking across experimental repeats, no statistically significant different abundance of COL1A1 or COL1A2 could be found between RA and OA, despite a marginally increased mean abundance in the former (Figure 5.9 D&F). COL6A1, COL6A2, and COL6A3 on the other hand were significantly more abundant in OA compared to RA (Figure 5.9 H,J&L). For both type I and type VI collagens, these differences were driven by degradation. COL6A1/2/3 in particular had significantly greater semi-tryptic to tryptic ratios in RA compared to OA (Figure 5.9 I,K&M). Briefly, this ratio is between peptides released by endogenous proteases, termed semi-tryptic, and peptides released through tissue processing. A positive ratio points to high levels of degradation in the tissue while a negative ratio points to lower levels. This effect was significantly starker for collagen VI than collagen I.

These data corroborate the body of evidence built throughout this thesis, confirming that degradation plays an active role in the regulation of collagen VI deposition. This deposition is reduced in active RA disease due to increased degradation. These degradation products are here confirmed to be detectable.

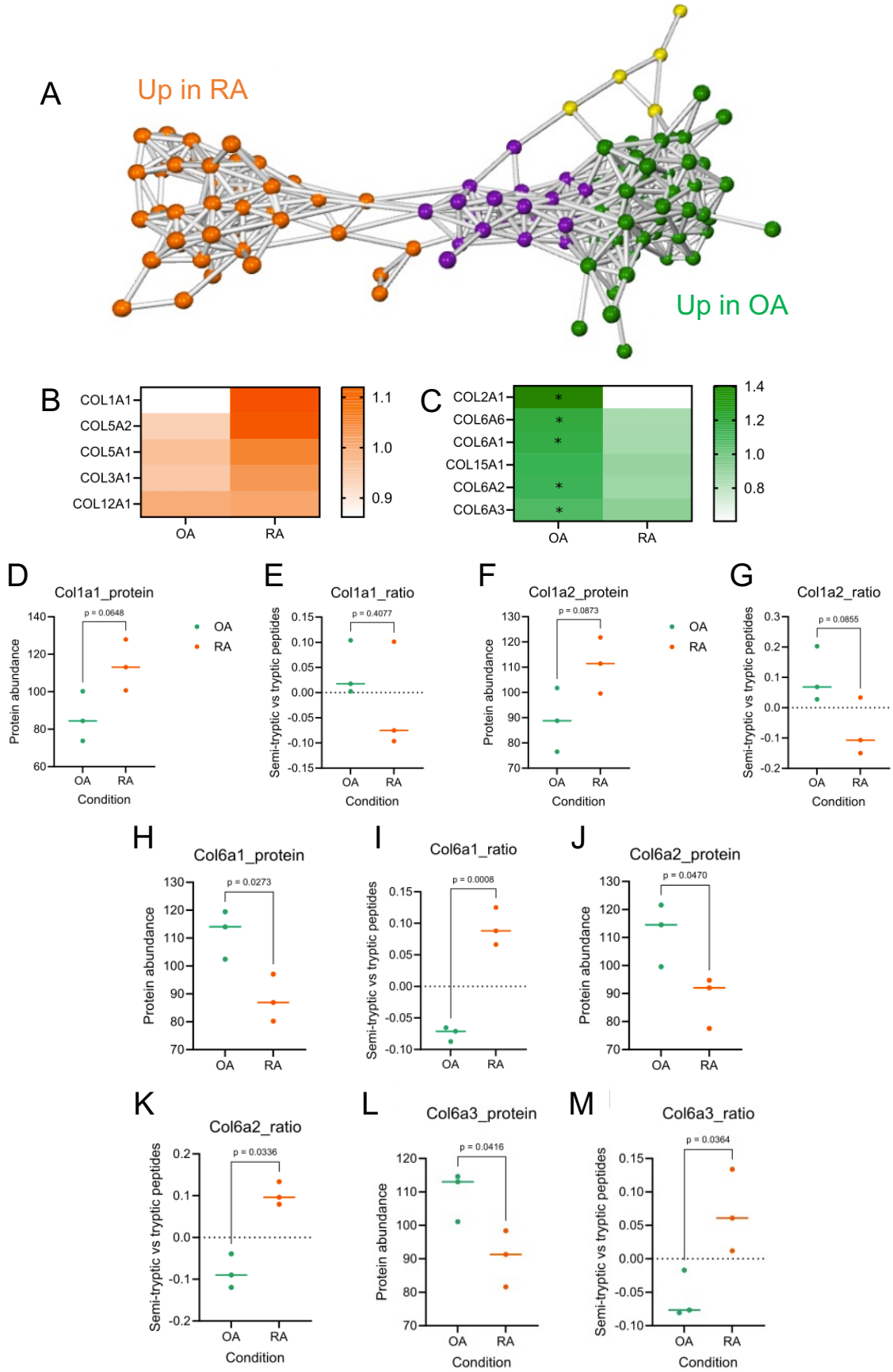


Figure 5.9 Mass spectrometry-based proteomics confirms the observed collagen VI degradation dynamics and its specificity. Published mass spectrometry dataset of synovial tissue from OA and RA patients was analysed by Dr. Anna Hoyle. Detected peptides were mapped to their parent protein and analyses were performed to assess protein abundance and semi-tryptic to tryptic peptide ratios. **A.** Network analysis of protein abundances and correlations in Graphia. **B.** Heatmap of protein abundance for proteins found to be overrepresented in RA samples, including COL1A1 and COL3A1. **C.** Heatmap of protein abundance for proteins found to be overrepresented in OA samples, including COL6A1/COL6A2/COL6A3. Stars denote a p-value < 0.05 through a t-test**. D-M.** Dot plots of mean protein abundances and mean semi-tryptic to tryptic peptide abundance ratios in OA and RA samples for COL1A1 (D&E), COL1A2 (F&G), COL6A1 (H&I), COL6A2 (J&K), COL6A3 (L&M). P-values for t-tests are labelled above each dot plot. Samples for 10 RA or 12 OA samples were pooled and split into 3 samples, each represented by a data point.

To complement these observations, we looked to map the peptide fragments detected to their proteins of origin, particularly collagen VI, to look for particular patterns of degradation in RA and OA (Figure 5.10). Figure 5.10 presents the abundance of degradation peptides mapped to specific 50 or 100 amino acid sequence sections of the COL6A1/2/3 protein structures, represented here as a bar chart. Green heatmaps show the frequency of MMP2 and MMP9 protease susceptibility sites in each section, the two canonical processors of collagen VI[305].

The sequence mapping revealed that few of the released and detectable degradation products stem from helical domains. This is particularly true for COL6A1 and COL6A2 (Figure 5.10 A&B), while some fragments are released from the helical domain of COL6A3, particularly the N-terminal part of the domain (Figure 5.10 C). RA samples were broadly richer in detected degradation products than OA samples across the three chains of collagen VI. While many of the released fragments were RA-specific, some OA-specific fragments were also detected in abundance stemming from the COL6A3 chain's 1450 and 1750 amino acid sections. Many of these degradation sites, particularly in the COL6A3 chain were found to map to experimentally validated MMP9 cleavage sites. One particular example which stood out is the example of the 2450 amino acid section of the COL6A3 chain. Peptides detected from this section were unique to RA in this dataset, experimentally validated as MMP9 cleavage sites, and scored highly in terms of protease susceptibility to MMP2 and MMP9 in this dataset. MMP9 is predominantly expressed by monocytes and neutrophils, two of the major infiltrating cells types associated with RA. Together with the association between monocytes and collagen VI degradation in tissue described in this thesis, these proteomic data solidify my proposed role for monocytes in the turnover of deposited collagen VI.

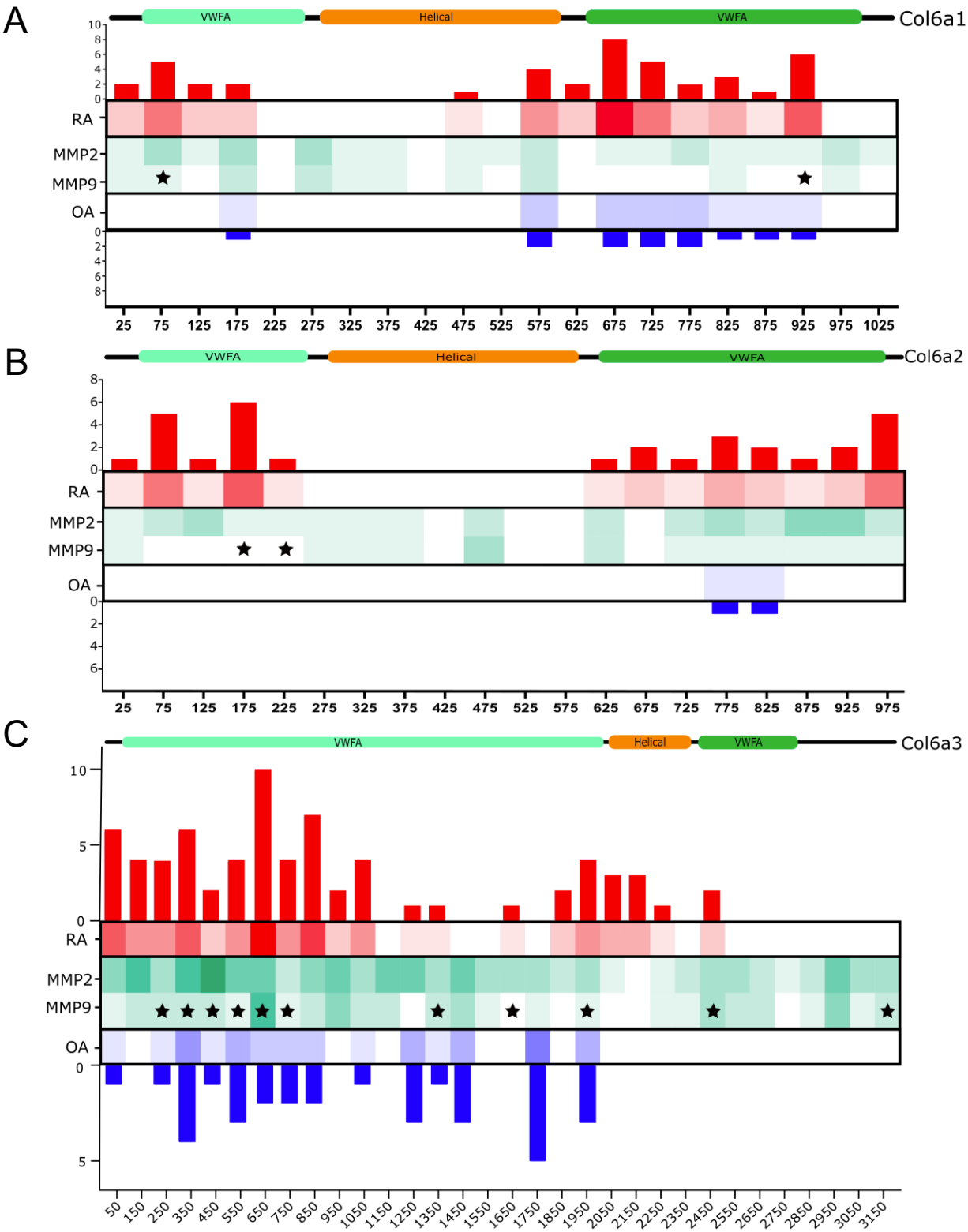


Figure 5.10 Proteomics points to RA-specific degradation sites and released fragments. Degraded collagen VI peptides were mapped back to 50-100 amino acid sections of the α1 (**A**), α2 (**B**), and α3 (**C**) chain structures. Peptide abundances in RA (red) and OA (blue) samples are represented by bar charts and their underlying heatmap. Protease susceptibility scores for MMP2 and MMP9 were assigned to each of the sections and are represented in the green heatmaps. Stars represent experimentally validated and reported MMP9 degradation sites.

Throughout this thesis, I identified collagen VI as a dysregulated ECM component of the RA synovium. Protein level analysis showed its deposition was associated with levels of remission-linked MerTK^{+} macrophages and with inflammation resolution, in human and mouse tissue respectively. Through this chapter specifically, I described an association between collagen VI deposition across tissues and immune cell restriction to domains termed *dark zones*. These dark zones were found to be specifically defined by collagen VI compared to collagen I. Most synovial immune cell subsets were found to be preferentially restricted to these zones, but not CD14^{+} monocytes. Infiltrating monocytes colocalised with CHP, a marker of active collagen degradation, creating a degradative front on the outer border of dark zones. Finally, degradation products from the cleavage of collagen VI, including by monocyte-specific protease MMP9, were found to be detectable in tissue through mass spectrometry.

With a couple of months remaining before the end of my thesis, I chose to start investigating the potential functional mechanisms underlying the described functions of collagen VI. The following sections describe this preliminary work and pave the way for future explorations.

5.2.9 Collagen VI is not adhesion permissive to THP1, Jurkat, and Daudi cell lines

To explore a potential mechanistic role for collagen VI in directing the spatial patterning of immune cells in tissue, particularly COL6A1-poor-restricted T-cells and B-cells, COL6A1-rich-enriched monocytes, I turned to in vitro systems and cell lines. THP1 monocytic cells, Jurkat T lymphocytic cells, and Daudi B-cell lines were chosen for this. 96-well plates were either left uncoated, or coated with fibronectin as a control adhesive ECM substrate, collagen VI, or fibronectin and collagen VI at a 1:1 concentration ratio. Wells were then coated with BSA to prevent adhesion of tested cells to uncoated plastic, and uncoated/BSA blocked wells were used as background controls. Cells were then plated on the substrates, left to adhere for 2h and washed gently to remove non-adherent cells. Percentages of adherent cells were then

read as confluency on the Incucyte and normalised to the experimental maximum (Figure 5.11). Daudi B-cells were unable to adhere to any of the substrates other than non-BSA-blocked plastic (Figure 5.11 A). Jurkat T-cells and THP1 monocytes were able to bind to fibronectin, acting here as a positive control for ECM adhesion (Figure 5.11 B&C). Both Jurkats and THP1s were found not to adhere to coated collagen VI. Added to fibronectin, collagen VI did not affect adhesion of Jurkats (Figure 5.11 B) but exerted an adhesion blocking effect for THP1s (Figure 5.11 C). This effect on THP1s was found to be dose-dependent, with decreasing ratios of collagen VI to fibronectin leading to attenuated inhibition (Figure 5.11 D). A one-to-one concentration ratio of collagen VI to fibronectin was sufficient for full blockade of THP1 adhesion.

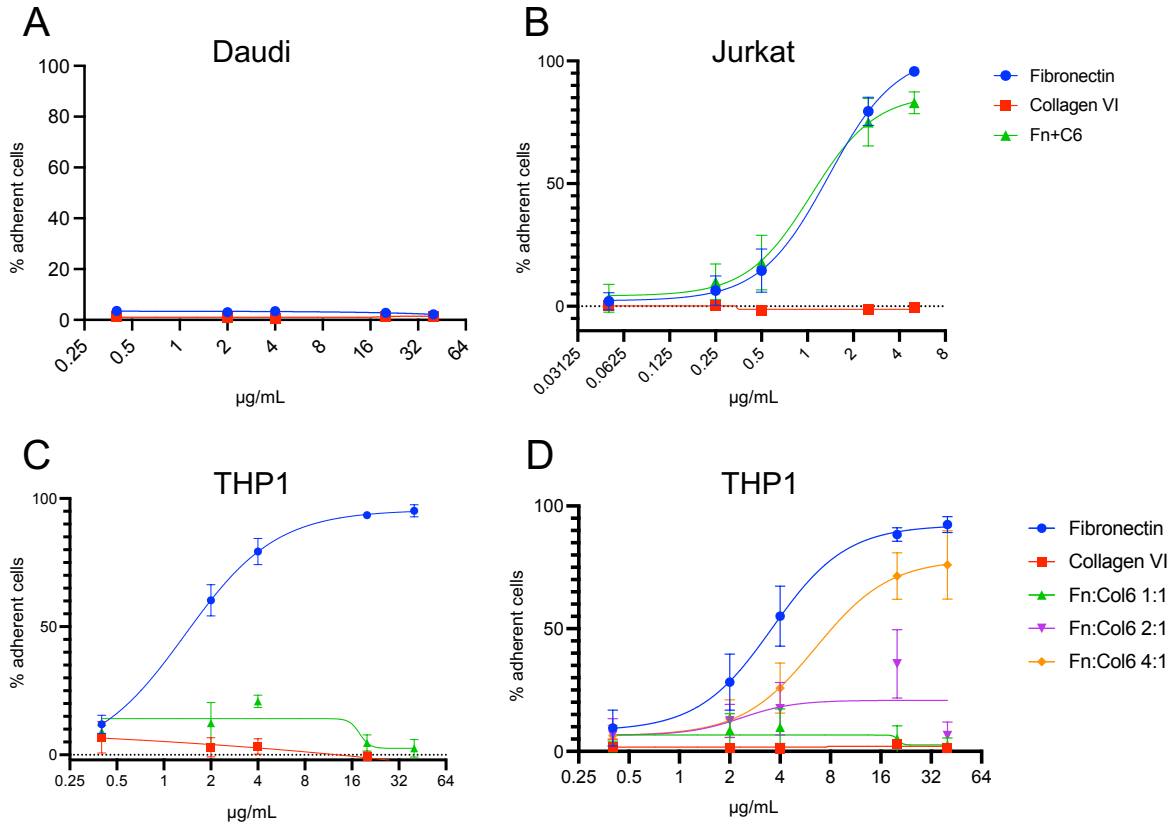


Figure 5.11 Collagen VI in not adhesion permissive and block THP1 adhesion to fibronectin. Adhesion of immune cells lines to human native intact collagen VI, fibronectin, or fibronectin and collagen VI was tested for Daudi (**A**), Jurkat T-cells (**B**), THP1 monocytes (**C**). Cells were starved for 24h in serum-free media while 96-well plates were coated with each ECM substrate condition. After BSA blocking, cells were left to adhere on the substrates for 2h, washed carefully to remove non-adherent cells and confluency was acquired on the Incucyte system. **D.** Adhesion assay run with the same methodology, testing for a dose-dependent effect of collagen VI in blocking THP1 adhesion to fibronectin. In each panel, each point represents the average confluency across 3 experiment repeats, and error bars are the SEM. Curves are a 4-parameter non-linear regression.

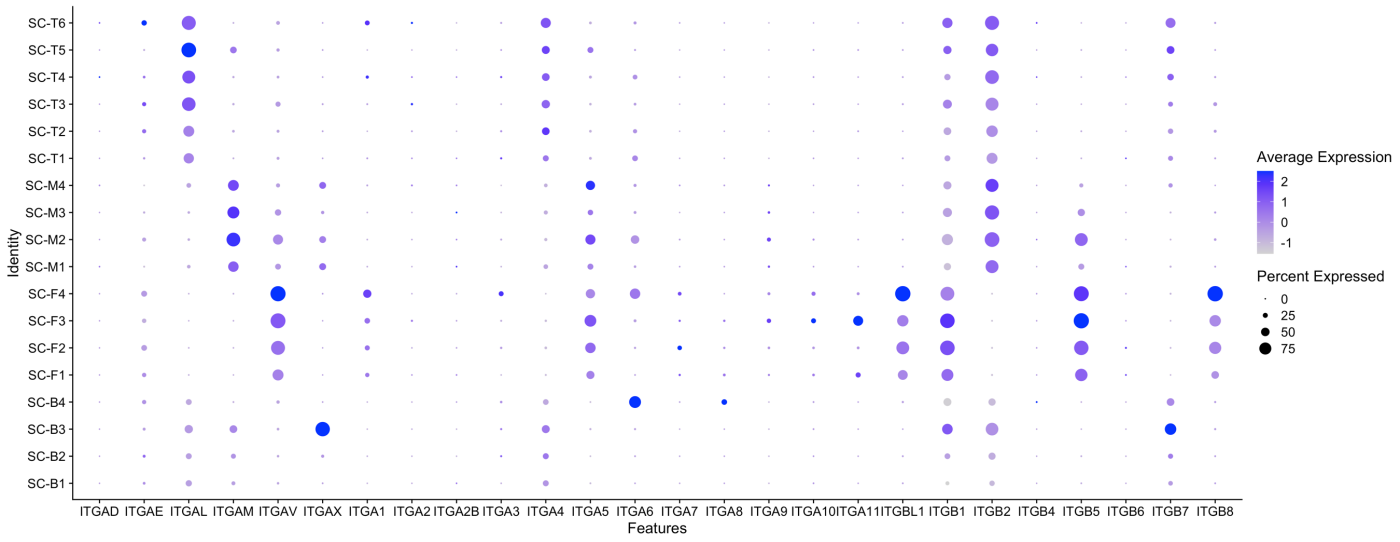
The insights from these data are two-fold. Firstly, Daudi, Jurkat, and THP1 cell lines do not adhere to collagen VI. Secondly, collagen VI modulates the adhesion of THP1 monocytes to other ECM substrates. This second point has potential implications for the spatial distribution of monocytes across tissues and was therefore chosen to be explored further.

5.2.10 Collagen VI acts as a molecular mimic obstructing cell adhesion

Through the previously described work, I showed that different synovial immune cell subsets have different predilections for COL6A1-rich and COL6A1-poor regions, and that collagen VI is able to block THP1 (but not Jurkat or Daudi) binding to fibronectin. Based on these cell binding observations, I hypothesised that this may be due to differential integrin usage between synovial immune cell subsets, and between my tested cell lines.

To assess this, I looked at integrin gene expression in the synovial cells of the AMP1 dataset (Figure 5.12 A). Integrins β1 and β2, encoded by *ITGB1* and *ITGB2*, were expressed across all the synovial immune cell subsets. B-cell subsets expressed *ITGAL, ITGAM, ITGAX,* and *ITGA4* at much lower levels and frequencies than the integrins expressed by T-cells and monocytes. SC-B3 cells, also known as autoimmune associated B-cells, expressed higher levels of these integrins than other B-cell subsets. T-cells were unique in expressing high levels of *ITGAL* and *ITGA4*. The latter, encoding the α4 integrin subunit, is known to bind to fibronectin and play a role in cell motility[306]. Finally, monocytes were unique in expressing high levels of *ITGAM*, also known as CD11b, another known fibronectin-binding integrin[307]. I sought to look at expression of these integrins in THP1, Jurkat, and Daudi by browsing bulk RNAseq data from the Cancer Cell Line Encyclopaedia (CCLE) on the EMBL-EBI website[308] (Figure 5.12 B). While data on *ITGA4* expression was not available in the database, *ITGB1* and *ITGB2* were confirmed to be expressed across the cell lines, while *ITGAM* (CD11b) was confirmed to have high levels of expression exclusively in the THP1 monocytes. *ITGAM* was notably not expressed in Jurkat (JKT) T-cells. Together these data are reassuring as to the relevance of the cell lines in studying integrin dependency of adhesion. These data also hint at a reliance on different integrins for fibronectin binding between different synovial cell subsets.

A



B

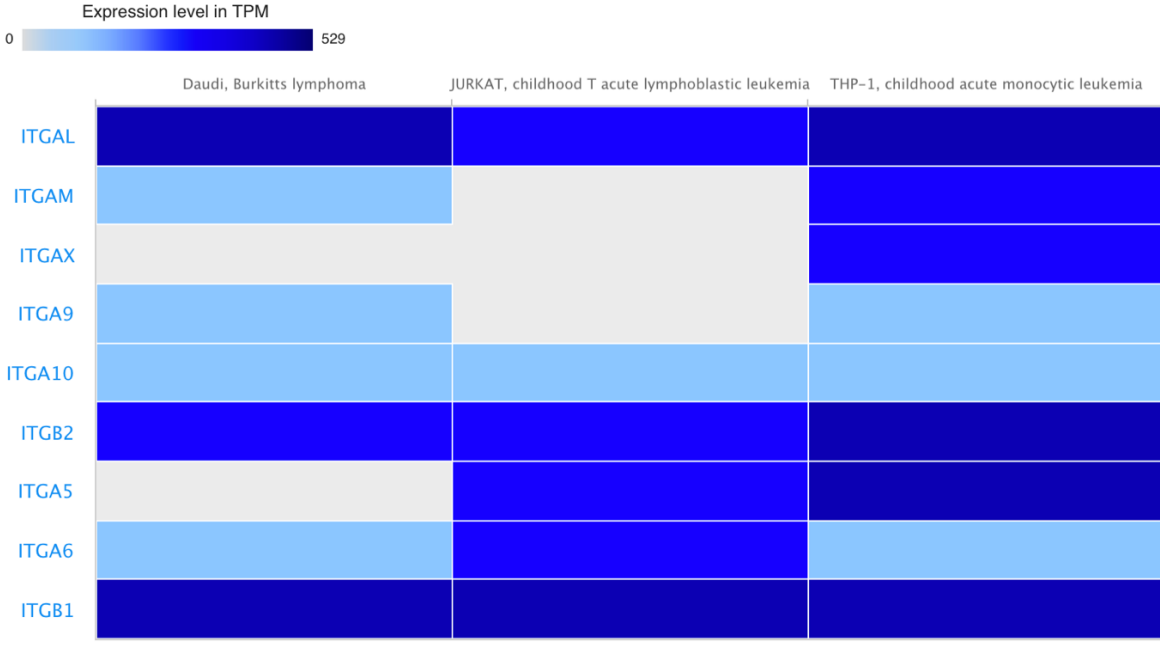


Figure 5.12 Integrin expression profiles of synovial cells and cell lines Sorted synovial cells from the AMP1 dataset were analysed for expression of all known α and β subunit genes. These genes were also used as input for the Cancer Cell Line Encyclopaedia (CCLE) browser on the EMBL-EBI website. **A.** Dot plot of Integrin expression by synovial T-cells and monocytes of the AMP dataset. Point size represents the percentage of each cell subset expressing each gene, and the blue colour intensity represents expression levels. **B.** Heatmap generated by the CCLE browser on the EMBL-EBI website. Blue colour intensity represents expression levels in transcripts per million (TPM).

To test this, I repeated the adhesion assay looking at THP1 and Jurkat binding to fibronectin, but this time incubating the cells with the small molecule BIO5192, a selective α4β1 inhibitor, or DMSO as a control (Figure 5.13 A&C). DMSO controls were found to not

affect adhesion of JKT or THP1 cells. As suspected however, BIO5192 was found to almost completely inhibit JKT cell adhesion, leading to an 89.02% adhesion reduction even at the lowest tested molarity of 10nM (Figure 5.13 A). This inhibitory effect was still somewhat observed in THP1 cells but was significantly weaker (adj. p-value = <0.0001), averaging a 34.32% reduction in adhesion (Figure 5.13 C). To test the role of CD11b in fibronectin binding, I also pre-incubated the fibronectin itself with the purified recombinant inserted (I) domain of CD11b, the ligand-binding domain of CD11b, generously donated by Dr. Anja Schwenzer. Fibronectin was therefore mixed with an increasing ratio (1x, 5x, 10x) of recombinant CD11b I-domain or I-domain denatured through boiling, as a control. The boiled CD11b I-domain was found to not affect the adhesion of either JKT T-cells or THP1 monocytes (Figure 5.13 B). The increasing intact CD11b I-domain ratios however did reduce the adhesion of THP1 cells (Figure 5.13 B). This inhibitory effect remained observable, even at the lowest tested ratio of 1:1, leading to a 41.31% adhesion reduction (Figure 5.13 C), suggesting CD11b may be one of the main fibronectin binding partners in THP1s. Unexpectedly however, the intact CD11b I-domain was also able to block adhesion of JKT cells, despite not expressing CD11b itself and therefore hinting at some unspecific mode of action. The inhibitory effect remained significantly stronger in CD11b-expressing THP1s however, with only 26.00% adhesion inhibition at a 1:1 ratio in JKT cells (adj. p-value = 0.0175) (Figure 5.13 C).

This discrepancy may partly be explained by the presence of integrin I-domains like the CD11b I-domain in β integrin subunits used by T-cells and JKT cells[309]. Coated CD11b I-domain binding to fibronectin could be obstructing I-domain binding sites, which the THP1 and monocyte αMβ2 integrin pair is doubly reliant on, while the JKT α4β1 integrin pair is only partly reliant on. These I-domains are structurally homologous to VWF[309]. This observed I-domain coating and blocking effect is also reminiscent of the observed collagen VI blocking effect on THP1-fibronectin binding. Collagen VI, and particularly the α3 chain of collagen VI, harbour the most VWF domains of all collagens[309]. Collagen VI VWF domains, homologous to integrin I-domains, could therefore by extension be blocking I-domain binding sites in fibronectin and competing with the CD11b presented by monocytes.

