Defining Fibroblast niches in Colorectal Cancer

Helen Elizabeth Jane Tyrrell



University of Oxford

Somerville College

Kennedy Institute of Rheumatology, NDORMS

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

Abstract

Colorectal cancer (CRC) is a heterogenous disease and high stromal content and specifically fibroblasts have been associated with advanced stage and poor prognosis. Fibroblasts are versatile cells that can alter from quiescent cells to myofibroblasts, that secrete extracellular matrix components and become contractile in situations such as wound healing, or alter to inflammatory fibroblasts, that secrete cytokines and chemokines to recruit immune cells. These phenotypes have been identified across multiple types of cancer through scRNAseq, and the key cytokines driving these clusters are hypothesised as TGFβ for myofibroblastic cancer associated fibroblasts (myCAFs) and IL-1 for inflammatory CAFs (iCAFs). However, few studies have identified them *in situ* in CRC and no fibroblast subsets have been conclusively associated with prognosis in this disease. It is also unknown whether the fibroblast clusters seen on scRNAseq represent permanent differentiation states or temporary responses to stimulation.

Here we use multiplex imaging of 60 rectal cancer sections to identify myCAF-like and iCAF-like stromal cells *in situ*. We have shown differences in their abundance with the Consensus Molecular subtypes (CMS), with iCAFs being found in CMS3 and myCAF-like stromal cells in CMS4. We have also stained adjacent slides with a panel to look for immune subsets and found greater numbers of both T-cells and neutrophils in CMS1 disease. We have found that the iCAF-like stromal cells are found in close association with immune cells, particularly neutrophils, consistent with the hypothesis of IL-1 driving this phenotype. We have also confirmed that treating primary stromal cells *in vitro* with IL-1β gives them an inflammatory phenotype that allows them to map more closely to iCAF clusters seen in scRNAseq. The myCAF-like stromal cells were identified deep within stromal tracts. We have also identified a CD56+ stromal subset, found surrounding the tumour nests and associated with CMS2 CRC.

We also hypothesised that fibroblasts from CRC tumours may be hyper-responsive to these key cytokines, TGFβ and IL-1β, compared to the matched normal adjacent fibroblasts, consolidating

pathological phenotypes. However, we found no evidence that once *in vitro* there were any differences in the stromal cells from tumour and normal adjacent, either at baseline or in terms of response to cytokine stimulation. This suggests that the CAF clusters seen are temporary activation states depending on the micro-environment the cells were exposed to immediately prior to sequencing.

Additionally, we hypothesised that TGFβ1, TGFβ2 and TGFβ3, may be having differential effects on fibroblasts. However, on bulk RNAseq we found no evidence that the transcriptional response to TGFβ stimulation is of a different character or magnitude between the three isoforms.

Acknowledgments

I would like to express my thanks to my primary supervisor, Prof Fiona Powrie, for giving me the opportunity to study for a PhD, and her guidance throughout this project. Also to my co-supervisor Prof Tim Maughan for his expertise in the colorectal cancer field and the opportunity to benefit from working with the S:CORT team.

I would also like to express my huge gratitude to my two post-doc supervisors, Dr Beth Mann and Dr Mathilde Pohin. Beth was always available to help me when I joined the lab, patiently teaching me all the techniques without which I never would have been able to get started. Mathilde stepped in when Beth left and has provided in depth understanding and direction in the fibroblast field. She has also looked through numerous drafts of this thesis!

I would like to thank my thesis committee, Prof Kim Midwood and Prof Eileen Parkes for their positive responses to my work and suggestions for improvement.

I would like to thank Dr Nick Ilott for spending so much of his time teaching me how to analyse my RNA sequencing results in R, a task that would have been impossible without his guidance. Also, Amit Halkhoree for performing cell deconvolution, again, something I could not have managed myself. Alistair Easton for his help and advice using HALO image analysis software. Jonathan Webber for teaching me how to use the LSR and to compensate properly.

I would also like to thank everyone in the Powrie lab for being such a supportive, helpful and fun team to work with: Claire, Alice, Ffion, Ekta, Declan, Rebecca, India, Alina, Sandi, Sam, Rachel, Avery, Alex, Elizete, Emily, Henry, Kelsey, Rahul, Yisu, Cornelia, Dehui, Matthias, Sumaiya, Tarrion, Xinyue, Juan, Caroline and Gemma.

I would like to thank CRUK for providing the funding for my PhD through the Clinical Research Training Fellowship.

Finally, I would like to thank my family. My husband Andrew for his constant love and support, and daughters Elsie and Jenny, who were only 3 and 10 months when I started this PhD, for bringing so much fun and happiness. Also my parents, for their unwavering faith that I could achieve this, and their hours of dedicated childcare so that I could work.

Contributions

- Dr Leticia Campo performed multiplex staining on cases from the S:CORT study.

- The Kennedy Histology Facility for paraffin embedding of CRC samples.

- Dr Nicholas Ilott performed initial processing and mapping of bulk RNAseq data. He also performed deconvolution of my bulk RNAseq data using the publicly available Pelka *et. al.* scRNAseq dataset using SingleR.

- Amit Halkhoree performed deconvolution of my bulk RNAseq data to the publicly available Pelka *et. al.* scRNAseq dataset using MuSiC.

Declaration of Authorship

I hereby declare that the data presented in this thesis is my own work, except where contributions made to the research are explicitly acknowledged, above or in the text of the thesis. To the best of my knowledge this thesis contains no materials published or written by another person, or which have been accepted for the award of any other degree or diploma at the University of Oxford, or any other educational institution.