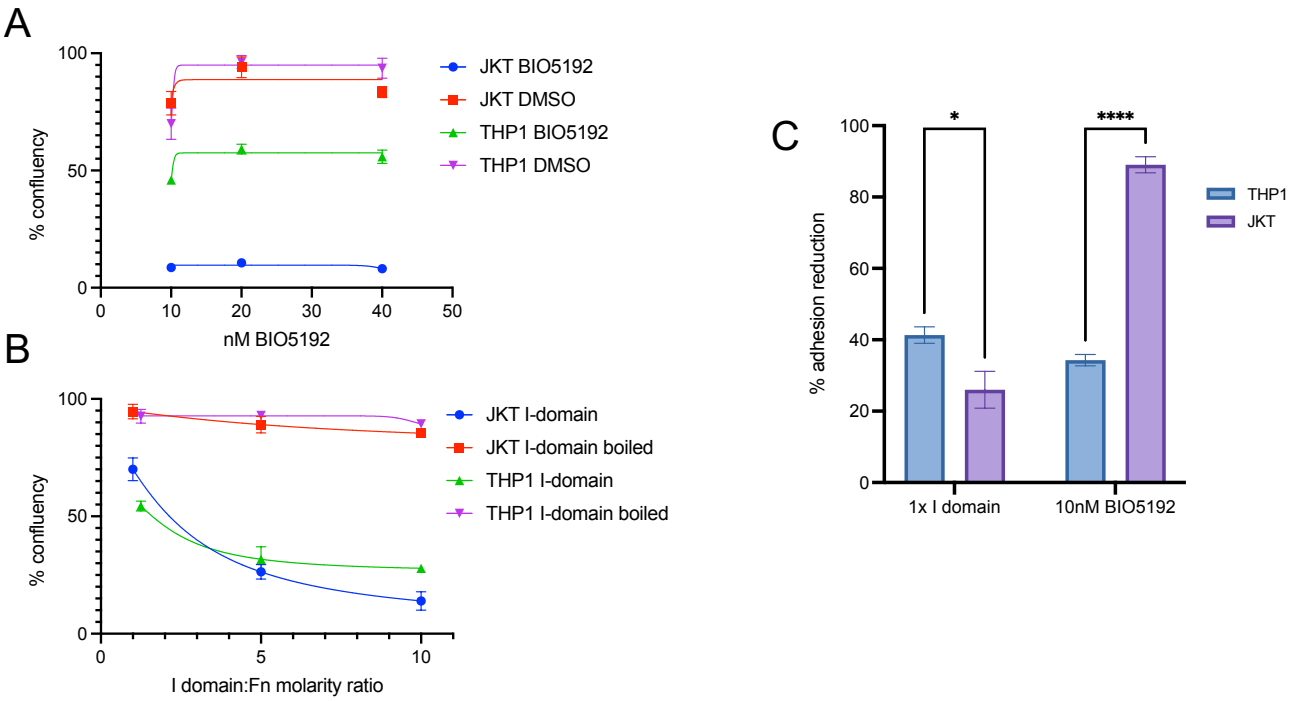


Figure 5.13 THP1-binding dynamics to fibronectin are integrin I-domain-dependent. A. Adhesion assay of THP1 and Jurkat (JKT) cells to fibronectin and incubated with α4β1 integrin inhibitor BIO5192 or DMSO. **B.** Adhesion assay of THP1 and Jurkat (JKT) cells to fibronectin coated with purified recombinant CD11b I-domain or the boiled denatured recombinant I-domain. Each point represents the average confluency across 3 biological replicates, and error bars are the SEM. Curves are a 4-parameter non-linear regression. **C.** Bar chart of the percentage adhesion reduction from each treatment condition and for both Jurkat and THP1 cells. Statistical test in a 2-way ANOVA with multiple comparisons. * = adj. p-value = 0.0175, **** = adj. p-value < 0.0001.

5.2.11 Collagen VI inhibits cell migration in preliminary in vitro experiments

With the observed differential adhesion dynamics between monocyte and T-cell lines, I sought to look for translation to migratory properties of cell subsets on collagen VI matrix. I therefore ran preliminary migration assays in parallel, one based on a Boyden chamber transwell system and a flow cytometry readout allowing for unbiased assessment of peripheral blood mononuclear cells (PBMCs) migration, and another based on an ECM-coated channel system and the THP1 and Jurkat cell lines.

The transwell system was operated by Zhi Yi Wong. Briefly, transwell inserts were coated overnight with ECM substrates and PBMCs were isolated from human blood cone donations (Figure 5.14 A). PBMCs were added to the transwell inserts, and CXCL12 was used as the chemoattractant because of its broad spectrum of targets[310]. Cells which successfully migrated through the transwell inserts were then phenotyped via a flow cytometry panel

including antibodies against CD45, CD3, CD4, CD8, CD19, and CD14. Cells could then be identified as $CD4^+$ T-cells, $CD8^+$ T-cells, $CD19^+$ B-cells, or $CD14^+$ monocytes. The abundance of migrated cells could then be normalised to a control condition with uncoated transwell inserts and no chemotactic gradient. As a negative control, pertussis toxin (PTX) was found to effectively block cell migration (Figure 5.14 B-E). CXCL12 alone induced 10-fold and 20-fold the control migration in CD4 and CD8 T-cells respectively (Figure 5.14 B&C). CXCL12 however induced weaker migration in B-cells and monocytes, achieving only a 4-fold and 3-fold increase respectively (Figure 5.14 D&E). No effect on migration could be seen from any of the ECM substrate coatings of the transwell inserts.

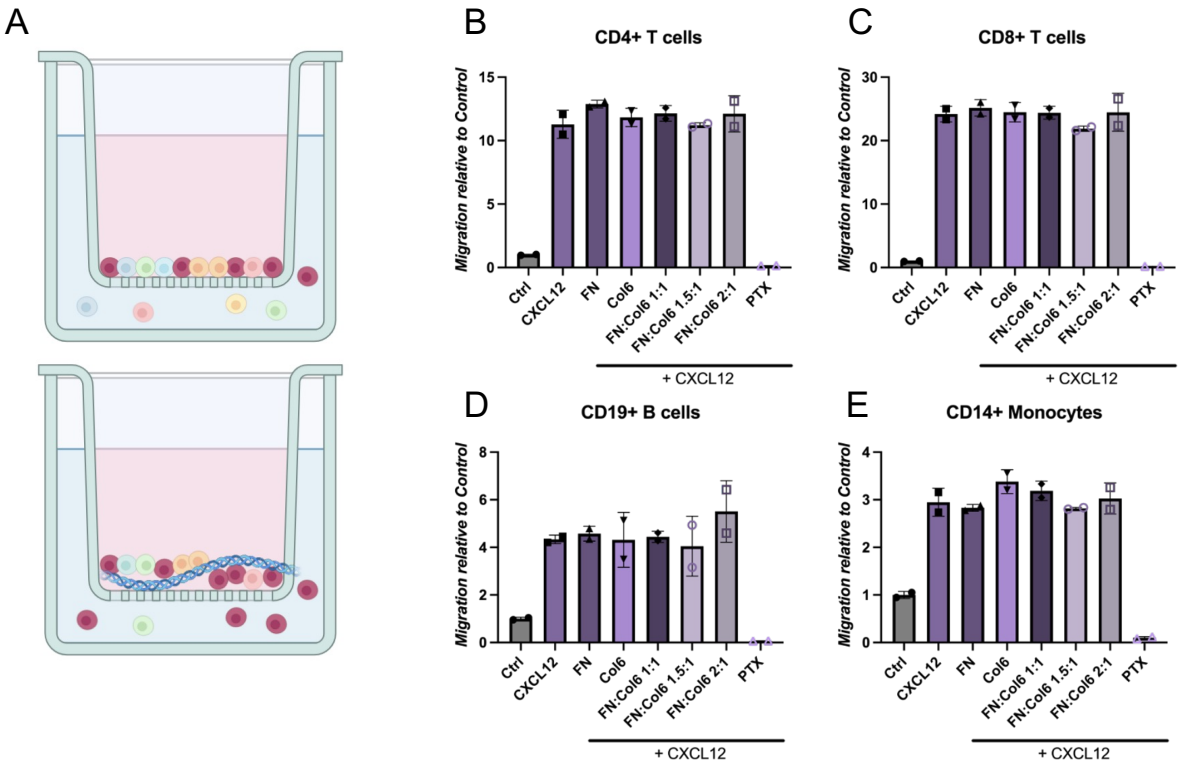


Figure 5.14 Transwell migration of peripheral blood mononuclear cells through ECM-coated membrane. A. Biorender schematic summarising the experimental set-up. PBMCs were isolated from blood cone donations and set to migrate through transwell inserts coated with either 40ug/mL fibronectin (FN), 40ug/mL collagen VI (Col6), or ratios of FN:Col6. 100ng/mL CXCL12 was used as the chemoattractant in the bottom chamber, and pertussis toxin (PTX) was used as a negative control. Cells were then phenotyped through flow cytometry. **B-E.** Bar plots of PBMC subtype migration into the bottom chamber relative to uncoated CXCL12-free control condition. Each point is a duplicate of a single donor, bars represent the mean of the duplicates and error bars the SD.

In parallel, the channel based system was also used. Briefly, Ibidi µ-Slide VI channels were coated with either fibronectin, collagen VI, or fibronectin and collagen VI. After washing the channel with serum-free media, cultured THP1 and JKT cells were added to the start well of the channel and left to adhere for 2h at 37°C, before adding FBS to the end well to start the migration process (Figure 5.15 A). Because of technical challenges detailed in the discussion of this chapter, the start wells of each channel were acquired to reflect migration. At 24h, both THP1 and JKT cells were found to leave the start wells effectively and enter the migration channel in fibronectin-coated channels (Figure 5.15 B). In collagen VI-coated channels however, THP1 and JKT cells were only able to exit the start well in smaller numbers, suggesting collagen VI is non-permissive to this type of migration. Collagen VI was also found to be able to reduce the migration of both cell types when added to fibronectin coated channels, suggesting a potential additional migration blocking effect.

These data, while preliminary, hint at a role for collagen VI that extends beyond adhesion and may impact migratory potential of immune cells in ways that require further characterisation.

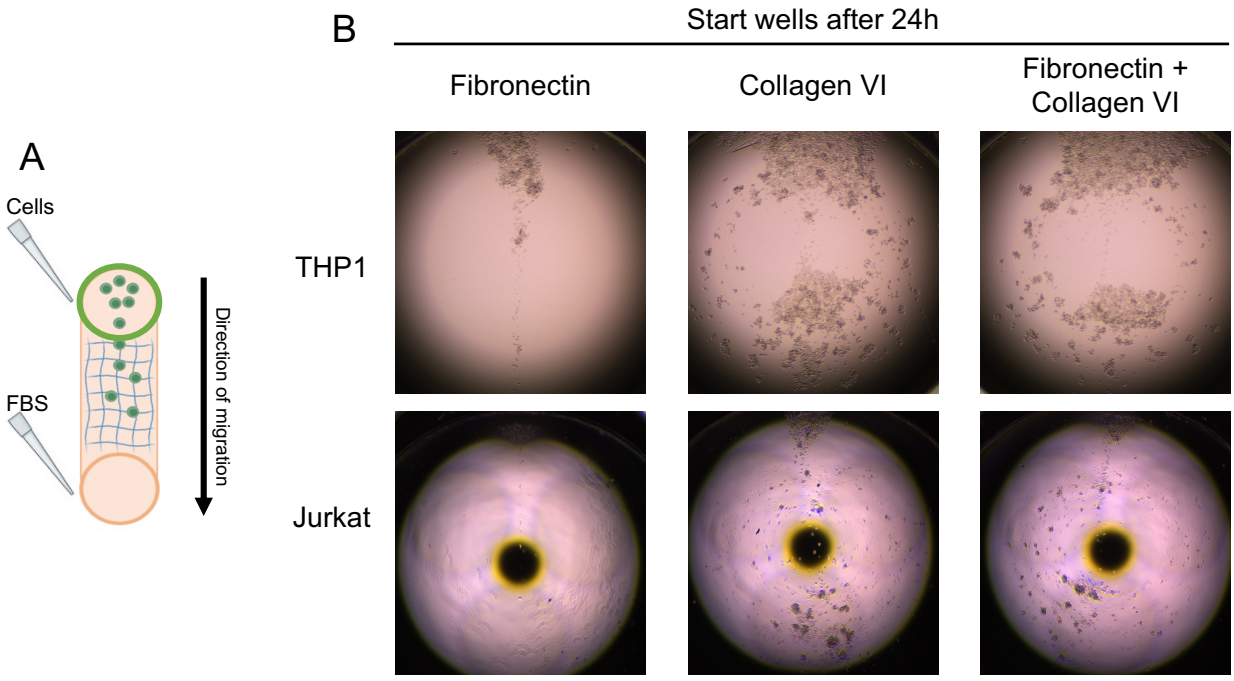


Figure 5.15 Ibidi channel-based migration of THP1 and Jurkat cells through ECM-coated channels. A. Biorender schematic summarising the experimental set-up. Ibidi µ-Slide VI channels were coated overnight with fibronectin, collagen VI, or fibronectin and collagen VI. THP1 or JKT cells were placed into the start well of the system and left to adhere for 2h before adding FBS as a chemoattractant to the end well. Cells were left to migrate for 24h and standard confocal microscopy pictures were taken of the start wells to visualise the cellular exodus. **B.** Standard confocal images of the start wells of the channel systems after 24h of migration through the different tested ECM coatings.

5.2.12 Collagen VI degradation products are bioactive

Finally, the proteomic data gathered in this chapter, coupled with the observation of monocyte localisation at active degradation sites, urged a closer study of the effects of released collagen VI degradation products on immune cells. As an exemplar, I chose an α3 chain fragment detected through the proteomics analysis, specific to RA over OA, and surrounded by experimentally validated N-terminal MMP9 and C-terminal furin cleavage sites, termed here D11 (Figure 5.16 A)[145,311]. The recombinant D11 amino acid sequence was cloned into an insect cell line by former Midwood lab member Dr. Jesse Coker, and expressed and purified in a first batch by Dr. Eleanor Williams and Korvus Wang. Adhesion assays, run as previously described, were then performed to assess the adhesive potential of D11 compared to the intact collagen VI. At 40ug/mL of coated substrate, THP1 cells were found to adhere to D11 to the same extent as peptide-9, a known CD11b-binding peptide of Tenascin-C used here as a positive control (Figure 5.16 B). Intact collagen VI, as previously described, was unable to support monocyte adhesion. This adhesive effect was confirmed to be dose-dependent through a serial dilution adhesion assay run by Linda Mies (Figure 5.16 C). Finally, I sought to start characterising the receptor for this monocyte binding. I chose to first test whether the binding was integrin dependent by pre-incubating cells with EDTA, which through its metal ion sequestering effect inhibits integrin function. Through this preliminary dose escalation of EDTA exposure, THP1 adhesion to D11 was found to be fully blocked by any tested EDTA concentration higher than 0mM. Peptide 9 adhesion on the other hand was abolished at a concentration between 1.25mM and 2.5mM of EDTA.

These data demonstrate the cleavage products released by monocytes through degradation of collagen VI, likely as part of their tissue ingress, are adhesion-permissive contrarily to intact collagen VI. This finding is discussed further in the discussion of this chapter.

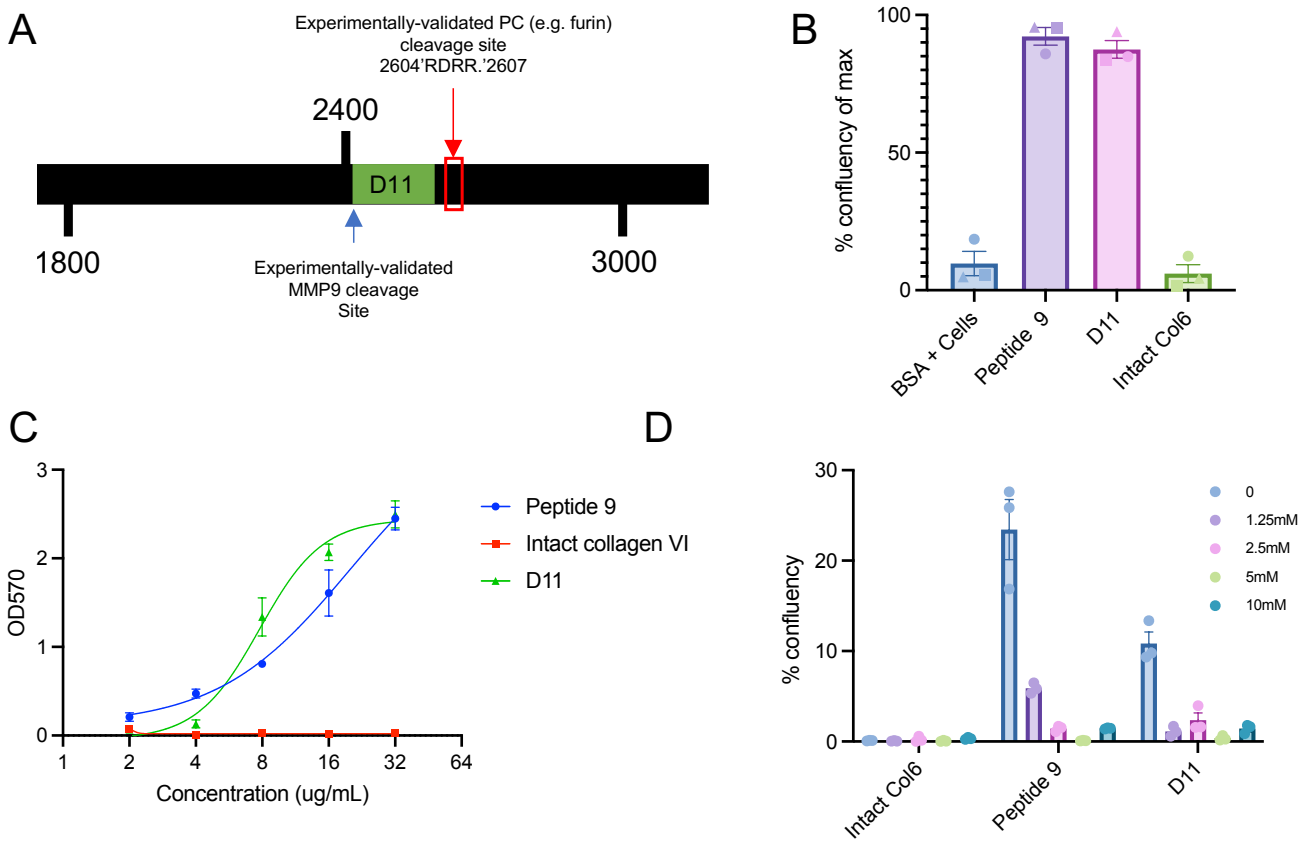


Figure 5.16 Cleavage product D11 in adhesion permissive for THP1 monocytes. The recombinant D11 amino acid sequenced was cloned into insect cell line by Dr. Jesse Coker, and purified by Dr. Eleanor Williams and Korvus Wang. Adhesion assays were then performed to start assessing its functions. **A.** COL6A3 sequence position of D11, along with surrounding experimentally validated protease degradation sites. **B.** Bar chart of % cell confluency normalised to experiment maximum as the output of three repeat adhesion assays. Each point represents the mean of 3 experiments. Peptide 9, D11, and intact collagen VI were all coated at 40ug/mL. Bars represent the mean of the 3 experiments, and error bars are SEM **C.** Serial dilution adhesion run by Linda Mies. Curve is a four-parameter non-linear regression. Each point represents the mean of 3 technical replicates, and error bars are SEM. **D.** Bar chart of the percentage confluency as a result of adhesion assay with dose escalation of integrin-inhibiting EDTA. Each point represents one of 3 technical replicates, the bar represents the mean of these replicates, and the error bars represent the SEM.

5.3 Technical discussion

5.3.1 Immunofluorescent imaging sample selection and analysis

One of the primary aims for this chapter was to shed new light on the collagen VI deposition patterns observed through my exploratory CellDIVE panel. To achieve this, 6 new sections from 3 previously stained OA and RA patient samples were stained for my validation panel, incorporating additional immune and matrix markers to establish immune subtype relationships and COL6A1 deposition pattern specificity. For this endeavour, samples were chosen based on the prior imaging performed in chapter 4 of this thesis. Samples containing a variety of immune infiltrate conformations and sizes were prioritised over others to provide a robust representation of possible collagen VI distributions. An obvious future refinement should be the addition of more samples to these explorations. The current set of 3 RA and 3 OA samples is identifying cross-disease features of synovial tissue dysregulation, but is not powered to support a comparison between OA and RA samples, which may hold interesting insights itself.

To study the compositional differences between collagen VI-rich and poor areas, fibre segmentation and local fibre density analysis were chosen. This selection was made to avoid some of the common pitfalls of immunofluorescent analysis, including the reliance of signal intensity metrics for quantification. As discussed in chapter 4, signal intensity is highly variable and dependent on tissue collection and processing. This makes the use of constant intensity thresholds for segmentation treacherous. In chapter 4, annotations were simply created through thresholds set by using the mean plus one standard deviation for the markers of each slide. This analysis was relevant to assess total collagen VI deposition area. Here instead, segmented fibres were assigned to collagen VI-rich and poor areas through local fibre density measurements. This analytical refinement enabled the defining of rich and poor areas, while preventing collagen VI-rich areas being defined as such by a singular deposited fibre.

5.3.2 In vitro functional assay refinements and limitations

The focus of the last sets of results in this chapter is the preliminary investigation of the potential functional underpinnings of collagen VI spatial restriction of immune cells. As described through the chapter, human cell lines were chosen over primary human cells to run the majority of these assays. This choice was made to provide the preliminary basis for reproducible assays, notably the adhesion assays, to test ECM substrate properties. A clear improvement to consider, with more time, should be the incorporation of primary cells and/or full PBMCs in these assays and their optimisation.

The migration assays presented here, both transwell-based and channel-based, are particularly preliminary and require further development. While transendothelial migration assays exist and are a workhorse assay for many studied systems[312,313], they hold little value here as the migration process of interest occurs just beyond the endothelial layer. Other improvements however should come in the form of ECM coating optimisation and chemoattractant selection. In the current experimental set-up trialled, simple human native protein coatings were chosen. These may not be sufficient and may require gel-like conformations or higher coating volumes to enact their migration-modulating effects. CXCL12 and were selected as chemoattractants for the transwell and channel migration assay respectively. This choice was made based on in house availability and the broad theoretical spectrum of targets of CXCL12[310]. In the transwell assay, while CXCL12 did induce migration of T-cells, B-cells, and monocytes, I found a significant bias towards the attraction of T-cells over B-cells and monocytes. One area of improvement should therefore be the incorporation of more chemokines into a chemoattractant cocktail to generate greater effect sizes. This may reveal currently undetected condition differences, particularly for monocytes. CCL2 should be the chemokine of choice for monocytes[314], and CXCL13 for B-cells[315].

The channel-based assay would also benefit from the incorporation of a proper chemotactic signal. Cells were found to not be attracted strongly enough by the FBS to reach the end well of the channel system, which explains the necessity to acquire images depicting their exit from the start well as a migration proxy. A stronger chemotactic signal may be enough

for cells to complete their migratory trajectory. Finally, tested cells, particularly THP1 cell lines may require activation to migrate through collagen VI, in particular if this migration is degradation-dependent. THP1 cells have indeed been describe to upregulate MMP expression in response to stimuli including PMA or TNF-α[316,317].

5.4 Biological discussion

5.4.1 MerTK regulation: new insights into an attractive yet elusive target

MerTK is a receptor tyrosine kinase member of the TAM family together with Tyro3 and Axl[318]. Like most tyrosine kinase receptors, MerTK is activated through ligand binding, which include Gas6. MerTK expression levels have been extensively associated with immune regulating macrophages as opposed to inflammatory macrophages. This is particularly evidenced by the higher levels of expression of MerTK on bone marrow-derived macrophages (BMDMs) polarised with IL-4 and IL-13 than BMDMs polarised with LPS and IFN-γ[319]. This upregulation was also found to be inducible in human monocytes through treatment with anti-inflammatory compounds like the glucocorticoid dexamethasone[320].

From this immunosuppressive association flows a notable interest in MerTK expression in IMIDs. Multiple groups have indeed confirmed functional associations through mouse work, with *Mertk* knockout mice being associated with Lupus-like diseases and excessive TNF-α production and septic shock in response to LPS exposure[321,322]. Its expression in RA synovial tissue, now extensively described throughout this thesis, supports its potentially critical clinical relevance. Again, MerTK+ macrophages are found predominantly in the lining layer and are lost gradually in phases of active disease, undergoing a compounding replacement at the hand of expanding and infiltrating monocyte-derived sublining populations[66,67]. In mouse models of IBD, double *Axl*/*Mertk* knockout mice suffer from increased severity in response to the DSS colitis model[323], while *Gas6* knockout mice are also more susceptible to DSS colitis[324]. In human Crohn's Disease (CD), MerTK positivity was found to be higher in control non-inflamed tissue regions, lower in inflamed tissue, and

highest in fibrotic tissue areas, pointing to a role for these macrophages in tissue repair processes and fibrosis[325]. Because of these anti-inflammatory implications, MerTK has become an attractive therapeutic target in cancer[326], and by extension may prove to be a promising IMID therapeutic target through either agonism or upregulation. These points are discussed further in the discussion section on clinical relevance.

Considering the context that precedes it, the data suggests a potential role for tissue microenvironment composition in driving MerTK upregulation and macrophage polarisation towards anti-inflammatory phenotypes by extension. While refuting the original hypothesis of collagen VI protein directly driving MerTK upregulation, the data does suggest that gel-like formulations of collagen VI, and other ECM substrates including Matrigel, may promote some degree of polarisation. This corroborates the existing body of evidence pointing to mechanical forces being involved in anti-inflammatory macrophage polarisation[327–329]. As with in any other tissue, synovial tissue composition is crucially designed to handle and respond to its specific mechanical forces. MerTK$^+$ macrophages sit in the lining layer of the synovium, a niche of the tissue exposed to its specific set of mechanical forces. These cells notably face the synovial fluid and the compressive forces put through the joint cavity, while also being tightly surrounded by the fibroblast-deposited basement-like membrane. One hypothesis worth exploring could be that the dysregulation of the mechanical properties of this tightly regulated niche leads to a non-permissive milieu for MerTK$^+$ macrophage maintenance, which leads to their inevitable replacement and loss of homeostatic control.

5.4.2 Collagen VI acts as a cellular sieve filtering immune ingress

The data presented as part of this chapter suggests that collagen VI deposited in anabolism-biased phases of tissue remodelling associated with remission-like tissue states, offers selecting specificity in retaining immune cell subsets. The localisation of this deposition on the outskirts of the perivascular niche, couple with this cell-retaining effect, puts it in a critical position to regulate immune cell ingress and establishment in the tissue. This tissue exclusion from collagen VI-rich regions was found to be most potently restricting to B-cells,

$CD4^+$ T-cells, and DCs, less so to $CD8^+$ T-cells and neutrophils, and least restricting to monocytes. This effect also appeared to be collagen VI specific. Indeed, type I collagen, which was characterised by a similar overexpression in RA over OA and active RA over remission as collagen VI, was found to have vastly different deposition patterns to collagen VI.

Different factors could underpin this observed specific restrictive effect. The first, and the object of preliminary investigations described in this chapter, is the role of collagen VI as a molecular mimic of integrin I domains. As touched on in the chapter, subsets of integrins including contain a structural inserted (I)-domain. This is the case of 9 of the 18 α subunits (1, 2, 10, 11, D, E, M, L, X) and all the β subunits of integrin pairings[309]. This inserted domain is in effect a VWF domain, and is structurally homologous to the 3 VWF domains present in the α1 and α2 chains of collagen VI, and the 12 domains present in α3. These collagen VI VWF domains even retain ECM adhesive properties of their homologous counterparts, as evidence by their binding to collagen type I[330]. Collagen VI VWF domains, through their structural homology to integrin I-domains, may therefore be obstructing integrin binding sites on other ECM substrates like fibronectin and modulating adhesion. This model is coherent with preliminary adhesion assay results pointing to a blocking effect of CD11b-expressing monocytes binding to fibronectin by collagen VI (Figure 5.11). It is also coherent with the complimentary results presented in Figure 5.12, demonstrating monocyte adhesion inhibition through coating of fibronectin with purified recombinant CD11b I-domains. Interestingly, I found CD11b I-domains also somewhat blocked T-cell adhesion to fibronectin, despite their likely major integrin α4 not containing an I-domain. This effect could notably be due to the presence of I-domains in the β subunits of these integrin pairings. Certain concentrations of collagen VI and their ratio to fibronectin or other ECM substrates may therefore create a “goldilocks” modulated adhesion, which is migration-permissive in tissue. By softening the adhesion dose-dependent slope, collagen VI may create a large window of opportunity for optimal monocyte migration, while T-cell adhesion remains sharp and almost binary between adhesion or no adhesion. If so, I would also hypothesise that cells with greater CD11b positivity would be broadly less restricted by collagen VI dark zones in tissue. This true for

some subsets, yet a bit more unclear for others. As previously described, $CD8^+$ T-cells were found to be less restricted to dark zones than CD4+ T-cells, and have been reported to express higher levels of CD11b, particularly following recent activation[331–333]. Neutrophils too are known to express CD11b, which is upregulated following activation[334]. DCs however can also express high levels of CD11b and were restricted to the dark zones, suggesting some other mechanisms including chemotactic gradients may be at play.

5.4.3 Monocyte ingress regulates the cellular sieve and modulates immune cell establishment in tissue

Throughout this thesis, collagen VI was found to be regulated through a fragile balance of degradation and deposition in RA, with hints at a similar process in other conditions such as IBD and head and neck cancer. Because of the almost universal expression of collagen VI in active RA by fibroblasts described in chapter 3, and its effective deposition being associated with remission-like tissue states as described in chapter 4, I hypothesised that the formation of dark zones would be degradation dependent. In this chapter, I identified dark zones as being active degradation sites, as highlighted by the rings of active degradation outlining them. These collagen degradation rings were lined with monocytes. When seen exiting the dark zones, deep tissue monocytes were also found to leave behind them paths of active collagen degradation. This was finally supported by proteomic analyses, which detected a broad array of MMP9-released cleavage products, a notable monocyte and neutrophil-associated protease[335]. My data therefore hints at monocytes being the primary implicated degraders of collagen VI, while fibroblasts are its depositors. In an inadvertent foreshadowing exercise, this was hinted at in the first figure of the first results chapter (Figure 3.1). Looking at matrisomal gene subcategory enrichment in the AMP dataset, fibroblasts were indeed found to be the clear predominant expressors of core matrisome genes while monocytes were found to contribute significantly to ECM regulator expression. With significant interest surrounding fibroblast-macrophage interactions in the synovium and in IMIDs, the deeper study of the

tandem of matrix depositors and degraders may provide the key to understanding the architectural narrative of the tissue in health and disease.

This partnership leads to a new working model of immune spatial patterning in RA, and possibly in other inflammatory diseases (Figure 5.17). In early inflammatory insults, neutrophils are known to be first on site and to respond[74] (Figure 5.17 A). Their tissue entry leads to inevitable tissue damage, partly due to the ROS and NET-dependent actions of neutrophils[336,337], and as evidence by the mouse AIA imaging done reported in chapter 4. Upon the resolution of this primary insult, fibroblast engage tissue repair mechanisms, including the deposition of tissue repair and fibrosis-associated collagen VI[148,169] (Figure 5.17 B). This deposited collagen VI, not present in mouse or human naïve tissue, restores tissue integrity while also cuffing the vascular niche to prevent further immune infiltration and the establishment of lymphocytes in the tissue. Upon further inflammatory insults or active disease flares, monocytes are recruited into the tissue and degrade their way through the deposited collagen VI layer (Figure 5.17 C). In forging their path, they create the observed *dark zones* and push their boundaries through the formation of the degradative front. This space clearance leads to the creation of a permissive niche for lymphocytes to settle in, starting with the formation of small aggregates which expand into large TLSs. Daring to speculate, this step-wise process could be an evolutionarily-inherited mechanism ensuring that the creation of adaptive immune outposts is restricted to tissues succumbing to repeat inflammatory insults. Interestingly finally, work on lymph nodes (LN) hints at a similar necessary disassembly of collagen VI to allow for LN expansion in inflammatory challenges[338].

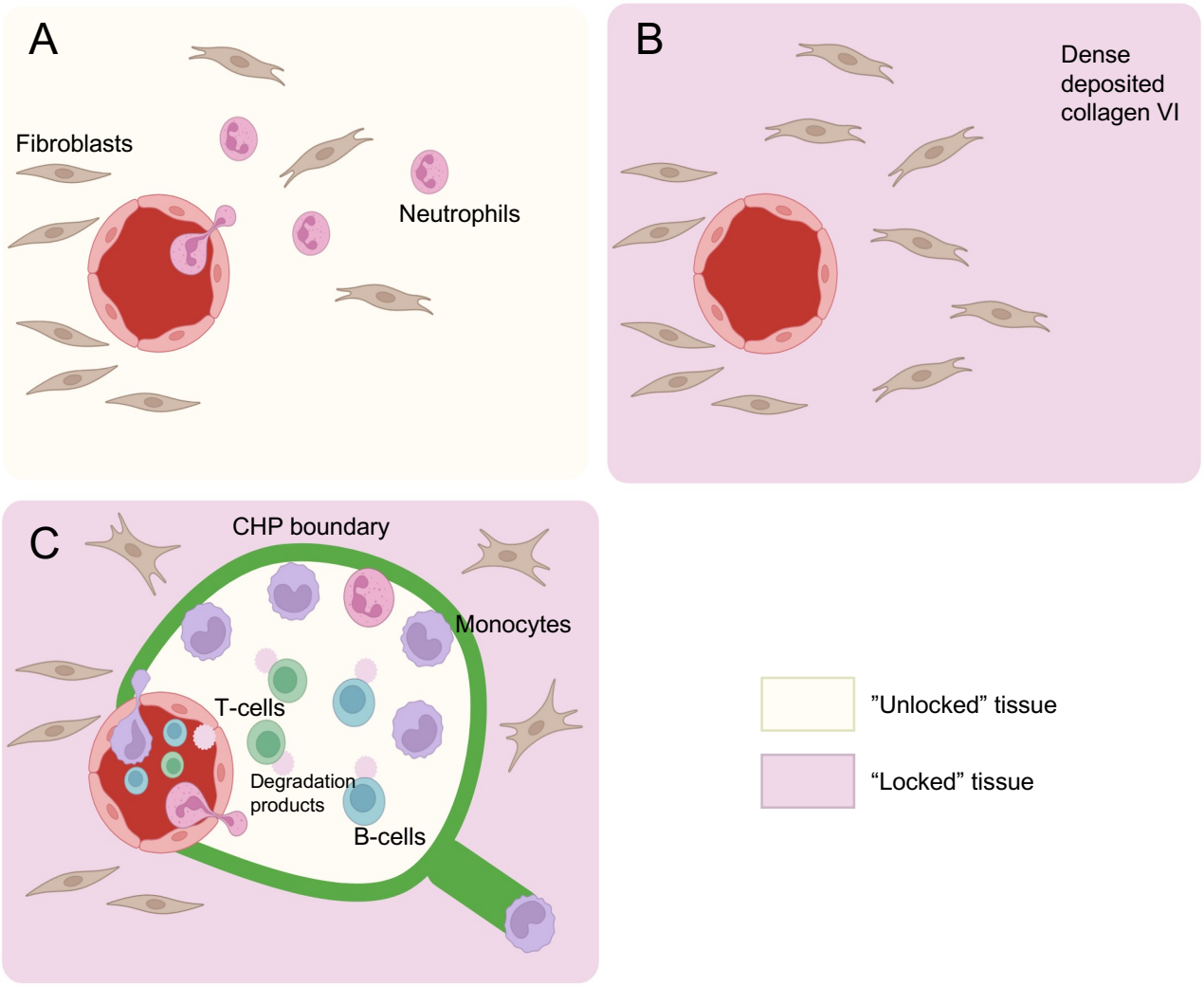


Figure 5.17 Working model of collagen VI deposition and immune regulation. A. Upon primary inflammatory insult, neutrophils enter the tissue and, through degranulation and NETosis, generate collateral tissue damage. **B.** In response to this tissue damage, synovial fibroblasts deposit ECM including collagen VI. This effective deposition occurs once the balance between TIMPs and MMPs shifts away from degradation and towards deposition. **C.** Upon repeat inflammatory insults, neutrophils and monocytes are recruited to the tissue by activated fibroblasts and macrophages. Neutrophils and monocytes have to degrade their way through the deposited collagen VI, creating the dark zones and the CHP-labelled degradative front that lines them. This expansion creates a permissive niche for lymphocytes to settle.

Chapter 6 | Final discussion

6.1 Conclusions

IMIDs, through their complexity and severity, have proved to be a challenging set of conditions to treat beyond mere symptom alleviation. Potentiated by modern technologies, the gradual uncovering of the cellular underpinnings of IMIDs is ushering in a new age therapeutic development. A conceptual shift from the study of cellular mechanisms to tissue-wide dynamics promises to bring context to already discovered pathological mechanisms and to reveal new patterns with targetable potential. Within this, a new-found appreciation for the extracellular matrix as more than a simple scaffold but a critical actor in the functional outcomes of a tissue promises to improve our understanding of IMIDs. Focusing on RA as a model IMID with hints towards ECM dysregulation[83], the aims of this work were therefore to characterise the expression of ECM genes and the spatial distribution of some of the proteins they encode, and to generate and test functional hypotheses surrounding the identified key ECM proteins.

6.1.1 A complex transcriptomic landscape with hints at dysregulation

Bioinformatic analyses reported in chapter 3 identified fibroblasts as the cell type most enriched for core matrisome gene expression, while monocyte gene expression was found to be enriched in matrisome-associated genes including proteases. Upon closer inspection, published functionally and spatially distinct fibroblast subsets were found to have specific and unique matrix expression profiles. De novo matrisome clustering applied to fibroblasts identified clusters with significant homology to classically clustered fibroblasts pointing to the critical role of these spatially and functionally distinct fibroblasts in expressing the ECM building blocks of distinct localised microarchitectures. Differential expression analysis applied to RA and frozen shoulder dataset fibroblasts revealed unique disease-specific matrix profiles, while also defining a cross-disease inflammation resolution ECM expression signature. Finally, a target prioritisation scheme was developed to identify key differentially expressed matrix genes and their place within broader dysregulated pathways in active RA disease states. The IGF signalling pathway, and the dysregulation of basement membrane

composition were found to be the two of such key pathways, and elected to be studied further at the protein level.

Together these data paint a complex picture of synovial tissue composition, where each cell type, particularly fibroblasts, macrophages, and monocytes, work together to define the tissular landscape. Within this landscape, fibroblasts are the core depositors of ECM, with different fibroblast subsets depositing different matrix components. Their expression profiles are dysregulated in RA, possibly leading to growth factor signalling imbalances and disassembly of basement membranes, critical structural components of the synovial architecture.

6.1.2 Key ECM proteins define subsynovial niches with functional implications

Multiplexed immunofluorescence was used to add spatial resolution and protein-level information to the transcriptomic data gathered surrounding the key matrix genes and proteins. IGFBP7, found in chapter 3 to be overexpressed in RA compared to OA and active RA compared to RA in remission across sublining fibroblasts, was here found to be specifically restricted to the vascular niche. This corroborated its published functions as angiomodulin, playing a synergistic angiogenic role with its partner VEGF-A[282], and being associated with increased vascular leakiness[283]. This mechanism is suspected to be at play in many IMIDs and could warrant further investigations[339–341]. IGFBP5 protein levels and distribution matched the transcriptomics, highlighting the synovial lining layer as its predominant niche. Given its overexpression in remission compared to active RA and its published associations with tissue repair promotion and fibrosis[274,288], one can hypothesise that IGFBP5 may be playing a role in shaping the ECM composition of the lining layer. IGFBP5 could for example be upregulated in remission as part of the tissue remodelling required to re-establish lining layer integrity after phases of inflammation. Prior to the spatial protein mapping, COL4A1 was found to be expressed in the lining layer of the synovium, likely contributing to the reported yet elusive basement-like membrane[247,248]. Staining with my antibody of choice unfortunately revealed COL4A1 to be exclusively localised to vascular basement membranes, possibly due to post

translational modifications or conformational peculiarities obstructing the binding site of the monoclonal antibody. COL6A1 finally was found to be differentially deposited at the protein level compared to its expression dynamics. COL6A1 deposition was indeed found to be associated with remission-like tissue states, likely owing to differences in degradation and deposition potentials between phases of active RA and RA in remission. Spatial patterns of collagen VI also hinted at immune hubs localising to areas devoid of collagen VI.

Together, the data collated in this chapter characterised the specific niches associated with the matrix proteins of interest. They also solidified collagen VI as a functionally interesting target, with associations with MerTK macrophage positivity and spatial localisation of immune cells.

6.1.3 Collagen VI is a remission-deposited regulator of immune ingress which is remodelled by infiltrating monocytes

A renewed focus was therefore placed on the tissular distribution of collagen VI and its relationships with immune cell subsets. I found collagen VI exerts a differential restricting effect on different immune subsets, with monocytes being the only whole unrestricted cell type. This effect was found to be specific to collagen VI, with collagen I exhibiting a vastly different deposition pattern despite converging expression dynamics. Looking into the functional underlying mechanisms of this restriction, I found collagen VI exerted an adhesion blocking action on monocytes specifically over tested T-cell and B-cell lines. One hypothesis to explain these discrepancies is that of molecular mimicry of the collagen VI VWF domains for integrin I-domains. Collagen VI deposited in remission-like tissue states may indeed outcompete integrins for the binding to other ECM substrates such as fibronectin. This is cohesive given the known binding of these same I-domains to collagen I. Another hypothesis to explain the lesser monocyte retention is one of degradation-dependent migration. Monocytes were found to localise to a degradative front forming the outer border of collagen VI dark zones, and their deeper tissue ingress was found to be degradation associated. Proteomics also confirmed the RA-associated release of MMP9 degradation products of collagen VI, a protease notably

secreted by monocytes[342]. Together, these data therefore point to collagen VI deposition by fibroblasts as more than just a restoration of tissue integrity, but also an immunomodulatory mechanism. Fibroblasts thereby restrict tissue access to the cellular orchestrators of adaptive immunity. A process regulated on a second level by the repeated infiltration of monocytes, which degrade their way through collagen VI, leading to an adaptive immune permissive domain expansion.

6.2 Clinical relevance

The work detailed throughout this thesis holds translational and potential clinical relevance which is worth exploring. These potential applications are particularly relevant to IMIDs, and specifically RA.

As extensively described throughout the literature, loss of the MerTK^{+} macrophage compartment is both associated and functionally involved in active RA disease[66,67]. Part of my work hints at the rebalancing of tissue microenvironment composition as a potential mechanism of interest for the induction of MerTK^{+} macrophages. Reestablishment of a structural homeostasis may therefore be necessary to induce the sustained recovery of this anti-inflammatory macrophage compartment in RA.

In addition to this loss of homeostatic tissue resident macrophages, RA pathology is driven by extensive dysregulated immune ingress[58]. On this front, my work sheds a new light on collagen VI deposition as an endogenous regulatory mechanism of this process. This novel mechanism could be targeted in multiple ways. Firstly, as evidenced through my work, effective collagen VI deposition is the result of a favourable balance between circulating MMP and TIMP levels. MMP inhibition, in particularly MMP9, could therefore be promising therapeutic avenues. The AIA mouse model supports these predictions, with MMP9 knockout mice displaying milder arthritis[343]. A phase 1b study has also shown Andecaliximab to be safe, well tolerated, and effective in binding to circulating MMP9[344], while the phase 2 trial was terminated for undisclosed reasons unrelated to safety (NCT02862574). The new mechanistic

insights provided by this study may therefore enable rational trial designs and patient stratification for re-consideration of MMP9 as a therapeutic target of interest.

Instead of biasing the microenvironment towards collagen VI deposition, the mechanism by which collagen VI permits monocyte ingress could also be targeted. As previously described, collagen VI creates a “goldilocks” adhesion level to potentiate monocyte migration in tissue by obstructing some of their binding sites to other ECM substrates. These collagen VI domains could be targeted to abolish this optimal migratory window. This rebalancing could be done through exogenous addition of purified collagen VI binding domains, or conversely through the inhibition of the competitively binding collagen VI domains, through steric hindrance of a large binding antibody for example.

Finally and most obviously, the identification of this novel mechanism strengthens and adds resolution to the use of collagen VI and its degradation products as biomarkers. In parallel to this thesis Nordic Bioscience has been harnessing collagen degradation products, including collagen VI, as biomarkers of disease activity status in fibrosis[305], other IMIDs like IBD[183], and cancer[345]. Interestingly, their work has recently culminated in the identification of decreases in levels of collagen VI degradation biomarker C6M as an effective biomarker of treatment effectiveness in RA patients treated with the IL-6 receptor antagonist Tocilizumab[184]. Within this context, this work sheds a light on the tissue dynamics associated with increased or decreased collagen VI degradation. Following further validation, such biomarkers could for example be used as proxies for the quantification of lymphocytic settlement potential in tissue.

This body of work also provides transcript and protein-level evidence for the pursual of other matrix targets of interest, which could not be explored fully in the span of a thesis. These notably include Netrin-4, IGFBP5, and IGFBP7.

6.3 Future work

The work presented in this thesis highlights collagen VI as an important remission-deposited regulator of immune ingress, regulated itself by a fragile balance of proteases and their inhibitors. Preliminary results suggest this migratory restriction is mediated by a

competitive molecular mimicry and proteolytic migration of monocytes, expanding lymphocyte-permissive niches and releasing bioactive degradation products. A number of important experiments should be carried out to validate and finish characterising the discovered mechanisms.

6.3.1 Ongoing and near future work

a. In vitro characterisation of collagen VI migration effects

Work should be performed to build on the preliminary migration assays reported in chapter 5. These assays are being run by new incoming DPhil student Leia Worthington. The work currently focuses on the optimisation of both the ECM coating of transwell inserts and the choice of chemoattractant. The aim of this work is to validate the observed tissue effects of collagen VI on migration in vitro. A reproducible migration assay will also enable the testing of the effect of the purified recombinant I-domain on cell migration.

b. Extending our multiplexed immunofluorescence validation panel

One area of improvement and expansion is the multiplexed immunofluorescence validation panel employed in chapter 5. While the panel was sufficient in highlighting the specific monocyte predilection for collagen VI-rich regions, OA and RA patient samples remain particularly variable. The additional staining for 3 more patients in each condition could be highly valuable in capturing this variability. The incorporation of more markers, particularly CD11b, could add significant value. A direct tissular assessment of CD11b positivity of cells escaping the dark zones would strengthen the suspected uncovered mechanism.

c. Collagen VI fragment-specific antibody development and early applications

Antibody generation company Biosynth were commissioned to generate novel antibodies specific to the MMP9 cleavages site-adjacent D11 fragment found to be detectable and specific to RA samples through proteomics. This antibody has begun to be validated through ELISA and Western blot by Molly Bower. Further interesting validation would include testing the suitability of the antibody for immunohistochemistry (IHC) or immunofluorescence

with an eye on future tissue mapping. With the antibody being validated for ELISAs, an interesting early application could be the quantification of D11 in the serum of RA and healthy control patients. This would provide a stepping stone for its use as a biomarker.

6.3.2 Medium term future work

In the medium term (3-6 months) D11 should be purified in larger quantities to support higher throughput functional investigations. Attempts were made to purify it in house, which led to high degrees of protein aggregation and unusable yields. Work will be done to optimise the purification process and potentially to unfold the aggregated purified protein into functionally active units. This purified protein will then be put through adhesion and migration assays to assess its functional similarities and dissimilarities to the full length native collagen VI. Finally cells will be isolated from D11 adhesion assays and subjected to bulk RNAseq to look for any additional bioactive properties of D11 through the identification of differentially regulated signalling pathways. If the D11 antibody is found to be usable for IHC and IF, work will then be done to map D11 distribution throughout human synovial tissue. Spatial mapping, as well as colocalisation with specific immune cell subsets should shed new light on the potential functions of these degradation products in tissue. Based on these early identified functional roles for intact collagen VI and D11, mouse models should also be run to assess the impact of these mechanisms on models of arthritis.

6.3.3 Long term project outlooks

This project has uncovered a novel regulator of cellular spatial distribution in immune-related conditions. This opens a number of questions which could be investigated in the future, including:

- Are these spatial restriction systems targetable therapeutically?
- Is this a conserved mechanism for the establishment of peripheral lymphoid aggregates across all IMIDs?
- Do other core ECM building blocks contribute to the effect in a complex interplay?

Appendix Figures

Supplementary Figure 1. Boxplots of the AUC expression enrichment scores (Y axis) for each AMP1 cell cluster (X axis) and for each matrisome subcategory

Supplementary Figure 2. Dot plot of the matrisome gene expression levels in each of the total synovial clusters annotated in the Alivernini dataset.

Supplementary Figure 3. Dot plot of the matrisome gene expression levels in each of the total synovial clusters annotated in the Alivernini dataset.

Supplementary Figure 4. Dot plot of the matrisome gene expression levels in each of the total synovial clusters annotated in the Alivernini dataset.

Supplementary Figure 5. AMP1 dataset differentially expressed matrisome genes between OA and RA fibroblasts.

Supplementary Figure 6. Alivernini dataset differentially expressed matrisome genes between remission and active RA fibroblasts.

Supplementary Figure 7. Wei dataset differentially expressed matrisome genes between control mouse and serum transfer-induced arthritis (STIA) mouse fibroblasts.

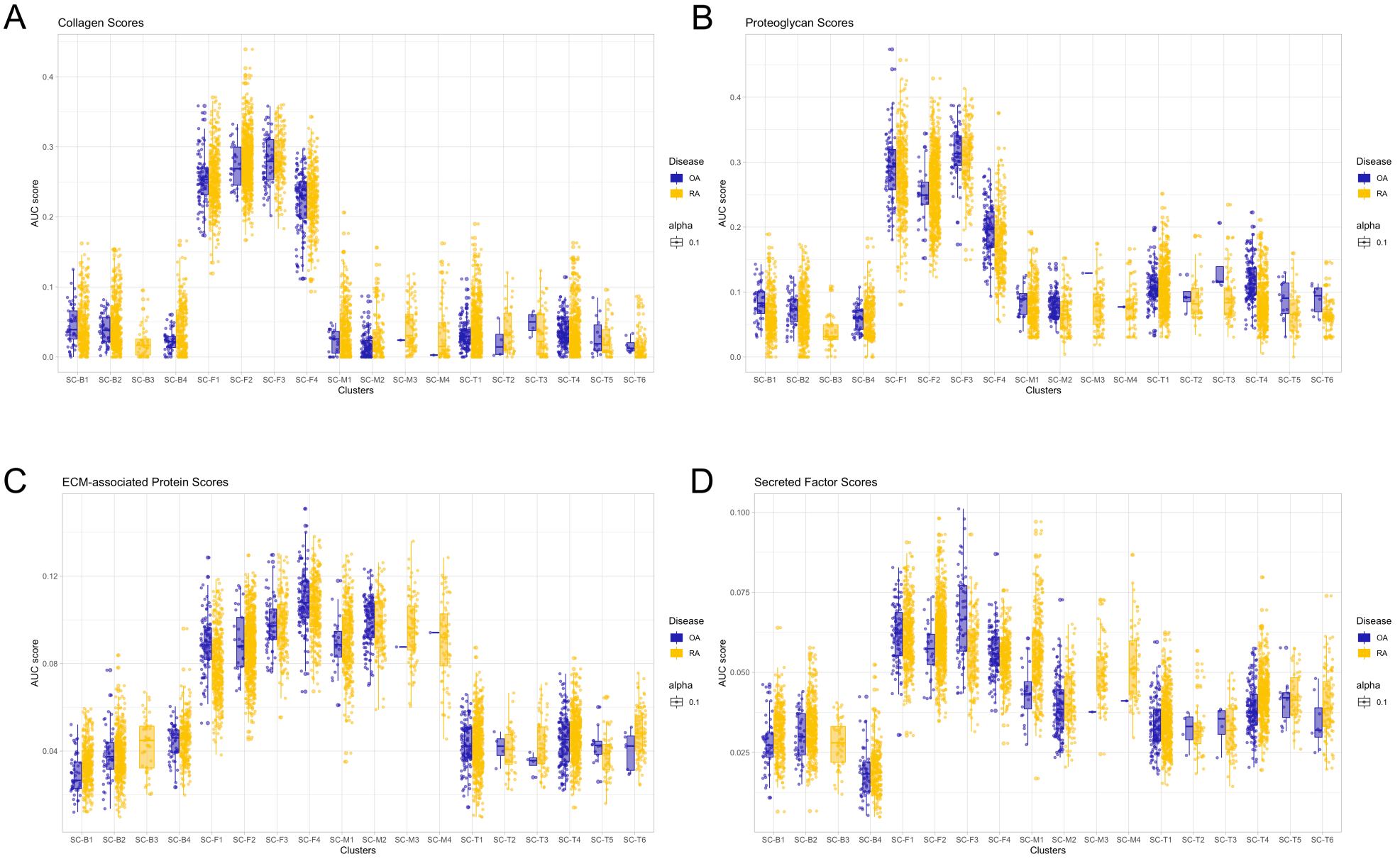
Supplementary Figure *8*. Individual antibody signals for each antibody in the multiplexed immunofluorescence exploratory panel.

Supplementary Figure *9*. COL4A1 and COL6A1 define two vascular endothelial subtypes, separate from LYVE1+ lymphatic endothelium

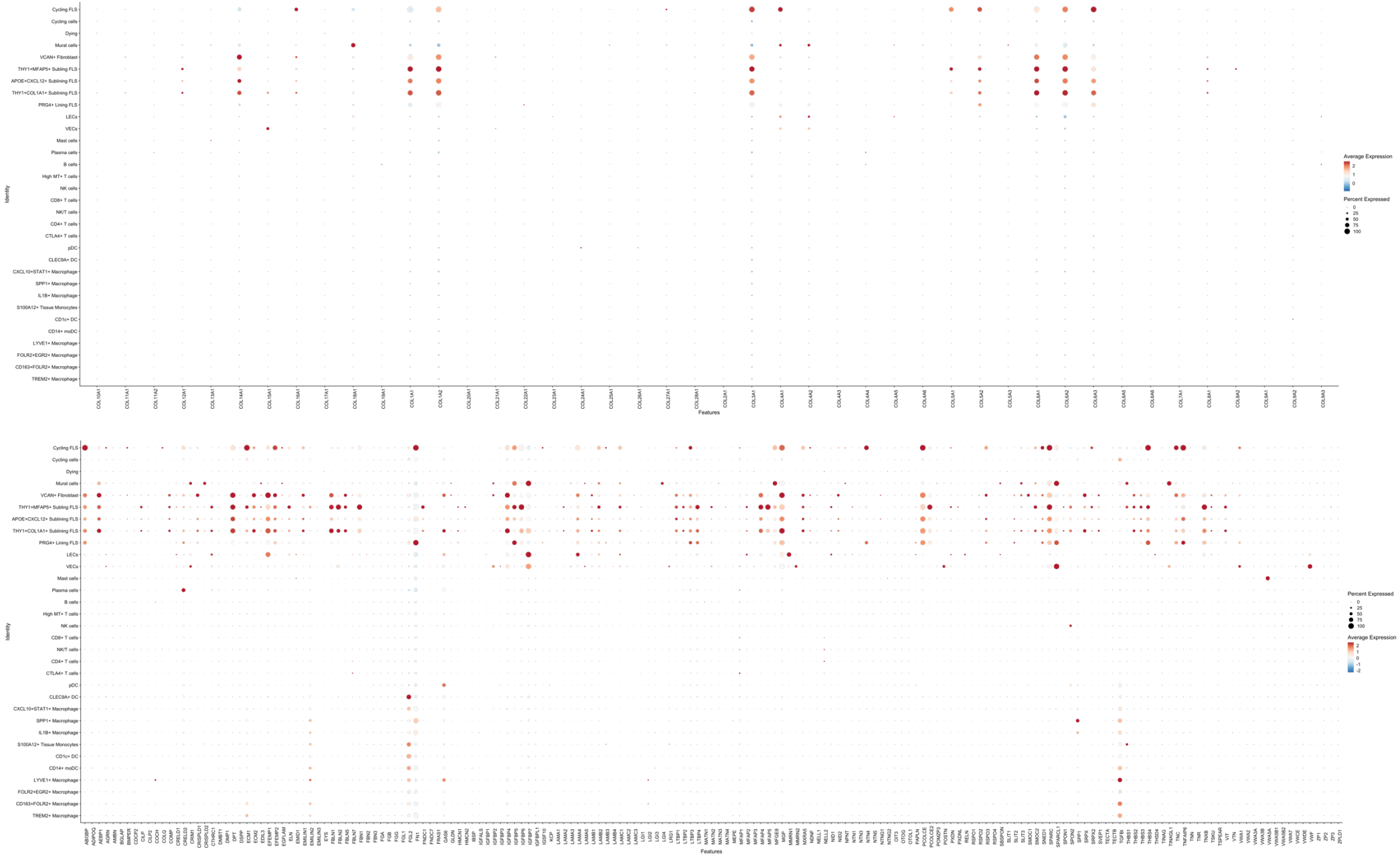
Supplementary Table 1. Curated matrisome excluding cytokines and chemokines.

Supplementary Table 2. Decision matrix for cell phenotyping as part of validation panel.

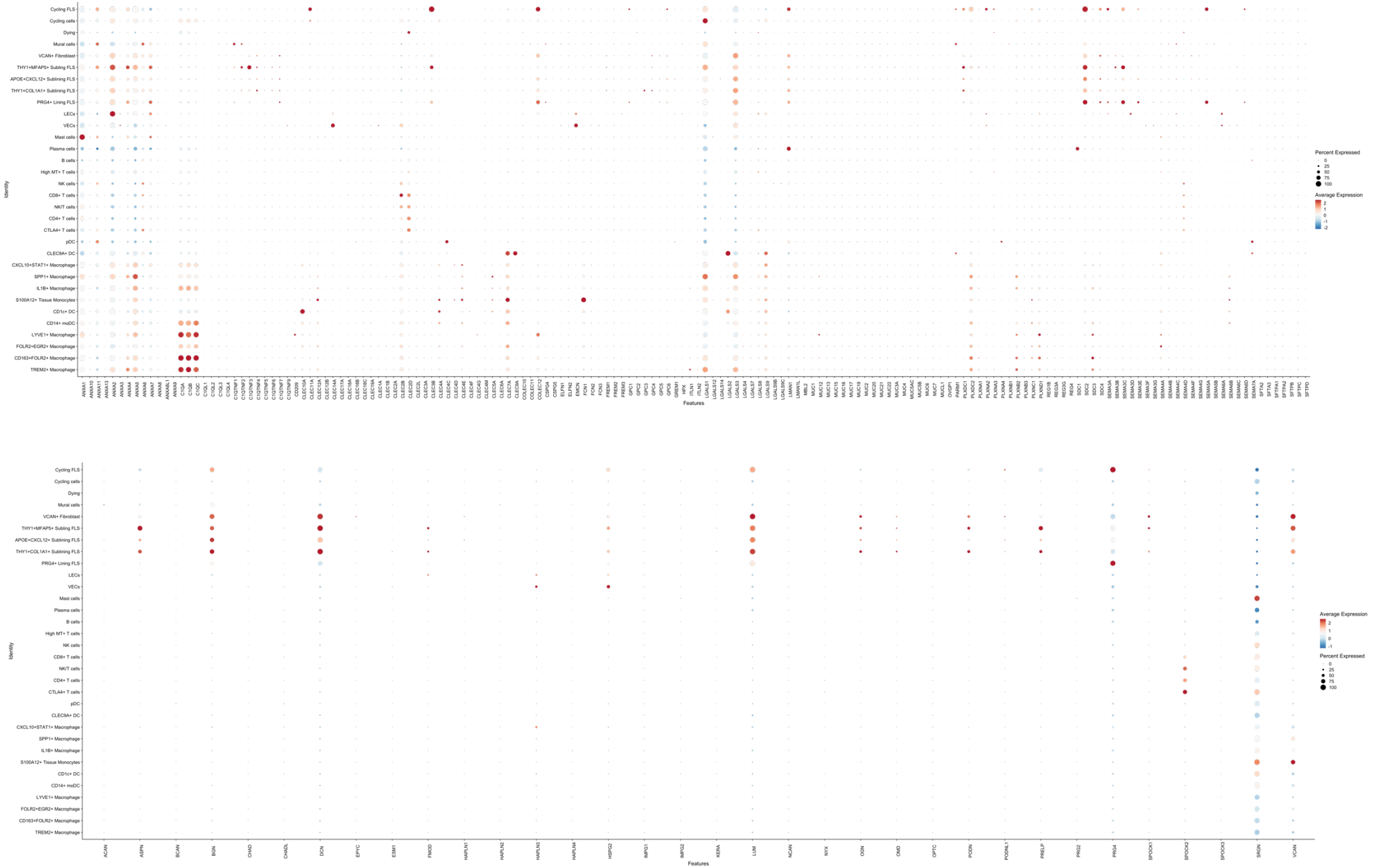
Supplementary Figure 1. Boxplots of the AUC expression enrichment scores (Y axis) for each AMP1 cell cluster (X axis) and for each matrisome subcategory: Collagens (**A**), Proteoglycans (**B**), ECM-associated proteins (**C**), Secreted Factors (**D**). Each point on the boxplot represents an individual cell of the dataset, split between OA (blue) and RA (yellow).



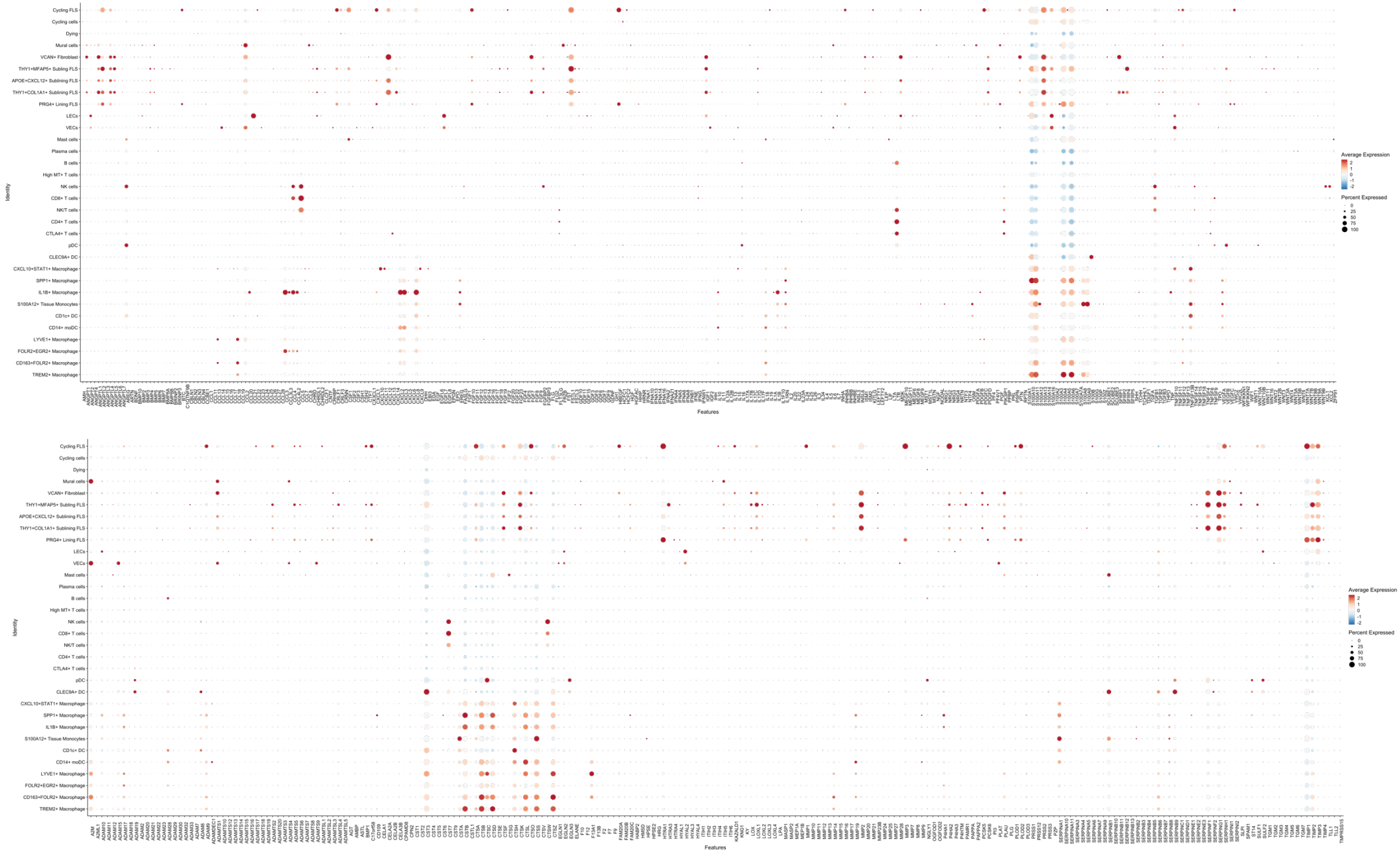
Supplementary Figure 2. Dot plot of the matrisome gene expression levels in each of the total synovial clusters annotated in the Alivernini dataset. Dot size represents the percentage of cells in the cluster expressing the gene of interest, and dot colour represents its level of expression. **A.** Dot plot of matrisome glycoproteins. **B.** Dot plot of matrisome collagens.

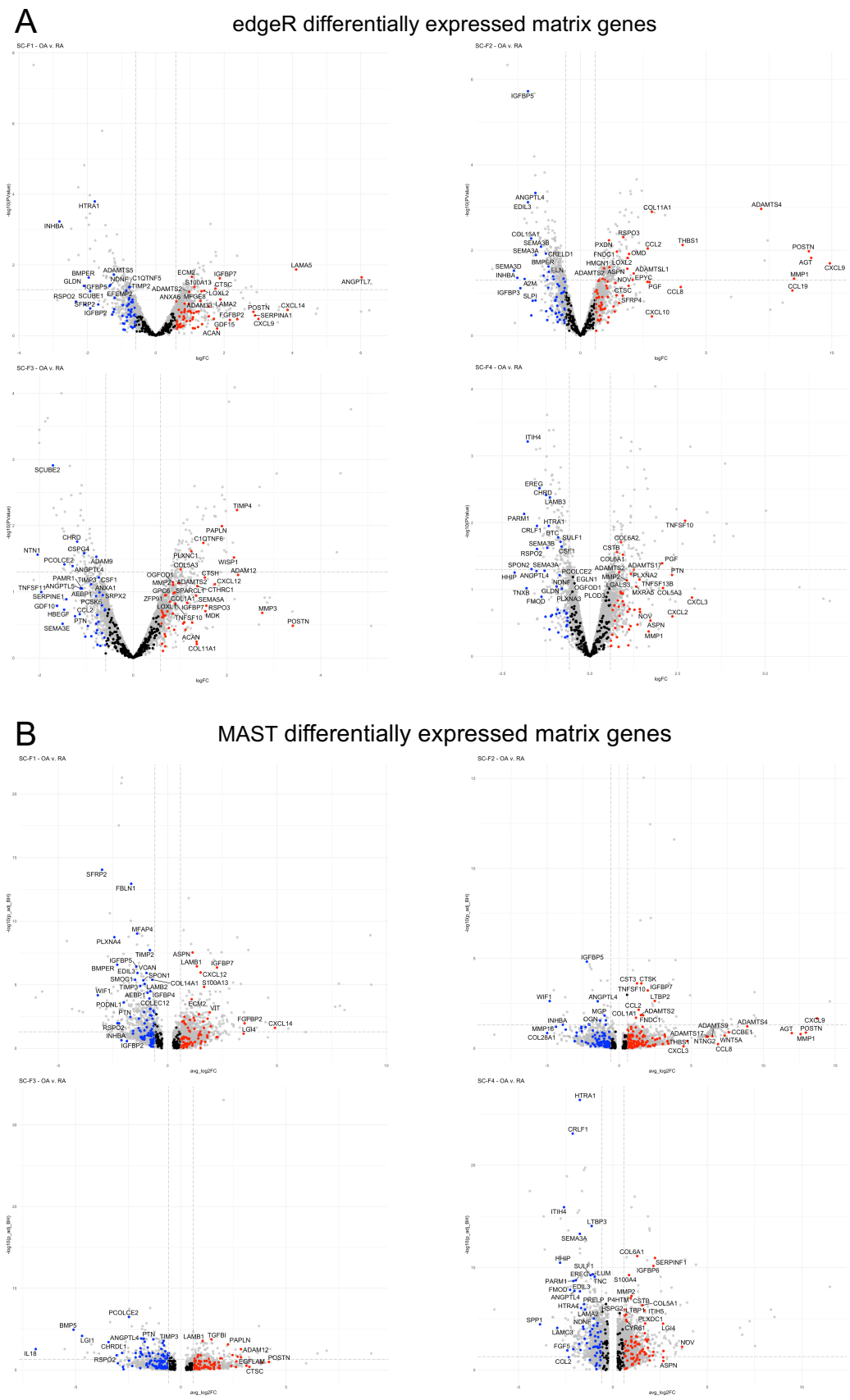


Supplementary Figure 3. Dot plot of the matrisome gene expression levels in each of the total synovial clusters annotated in the Alivernini dataset. Dot size represents the percentage of cells in the cluster expressing the gene of interest, and dot colour represents its level of expression. **A.** Dot plot of matrisome proteoglycans. **B.** Dot plot of matrisome ECM-affiliated genes.

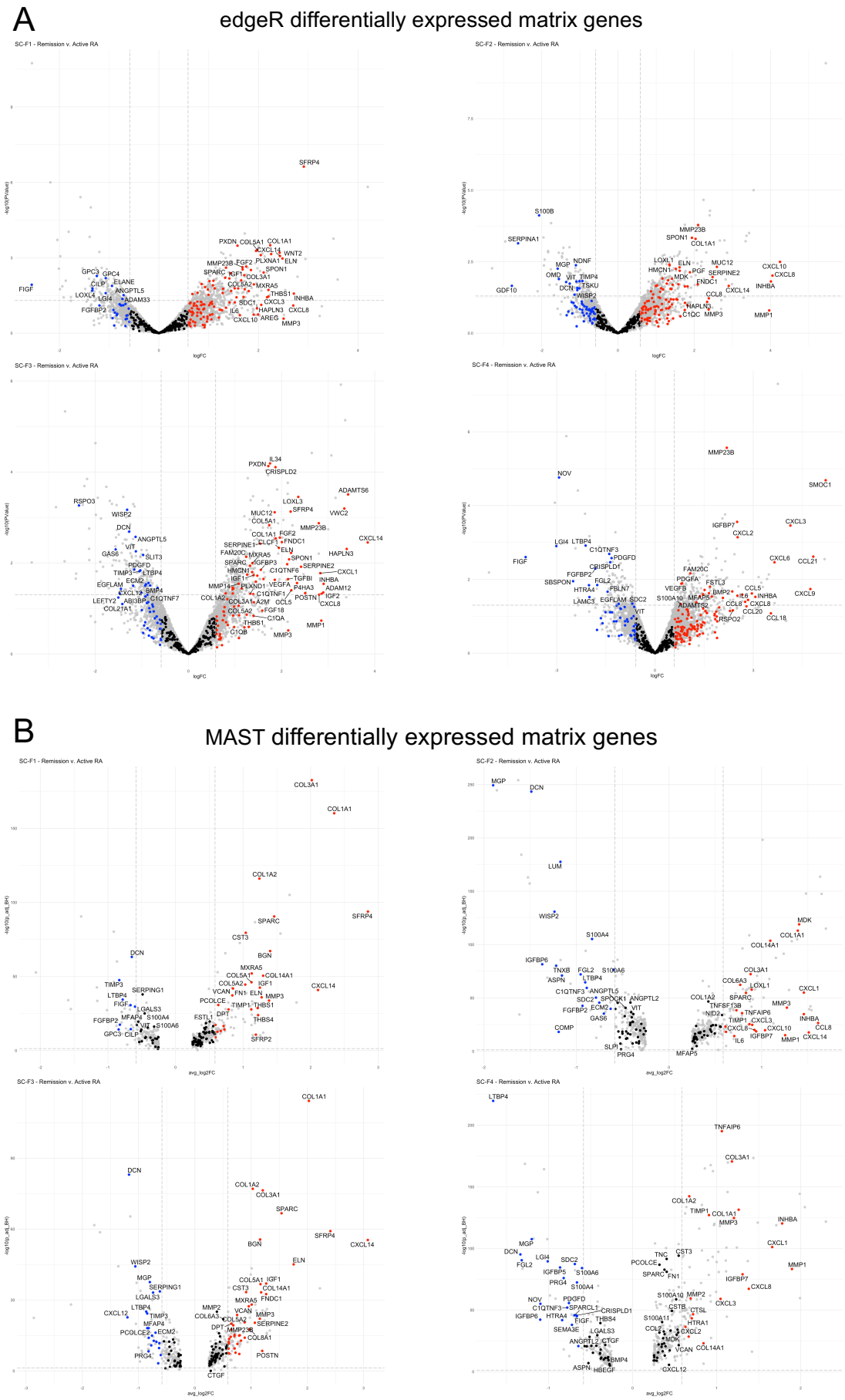


Supplementary Figure 4. Dot plot of the matrisome gene expression levels in each of the total synovial clusters annotated in the Alivernini dataset. Dot size represents the percentage of cells in the cluster expressing the gene of interest, and dot colour represents its level of expression. **A.** Dot plot of matrisome ECM regulators. **B.** Dot plot of matrisome secreted factor genes.



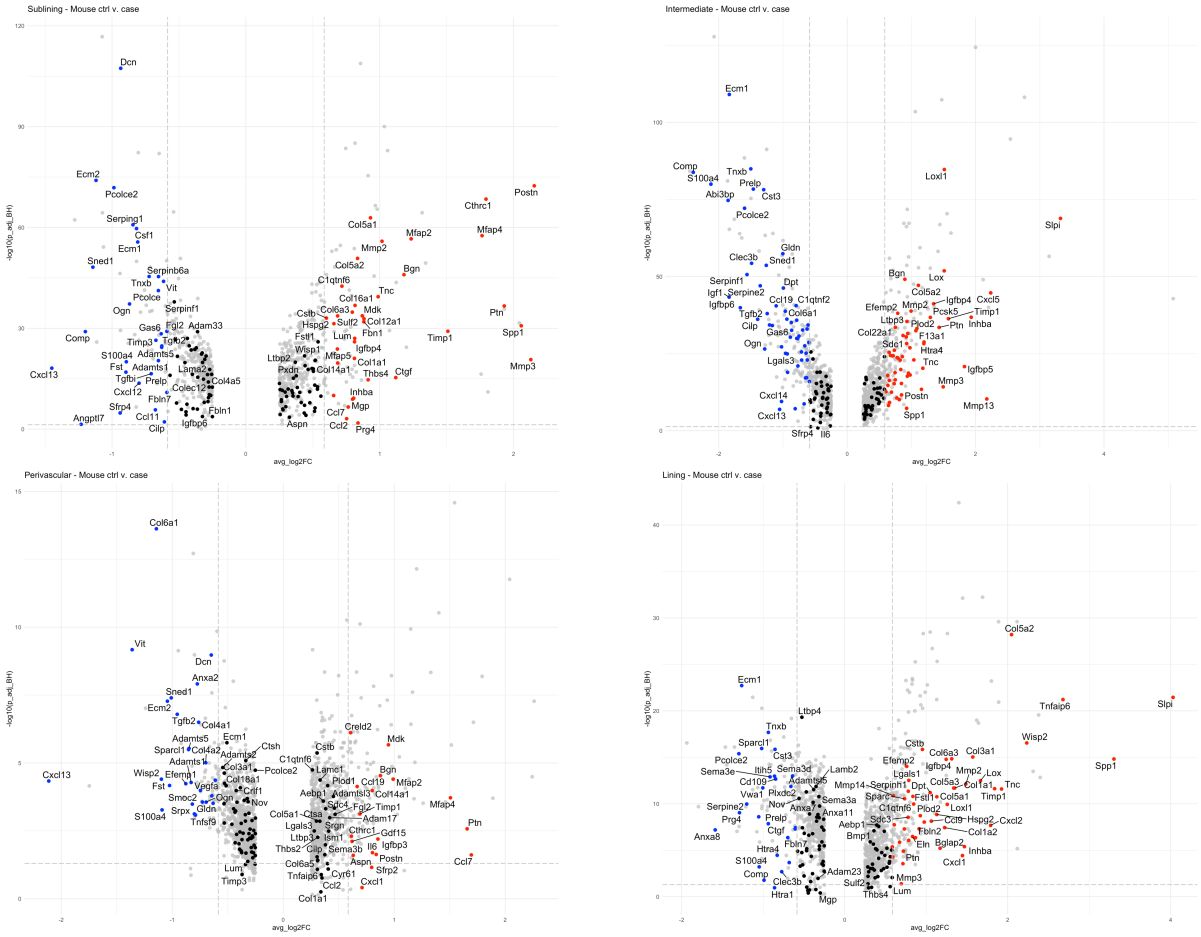


Supplementary Figure 5. AMP1 dataset differentially expressed matrisome genes between OA and RA fibroblasts. A. Volcano plots showing the output DEGs from the pseudobulk *edgeR* analysis. **B.** Volcano plots showing the output DEGs from the *MAST* analysis. Differential expression is here defined by a fold change >1.5 and a p-value <0.05. DEGs in blue are differentially expressed in OA and DEGs in red are differentially expressed in RA. Non-matrisome genes are greyed out.

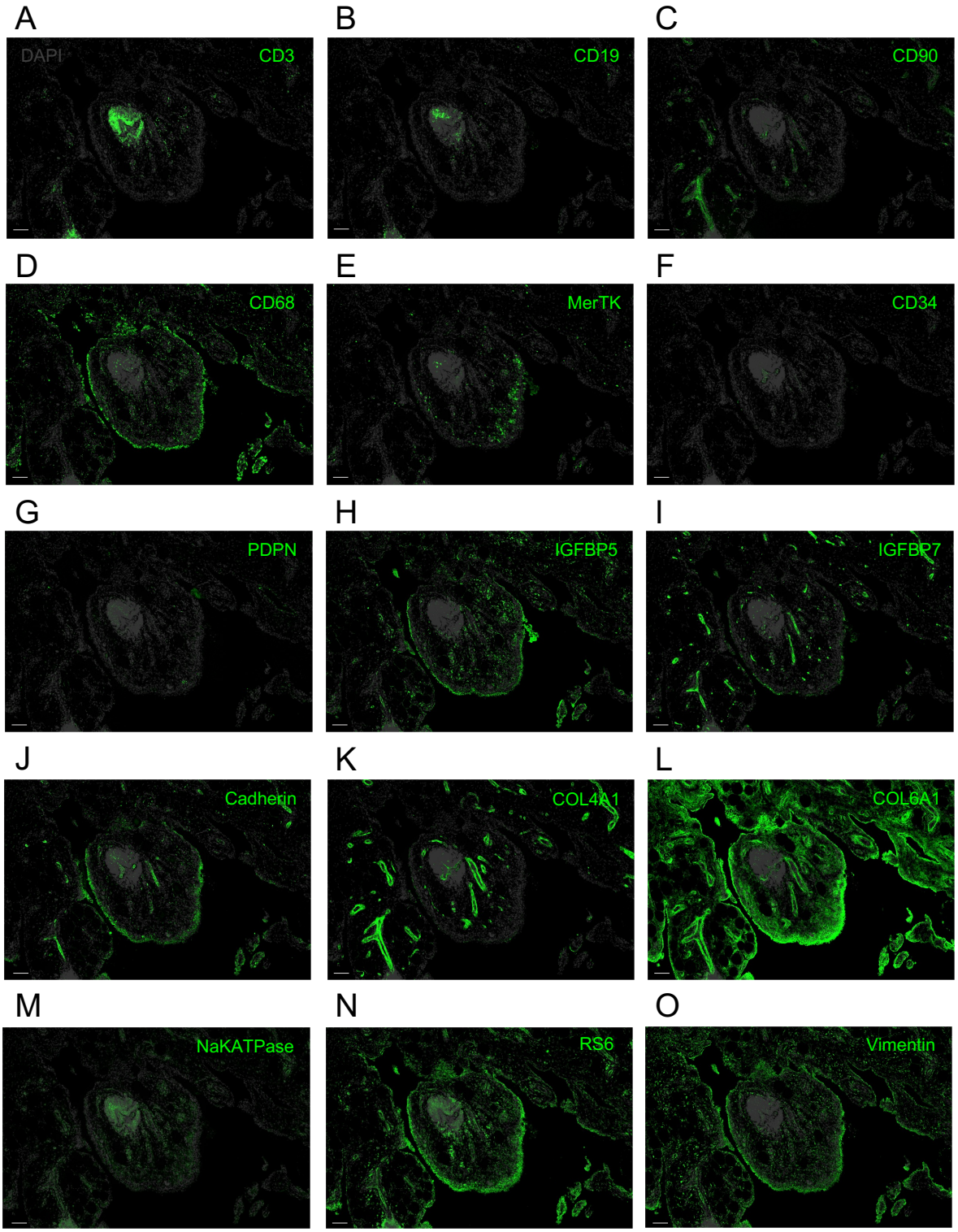


Supplementary Figure 6. Alivernini dataset differentially expressed matrisome genes between remission and active RA fibroblasts. A. Volcano plots showing the output DEGs from the pseudobulk *edgeR* analysis. **B.** Volcano plots showing the output DEGs from the *MAST* analysis. Differential expression is here defined by a fold change >1.5 and a p-value <0.05. DEGs in blue are differentially expressed in remission and DEGs in red are differentially expressed in active RA. Non-matrisome genes are greyed out.

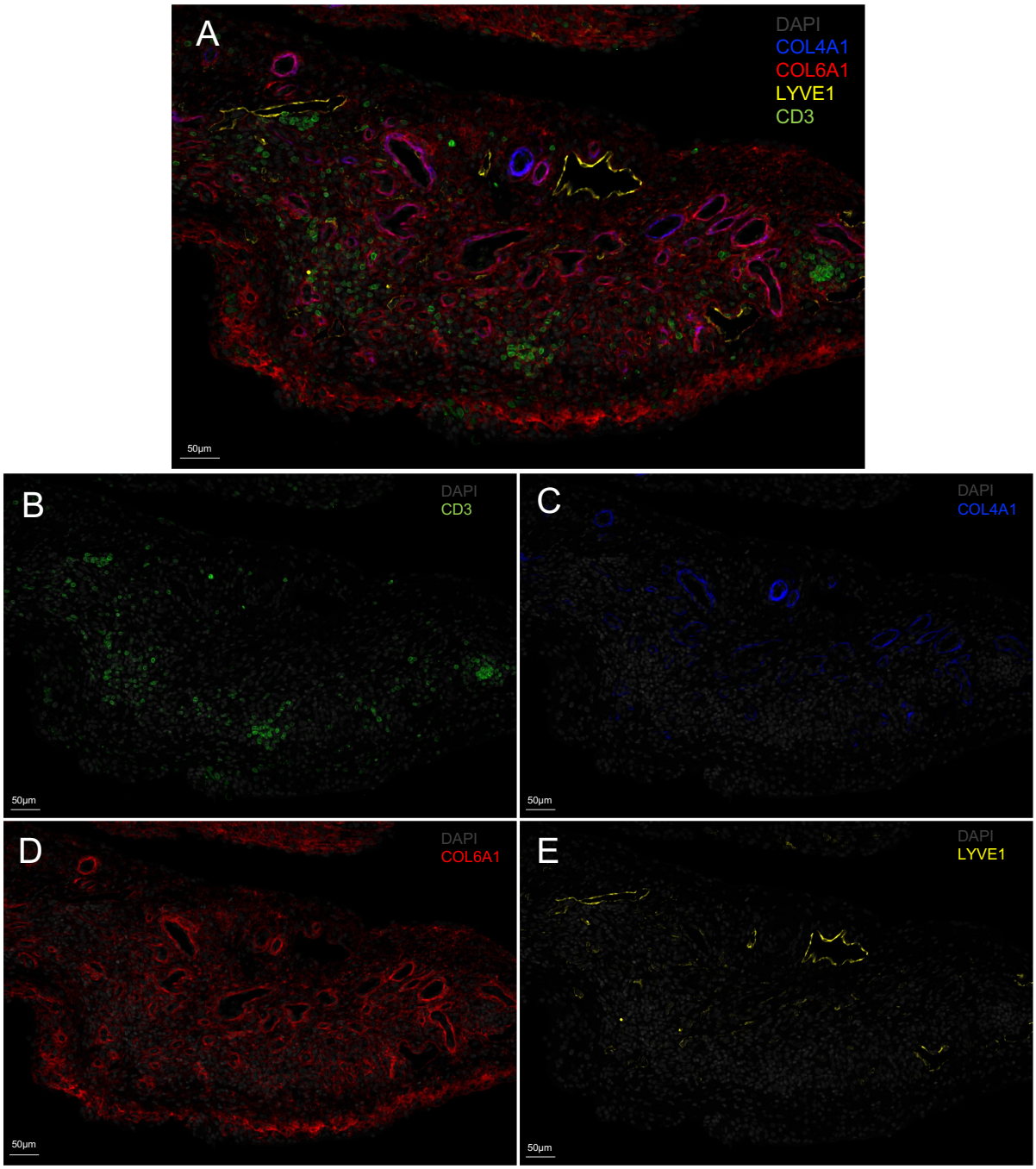
MAST differentially expressed matrix genes



Supplementary Figure 7. Wei dataset differentially expressed matrisome genes between control mouse and serum transfer-induced arthritis (STIA) mouse fibroblasts. Volcano plots showing the output DEGs from the *MAST* analysis. Differential expression is here defined by a fold change >1.5 and a p-value <0.05. DEGs in blue are differentially expressed in control mice and DEGs in red are differentially expressed in STIA mice. Non-matrisome genes are greyed out.



Supplementary Figure 8. Individual antibody signals for each antibody in the multiplexed immunofluorescence exploratory panel. 5µm thick sections were cut from synovial biopsies of 5 OA and 9 RA patients and antigen retrieved (Table 4.1). The sections were then cyclically stained for rounds of 3 antibodies at a time, described in Table 4.2, and imaged on the CellDIVE system. Visualisation was then performed in QuPath to remove background fluorescence and export the reported images. Error bars = 100µm.



Supplementary Figure 9. COL4A1 and COL6A1 define two vascular endothelial subtypes, separate from LYVE1+ lymphatic endothelium. 5µm thick serial section of JRP112 was cut from synovial biopsy antigen retrieved. The slide was then stained for DAPI, CD3 (**B**), COL4A1 (**C**), COL6A1 (**D**), LYVE1 (**E**). Error bars = 50µm.

Supplementary Table 1. Curated matrisome excluding cytokines and chemokines.

Division	Category	Subcategory	Gene Symbol
Core matrisome	ECM Glycoproteins		ABI3BP
Core matrisome	ECM Glycoproteins		ADIPOQ
Core matrisome	ECM Glycoproteins		AEBP1
Core matrisome	ECM Glycoproteins		AGRN
Core matrisome	ECM Glycoproteins		AMBN
Core matrisome	ECM Glycoproteins		AMELX
Core matrisome	ECM Glycoproteins		AMELY
Core matrisome	ECM Glycoproteins		BGLAP
Core matrisome	ECM Glycoproteins		BMPER
Core matrisome	ECM Glycoproteins		BSPH1
Core matrisome	ECM Glycoproteins		CDCP2
Core matrisome	ECM Glycoproteins		CILP
Core matrisome	ECM Glycoproteins		CILP2
Core matrisome	ECM Glycoproteins		COCH
Core matrisome	ECM Glycoproteins		COLQ
Core matrisome	ECM Glycoproteins		COMP
Core matrisome	ECM Glycoproteins		CRELD1
Core matrisome	ECM Glycoproteins		CRELD2
Core matrisome	ECM Glycoproteins		CRIM1
Core matrisome	ECM Glycoproteins		CRISPLD1
Core matrisome	ECM Glycoproteins		CRISPLD2
Core matrisome	ECM Glycoproteins		CTGF
Core matrisome	ECM Glycoproteins		CTHRC1
Core matrisome	ECM Glycoproteins		CYR61
Core matrisome	ECM Glycoproteins		DDX26B
Core matrisome	ECM Glycoproteins		DMBT1
Core matrisome	ECM Glycoproteins		DMP1
Core matrisome	ECM Glycoproteins		DPT
Core matrisome	ECM Glycoproteins		DSPP
Core matrisome	ECM Glycoproteins		ECM1
Core matrisome	ECM Glycoproteins		ECM2
Core matrisome	ECM Glycoproteins		EDIL3
Core matrisome	ECM Glycoproteins		EFEMP1
Core matrisome	ECM Glycoproteins		EFEMP2
Core matrisome	ECM Glycoproteins		EGFLAM
Core matrisome	ECM Glycoproteins		ELN
Core matrisome	ECM Glycoproteins		ELSPBP1
Core matrisome	ECM Glycoproteins		EMID1
Core matrisome	ECM Glycoproteins		EMILIN1
Core matrisome	ECM Glycoproteins		EMILIN2
Core matrisome	ECM Glycoproteins		EMILIN3
Core matrisome	ECM Glycoproteins		EYS
Core matrisome	ECM Glycoproteins		FBLN1
Core matrisome	ECM Glycoproteins		FBLN2
Core matrisome	ECM Glycoproteins		FBLN5
Core matrisome	ECM Glycoproteins		FBLN7
Core matrisome	ECM Glycoproteins		FBN1
Core matrisome	ECM Glycoproteins		FBN2
Core matrisome	ECM Glycoproteins		FBN3
Core matrisome	ECM Glycoproteins		FGA
Core matrisome	ECM Glycoproteins		FGB
Core matrisome	ECM Glycoproteins		FGG
Core matrisome	ECM Glycoproteins		FGL1
Core matrisome	ECM Glycoproteins		FGL2
Core matrisome	ECM Glycoproteins		FN1
Core matrisome	ECM Glycoproteins		FNDC1
Core matrisome	ECM Glycoproteins		FNDC7
Core matrisome	ECM Glycoproteins		FNDC8
Core matrisome	ECM Glycoproteins		FRAS1
Core matrisome	ECM Glycoproteins		GAS6
Core matrisome	ECM Glycoproteins		GLDN
Core matrisome	ECM Glycoproteins		HMCN1
Core matrisome	ECM Glycoproteins		HMCN2
Core matrisome	ECM Glycoproteins		IBSP
Core matrisome	ECM Glycoproteins		IGFALS
Core matrisome	ECM Glycoproteins		IGFBP1
Core matrisome	ECM Glycoproteins		IGFBP2
Core matrisome	ECM Glycoproteins		IGFBP3
Core matrisome	ECM Glycoproteins		IGFBP4
Core matrisome	ECM Glycoproteins		IGFBP5
Core matrisome	ECM Glycoproteins		IGFBP6
Core matrisome	ECM Glycoproteins		IGFBP7
Core matrisome	ECM Glycoproteins		IGFBPL1
Core matrisome	ECM Glycoproteins		IGSF10
Core matrisome	ECM Glycoproteins		KAL1
Core matrisome	ECM Glycoproteins		KCP
Core matrisome	ECM Glycoproteins		LAMA1
Core matrisome	ECM Glycoproteins		LAMA2
Core matrisome	ECM Glycoproteins		LAMA3
Core matrisome	ECM Glycoproteins		LAMA4
Core matrisome	ECM Glycoproteins		LAMA5
Core matrisome	ECM Glycoproteins		LAMB1
Core matrisome	ECM Glycoproteins		LAMB2
Core matrisome	ECM Glycoproteins		LAMB3
Core matrisome	ECM Glycoproteins		LAMB4
Core matrisome	ECM Glycoproteins		LAMC1
Core matrisome	ECM Glycoproteins		LAMC2
Core matrisome	ECM Glycoproteins		LAMC3
Core matrisome	ECM Glycoproteins		LGI1
Core matrisome	ECM Glycoproteins		LGI2
Core matrisome	ECM Glycoproteins		LGI3
Core matrisome	ECM Glycoproteins		LGI4
Core matrisome	ECM Glycoproteins		LRG1
Core matrisome	ECM Glycoproteins		LTBP1
Core matrisome	ECM Glycoproteins		LTBP2
Core matrisome	ECM Glycoproteins		LTBP3
Core matrisome	ECM Glycoproteins		LTBP4
Core matrisome	ECM Glycoproteins		MATN1
Core matrisome	ECM Glycoproteins		MATN2
Core matrisome	ECM Glycoproteins		MATN3
Core matrisome	ECM Glycoproteins		MATN4
Core matrisome	ECM Glycoproteins		MEPE
Core matrisome	ECM Glycoproteins		MFAP1
Core matrisome	ECM Glycoproteins		MFAP2
Core matrisome	ECM Glycoproteins		MFAP3
Core matrisome	ECM Glycoproteins		MFAP4
Core matrisome	ECM Glycoproteins		MFAP5
Core matrisome	ECM Glycoproteins		MFGE8
Core matrisome	ECM Glycoproteins		MGP
Core matrisome	ECM Glycoproteins		MMRN1
Core matrisome	ECM Glycoproteins		MMRN2
Core matrisome	ECM Glycoproteins		MXRA5
Core matrisome	ECM Glycoproteins		NDNF
Core matrisome	ECM Glycoproteins		NELL1
Core matrisome	ECM Glycoproteins		NELL2
Core matrisome	ECM Glycoproteins		NID1
Core matrisome	ECM Glycoproteins		NID2
Core matrisome	ECM Glycoproteins		NOV
Core matrisome	ECM Glycoproteins		NPNT
Core matrisome	ECM Glycoproteins		NTN1
Core matrisome	ECM Glycoproteins		NTN3
Core matrisome	ECM Glycoproteins		NTN4
Core matrisome	ECM Glycoproteins		NTN5
Core matrisome	ECM Glycoproteins		NTNG1
Core matrisome	ECM Glycoproteins		NTNG2
Core matrisome	ECM Glycoproteins		OIT3
Core matrisome	ECM Glycoproteins		OTOG
Core matrisome	ECM Glycoproteins		OTOL1
Core matrisome	ECM Glycoproteins		PAPLN
Core matrisome	ECM Glycoproteins		PCOLCE
Core matrisome	ECM Glycoproteins		PCOLCE2
Core matrisome	ECM Glycoproteins		POMZP3
Core matrisome	ECM Glycoproteins		POSTN
Core matrisome	ECM Glycoproteins		PXDN
Core matrisome	ECM Glycoproteins		PXDNL
Core matrisome	ECM Glycoproteins		RELN
Core matrisome	ECM Glycoproteins		RSPO1
Core matrisome	ECM Glycoproteins		RSPO2
Core matrisome	ECM Glycoproteins		RSPO3
Core matrisome	ECM Glycoproteins		RSPO4
Core matrisome	ECM Glycoproteins		SBSPON
Core matrisome	ECM Glycoproteins		SLIT1
Core matrisome	ECM Glycoproteins		SLIT2
Core matrisome	ECM Glycoproteins		SLIT3
Core matrisome	ECM Glycoproteins		SMOC1
Core matrisome	ECM Glycoproteins		SMOC2
Core matrisome	ECM Glycoproteins		SNED1
Core matrisome	ECM Glycoproteins		SPARC
Core matrisome	ECM Glycoproteins		SPARCL1
Core matrisome	ECM Glycoproteins		SPON1
Core matrisome	ECM Glycoproteins		SPON2
Core matrisome	ECM Glycoproteins		SPP1
Core matrisome	ECM Glycoproteins		SRPX
Core matrisome	ECM Glycoproteins		SRPX2
Core matrisome	ECM Glycoproteins		SSPO
Core matrisome	ECM Glycoproteins		SVEP1
Core matrisome	ECM Glycoproteins		TECTA
Core matrisome	ECM Glycoproteins		TECTB
Core matrisome	ECM Glycoproteins		TGFBI
Core matrisome	ECM Glycoproteins		THBS1
Core matrisome	ECM Glycoproteins		THBS2
Core matrisome	ECM Glycoproteins		THBS3
Core matrisome	ECM Glycoproteins		THBS4
Core matrisome	ECM Glycoproteins		THSD4
Core matrisome	ECM Glycoproteins		TINAG
Core matrisome	ECM Glycoproteins		TINAGL1
Core matrisome	ECM Glycoproteins		TNC
Core matrisome	ECM Glycoproteins		TNFAIP6
Core matrisome	ECM Glycoproteins		TNN
Core matrisome	ECM Glycoproteins		TNR
Core matrisome	ECM Glycoproteins		TNXB
Core matrisome	ECM Glycoproteins		TSKU
Core matrisome	ECM Glycoproteins		TSPEAR
Core matrisome	ECM Glycoproteins		VIT
Core matrisome	ECM Glycoproteins		VTN
Core matrisome	ECM Glycoproteins		VWA1
Core matrisome	ECM Glycoproteins		VWA2
Core matrisome	ECM Glycoproteins		VWA3A
Core matrisome	ECM Glycoproteins		VWA3B
Core matrisome	ECM Glycoproteins		VWA5A
Core matrisome	ECM Glycoproteins		VWA5B1
Core matrisome	ECM Glycoproteins		VWA5B2
Core matrisome	ECM Glycoproteins		VWA7
Core matrisome	ECM Glycoproteins		VWA9
Core matrisome	ECM Glycoproteins		VWCE
Core matrisome	ECM Glycoproteins		VWDE
Core matrisome	ECM Glycoproteins		VWF
Core matrisome	ECM Glycoproteins		WISP1
Core matrisome	ECM Glycoproteins		WISP2
Core matrisome	ECM Glycoproteins		WISP3
Core matrisome	ECM Glycoproteins		ZP1
Core matrisome	ECM Glycoproteins		ZP2
Core matrisome	ECM Glycoproteins		ZP3
Core matrisome	ECM Glycoproteins		ZP4
Core matrisome	ECM Glycoproteins		ZPLD1

Core matrisome	Collagens		COL10A1
Core matrisome	Collagens		COL11A1
Core matrisome	Collagens		COL11A2
Core matrisome	Collagens		COL12A1
Core matrisome	Collagens		COL13A1
Core matrisome	Collagens		COL14A1
Core matrisome	Collagens		COL15A1
Core matrisome	Collagens		COL16A1
Core matrisome	Collagens		COL17A1
Core matrisome	Collagens		COL18A1
Core matrisome	Collagens		COL19A1
Core matrisome	Collagens		COL1A1
Core matrisome	Collagens		COL1A2
Core matrisome	Collagens		COL20A1
Core matrisome	Collagens		COL21A1
Core matrisome	Collagens		COL22A1
Core matrisome	Collagens		COL23A1
Core matrisome	Collagens		COL24A1
Core matrisome	Collagens		COL25A1
Core matrisome	Collagens		COL26A1
Core matrisome	Collagens		COL27A1
Core matrisome	Collagens		COL28A1
Core matrisome	Collagens		COL2A1
Core matrisome	Collagens		COL3A1
Core matrisome	Collagens		COL4A1
Core matrisome	Collagens		COL4A2
Core matrisome	Collagens		COL4A3
Core matrisome	Collagens		COL4A4
Core matrisome	Collagens		COL4A5
Core matrisome	Collagens		COL4A6
Core matrisome	Collagens		COL5A1
Core matrisome	Collagens		COL5A2
Core matrisome	Collagens		COL5A3
Core matrisome	Collagens		COL6A1
Core matrisome	Collagens		COL6A2
Core matrisome	Collagens		COL6A3
Core matrisome	Collagens		COL6A5
Core matrisome	Collagens		COL6A6
Core matrisome	Collagens		COL7A1
Core matrisome	Collagens		COL8A1
Core matrisome	Collagens		COL8A2
Core matrisome	Collagens		COL9A1
Core matrisome	Collagens		COL9A2
Core matrisome	Collagens		COL9A3
Core matrisome	Proteoglycans		ACAN
Core matrisome	Proteoglycans		ASPN
Core matrisome	Proteoglycans		BCAN
Core matrisome	Proteoglycans		BGN
Core matrisome	Proteoglycans		CHAD
Core matrisome	Proteoglycans		CHADL
Core matrisome	Proteoglycans		DCN
Core matrisome	Proteoglycans		EPYC
Core matrisome	Proteoglycans		ESM1
Core matrisome	Proteoglycans		FMOD
Core matrisome	Proteoglycans		HAPLN1
Core matrisome	Proteoglycans		HAPLN2
Core matrisome	Proteoglycans		HAPLN3
Core matrisome	Proteoglycans		HAPLN4
Core matrisome	Proteoglycans		HSPG2
Core matrisome	Proteoglycans		IMPG1
Core matrisome	Proteoglycans		IMPG2
Core matrisome	Proteoglycans		KERA
Core matrisome	Proteoglycans		LUM
Core matrisome	Proteoglycans		NCAN
Core matrisome	Proteoglycans		NYX
Core matrisome	Proteoglycans		OGN
Core matrisome	Proteoglycans		OMD
Core matrisome	Proteoglycans		OPTC
Core matrisome	Proteoglycans		PODN
Core matrisome	Proteoglycans		PODNL1
Core matrisome	Proteoglycans		PRELP
Core matrisome	Proteoglycans		PRG2
Core matrisome	Proteoglycans		PRG3
Core matrisome	Proteoglycans		PRG4
Core matrisome	Proteoglycans		SPOCK1
Core matrisome	Proteoglycans		SPOCK2
Core matrisome	Proteoglycans		SPOCK3
Core matrisome	Proteoglycans		SRGN
Core matrisome	Proteoglycans		VCAN

Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA1
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA10
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA11
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA13
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA2
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA3
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA4
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA5
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA6
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA7
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA8
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA8L1
Matrisome-associated	ECM-affiliated Proteins	annexin	ANXA9
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QA
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QB
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QC
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QL1
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QL2
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QL3
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QL4
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF1
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF2
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF3
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF4
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF5
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF6
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF7
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF8
Matrisome-associated	ECM-affiliated Proteins	complement-associated	C1QTNF9
Matrisome-associated	ECM-affiliated Proteins	other	CD209
Matrisome-associated	ECM-affiliated Proteins	other	CLC
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC10A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC11A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC12A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC12B
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC14A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC17A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC18A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC18B
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC18C
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC19A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC1A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC1B
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC2A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC2B
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC2D
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC2L
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC3A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC3B
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4C
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4D
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4E
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4F
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4G
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC4M
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC5A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC6A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC7A
Matrisome-associated	ECM-affiliated Proteins	lectin	CLEC9A
Matrisome-associated	ECM-affiliated Proteins	lectin	COLEC10
Matrisome-associated	ECM-affiliated Proteins	lectin	COLEC11
Matrisome-associated	ECM-affiliated Proteins	lectin	COLEC12
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	CSPG4
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	CSPG5
Matrisome-associated	ECM-affiliated Proteins	other	ELFN1
Matrisome-associated	ECM-affiliated Proteins	other	ELFN2
Matrisome-associated	ECM-affiliated Proteins	mucin	EMCN
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	FCN1
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	FCN2
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	FCN3
Matrisome-associated	ECM-affiliated Proteins	other	FREM1
Matrisome-associated	ECM-affiliated Proteins	other	FREM2
Matrisome-associated	ECM-affiliated Proteins	other	FREM3
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	GPC1
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	GPC2
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	GPC3
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	GPC4
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	GPC5
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	GPC6
Matrisome-associated	ECM-affiliated Proteins	other	GREM1
Matrisome-associated	ECM-affiliated Proteins	lectin	GRIFIN
Matrisome-associated	ECM-affiliated Proteins	other	HPX
Matrisome-associated	ECM-affiliated Proteins	lectin	HSPC159
Matrisome-associated	ECM-affiliated Proteins	lectin	ITLN1
Matrisome-associated	ECM-affiliated Proteins	lectin	ITLN2
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS1
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS12
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS13
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS14
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS16
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS2
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS3
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS4
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS7
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS8
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS9
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS9B
Matrisome-associated	ECM-affiliated Proteins	lectin	LGALS9C
Matrisome-associated	ECM-affiliated Proteins	lectin	LMAN1
Matrisome-associated	ECM-affiliated Proteins	lectin	LMAN1L
Matrisome-associated	ECM-affiliated Proteins	lectin	MBL2
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC1
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC12
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC13
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC15
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC16
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC17
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC19
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC2
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC20
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC21
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC22
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC3A
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC4
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC5AC
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC5B
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC6
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC7
Matrisome-associated	ECM-affiliated Proteins	mucin	MUC8
Matrisome-associated	ECM-affiliated Proteins	mucin	MUCL1
Matrisome-associated	ECM-affiliated Proteins	mucin	OVGP1
Matrisome-associated	ECM-affiliated Proteins	mucin	PARM1
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXDC1
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXDC2
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNA1
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNA2
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNA3
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNA4
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNB1
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNB2
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNB3
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXNC1
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	PLXND1
Matrisome-associated	ECM-affiliated Proteins	other	PROL1
Matrisome-associated	ECM-affiliated Proteins	other	REG1A
Matrisome-associated	ECM-affiliated Proteins	other	REG1B
Matrisome-associated	ECM-affiliated Proteins	other	REG3A
Matrisome-associated	ECM-affiliated Proteins	other	REG3G
Matrisome-associated	ECM-affiliated Proteins	other	REG4
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	SDC1
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	SDC2
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	SDC3
Matrisome-associated	ECM-affiliated Proteins	proteoglycan	SDC4
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3A
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3B
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3C
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3D
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3E
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3F
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA3G
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA4A
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA4B
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA4C
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA4D
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA4F
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA4G
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA5A
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA5B
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA6A
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA6B
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA6C
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA6D
Matrisome-associated	ECM-affiliated Proteins	semaphorin signalling	SEMA7A
Matrisome-associated	ECM-affiliated Proteins	other	SFTA2
Matrisome-associated	ECM-affiliated Proteins	other	SFTA3
Matrisome-associated	ECM-affiliated Proteins	other	SFTPA1
Matrisome-associated	ECM-affiliated Proteins	other	SFTPA2
Matrisome-associated	ECM-affiliated Proteins	other	SFTPB
Matrisome-associated	ECM-affiliated Proteins	other	SFTPC
Matrisome-associated	ECM-affiliated Proteins	other	SFTPD

Matrisome-associated	ECM Regulators	peptidase inhibitor	A2M
Matrisome-associated	ECM Regulators	peptidase inhibitor	A2ML1
Matrisome-associated	ECM Regulators	peptidase	ADAM10
Matrisome-associated	ECM Regulators	peptidase	ADAM11
Matrisome-associated	ECM Regulators	peptidase	ADAM12
Matrisome-associated	ECM Regulators	peptidase	ADAM15
Matrisome-associated	ECM Regulators	peptidase	ADAM17
Matrisome-associated	ECM Regulators	peptidase	ADAM18
Matrisome-associated	ECM Regulators	peptidase	ADAM19
Matrisome-associated	ECM Regulators	peptidase	ADAM2
Matrisome-associated	ECM Regulators	peptidase	ADAM20
Matrisome-associated	ECM Regulators	peptidase	ADAM21
Matrisome-associated	ECM Regulators	peptidase	ADAM22
Matrisome-associated	ECM Regulators	peptidase	ADAM23
Matrisome-associated	ECM Regulators	peptidase	ADAM28
Matrisome-associated	ECM Regulators	peptidase	ADAM29
Matrisome-associated	ECM Regulators	peptidase	ADAM30
Matrisome-associated	ECM Regulators	peptidase	ADAM32
Matrisome-associated	ECM Regulators	peptidase	ADAM33
Matrisome-associated	ECM Regulators	peptidase	ADAM7
Matrisome-associated	ECM Regulators	peptidase	ADAM8
Matrisome-associated	ECM Regulators	peptidase	ADAM9
Matrisome-associated	ECM Regulators	peptidase	ADAMDEC1
Matrisome-associated	ECM Regulators	peptidase	ADAMTS1
Matrisome-associated	ECM Regulators	peptidase	ADAMTS10
Matrisome-associated	ECM Regulators	peptidase	ADAMTS12
Matrisome-associated	ECM Regulators	peptidase	ADAMTS13
Matrisome-associated	ECM Regulators	peptidase	ADAMTS14
Matrisome-associated	ECM Regulators	peptidase	ADAMTS15
Matrisome-associated	ECM Regulators	peptidase	ADAMTS16
Matrisome-associated	ECM Regulators	peptidase	ADAMTS17
Matrisome-associated	ECM Regulators	peptidase	ADAMTS18
Matrisome-associated	ECM Regulators	peptidase	ADAMTS19
Matrisome-associated	ECM Regulators	peptidase	ADAMTS2
Matrisome-associated	ECM Regulators	peptidase	ADAMTS20
Matrisome-associated	ECM Regulators	peptidase	ADAMTS3
Matrisome-associated	ECM Regulators	peptidase	ADAMTS4
Matrisome-associated	ECM Regulators	peptidase	ADAMTS5
Matrisome-associated	ECM Regulators	peptidase	ADAMTS6
Matrisome-associated	ECM Regulators	peptidase	ADAMTS7
Matrisome-associated	ECM Regulators	peptidase	ADAMTS8
Matrisome-associated	ECM Regulators	peptidase	ADAMTS9
Matrisome-associated	ECM Regulators	peptidase	ADAMTSL1
Matrisome-associated	ECM Regulators	peptidase	ADAMTSL2
Matrisome-associated	ECM Regulators	peptidase	ADAMTSL3
Matrisome-associated	ECM Regulators	peptidase	ADAMTSL4
Matrisome-associated	ECM Regulators	peptidase	ADAMTSL5
Matrisome-associated	ECM Regulators	peptidase inhibitor	AGT
Matrisome-associated	ECM Regulators	peptidase inhibitor	AMBP
Matrisome-associated	ECM Regulators	peptidase	ASTL
Matrisome-associated	ECM Regulators	peptidase	BMP1
Matrisome-associated	ECM Regulators	other	C17orf58
Matrisome-associated	ECM Regulators	other	CD109
Matrisome-associated	ECM Regulators	peptidase	CELA1
Matrisome-associated	ECM Regulators	peptidase	CELA2A
Matrisome-associated	ECM Regulators	peptidase	CELA2B
Matrisome-associated	ECM Regulators	peptidase	CELA3A
Matrisome-associated	ECM Regulators	peptidase	CELA3B
Matrisome-associated	ECM Regulators	peptidase inhibitor	CPAMD8
Matrisome-associated	ECM Regulators	peptidase	CPN2
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST1
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST11
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST2
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST3
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST4
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST5
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST6
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST7
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST8
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST9
Matrisome-associated	ECM Regulators	peptidase inhibitor	CST9L
Matrisome-associated	ECM Regulators	peptidase inhibitor	CSTA
Matrisome-associated	ECM Regulators	peptidase inhibitor	CSTB
Matrisome-associated	ECM Regulators	peptidase inhibitor	CSTL1
Matrisome-associated	ECM Regulators	peptidase	CTSA
Matrisome-associated	ECM Regulators	peptidase	CTSB
Matrisome-associated	ECM Regulators	peptidase	CTSC
Matrisome-associated	ECM Regulators	peptidase	CTSD
Matrisome-associated	ECM Regulators	peptidase	CTSE
Matrisome-associated	ECM Regulators	peptidase	CTSF
Matrisome-associated	ECM Regulators	peptidase	CTSG
Matrisome-associated	ECM Regulators	peptidase	CTSH
Matrisome-associated	ECM Regulators	peptidase	CTSK
Matrisome-associated	ECM Regulators	peptidase	CTSL
Matrisome-associated	ECM Regulators	peptidase	CTSO
Matrisome-associated	ECM Regulators	peptidase	CTSS
Matrisome-associated	ECM Regulators	peptidase	CTSV
Matrisome-associated	ECM Regulators	peptidase	CTSW
Matrisome-associated	ECM Regulators	peptidase	CTSZ
Matrisome-associated	ECM Regulators	other	EGLN1
Matrisome-associated	ECM Regulators	other	EGLN2
Matrisome-associated	ECM Regulators	other	EGLN3
Matrisome-associated	ECM Regulators	peptidase	ELANE
Matrisome-associated	ECM Regulators	peptidase	F10
Matrisome-associated	ECM Regulators	peptidase	F12
Matrisome-associated	ECM Regulators	peptidase	F13A1
Matrisome-associated	ECM Regulators	peptidase	F13B
Matrisome-associated	ECM Regulators	peptidase	F2
Matrisome-associated	ECM Regulators	peptidase	F7
Matrisome-associated	ECM Regulators	peptidase	F9
Matrisome-associated	ECM Regulators	other	FAM20A
Matrisome-associated	ECM Regulators	other	FAM20B
Matrisome-associated	ECM Regulators	other	FAM20C
Matrisome-associated	ECM Regulators	crosslinker	HABP2
Matrisome-associated	ECM Regulators	peptidase inhibitor	HMSD
Matrisome-associated	ECM Regulators	peptidase	HPSE
Matrisome-associated	ECM Regulators	peptidase	HPSE2
Matrisome-associated	ECM Regulators	peptidase	HRG
Matrisome-associated	ECM Regulators	peptidase	HTRA1
Matrisome-associated	ECM Regulators	peptidase	HTRA3
Matrisome-associated	ECM Regulators	peptidase	HTRA4
Matrisome-associated	ECM Regulators	peptidase	HYAL1
Matrisome-associated	ECM Regulators	peptidase	HYAL2
Matrisome-associated	ECM Regulators	peptidase	HYAL3
Matrisome-associated	ECM Regulators	peptidase	HYAL4
Matrisome-associated	ECM Regulators	peptidase inhibitor	ITIH1
Matrisome-associated	ECM Regulators	peptidase inhibitor	ITIH2
Matrisome-associated	ECM Regulators	peptidase inhibitor	ITIH3
Matrisome-associated	ECM Regulators	peptidase inhibitor	ITIH4
Matrisome-associated	ECM Regulators	peptidase inhibitor	ITIH5
Matrisome-associated	ECM Regulators	peptidase inhibitor	ITIH6
Matrisome-associated	ECM Regulators	peptidase inhibitor	KAZALD1
Matrisome-associated	ECM Regulators	peptidase inhibitor	KNG1
Matrisome-associated	ECM Regulators	peptidase	KY
Matrisome-associated	ECM Regulators	crosslinker	LEPRE1
Matrisome-associated	ECM Regulators	crosslinker	LEPREL1
Matrisome-associated	ECM Regulators	crosslinker	LEPREL2
Matrisome-associated	ECM Regulators	crosslinker	LOX
Matrisome-associated	ECM Regulators	crosslinker	LOXL1
Matrisome-associated	ECM Regulators	crosslinker	LOXL2
Matrisome-associated	ECM Regulators	crosslinker	LOXL3
Matrisome-associated	ECM Regulators	crosslinker	LOXL4
Matrisome-associated	ECM Regulators	other	LPA
Matrisome-associated	ECM Regulators	peptidase	MASP1
Matrisome-associated	ECM Regulators	peptidase	MASP2
Matrisome-associated	ECM Regulators	peptidase	MEP1A
Matrisome-associated	ECM Regulators	peptidase	MEP1B
Matrisome-associated	ECM Regulators	peptidase	MMP1
Matrisome-associated	ECM Regulators	peptidase	MMP10
Matrisome-associated	ECM Regulators	peptidase	MMP11
Matrisome-associated	ECM Regulators	peptidase	MMP12
Matrisome-associated	ECM Regulators	peptidase	MMP13
Matrisome-associated	ECM Regulators	peptidase	MMP14
Matrisome-associated	ECM Regulators	peptidase	MMP15
Matrisome-associated	ECM Regulators	peptidase	MMP16
Matrisome-associated	ECM Regulators	peptidase	MMP17
Matrisome-associated	ECM Regulators	peptidase	MMP19
Matrisome-associated	ECM Regulators	peptidase	MMP2
Matrisome-associated	ECM Regulators	peptidase	MMP20
Matrisome-associated	ECM Regulators	peptidase	MMP21
Matrisome-associated	ECM Regulators	peptidase	MMP23B
Matrisome-associated	ECM Regulators	peptidase	MMP24
Matrisome-associated	ECM Regulators	peptidase	MMP25
Matrisome-associated	ECM Regulators	peptidase	MMP26
Matrisome-associated	ECM Regulators	peptidase	MMP27
Matrisome-associated	ECM Regulators	peptidase	MMP28
Matrisome-associated	ECM Regulators	peptidase	MMP3
Matrisome-associated	ECM Regulators	peptidase	MMP7
Matrisome-associated	ECM Regulators	peptidase	MMP8
Matrisome-associated	ECM Regulators	peptidase	MMP9
Matrisome-associated	ECM Regulators	peptidase	NGLY1
Matrisome-associated	ECM Regulators	other	OGFOD1
Matrisome-associated	ECM Regulators	other	OGFOD2
Matrisome-associated	ECM Regulators	other	P4HA1
Matrisome-associated	ECM Regulators	other	P4HA2
Matrisome-associated	ECM Regulators	other	P4HA3
Matrisome-associated	ECM Regulators	other	P4HTM
Matrisome-associated	ECM Regulators	peptidase	PAMR1
Matrisome-associated	ECM Regulators	peptidase	PAPPA
Matrisome-associated	ECM Regulators	peptidase	PAPPA2
Matrisome-associated	ECM Regulators	peptidase	PCSK5
Matrisome-associated	ECM Regulators	peptidase	PCSK6
Matrisome-associated	ECM Regulators	peptidase inhibitor	PI3
Matrisome-associated	ECM Regulators	peptidase	PLAT
Matrisome-associated	ECM Regulators	peptidase	PLAU
Matrisome-associated	ECM Regulators	peptidase	PLG
Matrisome-associated	ECM Regulators	crosslinker	PLOD1
Matrisome-associated	ECM Regulators	crosslinker	PLOD2
Matrisome-associated	ECM Regulators	crosslinker	PLOD3
Matrisome-associated	ECM Regulators	peptidase	PRSS1
Matrisome-associated	ECM Regulators	peptidase	PRSS12
Matrisome-associated	ECM Regulators	peptidase	PRSS2
Matrisome-associated	ECM Regulators	peptidase	PRSS3
Matrisome-associated	ECM Regulators	peptidase inhibitor	PZP
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA10
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA11
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA12
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA2
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA3
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA4
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA5
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA6
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA7
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINA9
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB10
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB11
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB12
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB13
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB2
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB3
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB4
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB5
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB6
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB7
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB8
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINB9
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINC1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPIND1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINE1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINE2
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINE3
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINF1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINF2
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPING1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINH1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINI1
Matrisome-associated	ECM Regulators	peptidase inhibitor	SERPINI2
Matrisome-associated	ECM Regulators	peptidase inhibitor	SLPI
Matrisome-associated	ECM Regulators	peptidase	SPAM1
Matrisome-associated	ECM Regulators	peptidase	ST14
Matrisome-associated	ECM Regulators	other	SULF1
Matrisome-associated	ECM Regulators	other	SULF2
Matrisome-associated	ECM Regulators	crosslinker	TGM1
Matrisome-associated	ECM Regulators	crosslinker	TGM2
Matrisome-associated	ECM Regulators	crosslinker	TGM3
Matrisome-associated	ECM Regulators	crosslinker	TGM4
Matrisome-associated	ECM Regulators	crosslinker	TGM5
Matrisome-associated	ECM Regulators	crosslinker	TGM6
Matrisome-associated	ECM Regulators	crosslinker	TGM7
Matrisome-associated	ECM Regulators	peptidase inhibitor	TIMP1
Matrisome-associated	ECM Regulators	peptidase inhibitor	TIMP2
Matrisome-associated	ECM Regulators	peptidase inhibitor	TIMP3
Matrisome-associated	ECM Regulators	peptidase inhibitor	TIMP4
Matrisome-associated	ECM Regulators	peptidase	TLL1
Matrisome-associated	ECM Regulators	peptidase	TLL2
Matrisome-associated	ECM Regulators	peptidase	TMPRSS15

Matrisome-associated	Secreted Factors	hormone	AMH
Matrisome-associated	Secreted Factors	growth factor	ANGPT1
Matrisome-associated	Secreted Factors	growth factor	ANGPT2
Matrisome-associated	Secreted Factors	growth factor	ANGPT4
Matrisome-associated	Secreted Factors	growth factor	ANGPTL1
Matrisome-associated	Secreted Factors	growth factor	ANGPTL2
Matrisome-associated	Secreted Factors	growth factor	ANGPTL3
Matrisome-associated	Secreted Factors	growth factor	ANGPTL4
Matrisome-associated	Secreted Factors	growth factor	ANGPTL5
Matrisome-associated	Secreted Factors	growth factor	ANGPTL6
Matrisome-associated	Secreted Factors	growth factor	ANGPTL7
Matrisome-associated	Secreted Factors	growth factor	AREG
Matrisome-associated	Secreted Factors	growth factor	ARTN
Matrisome-associated	Secreted Factors	growth factor	BDNF
Matrisome-associated	Secreted Factors	growth factor	BMP10
Matrisome-associated	Secreted Factors	growth factor	BMP15
Matrisome-associated	Secreted Factors	growth factor	BMP2
Matrisome-associated	Secreted Factors	growth factor	BMP3
Matrisome-associated	Secreted Factors	growth factor	BMP4
Matrisome-associated	Secreted Factors	growth factor	BMP5
Matrisome-associated	Secreted Factors	growth factor	BMP6
Matrisome-associated	Secreted Factors	growth factor	BMP7
Matrisome-associated	Secreted Factors	growth factor	BMP8A
Matrisome-associated	Secreted Factors	growth factor	BMP8B
Matrisome-associated	Secreted Factors	other	BRINP2
Matrisome-associated	Secreted Factors	other	BRINP3
Matrisome-associated	Secreted Factors	other	BTC
Matrisome-associated	Secreted Factors	other	CBLN1
Matrisome-associated	Secreted Factors	other	CBLN2
Matrisome-associated	Secreted Factors	other	CBLN3
Matrisome-associated	Secreted Factors	other	CBLN4
Matrisome-associated	Secreted Factors	growth factor	CCBE1
Matrisome-associated	Secreted Factors	growth factor	CFC1
Matrisome-associated	Secreted Factors	growth factor	CFC1B
Matrisome-associated	Secreted Factors	growth factor inhibitor	CHRD
Matrisome-associated	Secreted Factors	growth factor inhibitor	CHRDL1
Matrisome-associated	Secreted Factors	growth factor inhibitor	CHRDL2
Matrisome-associated	Secreted Factors	growth factor	CNTF
Matrisome-associated	Secreted Factors	hormone	CRHBP
Matrisome-associated	Secreted Factors	other	CRLF1
Matrisome-associated	Secreted Factors	other	CRLF3
Matrisome-associated	Secreted Factors	calcium-binding	CRNN
Matrisome-associated	Secreted Factors	growth factor	CSF1
Matrisome-associated	Secreted Factors	growth factor	CSF2
Matrisome-associated	Secreted Factors	growth factor	CSF3
Matrisome-associated	Secreted Factors	hormone	CSH1
Matrisome-associated	Secreted Factors	hormone	CSH2
Matrisome-associated	Secreted Factors	hormone	CSHL1
Matrisome-associated	Secreted Factors	other	DHH
Matrisome-associated	Secreted Factors	growth factor	EGF
Matrisome-associated	Secreted Factors	growth factor	EGFL6
Matrisome-associated	Secreted Factors	growth factor	EGFL7
Matrisome-associated	Secreted Factors	growth factor	EGFL8
Matrisome-associated	Secreted Factors	growth factor	EPGN
Matrisome-associated	Secreted Factors	growth factor	EPO
Matrisome-associated	Secreted Factors	growth factor	EREG
Matrisome-associated	Secreted Factors	growth factor	FGF1
Matrisome-associated	Secreted Factors	growth factor	FGF10
Matrisome-associated	Secreted Factors	growth factor	FGF11
Matrisome-associated	Secreted Factors	growth factor	FGF12
Matrisome-associated	Secreted Factors	growth factor	FGF13
Matrisome-associated	Secreted Factors	growth factor	FGF14
Matrisome-associated	Secreted Factors	growth factor	FGF16
Matrisome-associated	Secreted Factors	growth factor	FGF17
Matrisome-associated	Secreted Factors	growth factor	FGF18
Matrisome-associated	Secreted Factors	growth factor	FGF19
Matrisome-associated	Secreted Factors	growth factor	FGF2
Matrisome-associated	Secreted Factors	growth factor	FGF20
Matrisome-associated	Secreted Factors	growth factor	FGF21
Matrisome-associated	Secreted Factors	growth factor	FGF22
Matrisome-associated	Secreted Factors	growth factor	FGF23
Matrisome-associated	Secreted Factors	growth factor	FGF3
Matrisome-associated	Secreted Factors	growth factor	FGF4
Matrisome-associated	Secreted Factors	growth factor	FGF5
Matrisome-associated	Secreted Factors	growth factor	FGF6
Matrisome-associated	Secreted Factors	growth factor	FGF7
Matrisome-associated	Secreted Factors	growth factor	FGF8
Matrisome-associated	Secreted Factors	growth factor	FGF9
Matrisome-associated	Secreted Factors	growth factor	FGFBP1
Matrisome-associated	Secreted Factors	growth factor	FGFBP2
Matrisome-associated	Secreted Factors	growth factor	FGFBP3
Matrisome-associated	Secreted Factors	growth factor	FIGF
Matrisome-associated	Secreted Factors	calcium-binding	FLG
Matrisome-associated	Secreted Factors	calcium-binding	FLG2
Matrisome-associated	Secreted Factors	other	FRZB
Matrisome-associated	Secreted Factors	growth factor inhibitor	FST
Matrisome-associated	Secreted Factors	growth factor inhibitor	FSTL1
Matrisome-associated	Secreted Factors	growth factor inhibitor	FSTL3
Matrisome-associated	Secreted Factors	growth factor	GDF1
Matrisome-associated	Secreted Factors	growth factor	GDF10
Matrisome-associated	Secreted Factors	growth factor	GDF11
Matrisome-associated	Secreted Factors	growth factor	GDF15
Matrisome-associated	Secreted Factors	growth factor	GDF2
Matrisome-associated	Secreted Factors	growth factor	GDF3
Matrisome-associated	Secreted Factors	growth factor	GDF5
Matrisome-associated	Secreted Factors	growth factor	GDF6
Matrisome-associated	Secreted Factors	growth factor	GDF7
Matrisome-associated	Secreted Factors	growth factor	GDF9
Matrisome-associated	Secreted Factors	growth factor	GDNF
Matrisome-associated	Secreted Factors	hormone	GH1
Matrisome-associated	Secreted Factors	hormone	GH2
Matrisome-associated	Secreted Factors	growth factor	HBEGF
Matrisome-associated	Secreted Factors	other	HCFC1
Matrisome-associated	Secreted Factors	other	HCFC2
Matrisome-associated	Secreted Factors	growth factor	HGF
Matrisome-associated	Secreted Factors	growth factor	HGFAC
Matrisome-associated	Secreted Factors	other	HHIP
Matrisome-associated	Secreted Factors	calcium-binding	HRNR
Matrisome-associated	Secreted Factors	growth factor	IGF1
Matrisome-associated	Secreted Factors	growth factor	IGF2
Matrisome-associated	Secreted Factors	other	IHH
Matrisome-associated	Secreted Factors	growth factor inhibitor	INHA
Matrisome-associated	Secreted Factors	growth factor inhibitor	INHBA
Matrisome-associated	Secreted Factors	growth factor inhibitor	INHBB
Matrisome-associated	Secreted Factors	growth factor inhibitor	INHBC
Matrisome-associated	Secreted Factors	growth factor inhibitor	INHBE
Matrisome-associated	Secreted Factors	other	INS
Matrisome-associated	Secreted Factors	other	INS-IGF2
Matrisome-associated	Secreted Factors	other	INSL3
Matrisome-associated	Secreted Factors	other	INSL5
Matrisome-associated	Secreted Factors	other	INSL6
Matrisome-associated	Secreted Factors	growth factor inhibitor	ISM1
Matrisome-associated	Secreted Factors	growth factor inhibitor	ISM2
Matrisome-associated	Secreted Factors	growth factor	LEFTY1
Matrisome-associated	Secreted Factors	growth factor	LEFTY2
Matrisome-associated	Secreted Factors	hormone	LEP
Matrisome-associated	Secreted Factors	growth factor	MDK
Matrisome-associated	Secreted Factors	growth factor	MEGF10
Matrisome-associated	Secreted Factors	growth factor	MEGF11
Matrisome-associated	Secreted Factors	growth factor	MEGF6
Matrisome-associated	Secreted Factors	growth factor	MEGF8
Matrisome-associated	Secreted Factors	growth factor	MEGF9
Matrisome-associated	Secreted Factors	growth factor	MST1
Matrisome-associated	Secreted Factors	growth factor	MST1L
Matrisome-associated	Secreted Factors	growth factor	MSTN
Matrisome-associated	Secreted Factors	growth factor	NGF
Matrisome-associated	Secreted Factors	growth factor	NODAL
Matrisome-associated	Secreted Factors	other	NRG1
Matrisome-associated	Secreted Factors	other	NRG2
Matrisome-associated	Secreted Factors	other	NRG3
Matrisome-associated	Secreted Factors	other	NRG4
Matrisome-associated	Secreted Factors	other	NRTN
Matrisome-associated	Secreted Factors	other	NTF3
Matrisome-associated	Secreted Factors	other	NTF4
Matrisome-associated	Secreted Factors	growth factor	PDGFA
Matrisome-associated	Secreted Factors	growth factor	PDGFB
Matrisome-associated	Secreted Factors	growth factor	PDGFC
Matrisome-associated	Secreted Factors	growth factor	PDGFD
Matrisome-associated	Secreted Factors	growth factor	PF4
Matrisome-associated	Secreted Factors	growth factor	PF4V1
Matrisome-associated	Secreted Factors	growth factor	PGF
Matrisome-associated	Secreted Factors	other	PIK3IP1
Matrisome-associated	Secreted Factors	other	PRL
Matrisome-associated	Secreted Factors	growth factor	PSPN
Matrisome-associated	Secreted Factors	growth factor	PTN
Matrisome-associated	Secreted Factors	other	RPTN
Matrisome-associated	Secreted Factors	calcium-binding	S100A1
Matrisome-associated	Secreted Factors	calcium-binding	S100A10
Matrisome-associated	Secreted Factors	calcium-binding	S100A11
Matrisome-associated	Secreted Factors	calcium-binding	S100A12
Matrisome-associated	Secreted Factors	calcium-binding	S100A13
Matrisome-associated	Secreted Factors	calcium-binding	S100A14
Matrisome-associated	Secreted Factors	calcium-binding	S100A16
Matrisome-associated	Secreted Factors	calcium-binding	S100A2
Matrisome-associated	Secreted Factors	calcium-binding	S100A3
Matrisome-associated	Secreted Factors	calcium-binding	S100A4
Matrisome-associated	Secreted Factors	calcium-binding	S100A5
Matrisome-associated	Secreted Factors	calcium-binding	S100A6
Matrisome-associated	Secreted Factors	calcium-binding	S100A7
Matrisome-associated	Secreted Factors	calcium-binding	S100A7A
Matrisome-associated	Secreted Factors	calcium-binding	S100A7L2
Matrisome-associated	Secreted Factors	calcium-binding	S100A8
Matrisome-associated	Secreted Factors	calcium-binding	S100A9
Matrisome-associated	Secreted Factors	calcium-binding	S100B
Matrisome-associated	Secreted Factors	calcium-binding	S100G
Matrisome-associated	Secreted Factors	calcium-binding	S100P
Matrisome-associated	Secreted Factors	calcium-binding	S100Z
Matrisome-associated	Secreted Factors	other	SCUBE1
Matrisome-associated	Secreted Factors	other	SCUBE2
Matrisome-associated	Secreted Factors	other	SCUBE3
Matrisome-associated	Secreted Factors	wnt pathway	SFRP1
Matrisome-associated	Secreted Factors	wnt pathway	SFRP2
Matrisome-associated	Secreted Factors	wnt pathway	SFRP4
Matrisome-associated	Secreted Factors	wnt pathway	SFRP5
Matrisome-associated	Secreted Factors	growth factor	SHH
Matrisome-associated	Secreted Factors	calcium-binding	TCHH
Matrisome-associated	Secreted Factors	calcium-binding	TCHHL1
Matrisome-associated	Secreted Factors	growth factor	TDGF1
Matrisome-associated	Secreted Factors	growth factor	TGFA
Matrisome-associated	Secreted Factors	growth factor	TGFB1
Matrisome-associated	Secreted Factors	growth factor	TGFB2
Matrisome-associated	Secreted Factors	growth factor	TGFB3
Matrisome-associated	Secreted Factors	other	THPO
Matrisome-associated	Secreted Factors	other	TPO
Matrisome-associated	Secreted Factors	growth factor	VEGFA
Matrisome-associated	Secreted Factors	growth factor	VEGFB
Matrisome-associated	Secreted Factors	growth factor	VEGFC
Matrisome-associated	Secreted Factors	growth factor inhibitor	VWC2
Matrisome-associated	Secreted Factors	growth factor inhibitor	VWC2L
Matrisome-associated	Secreted Factors	other	WFIKKN1
Matrisome-associated	Secreted Factors	other	WFIKKN2
Matrisome-associated	Secreted Factors	other	WIF1
Matrisome-associated	Secreted Factors	wnt pathway	WNT1
Matrisome-associated	Secreted Factors	wnt pathway	WNT10A
Matrisome-associated	Secreted Factors	wnt pathway	WNT10B
Matrisome-associated	Secreted Factors	wnt pathway	WNT11
Matrisome-associated	Secreted Factors	wnt pathway	WNT16
Matrisome-associated	Secreted Factors	wnt pathway	WNT2
Matrisome-associated	Secreted Factors	wnt pathway	WNT2B
Matrisome-associated	Secreted Factors	wnt pathway	WNT3
Matrisome-associated	Secreted Factors	wnt pathway	WNT3A
Matrisome-associated	Secreted Factors	wnt pathway	WNT4
Matrisome-associated	Secreted Factors	wnt pathway	WNT5A
Matrisome-associated	Secreted Factors	wnt pathway	WNT5B
Matrisome-associated	Secreted Factors	wnt pathway	WNT6
Matrisome-associated	Secreted Factors	wnt pathway	WNT7A
Matrisome-associated	Secreted Factors	wnt pathway	WNT7B
Matrisome-associated	Secreted Factors	wnt pathway	WNT8A
Matrisome-associated	Secreted Factors	wnt pathway	WNT8B
Matrisome-associated	Secreted Factors	wnt pathway	WNT9A
Matrisome-associated	Secreted Factors	wnt pathway	WNT9B
Matrisome-associated	Secreted Factors	other	ZFP91
Retired	n/a		ADAM1A
Retired	n/a		ADAM21P1
Retired	n/a		ADAM3A
Retired	n/a		ADAM3B
Retired	n/a		ADAM5
Retired	n/a		ADAM6
Retired	n/a		ANXA2P2
Retired	n/a		ANXA8L2
Retired	n/a		BPIFA4P
Retired	n/a		C17orf101
Retired	n/a		COL6A4P1
Retired	n/a		COL6A4P2
Retired	n/a		CST9L2
Retired	n/a		CTF2P
Retired	n/a		CTSL3P
Retired	n/a		CTSLP3
Retired	n/a		CTSLP6
Retired	n/a		CTSLP7
Retired	n/a		DMBT1P1
Retired	n/a		EGFEM1P
Retired	n/a		FGF7P2
Retired	n/a		gene_A6NLB4
Retired	n/a		HYALP1
Retired	n/a		KGFLP1
Retired	n/a		KGFLP2
Retired	n/a		LOC400696
Retired	n/a		LOC728715
Retired	n/a		LPAL2
Retired	n/a		MBL1P
Retired	n/a		MST1P2
Retired	n/a		NCRNA00083
Retired	n/a		NEPNP
Retired	n/a		NTF6B
Retired	n/a		OVOS1
Retired	n/a		OVOS2
Retired	n/a		PPBPP1
Retired	n/a		PRSS3P1
Retired	n/a		SERPINA13P
Retired	n/a		TNFSF12-TNFSF13
Retired	n/a		TNXA
Retired	n/a		UNQ5830

Supplementary Table 2. Decision matrix for cell phenotyping as part of validation panel.
Pos = positive, neg = negative.

		MMP2	CHP	CD66b	COL3A1	MMP9	CD14	COL1A1	COL12A1	COL6A1	CD3	FN	CD20	MERTK	CD4	COL4A1	CD45	CD68	CD8	hIgG	CD11c	MZB1
all	Other Immune cells			anypos			anypos				anypos		anypos		anypos		anypos	anypos	anypos		anypos	anypos
all	Non-immune cells			allneg			allneg				allneg		allneg		allneg		allneg	allneg	allneg		allneg	allneg
Non-immune cells	Vascular cells															pos						
Other Immune cells	Lymphoid cells										anypos		anypos		anypos				anypos			anypos
Other Immune cells	Myeloid cells			anypos		anypos	anypos											anypos			anypos	
Lymphoid cells	T cells										anypos				anypos				anypos			
Lymphoid cells	B cells												pos									
Lymphoid cells	Plasma cells																					pos
T cells	CD8 T cells																		anypos			
T cells	CD4 T cells														anypos							
Myeloid cells	Neutrophils			anypos		anypos	neg															
Myeloid cells	Monocytes						anypos											neg				
Myeloid cells	Macrophages						neg											anypos			neg	
Myeloid cells	Dendritic cells						allneg											allneg			anypos	

Bibliography

1. Kuek, A., Hazleman, B. L. & Östör, A. J. K. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. *Postgraduate Medical Journal* **83**, 251–260 (2007).
2. Wu, D. *et al.* Global, regional, and national incidence of six major immune-mediated inflammatory diseases: findings from the global burden of disease study 2019. *eClinicalMedicine* **64**, (2023).
3. García, M. J. *et al.* Impact of immune-mediated diseases in inflammatory bowel disease and implications in therapeutic approach. *Sci Rep* **10**, 10731 (2020).
4. Raine, T. *et al.* ECCO Topical Review: Refractory Inflammatory Bowel Disease. *Journal of Crohn's and Colitis* **15**, 1605–1620 (2021).
5. Melville, A. R., Kearsley-Fleet, L., Buch, M. H. & Hyrich, K. L. Understanding Refractory Rheumatoid Arthritis: Implications for a Therapeutic Approach. *Drugs* **80**, 849–857 (2020).
6. Burau, V. *et al.* Post-COVID health policy responses to healthcare workforce capacities: A comparative analysis of health system resilience in six European countries. *Health Policy* **139**, 104962 (2024).
7. Voulgari, P. V. & Drosos, A. A. Adalimumab for rheumatoid arthritis. *Expert Opinion on Biological Therapy* **6**, 1349–1360 (2006).
8. Zhao, S., Mysler, E. & Moots, R. J. Etanercept for the Treatment of Rheumatoid Arthritis. *Immunotherapy* **10**, 433–445 (2018).
9. Ben-Horin, S. & Chowers, Y. Tailoring anti-TNF therapy in IBD: drug levels and disease activity. *Nat Rev Gastroenterol Hepatol* **11**, 243–255 (2014).
10. McLornan, D. P., Pope, J. E., Gotlib, J. & Harrison, C. N. Current and future status of JAK inhibitors. *The Lancet* **398**, 803–816 (2021).
11. Scribano, M. L. Vedolizumab for inflammatory bowel disease: From randomized controlled trials to real-life evidence. *World J Gastroenterol* **24**, 2457–2467 (2018).
12. Sullivan, L. H. *The Tall Office Building Artistically Considered*. (1922).
13. Hatton, I. A. *et al.* The human cell count and size distribution. *Proceedings of the National Academy of Sciences* **120**, e2303077120 (2023).

14. Sutherland, T. E., Dyer, D. P. & Allen, J. E. The extracellular matrix and the immune system: A mutually dependent relationship. *Science* **379**, eabp8964 (2023).
15. Wang, Y. *et al.* Impact of tumor-parenchyma biomechanics on liver metastatic progression: a multi-model approach. *Sci Rep* **11**, 1710 (2021).
16. Shao, X., Taha, I. N., Clauser, K. R., Gao, Y. (Tom) & Naba, A. MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Research* **48**, D1136–D1144 (2020).
17. Naba, A. *et al.* The extracellular matrix: Tools and insights for the 'omics' era. *Matrix Biol* **49**, 10–24 (2016).
18. Lin, X., Patil, S., Gao, Y.-G. & Qian, A. The Bone Extracellular Matrix in Bone Formation and Regeneration. *Front. Pharmacol.* **11**, (2020).
19. Saito, M. & Marumo, K. Effects of Collagen Crosslinking on Bone Material Properties in Health and Disease. *Calcif Tissue Int* **97**, 242–261 (2015).
20. Moorehead, C., Prudnikova, K. & Marcolongo, M. The regulatory effects of proteoglycans on collagen fibrillogenesis and morphology investigated using biomimetic proteoglycans. *Journal of Structural Biology* **206**, 204–215 (2019).
21. Coulson-Thomas, Y. M. *et al.* The Identification of Proteoglycans and Glycosaminoglycans in Archaeological Human Bones and Teeth. *PLOS ONE* **10**, e0131105 (2015).
22. Rosset, E. M. & Bradshaw, A. D. SPARC/osteonectin in mineralized tissue. *Matrix Biology* **52–54**, 78–87 (2016).
23. Vincent, T. L., McClurg, O. & Troeberg, L. The Extracellular Matrix of Articular Cartilage Controls the Bioavailability of Pericellular Matrix-Bound Growth Factors to Drive Tissue Homeostasis and Repair. *Int J Mol Sci* **23**, 6003 (2022).
24. Meizlish, M. L. *et al.* Mechanosensing regulates tissue repair program in macrophages. *Science Advances* **10**, eadk6906 (2024).
25. Guenat, O. T. & Berthiaume, F. Incorporating mechanical strain in organs-on-a-chip: Lung and skin. *Biomicrofluidics* **12**, 042207 (2018).
26. Atkinson, J.-A. M., McKenna, K. T., Barnett, A. G., McGrath, D. J. & Rudd, M. A Randomized, Controlled Trial to Determine the Efficacy of Paper Tape in Preventing Hypertrophic Scar Formation in Surgical Incisions that Traverse Langer's Skin Tension Lines. *Plastic and Reconstructive Surgery* **116**, 1648 (2005).

27. Waldeland, J. O. & Evje, S. A multiphase model for exploring tumor cell migration driven by autologous chemotaxis. *Chemical Engineering Science* **191**, 268–287 (2018).
28. Robinson, J. P., Ostafe, R., Iyengar, S. N., Rajwa, B. & Fischer, R. Flow Cytometry: The Next Revolution. *Cells* **12**, 1875 (2023).
29. Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* **18**, 35–45 (2018).
30. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**, eabq4964 (2023).
31. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
32. Rao, N., Clark, S. & Habern, O. Bridging Genomics and Tissue Pathology. *Genetic Engineering & Biotechnology News* **40**, 50–51 (2020).
33. Rosenbloom, A. *et al.* A complete pipeline for high-plex spatial proteomic profiling and analysis of neural cell phenotypes on the CosMx™ Spatial Molecular Imager and AtoMx™ Spatial Informatics Platform. *Alzheimer's & Dementia* **19**, e076665 (2023).
34. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
35. Oliveira, M. F. *et al.* Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling. 2024.06.04.597233 Preprint at https://doi.org/10.1101/2024.06.04.597233 (2024).
36. Prabahar, A. *et al.* Unraveling the complex relationship between mRNA and protein abundances: a machine learning-based approach for imputing protein levels from RNA-seq data. *NAR Genomics and Bioinformatics* **6**, lqae019 (2024).
37. Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol* **40**, 1231–1240 (2022).
38. Rochais, M. L., Hemon, P., Pers, J.-O. & Uguen, A. Application of High-Throughput Imaging Mass Cytometry Hyperion in Cancer Research. *Frontiers in Immunology* **13**, (2022).

39. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nature protocols* **16**, 3802 (2021).
40. Ma, F. *et al.* The cellular architecture of the antimicrobial response network in human leprosy granulomas. *Nat Immunol* **22**, 839–850 (2021).
41. Pachynski, R. K. *et al.* Single-cell Spatial Proteomic Revelations on the Multiparametric MRI Heterogeneity of Clinically Significant Prostate Cancer. *Clinical Cancer Research* **27**, 3478–3490 (2021).
42. Khan, I. M. *et al.* The Development of Synovial Joints. in *Current Topics in Developmental Biology* vol. 79 1–36 (Academic Press, 2007).
43. McKee, T. J., Perlman, G., Morris, M. & Komarova, S. V. Extracellular matrix composition of connective tissues: a systematic review and meta-analysis. *Sci Rep* **9**, 10542 (2019).
44. Tabbaa, S. M., Görtz, S. & Lattermann, C. 21 - A Basic Science Understanding of Cartilage Repair and Its Clinical Application. in *Evidence-Based Management of Complex Knee Injuries* (eds. LaPrade, R. F. & Chahla, J.) 241–252 (Elsevier, Philadelphia, 2022). doi:10.1016/B978-0-323-71310-8.00021-9.
45. Greene, G. W. *et al.* Lubricin: A versatile, biological anti-adhesive with properties comparable to polyethylene glycol. *Biomaterials* **53**, 127–136 (2015).
46. Lavelle, W. F., Lavelle, E. D. & Lavelle, L. A. Intra-articular injections. in *Current Therapy in Pain* (ed. Smith, H. S.) 591–596 (W.B. Saunders, Philadelphia, 2009). doi:10.1016/B978-1-4160-4836-7.00081-X.
47. Levick, J. R. Microvascular Architecture and Exchange in Synovial Joints. *Microcirculation* **2**, 217–233 (1995).
48. Smith, M. D. The Normal Synovium. *Open Rheumatol J* **5**, 100–106 (2011).
49. Konttinen, Y. T. *et al.* Synovial Fibroblasts. *Scandinavian Journal of Rheumatology* **17**, 95–103 (1988).
50. Gulati, M., Farah, Z. & Mouyis, M. Clinical features of rheumatoid arthritis. *Medicine* **46**, 211–215 (2018).
51. Aviña-Zubieta, J. A. *et al.* Risk of cardiovascular mortality in patients with rheumatoid arthritis: A meta-analysis of observational studies. *Arthritis Care & Research* **59**, 1690–1697 (2008).
52. Marrie, R. A. *et al.* Increased Burden of Psychiatric Disorders in Rheumatoid Arthritis. *Arthritis Care & Research* **70**, 970–978 (2018).

53. Matcham, F. *et al.* The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: A systematic review and meta-analysis. *Seminars in Arthritis and Rheumatism* **44**, 123–130 (2014).
54. Kay, J. & Upchurch, K. S. ACR/EULAR 2010 rheumatoid arthritis classification criteria. *Rheumatology* **51**, vi5–vi9 (2012).
55. Guthridge, J. M., Wagner, C. A. & James, J. A. The promise of precision medicine in rheumatology. *Nat Med* **28**, 1363–1371 (2022).
56. Myasoedova, E., Crowson, C. S., Kremers, H. M., Therneau, T. M. & Gabriel, S. E. Is the incidence of rheumatoid arthritis rising?: Results from Olmsted County, Minnesota, 1955–2007. *Arthritis & Rheumatism* **62**, 1576–1582 (2010).
57. Malemba, J. J. *et al.* The epidemiology of rheumatoid arthritis in Kinshasa, Democratic Republic of Congo—a population-based study. *Rheumatology* **51**, 1644–1647 (2012).
58. McInnes, I. B. & Schett, G. The Pathogenesis of Rheumatoid Arthritis. *New England Journal of Medicine* **365**, 2205–2219 (2011).
59. Nguyen, H. N. *et al.* Autocrine Loop Involving IL-6 Family Member LIF, LIF Receptor, and STAT4 Drives Sustained Fibroblast Production of Inflammatory Mediators. *Immunity* **46**, 220–232 (2017).
60. Hashizume, M., Hayakawa, N. & Mihara, M. IL-6 trans-signalling directly induces RANKL on fibroblast-like synovial cells and is involved in RANKL induction by TNF-α and IL-17. *Rheumatology* **47**, 1635–1640 (2008).
61. Bottini, N. & Firestein, G. S. Duality of fibroblast-like synoviocytes in RA: passive responders and imprinted aggressors. *Nat Rev Rheumatol* **9**, 24–33 (2013).
62. Croft, A. P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**, 246–251 (2019).
63. Zhang, F. *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* **20**, 928–942 (2019).
64. Wei, K. *et al.* Notch signalling drives synovial fibroblast identity and arthritis pathology. *Nature* **582**, 259–264 (2020).
65. Friščić, J. *et al.* The complement system drives local inflammatory tissue priming by metabolic reprogramming of synovial fibroblasts. *Immunity* **54**, 1002-1021.e10 (2021).

66. Kurowska-Stolarska, M. & Alivernini, S. Synovial tissue macrophages in joint homeostasis, rheumatoid arthritis and disease remission. *Nat Rev Rheumatol* **18**, 384–397 (2022).
67. Alivernini, S. *et al.* Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat Med* **26**, 1295–1306 (2020).
68. Misharin, A. V. *et al.* Nonclassical Ly6C(-) monocytes drive the development of inflammatory arthritis in mice. *Cell Rep* **9**, 591–604 (2014).
69. Rossol, M., Kraus, S., Pierer, M., Baerwald, C. & Wagner, U. The CD14brightCD16+ monocyte subset is expanded in rheumatoid arthritis and promotes expansion of the Th17 cell population. *Arthritis & Rheumatism* **64**, 671–677 (2012).
70. McGarry, T. *et al.* Rheumatoid arthritis CD14+ monocytes display metabolic and inflammatory dysfunction, a phenotype that precedes clinical manifestation of disease. *Clinical & Translational Immunology* **10**, e1237 (2021).
71. Eggleton, P., Wang, L., Penhallow, J., Crawford, N. & Brown, K. A. Differences in oxidative response of subpopulations of neutrophils from healthy subjects and patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases* **54**, 916–923 (1995).
72. Wright, H. L., Chikura, B., Bucknall, R. C., Moots, R. J. & Edwards, S. W. Changes in expression of membrane TNF, NF-{kappa}B activation and neutrophil apoptosis during active and resolved inflammation. *Ann Rheum Dis* **70**, 537–543 (2011).
73. Khandpur, R. *et al.* NETs Are a Source of Citrullinated Autoantigens and Stimulate Inflammatory Responses in Rheumatoid Arthritis. *Science Translational Medicine* **5**, 178ra40-178ra40 (2013).
74. Zec, K. *et al.* Macrophages in the synovial lining niche initiate neutrophil recruitment and articular inflammation. *Journal of Experimental Medicine* **220**, e20220595 (2023).
75. Suwa, Y., Nagafuchi, Y., Yamada, S. & Fujio, K. The role of dendritic cells and their immunometabolism in rheumatoid arthritis. *Front Immunol* **14**, 1161148 (2023).
76. Canavan, M. *et al.* Functionally Mature CD1c+ Dendritic Cells Preferentially Accumulate in the Inflammatory Arthritis Synovium. *Front. Immunol.* **12**, (2021).

77. Benson, R. A. *et al.* Identifying the Cells Breaching Self-Tolerance in Autoimmunity. *The Journal of Immunology* **184**, 6378–6385 (2010).
78. Klarenbeek, P. L. *et al.* Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Annals of the Rheumatic Diseases* **71**, 1088–1093 (2012).
79. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* **44**, 291–296 (2012).
80. Rao, D. A. *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* **542**, 110 (2017).
81. Bombardieri, M. *et al.* A BAFF/APRIL-dependent TLR3-stimulated pathway enhances the capacity of rheumatoid synovial fibroblasts to induce AID expression and Ig class-switching in B cells. *Ann Rheum Dis* **70**, 1857–1865 (2011).
82. Pitzalis, C., Jones, G. W., Bombardieri, M. & Jones, S. A. Ectopic lymphoid-like structures in infection, cancer and autoimmunity. *Nat Rev Immunol* **14**, 447–462 (2014).
83. Buckley, C. D., Ospelt, C., Gay, S. & Midwood, K. S. Location, location, location: how the tissue microenvironment affects inflammation in RA. *Nat Rev Rheumatol* **17**, 195–212 (2021).
84. Middleton, J., Patterson, A. M., Gardner, L., Schmutz, C. & Ashton, B. A. Leukocyte extravasation: chemokine transport and presentation by the endothelium. *Blood* **100**, 3853–3860 (2002).
85. Midwood, K. *et al.* Tenascin-C is an endogenous activator of Toll-like receptor 4 that is essential for maintaining inflammation in arthritic joint disease. *Nat Med* **15**, 774–780 (2009).
86. Aungier, S. R. *et al.* Targeting early changes in the synovial microenvironment: a new class of immunomodulatory therapy? *Ann Rheum Dis* **78**, 186–191 (2019).
87. Smolen, J. S. & Steiner, G. Therapeutic strategies for rheumatoid arthritis. *Nat Rev Drug Discov* **2**, 473–488 (2003).
88. Fleischmann, R. Don’t forget traditional DMARDs. *Rheumatology* **50**, 429–430 (2011).

89. Ramamoorthy, S. & Cidlowski, J. A. Corticosteroids-Mechanisms of Action in Health and Disease. *Rheum Dis Clin North Am* **42**, 15–31 (2016).
90. Boers, M. *et al.* Low dose, add-on prednisolone in patients with rheumatoid arthritis aged 65+: the pragmatic randomised, double-blind placebo-controlled GLORIA trial. *Annals of the Rheumatic Diseases* **81**, 925–936 (2022).
91. Strehl, C. *et al.* Defining conditions where long-term glucocorticoid treatment has an acceptably low level of harm to facilitate implementation of existing recommendations: viewpoints from an EULAR task force. *Ann Rheum Dis* **75**, 952–957 (2016).
92. Hetland, M. L. *et al.* Active conventional treatment and three different biological treatments in early rheumatoid arthritis: phase IV investigator initiated, randomised, observer blinded clinical trial. *BMJ* **371**, m4328 (2020).
93. Gellad, W. F. & Good, C. B. Adalimumab and the Challenges for Biosimilars. *JAMA* **322**, 2171–2172 (2019).
94. Goffe, B. & Cather, J. C. Etanercept: An overview. *Journal of the American Academy of Dermatology* **49**, 105–111 (2003).
95. Smolen, J. S. & Emery, P. Infliximab: 12 years of experience. *Arthritis Res Ther* **13**, S2 (2011).
96. Ma, X. & Xu, S. TNF inhibitor therapy for rheumatoid arthritis. *Biomed Rep* **1**, 177–184 (2013).
97. Cohen, S. *et al.* Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four-week, multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* **46**, 614–624 (2002).
98. Weinblatt, M. E. *et al.* Tocilizumab as monotherapy or in combination with nonbiologic disease-modifying antirheumatic drugs: twenty-four-week results of an open-label, clinical practice study. *Arthritis Care Res (Hoboken)* **65**, 362–371 (2013).
99. Westhovens, R. & Verschueren, P. The Efficacy and Safety of Abatacept in Rheumatoid Arthritis. *Ther Adv Musculoskelet Dis* **2**, 89–94 (2010).
100. Onrust, S. V., Lamb, H. M. & Barman Balfour, J. A. Rituximab. *Drugs* **58**, 79–88 (1999).

101. Lopez-Olivo, M. A., Urruela, M. A., McGahan, L., Pollono, E. N. & Suarez-Almazor, M. E. Rituximab for rheumatoid arthritis. *Cochrane Database of Systematic Reviews* (2015) doi:10.1002/14651858.CD007356.pub2.
102. Lee, E. B. *et al.* Tofacitinib versus Methotrexate in Rheumatoid Arthritis. *New England Journal of Medicine* **370**, 2377–2386 (2014).
103. Chiu, Y.-M. & Chen, D.-Y. Infection risk in patients undergoing treatment for inflammatory arthritis: non-biologics versus biologics. *Expert Review of Clinical Immunology* **16**, 207–228 (2020).
104. Morante-Palacios, O., Fondelli, F., Ballestar, E. & Martínez-Cáceres, E. M. Tolerogenic Dendritic Cells in Autoimmunity and Inflammatory Diseases. *Trends in Immunology* **42**, 59–75 (2021).
105. Bell, G. M. *et al.* Autologous tolerogenic dendritic cells for rheumatoid and inflammatory arthritis. *Annals of the Rheumatic Diseases* **76**, 227–234 (2017).
106. Zhang, F. *et al.* Cellular deconstruction of inflamed synovium defines diverse inflammatory phenotypes in rheumatoid arthritis. 2022.02.25.481990 Preprint at https://doi.org/10.1101/2022.02.25.481990 (2022).
107. Pratt, A. G. *et al.* Targeting synovial fibroblast proliferation in rheumatoid arthritis (TRAFIC): an open-label, dose-finding, phase 1b trial. *The Lancet Rheumatology* **3**, e337–e346 (2021).
108. Cope, A. P. *et al.* Abatacept in individuals at high risk of rheumatoid arthritis (APIPPRA): a randomised, double-blind, multicentre, parallel, placebo-controlled, phase 2b clinical trial. *The Lancet* **403**, 838–849 (2024).
109. Lewis, M. J. *et al.* Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep* **28**, 2455-2470.e5 (2019).
110. Sharma, S., Dwivedi, S., Chandra, S., Srivastava, A. & Vijay, P. Collagen: A Brief Analysis. *Oral & Maxillofacial Pathology Journal* **10**, (2019).
111. Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annu Rev Biochem* **78**, 929–958 (2009).
112. Ramshaw, J. A. M., Shah, N. K. & Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host–Guest Triple-Helical Peptides. *Journal of Structural Biology* **122**, 86–91 (1998).

113. Bella, J. & Hulmes, D. J. S. Fibrillar Collagens. in *Fibrous Proteins: Structures and Mechanisms* (eds. Parry, D. A. D. & Squire, J. M.) 457–490 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-49674-0_14.
114. Hulmes, D. J., Jesior, J. C., Miller, A., Berthet-Colominas, C. & Wolff, C. Electron microscopy shows periodic structure in collagen fibril cross sections. *Proceedings of the National Academy of Sciences* **78**, 3567–3571 (1981).
115. Revell, C. K. *et al.* Collagen fibril assembly: New approaches to unanswered questions. *Matrix Biology Plus* **12**, 100079 (2021).
116. Buehler, M. J. Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *Proceedings of the National Academy of Sciences* **103**, 12285–12290 (2006).
117. Fung, YC ,. Author & Huston, RL ,. Reviewer. Introduction to Bioengineering. *Applied Mechanics Reviews* **55**, B39 (2002).
118. Lauer-Fields, J. L., Juska, D. & Fields, G. B. Matrix metalloproteinases and collagen catabolism. *Peptide Science* **66**, 19–32 (2002).
119. Cawston, T. *et al.* The Regulation of MMPs and TIMPs in Cartilage Turnover. *Annals of the New York Academy of Sciences* **878**, 120–129 (1999).
120. Kk, N. & Ja, M. Mitochondrial redox control of matrix metalloproteinases. *Free radical biology & medicine* **37**, (2004).
121. Kong, W., Lyu, C., Liao, H. & Du, Y. Collagen crosslinking: effect on structure, mechanics and fibrosis progression. *Biomed. Mater.* **16**, 062005 (2021).
122. Chang, J. *et al.* Circadian control of the secretory pathway maintains collagen homeostasis. *Nat Cell Biol* **22**, 74–86 (2020).
123. Forlino, A., Cabral, W. A., Barnes, A. M. & Marini, J. C. New Perspectives on Osteogenesis Imperfecta. *Nat Rev Endocrinol* **7**, 540–557 (2011).
124. Marini, J. C. *et al.* Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Human Mutation* **28**, 209–221 (2007).
125. Burgstaller, G. *et al.* The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. *European Respiratory Journal* **50**, (2017).
126. Tjin, G. *et al.* Lysyl oxidases regulate fibrillar collagen remodelling in idiopathic pulmonary fibrosis. *Disease Models & Mechanisms* **10**, 1301–1312 (2017).

127. Chen, Y. *et al.* Oncogenic collagen I homotrimers from cancer cells bind to α3β1 integrin and impact tumor microbiome and immunity to promote pancreatic cancer. *Cancer Cell* **40**, 818-834.e9 (2022).
128. Pickup, M. W. *et al.* Stromally Derived Lysyl Oxidase Promotes Metastasis of Transforming Growth Factor-β–Deficient Mouse Mammary Carcinomas. *Cancer Research* **73**, 5336–5346 (2013).
129. Bougherara, H. *et al.* Real-Time Imaging of Resident T Cells in Human Lung and Ovarian Carcinomas Reveals How Different Tumor Microenvironments Control T Lymphocyte Migration. *Front. Immunol.* **6**, (2015).
130. Badalamente, M. A., Hurst, L. C., Benhaim, P. & Cohen, B. M. Efficacy and Safety of Collagenase Clostridium Histolyticum in the Treatment of Proximal Interphalangeal Joints in Dupuytren Contracture: Combined Analysis of 4 Phase 3 Clinical Trials. *The Journal of Hand Surgery* **40**, 975–983 (2015).
131. Sasaki, K. *et al.* Engineered collagen-binding serum albumin as a drug conjugate carrier for cancer therapy. *Science Advances* **5**, eaaw6081 (2019).
132. Ishihara, J. *et al.* Targeted antibody and cytokine cancer immunotherapies through collagen affinity. *Science Translational Medicine* **11**, eaau3259 (2019).
133. Cescon, M., Gattazzo, F., Chen, P. & Bonaldo, P. Collagen VI at a glance. *Journal of Cell Science* **128**, 3525–3531 (2015).
134. Fitzgerald, J., Rich, C., Zhou, F. H. & Hansen, U. Three Novel Collagen VI Chains, α4(VI), α5(VI), and α6(VI) *. *Journal of Biological Chemistry* **283**, 20170–20180 (2008).
135. Gara, S. K. *et al.* Three Novel Collagen VI Chains with High Homology to the α3 Chain *. *Journal of Biological Chemistry* **283**, 10658–10670 (2008).
136. Bonaldo, P., Russo, V., Bucciotti, F., Bressan, G. M. & Colombatti, A. α1 Chain of Chick Type VI Collagen. *Journal of Biological Chemistry* **264**, 5575–5580 (1989).
137. Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R. & Colombatti, A. Structural and functional features of the .alpha.3 chain indicate a bridging role for chicken collagen VI in connective tissues. *ACS Publications* https://pubs.acs.org/doi/pdf/10.1021/bi00457a021 (2002) doi:10.1021/bi00457a021.

138. Doliana, R., Bonaldo, P. & Colombatti, A. Multiple forms of chicken alpha 3(VI) collagen chain generated by alternative splicing in type A repeated domains. *Journal of Cell Biology* **111**, 2197–2205 (1990).
139. Dziadek, M. *et al.* Alternative splicing of transcripts for the alpha3 chain of mouse collagen VI: identification of an abundant isoform lacking domains N7–N10 in mouse and human. *Matrix Biology* **21**, 227–241 (2002).
140. Saitta, B., Stokes, D. G., Vissing, H., Timpl, R. & Chu, M. L. Alternative splicing of the human alpha 2(VI) collagen gene generates multiple mRNA transcripts which predict three protein variants with distinct carboxyl termini. *Journal of Biological Chemistry* **265**, 6473–6480 (1990).
141. Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E. & Engel, J. Electron-microscopical approach to a structural model of intima collagen. *Biochemical Journal* **211**, 303–311 (1983).
142. Knupp, C. *et al.* Structural correlation between collagen VI microfibrils and collagen VI banded aggregates. *Journal of Structural Biology* **154**, 312–326 (2006).
143. Bernardi, P. & Bonaldo, P. Dysfunction of Mitochondria and Sarcoplasmic Reticulum in the Pathogenesis of Collagen VI Muscular Dystrophies. *Annals of the New York Academy of Sciences* **1147**, 303–311 (2008).
144. Aigner, T., Hambach, L., Söder, S., Schlötzer-Schrehardt, U. & Pöschl, E. The C5 domain of Col6A3 is cleaved off from the Col6 fibrils immediately after secretion. *Biochem Biophys Res Commun* **290**, 743–748 (2002).
145. Heumüller, S. E. *et al.* C-terminal proteolysis of the collagen VI α3 chain by BMP-1 and proprotein convertase(s) releases endotrophin in fragments of different sizes. *The Journal of Biological Chemistry* **294**, 13769 (2019).
146. Park, J. & Scherer, P. E. Adipocyte-derived endotrophin promotes malignant tumor progression. *J Clin Invest* **122**, 4243–4256 (2012).
147. Willumsen, N., Bager, C. & Karsdal, M. A. Matrix Metalloprotease Generated Fragments of Type VI Collagen Have Serum Biomarker Potential in Cancer – A Proof of Concept Study. *Transl Oncol* **12**, 693–698 (2019).
148. Williams, L. M. *et al.* Identifying collagen VI as a target of fibrotic diseases regulated by CREBBP/EP300. *Proc Natl Acad Sci U S A* **117**, 20753–20763 (2020).

149. Zou, Y., Zhang, R.-Z., Sabatelli, P., Chu, M.-L. & Bönnemann, C. G. Muscle interstitial fibroblasts are the main source of collagen VI synthesis in skeletal muscle: implications for congenital muscular dystrophy types Ullrich and Bethlem. *J Neuropathol Exp Neurol* **67**, 144–154 (2008).
150. Ritty, T. M., Roth, R. & Heuser, J. E. Tendon Cell Array Isolation Reveals a Previously Unknown Fibrillin-2-Containing Macromolecular Assembly. *Structure* **11**, 1179–1188 (2003).
151. Sievers, J., Pehlemann, F. W., Gude, S. & Berry, M. Meningeal cells organize the superficial glia limitans of the cerebellum and produce components of both the interstitial matrix and the basement membrane. *J Neurocytol* **23**, 135–149 (1994).
152. Vitale, P., Braghetta, P., Volpin, D., Bonaldo, P. & Bressan, G. M. Mechanisms of transcriptional activation of the *col6a1* gene during Schwann cell differentiation. *Mechanisms of Development* **102**, 145–156 (2001).
153. Pasarica, M. *et al.* Adipose Tissue Collagen VI in Obesity. *The Journal of Clinical Endocrinology & Metabolism* **94**, 5155–5162 (2009).
154. Khan, T. *et al.* Metabolic Dysregulation and Adipose Tissue Fibrosis: Role of Collagen VI. *Molecular and Cellular Biology* **29**, 1575–1591 (2009).
155. Alexopoulos, L. G., Youn, I., Bonaldo, P. & Guilak, F. Developmental and osteoarthritic changes in Col6a1-knockout mice: Biomechanics of type VI collagen in the cartilage pericellular matrix. *Arthritis & Rheumatism* **60**, 771–779 (2009).
156. Lettmann, S. *et al.* Col6a1 Null Mice as a Model to Study Skin Phenotypes in Patients with Collagen VI Related Myopathies: Expression of Classical and Novel Collagen VI Variants during Wound Healing. *PLOS ONE* **9**, e105686 (2014).
157. Jöbsis, G. J., Boers, J. M., Barth, P. G. & de Visser, M. Bethlem myopathy: a slowly progressive congenital muscular dystrophy with contractures. *Brain* **122**, 649–655 (1999).
158. Bonnemann, C. G. Congenital Muscular Dystrophy. in *Encyclopedia of Neuroscience* (ed. Squire, L. R.) 67–74 (Academic Press, Oxford, 2009). doi:10.1016/B978-008045046-9.01520-5.
159. Bönnemann, C. G. The collagen VI-related myopathies: muscle meets its matrix. *Nat Rev Neurol* **7**, 379–390 (2011).

160. Lampe, A. K. & Bushby, K. M. D. Collagen VI related muscle disorders. *Journal of Medical Genetics* **42**, 673–685 (2005).
161. Butterfield, R. J. *et al.* Transcriptome profiling identifies regulators of pathogenesis in collagen VI related muscular dystrophy. *PLoS One* **12**, e0189664 (2017).
162. Kwong, A. K. *et al.* Collagen VI-related myopathies: clinical variability, phenotype-genotype correlation and exploratory transcriptome study. *Neuromuscular Disorders* **33**, 371–381 (2023).
163. Theocharidis, G. *et al.* Type VI Collagen Regulates Dermal Matrix Assembly and Fibroblast Motility. *J Invest Dermatol* **136**, 74–83 (2016).
164. Peltonen, J. *et al.* Activation of Collagen Gene Expression in Keloids: Co-Localization of Type I and VI Collagen and Transforming Growth Factor-β1 mRNA. *Journal of Investigative Dermatology* **97**, 240–248 (1991).
165. Peltonen, J., Kähäri, L., Uitto, J. & Jimenez, S. A. Increased expression of type VI collagen genes in systemic sclerosis. *Arthritis & Rheumatism* **33**, 1829–1835 (1990).
166. Takasaki, S., Fujiwara, S., Shinkai, H. & Ooshima, A. Human Type VI Collagen: Purification from Human Subcutaneous Fat Tissue and an Immunohistochemical Study of Morphea and Systemic Sclerosis. *The Journal of Dermatology* **22**, 480–485 (1995).
167. Magro, G., Grasso, S., Colombatti, A. & Lopes, M. Immunohistochemical distribution of type VI collagen in developing human kidney. *Histochem J* **28**, 385–390 (1996).
168. Vleming, L. J. *et al.* Progression of chronic renal disease in humans is associated with the deposition of basement membrane components and decorin in the interstitial extracellular matrix. *Clin Nephrol* **44**, 211–219 (1995).
169. Specks, U., Nerlich, A., Colby, T. V., Wiest, I. & Timpl, R. Increased expression of type VI collagen in lung fibrosis. *Am J Respir Crit Care Med* **151**, 1956–1964 (1995).
170. Ucero, A. C. *et al.* Fra-2–expressing macrophages promote lung fibrosis. *J Clin Invest* **129**, 3293–3309 (2019).
171. Herrera, J. *et al.* Registration of the extracellular matrix components constituting the fibroblastic focus in idiopathic pulmonary fibrosis. *JCI Insight* **4**, e125185.

172. Barron, A. M. S., Fabre, T. & De, S. Distinct fibroblast functions associated with fibrotic and immune-mediated inflammatory diseases and their implications for therapeutic development. Preprint at https://doi.org/10.12688/f1000research.143472.1 (2024).
173. Nanda, A. *et al.* TEM8 Interacts with the Cleaved C5 Domain of Collagen α3(VI). *Cancer Research* **64**, 817–820 (2004).
174. Arafat, H. *et al.* Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer. *Surgery* **150**, 306–315 (2011).
175. Motrescu, E. R. *et al.* Matrix metalloproteinase-11/stromelysin-3 exhibits collagenolytic function against collagen VI under normal and malignant conditions. *Oncogene* **27**, 6347–6355 (2008).
176. Zhang, X. *et al.* High expression of COL6A1 predicts poor prognosis and response to immunotherapy in bladder cancer. *Cell Cycle* **0**, 1–9 (2022).
177. Hou, T. *et al.* Expression of COL6A1 predicts prognosis in cervical cancer patients. *Am J Transl Res* **8**, 2838–2844 (2016).
178. Owusu-Ansah, K. G. *et al.* COL6A1 promotes metastasis and predicts poor prognosis in patients with pancreatic cancer. *Int J Oncol* **55**, 391–404 (2019).
179. Zhang, L., Wang, L., Yang, H., Li, C. & Fang, C. Identification of potential genes related to breast cancer brain metastasis in breast cancer patients. *Bioscience Reports* **41**, BSR20211615 (2021).
180. Cescon, M. *et al.* Collagen VI sustains cell stemness and chemotherapy resistance in glioblastoma. *Cell. Mol. Life Sci.* **80**, 233 (2023).
181. Park, J., Morley, T. S. & Scherer, P. E. Inhibition of endotrophin, a cleavage product of collagen VI, confers cisplatin sensitivity to tumours. *EMBO Mol Med* **5**, 935–948 (2013).
182. Holm Nielsen, S. *et al.* A Fragment of Collagen Type VI alpha-3 chain is Elevated in Serum from Patients with Gastrointestinal Disorders. *Sci Rep* **10**, 5910 (2020).
183. Lindholm, M. *et al.* Endotrophin and C6Ma3, serological biomarkers of type VI collagen remodelling, reflect endoscopic and clinical disease activity in IBD. *Sci Rep* **11**, 14713 (2021).
184. Thudium, C. S., Frederiksen, P., Karsdal, M. A. & Bay-Jensen, A.-C. Changes in type VI collagen degradation reflect clinical response to treatment in

rheumatoid arthritis patients treated with tocilizumab. *Arthritis Research & Therapy* **26**, 3 (2024).

185. Ng, M. T. H. *et al.* A single cell atlas of frozen shoulder capsule identifies features associated with inflammatory fibrosis resolution. *Nat Commun* **15**, 1394 (2024).
186. Aibar, S. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).
187. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
188. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
189. Kolde, R. Pheatmap: pretty heatmaps. *R package version* **1**, 726 (2019).
190. Brunson, J. C. ggalluvial: Layered Grammar for Alluvial Plots. *Journal of Open Source Software* **5**, 2017 (2020).
191. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
192. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience* **7**, giy083 (2018).
193. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
194. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16**, 278 (2015).
195. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology* **15**, e8746 (2019).
196. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
197. Ren, X. *et al.* Quantitative Proteomic Analysis of Synovial Tissue Reveals That Upregulated OLFM4 Aggravates Inflammation in Rheumatoid Arthritis. *J. Proteome Res.* **20**, 4746–4757 (2021).

198. Freeman, T. C. *et al.* Graphia: A platform for the graph-based visualisation and analysis of high dimensional data. *PLOS Computational Biology* **18**, e1010310 (2022).
199. Nirmal, A. J. & Sorger, P. K. SCIMAP: A Python Toolkit for Integrated Spatial Analysis of Multiplexed Imaging Data. *Journal of Open Source Software* **9**, 6604 (2024).
200. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).
201. Primorac, D. *et al.* Knee Osteoarthritis: A Review of Pathogenesis and State-Of-The-Art Non-Operative Therapeutic Considerations. *Genes (Basel)* **11**, 854 (2020).
202. Rims, C. *et al.* Citrullinated Aggrecan Epitopes as Targets of Auto-reactive CD4+ T cells in Patients with Rheumatoid Arthritis. *Arthritis Rheumatol* **71**, 518–528 (2019).
203. Pieper, J. *et al.* HLA-DRB1*04:01 restricted T cells in blood and synovial fluid of RA patients. *J Autoimmun* **92**, 47–56 (2018).
204. Schwenzer, A. *et al.* Identification of an immunodominant peptide from citrullinated tenascin-C as a major target for autoantibodies in rheumatoid arthritis. *Annals of the Rheumatic Diseases* **75**, 1876–1883 (2016).
205. *Matrix Metalloproteinases and Tissue Remodeling in Health and Disease: Target Tissues and Therapy*. (Academic Press, 2017).
206. Zhang, F. *et al.* Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes. *Nature* **623**, 616–624 (2023).
207. Culemann, S. *et al.* Locally renewing resident synovial macrophages provide a protective barrier for the joint. *Nature* **572**, 670–675 (2019).
208. Rodriguez-Lopez, J., Perez-Pampin, E., Gomez-Reino, J. J. & Gonzalez, A. Regulatory polymorphisms in extracellular matrix protease genes and susceptibility to rheumatoid arthritis: a case-control study. *Arthritis Res Ther* **8**, R1 (2006).
209. Foroughi, F. *et al.* Association between matrix metalloproteinase-9-1562C/T gene polymorphism and MMP-9 serum level in rheumatoid arthritis. *Journal of Immunoassay and Immunochemistry* **45**, 362–381 (2024).
210. Mizoguchi, F. *et al.* Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat Commun* **9**, 789 (2018).

211. Zhou, S., Chen, S., Pei, Y. A. & Pei, M. Nidogen: A matrix protein with potential roles in musculoskeletal tissue regeneration. *Genes Dis* **9**, 598–609 (2021).
212. Chen, J., Zhang, J., Hong, L. & Zhou, Y. EGFLAM correlates with cell proliferation, migration, invasion and poor prognosis in glioblastoma. *Cancer Biomark* **24**, 343–350 (2019).
213. Lu, Y. *et al.* Role of Pericytes in Cardiomyopathy-Associated Myocardial Infarction Revealed by Multiple Single-Cell Sequencing Analysis. *Biomedicines* **11**, 2896 (2023).
214. Denninger, K. C. *et al.* Kinetics of gene expression and bone remodelling in the clinical phase of collagen-induced arthritis. *Arthritis Research & Therapy* **17**, 43 (2015).
215. Andersson, K. M. E. *et al.* Inflammation in the hippocampus affects IGF1 receptor signaling and contributes to neurological sequelae in rheumatoid arthritis. *Proceedings of the National Academy of Sciences* **115**, E12063–E12072 (2018).
216. Evdokimova, V. *et al.* IGFBP7 Binds to the IGF-1 Receptor and Blocks Its Activation by Insulin-Like Growth Factors. *Science Signaling* **5**, ra92–ra92 (2012).
217. Wang, W.-M. *et al.* Transforming Growth Factor-β Induces Secretion of Activated ADAMTS-2: A PROCOLLAGEN III N-PROTEINASE*. *Journal of Biological Chemistry* **278**, 19549–19557 (2003).
218. Cain, S. A., Woods, S., Singh, M., Kimber, S. J. & Baldock, C. ADAMTS6 cleaves the large latent TGFβ complex and increases the mechanotension of cells to activate TGFβ. *Matrix Biology* **114**, 18–34 (2022).
219. Gomes, L. R., Terra, L. F., Wailemann, R. A., Labriola, L. & Sogayar, M. C. TGF-β1 modulates the homeostasis between MMPs and MMP inhibitors through p38 MAPK and ERK1/2 in highly invasive breast cancer cells. *BMC Cancer* **12**, 26 (2012).
220. Kim, H.-S., Shang, T., Chen, Z., Pflugfelder, S. C. & Li, D.-Q. TGF-β1 stimulates production of gelatinase (MMP-9), collagenases (MMP-1, -13) and stromelysins (MMP-3, -10, -11) by human corneal epithelial cells. *Experimental Eye Research* **79**, 263–274 (2004).

221. Li, X. *et al.* Overexpression of SerpinE2/protease nexin-1 Contribute to Pathological Cardiac Fibrosis via increasing Collagen Deposition. *Sci Rep* **6**, 37635 (2016).
222. Park, S.-A. *et al.* TIMP-1 mediates TGF-β-dependent crosstalk between hepatic stellate and cancer cells via FAK signaling. *Sci Rep* **5**, 16492 (2015).
223. Akiel, M. *et al.* IGFBP7 Deletion Promotes Hepatocellular Carcinoma. *Cancer Research* **77**, 4014–4025 (2017).
224. Heydemann, A. *et al.* Latent TGF-β–binding protein 4 modifies muscular dystrophy in mice. *J Clin Invest* **119**, 3703–3712 (2009).
225. Wang, Y. & Nie, G. Overview of Human HtrA Family Proteases and Their Distinctive Physiological Roles and Unique Involvement in Diseases, Especially Cancer and Pregnancy Complications. *Int J Mol Sci* **22**, 10756 (2021).
226. Johnson, K., Farley, D., Hu, S.-I. & Terkeltaub, R. One of two chondrocyte-expressed isoforms of cartilage intermediate-layer protein functions as an insulin-like growth factor 1 antagonist. *Arthritis Rheum* **48**, 1302–1314 (2003).
227. Ding, H. & Wu, T. Insulin-Like Growth Factor Binding Proteins in Autoimmune Diseases. *Front. Endocrinol.* **9**, (2018).
228. Yan, J. *et al.* FGL2 promotes tumor progression in the CNS by suppressing CD103+ dendritic cell differentiation. *Nat Commun* **10**, 448 (2019).
229. Boström, E. A. *et al.* Resistin and insulin/insulin-like growth factor signaling in rheumatoid arthritis. *Arthritis Rheum* **63**, 2894–2904 (2011).
230. Sekiguchi, R. & Yamada, K. M. Basement membranes in development and disease. *Curr Top Dev Biol* **130**, 143–191 (2018).
231. Hynes, R. O. & Naba, A. Overview of the Matrisome—An Inventory of Extracellular Matrix Constituents and Functions. *Cold Spring Harb Perspect Biol* **4**, a004903 (2012).
232. Zheng, X., Wu, S.-L., Hincapie, M. & Hancock, W. S. Study of the human plasma proteome of rheumatoid arthritis. *J Chromatogr A* **1216**, 3538–3545 (2009).
233. Jin, E.-H., Shim, S.-C., Kim, H.-G., Chae, S.-C. & Chung, H.-T. Polymorphisms of COTL1 gene identified by proteomic approach and their association with autoimmune disorders. *Exp Mol Med* **41**, 354–361 (2009).

234. Kastrinaki, M.-C. *et al.* Functional, molecular and proteomic characterisation of bone marrow mesenchymal stem cells in rheumatoid arthritis. *Ann Rheum Dis* **67**, 741–749 (2008).
235. Bo, G.-P. *et al.* Analyses of differential proteome of human synovial fibroblasts obtained from arthritis. *Clin Rheumatol* **28**, 191–199 (2009).
236. Zhang, H., Fan, L. Y., Zong, M., Sun, L. S. & Lu, L. Proteins related to the functions of fibroblast-like synoviocytes identified by proteomic analysis. *Clin Exp Rheumatol* **30**, 213–221 (2012).
237. Katano, M. *et al.* Implication of granulocyte-macrophage colony-stimulating factor induced neutrophil gelatinase-associated lipocalin in pathogenesis of rheumatoid arthritis revealed by proteome analysis. *Arthritis Res Ther* **11**, R3 (2009).
238. Dotzlaw, H., Schulz, M., Eggert, M. & Neeck, G. A pattern of protein expression in peripheral blood mononuclear cells distinguishes rheumatoid arthritis patients from healthy individuals. *Biochim Biophys Acta* **1696**, 121–129 (2004).
239. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
240. Rich, J. M. *et al.* The impact of package selection and versioning on single-cell RNA-seq analysis. 2024.04.04.588111 Preprint at https://doi.org/10.1101/2024.04.04.588111 (2024).
241. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578 (2018).
242. Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S. & Marioni, J. C. Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nat Methods* **14**, 565–571 (2017).
243. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods* **15**, 255–261 (2018).
244. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat Commun* **12**, 5692 (2021).
245. Pompili, S., Latella, G., Gaudio, E., Sferra, R. & Vetuschi, A. The Charming World of the Extracellular Matrix: A Dynamic and Protective Network of the Intestinal Wall. *Front. Med.* **8**, (2021).

246. Halfter, W. *et al.* New concepts in basement membrane biology. *The FEBS Journal* **282**, 4466–4479 (2015).
247. Schneider, M. *et al.* Basement membrane proteins in synovial membrane: Distribution in rheumatoid arthritis and synthesis by fibroblast-like cells. *Clin Rheumatol* **13**, 90–97 (1994).
248. Steenvoorden, M. M. *et al.* Transition of healthy to diseased synovial tissue in rheumatoid arthritis is associated with gain of mesenchymal/fibrotic characteristics. *Arthritis Research & Therapy* **8**, R165 (2006).
249. Poduval, P., T. Sillat, A. Beklen, V. P. Kouri, I. Virtanen, and Y. T. Konttinen. Type IV Collagen Alpha-Chain Composition in Synovial Lining from Trauma Patients and Patients with Rheumatoid Arthritis. *Arthritis and Rheumatism* 56, no. 12 (December 2007): 3959–67.
250. Reuten, R. *et al.* Basement membrane stiffness determines metastases formation. *Nat. Mater.* **20**, 892–903 (2021).
251. Jay, G. D., Torres, J. R., Warman, M. L., Laderer, M. C. & Breuer, K. S. The role of lubricin in the mechanical behavior of synovial fluid. *Proceedings of the National Academy of Sciences* **104**, 6194–6199 (2007).
252. Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
253. Korsunsky, I. *et al.* Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. *Med* **3**, 481-518.e14 (2022).
254. Kalus, W. *et al.* Structure of the IGF-binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5): implications for IGF and IGF-I receptor interactions. *EMBO J* **17**, 6558–6572 (1998).
255. Han, N. *et al.* Local application of IGFBP5 protein enhanced periodontal tissue regeneration via increasing the migration, cell proliferation and osteo/dentinogenic differentiation of mesenchymal stem cells in an inflammatory niche. *Stem Cell Res Ther* **8**, 210 (2017).
256. Nguyen, X.-X., Renaud, L. & Feghali-Bostwick, C. Identification of Impacted Pathways and Transcriptomic Markers as Potential Mediators of Pulmonary Fibrosis in Transgenic Mice Expressing Human IGFBP5. *Int J Mol Sci* **22**, 12609 (2021).

257. Hand, C., Clipsham, K., Rees, J. L. & Carr, A. J. Long-term outcome of frozen shoulder. *Journal of Shoulder and Elbow Surgery* **17**, 231–236 (2008).
258. Erlandsson, M. C. *et al.* IGF-1R signalling contributes to IL-6 production and T cell dependent inflammation in rheumatoid arthritis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1863**, 2158–2170 (2017).
259. Zhang, Y., Yan, N., Wang, X., Chang, Y. & Wang, Y. MiR-129-5p regulates cell proliferation and apoptosis via IGF-1R/Src/ERK/Egr-1 pathway in RA-fibroblast-like synoviocytes. *Biosci Rep* **39**, BSR20192009 (2019).
260. Xie, Y. *et al.* Regulation of insulin-like growth factor signaling by metformin in endometrial cancer cells. *Oncology Letters* **8**, 1993–1999 (2014).
261. Douglas Raymond S. *et al.* Teprotumumab for the Treatment of Active Thyroid Eye Disease. *New England Journal of Medicine* **382**, 341–352 (2020).
262. Sakuma, M. *et al.* TGF-β type I receptor kinase inhibitor down-regulates rheumatoid synoviocytes and prevents the arthritis induced by type II collagen antibody. *International Immunology* **19**, 117–126 (2007).
263. Aarts, J. *et al.* Local inhibition of TGF-β1 signaling improves Th17/Treg balance but not joint pathology during experimental arthritis. *Sci Rep* **12**, 3182 (2022).
264. Dahlin, J. L., Inglese, J. & Walters, M. A. Mitigating risk in academic preclinical drug discovery. *Nat Rev Drug Discov* **14**, 279–294 (2015).
265. Yamaguchi, Y., Mann, D. M. & Ruoslahti, E. Negative regulation of transforming growth factor-beta by the proteoglycan decorin. *Nature* **346**, 281–284 (1990).
266. Frey, H., Schroeder, N., Manon-Jensen, T., Iozzo, R. V. & Schaefer, L. Biological interplay between proteoglycans and their innate immune receptors in inflammation. *FEBS J* **280**, 2165–2179 (2013).
267. Polgár, A. *et al.* Elevated levels of synovial fluid antibodies reactive with the small proteoglycans biglycan and decorin in patients with rheumatoid arthritis or other joint diseases. *Rheumatology (Oxford)* **42**, 522–527 (2003).
268. Ly, M., Laremore, T. N. & Linhardt, R. J. Proteoglycomics: Recent Progress and Future Challenges. *OMICS* **14**, 389–399 (2010).
269. Rinaldi, N. *et al.* Loss of collagen type IV in rheumatoid synovia and cytokine effect on the collagen type-IV gene expression in fibroblast-like synoviocytes from rheumatoid arthritis. *Virchows Arch* **439**, 675–682 (2001).
270. Henderson, B., Revell, P. A. & Edwards, J. C. Synovial lining cell hyperplasia in rheumatoid arthritis: dogma and fact. *Ann Rheum Dis* **47**, 348–349 (1988).

271. Severino, V. *et al.* Insulin-like growth factor binding proteins 4 and 7 released by senescent cells promote premature senescence in mesenchymal stem cells. *Cell Death Dis* **4**, e911 (2013).
272. Akaogi, K. *et al.* Specific accumulation of tumor-derived adhesion factor in tumor blood vessels and in capillary tube-like structures of cultured vascular endothelial cells. *Proc Natl Acad Sci U S A* **93**, 8384–8389 (1996).
273. Parigi, S. M. *et al.* The spatial transcriptomic landscape of the healing mouse intestine following damage. *Nat Commun* **13**, 828 (2022).
274. Pilewski, J. M., Liu, L., Henry, A. C., Knauer, A. V. & Feghali-Bostwick, C. A. Insulin-Like Growth Factor Binding Proteins 3 and 5 Are Overexpressed in Idiopathic Pulmonary Fibrosis and Contribute to Extracellular Matrix Deposition. *The American Journal of Pathology* **166**, 399–407 (2005).
275. Mehta, B. *et al.* Op0223 Distinguishing Osteoarthritis and Rheumatoid Arthritis Synovium with Machine Learning Using Automated Cell Density and Pathologist Scores. *Annals of the Rheumatic Diseases* **81**, 146–147 (2022).
276. Trentham, D. E. Collagen arthritis as a relevant model for rheumatoid arthritis. evidence pro and con. *Arthritis & Rheumatism* **25**, 911–916 (1982).
277. Christensen, A. D., Haase, C., Cook, A. D. & Hamilton, J. A. K/BxN Serum-Transfer Arthritis as a Model for Human Inflammatory Arthritis. *Front. Immunol.* **7**, (2016).
278. Brackertz, D., Mitchell, G. F. & Mackay, I. R. Antigen-induced arthritis in mice. *Arthritis & Rheumatism* **20**, 841–850 (1977).
279. Sheng, W. *et al.* Multiplex Immunofluorescence: A Powerful Tool in Cancer Immunotherapy. *International Journal of Molecular Sciences* **24**, (2023).
280. Zahn-Zabal, M. *et al.* The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research* **48**, D328–D334 (2020).
281. Uray, K., Mizuno, M., Inazu, T., Goto, K. & Hudecz, F. The effect of glycosylation on the antibody recognition of a MUC2 mucin epitope. *Peptide Science* **102**, 390–395 (2014).
282. Hooper, A. T. *et al.* Angiomodulin is a specific marker of vasculature and regulates VEGF-A dependent neo-angiogenesis. *Circ Res* **105**, 201–208 (2009).

283. Komiya, E. *et al.* Angiomodulin, a marker of cancer vasculature, is upregulated by vascular endothelial growth factor and increases vascular permeability as a ligand of integrin αvβ3. *Cancer Med* **3**, 537–549 (2014).
284. Brooks, P. C., Clark, R. A. F. & Cheresh, D. A. Requirement of Vascular Integrin αvβ3 for Angiogenesis. *Science* **264**, 569–571 (1994).
285. Sarrazy, V. *et al.* Integrins αvβ5 and αvβ3 promote latent TGF-β1 activation by human cardiac fibroblast contraction. *Cardiovasc Res* **102**, 407–417 (2014).
286. Antonov, A. S. *et al.* αVβ3 Integrin Regulates Macrophage Inflammatory Responses via PI3 Kinase/Akt-Dependent NF-κB Activation. *J Cell Physiol* **226**, 469–476 (2011).
287. Insulin-like growth factor binding protein 5 induces skin fibrosis: A novel murine model for dermal fibrosis. doi:10.1002/art.22084.
288. Cecchini, M. J., Hosein, K., Howlett, C. J., Joseph, M. & Mura, M. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. *Respiratory Research* **19**, 153 (2018).
289. Kim, K. S. *et al.* Induction of Cellular Senescence by Insulin-like Growth Factor Binding Protein-5 through a p53-dependent Mechanism. *MBoC* **18**, 4543–4552 (2007).
290. Yasuoka, H. *et al.* Insulin-like growth factor-binding protein-5 induces pulmonary fibrosis and triggers mononuclear cellular infiltration. *Am J Pathol* **169**, 1633–1642 (2006).
291. Yasuoka, H. *et al.* The Fibrotic Phenotype Induced by IGFBP-5 Is Regulated by MAPK Activation and Egr-1-Dependent and -Independent Mechanisms. *The American Journal of Pathology* **175**, 605–615 (2009).
292. Kuo, H.-J., Maslen, C. L., Keene, D. R. & Glanville, R. W. Type VI Collagen Anchors Endothelial Basement Membranes by Interacting with Type IV Collagen *. *Journal of Biological Chemistry* **272**, 26522–26529 (1997).
293. Mereness, J. A. *et al.* Collagen VI Deficiency Results in Structural Abnormalities in the Mouse Lung. *Am J Pathol* **190**, 426–441 (2020).
294. Sabatelli, P. *et al.* Collagen VI deficiency affects the organization of fibronectin in the extracellular matrix of cultured fibroblasts. *Matrix Biology* **20**, 475–486 (2001).

295. Fernandes, N. R. J. *et al.* CD4+ T Cell Interstitial Migration Controlled by Fibronectin in the Inflamed Skin. *Front Immunol* **11**, 1501 (2020).
296. Wolf, K., Hu, H., Isaji, T. & Dardik, A. Molecular identity of arteries, veins and lymphatics. *J Vasc Surg* **69**, 253–262 (2019).
297. Wolf, J. & Carsons, S. E. Distribution of type VI collagen expression in synovial tissue and cultured synoviocytes: relation to fibronectin expression. *Ann Rheum Dis* **50**, 493–496 (1991).
298. Chen, P. *et al.* Collagen VI regulates peripheral nerve regeneration by modulating macrophage recruitment and polarization. *Acta Neuropathol* **129**, 97–113 (2015).
299. Wishart, A. L. *et al.* Decellularized extracellular matrix scaffolds identify full-length collagen VI as a driver of breast cancer cell invasion in obesity and metastasis. *Science Advances* **6**, eabc3175 (2020).
300. Lutter, L. *et al.* Human regulatory T cells locally differentiate and are functionally heterogeneous within the inflamed arthritic joint. *Clin Transl Immunology* **11**, e1420 (2022).
301. Cai, B. *et al.* MerTK cleavage limits proresolving mediator biosynthesis and exacerbates tissue inflammation. *Proc Natl Acad Sci U S A* **113**, 6526–6531 (2016).
302. Zizzo, G., Hilliard, B. A., Monestier, M. & Cohen, P. L. Efficient Clearance of Early Apoptotic Cells by Human Macrophages Requires M2c Polarization and MerTK Induction. *The Journal of Immunology* **189**, 3508–3520 (2012).
303. Choi, J.-Y. *et al.* Upregulation of Mer receptor tyrosine kinase signaling attenuated lipopolysaccharide-induced lung inflammation. *J Pharmacol Exp Ther* **344**, 447–458 (2013).
304. Domislovic, V. *et al.* Inflammatory Biomarkers of Extracellular Matrix Remodeling and Disease Activity in Crohn's Disease and Ulcerative Colitis. *J Clin Med* **11**, 5907 (2022).
305. Veidal, S. S. *et al.* MMP Mediated Degradation of Type VI Collagen Is Highly Associated with Liver Fibrosis – Identification and Validation of a Novel Biochemical Marker Assay. *PLoS One* **6**, e24753 (2011).
306. Mittelbrunn, M. & Sánchez-Madrid, F. Integrin Alpha 4 (Itga 4). in *Encyclopedia of Signaling Molecules* (ed. Choi, S.) 945–949 (Springer, New York, NY, 2012). doi:10.1007/978-1-4419-0461-4_143.

307. Arnaout, M. A. Biology and structure of leukocyte β 2 integrins and their role in inflammation. *F1000Res* **5**, F1000 Faculty Rev-2433 (2016).
308. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
309. Whittaker, C. A. & Hynes, R. O. Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. *Mol Biol Cell* **13**, 3369–3387 (2002).
310. Janssens, R., Struyf, S. & Proost, P. The unique structural and functional features of CXCL12. *Cell Mol Immunol* **15**, 299–311 (2018).
311. Ozols, M. *et al.* Predicting Proteolysis in Complex Proteomes Using Deep Learning. *Int J Mol Sci* **22**, 3071 (2021).
312. Buckley, C. D. *et al.* Identification of a phenotypically and functionally distinct population of long-lived neutrophils in a model of reverse endothelial migration. *Journal of Leukocyte Biology* **79**, 303–311 (2006).
313. Chavakis, T. *et al.* The Junctional Adhesion Molecule-C Promotes Neutrophil Transendothelial Migration in Vitro and in Vivo*. *Journal of Biological Chemistry* **279**, 55602–55608 (2004).
314. Qian, B.-Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* **475**, 222–225 (2011).
315. Bugatti, S. *et al.* High expression levels of the B cell chemoattractant CXCL13 in rheumatoid synovium are a marker of severe disease. *Rheumatology* **53**, 1886–1895 (2014).
316. Worley, J. R. *et al.* Metalloproteinase Expression in PMA-stimulated THP-1 Cells: Effects of peroxisome proliferator-activated receptor-γ (PPARγ) agonists and 9-cis-retinoic acid. *Journal of Biological Chemistry* **278**, 51340–51346 (2003).
317. Heidinger, M., Kolb, H., Krell, H.-W., Jochum, M. & Ries, C. Modulation of autocrine TNF-α-stimulated matrix metalloproteinase 9 (MMP-9) expression by mitogen-activated protein kinases in THP-1 monocytic cells. **387**, 69–78 (2006).
318. Myers, K. V., Amend, S. R. & Pienta, K. J. Targeting Tyro3, Axl and MerTK (TAM receptors): implications for macrophages in the tumor microenvironment. *Mol Cancer* **18**, 94 (2019).

319. Shibata, T. *et al.* Axl receptor blockade ameliorates pulmonary pathology resulting from primary viral infection and viral exacerbation of asthma. *J Immunol* **192**, 3569–3581 (2014).
320. Zahuczky, G., Kristóf, E., Majai, G. & Fésüs, L. Differentiation and Glucocorticoid Regulated Apopto-Phagocytic Gene Expression Patterns in Human Macrophages. Role of Mertk in Enhanced Phagocytosis. *PLOS ONE* **6**, e21349 (2011).
321. Cohen, P. L. *et al.* Delayed apoptotic cell clearance and lupus-like autoimmunity in mice lacking the c-mer membrane tyrosine kinase. *J Exp Med* **196**, 135–140 (2002).
322. Camenisch, T. D., Koller, B. H., Earp, H. S. & Matsushima, G. K. A novel receptor tyrosine kinase, Mer, inhibits TNF-alpha production and lipopolysaccharide-induced endotoxic shock. *J Immunol* **162**, 3498–3503 (1999).
323. Bosurgi, L. *et al.* Paradoxical role of the proto-oncogene Axl and Mer receptor tyrosine kinases in colon cancer. *Proceedings of the National Academy of Sciences* **110**, 13091–13096 (2013).
324. Akitake-Kawano, R. *et al.* Inhibitory role of Gas6 in intestinal tumorigenesis. *Carcinogenesis* **34**, 1567–1574 (2013).
325. Liu, J. *et al.* Osteopontin regulation of MerTK+ macrophages promotes Crohn's disease intestinal fibrosis. *iScience* **27**, (2024).
326. Tanim, K. M. *et al.* MERTK Inhibition as a Targeted Novel Cancer Therapy. *Int J Mol Sci* **25**, 7660 (2024).
327. Xu, Y., Ying, L., Lang, J. K., Hinz, B. & Zhao, R. Modeling mechanical activation of macrophages during pulmonary fibrogenesis for targeted anti-fibrosis therapy. *Science Advances* **10**, eadj9559 (2024).
328. Liu, Y. *et al.* Mechanical stimulation improves rotator cuff tendon-bone healing via activating IL-4/JAK/STAT signaling pathway mediated macrophage M2 polarization. *Journal of Orthopaedic Translation* **37**, 78–88 (2022).
329. Cai, G. *et al.* Piezo1-mediated M2 macrophage mechanotransduction enhances bone formation through secretion and activation of transforming growth factor-β1. *Cell Prolif* **56**, e13440 (2023).

330. Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R. & Colombatti, A. Structural and functional features of the alpha 3 chain indicate a bridging role for chicken collagen VI in connective tissues. *Biochemistry* **29**, 1245–1254 (1990).
331. McFarland, H. I., Nahill, S. R., Maciaszek, J. W. & Welsh, R. M. CD11b (Mac-1): a marker for CD8+ cytotoxic T cell activation and memory in virus infection. *J Immunol* **149**, 1326–1333 (1992).
332. Christensen, J. E., Andreasen, S. Ø., Christensen, J. P. & Thomsen, A. R. CD11b expression as a marker to distinguish between recently activated effector CD8+ T cells and memory cells. *International Immunology* **13**, 593–600 (2001).
333. Fiorentini, S. *et al.* CD11b Expression Identifies CD8+CD28+ T Lymphocytes with Phenotype and Function of Both Naive/Memory and Effector Cells1. *The Journal of Immunology* **166**, 900–907 (2001).
334. Abdel-Salam, B. K. A. & Ebaid, H. Expression of CD11b and CD18 on polymorphonuclear neutrophils stimulated with interleukin-2. *Cent Eur J Immunol* **39**, 209–215 (2014).
335. Al-Roub, A. *et al.* TNFα induces matrix metalloproteinase-9 expression in monocytic cells through ACSL1/JNK/ERK/NF-kB signaling pathways. *Sci Rep* **13**, 14351 (2023).
336. Laforge, M. *et al.* Tissue damage from neutrophil-induced oxidative stress in COVID-19. *Nat Rev Immunol* **20**, 515–516 (2020).
337. Cahilog, Z. *et al.* The Role of Neutrophil NETosis in Organ Injury: Novel Inflammatory Cell Death Mechanisms. *Inflammation* **43**, 2021–2032 (2020).
338. Martinez, V. G. *et al.* Fibroblastic Reticular Cells Control Conduit Matrix Deposition during Lymph Node Expansion. *Cell Rep* **29**, 2810-2822.e5 (2019).
339. Britzen-Laurent, N., Weidinger, C. & Stürzl, M. Contribution of Blood Vessel Activation, Remodeling and Barrier Function to Inflammatory Bowel Diseases. *Int J Mol Sci* **24**, 5517 (2023).
340. Paleolog, E. M. The vasculature in rheumatoid arthritis: cause or consequence? *Int J Exp Pathol* **90**, 249–261 (2009).
341. Prechl, J. & Czirják, L. The endothelial deprotection hypothesis for lupus pathogenesis: the dual role of C1q as a mediator of clearance and regulator of endothelial permeability. *F1000Res* **4**, 24 (2015).

342. Watanabe, R. *et al.* Matrix Metalloprotease-9 (MMP-9)-Producing Monocytes Enable T Cells to Invade the Vessel Wall and Cause Vasculitis. *Circ Res* **123**, 700–715 (2018).
343. Itoh, T. *et al.* The Role of Matrix Metalloproteinase-2 and Matrix Metalloproteinase-9 in Antibody-Induced Arthritis. *The Journal of Immunology* **169**, 2643–2647 (2002).
344. Gossage, D. L. *et al.* Phase 1b Study of the Safety, Pharmacokinetics, and Disease-related Outcomes of the Matrix Metalloproteinase-9 Inhibitor Andecaliximab in Patients With Rheumatoid Arthritis. *Clinical Therapeutics* **40**, 156-165.e5 (2018).
345. Leeming, D. J. *et al.* Endotrophin, a pro-peptide of Type VI collagen, is a biomarker of survival in cirrhotic patients with hepatocellular carcinoma. *Hepat Oncol* **8**, HEP32 (2020).