Table of Contents

1 Introduction 1

1.1 Colorectal Cancer Overview 1

1.1.1 Incidence and Risk Factors 1

1.1.2 The healthy GI tract 1

1.1.3 CRC Staging 2

1.1.4 Genetics of CRC 3

1.1.4.1 Pathways to CRC development 3

1.1.4.2 Common sporadic mutations in CRC 4

1.1.4.3 Germline mutations leading to CRC 6

1.1.5 Stratification of CRC and CMS subtypes 6

1.1.5.1 Stratification using bulk RNA sequencing data 6

1.1.5.2 Stratification using bulk RNA sequencing of the neoplastic cells 8

1.1.5.3 Histological stratification of CRC 10

1.1.6 CRC Treatment 11

1.1.7 The TME in CRC 13

1.2 The Immune System in Colorectal Cancer 14

1.2.1 The conflicting role of the Immune System in Colorectal cancer 14

1.2.2 The Innate Immune system in Colorectal cancer 14

1.2.3 The adaptive immune system 16

1.2.3.1 B and T Lymphocytes 16

1.2.3.2 Co-stimulatory and co-inhibitory receptors 16

1.2.3.3 CD4+ T-cell subsets 18

1.3 Fibroblasts in CRC 19

1.3.1 CMS4 is a stromal signature 19

1.3.2 Subsets of fibroblasts in health 20

1.3.3 Stromal subsets in health beyond fibroblasts 21

1.3.4 Subsets of fibroblasts in disease 22

1.3.4.1 The Pelka et al CRC scRNAseq dataset 26

1.3.5 Identifying fibroblasts at the protein level 30

1.3.5.1 Interactions 31

1.3.6 Drivers of the fibroblast phenotypes 33

1.3.6.1 IL-1β production 33

1.3.6.2 IL-1β signalling 34

1.3.6.3 TGFβ expression in Colorectal Cancer 35

1.3.6.4 TGFβ signalling 35

1.3.6.5 TGFβ effects in Cancer ... *37*
1.3.6.6 Targeting TGFβ for cancer treatment ... *38*
1.3.6.7 The three Isoforms of TGFβ ... *39*
1.3.6.8 Fine-tuning the fibroblast phenotypes ... *42*
1.3.7 Fibroblasts across disease ... 43
1.3.7.1 Permanency of the Fibroblast subsets ... *43*
1.3.7.2 Fibroblast subsets across disease stage ... *46*
1.3.7.3 Fibroblast subsets and Prognosis ... *47*
1.4 Research hypotheses and experimental aims ... **50**
2 Materials and Methods ... **52**
2.1 Ethical approval ... **52**
2.2 Cell Isolation ... **52**
2.2.1 Digestion and expansion of fibroblasts from resection tissue ... 52
2.2.2 Extraction of PBMCs from cones ... 53
2.3 Cell line culture ... **53**
2.3.1 Cytokines and pharmacological inhibitors ... 53
2.3.2 Maintenance and stimulation of PBMCs ... 53
2.3.3 Maintenance and stimulation of fibroblasts and primary stromal cells ... 54
2.3.4 Maintenance and stimulation of hPL-PC pericytes ... 54
2.3.5 MTT test ... 54
2.4 Gene expression analysis ... **55**
2.4.1 RNA extraction and cDNA synthesis ... 55
2.4.2 RT-qPCR ... 55
2.4.3 RNA sequencing ... 56
2.5 Protein expression analysis ... **57**
2.5.1 BCA Assay ... 57
2.5.2 Western blotting ... 57
2.5.3 ELISA ... 58
2.5.4 Flow cytometry ... 58
2.6 In situ analysis ... **59**
2.6.1 IHC on FFPE tissue ... 59
2.6.2 Immunofluorescence ... 60
2.6.3 Immunofluorescence on Vectar Polaris ... 61
2.7 Statistical analysis ... **62**
3 Identifying stromal cell location within CRC ... **63**
3.1 Introduction ... **63**
3.2 Stromal subsets panel on the Grampian dataset ... **65**

3.2.1 Demographics of the 60 Grampian cases 65
3.2.2 Markers to Identify myCAFs and iCAFs 69
3.2.3 Image Analysis of the Fibroblast panel by Cell Segmentation 72
3.2.4 Repeat analysis by cell segmentation, removing PDPN 83
3.2.5 Image Analysis of the Stromal panel by Area Quantification 94
3.3 The Immune panel on the Grampian dataset 101
3.3.1 Image Analysis of the Immune Panel 101
3.3.2 Correlation of Immune panel with metadata 103
3.4 Correlations between the Fibroblast and Immune Panels 106
3.4.1 Correlations across the entire slide 106
3.4.2 Proximity analysis between immune cell subsets and fibroblast subsets 107
3.4.3 CD56 staining around the Tumour Nests 109
3.5 Conclusions from Grampian Chapter 112
3.6 Limitations and Future work 114
4 Cytokine response in Tumour and Normal Adjacent Stromal cells . 117
4.1 Introduction to Chapter 117
4.2 Characterisation of the Fibroblast and Pericyte cell lines 119
4.2.1 Characterisation of the Unstimulated cell lines 119
4.2.2 Characterisation of the cytokine response in fibroblast and pericyte cell lines 125
4.2.2.1 Dose Response and Time-Course 125
4.2.2.2 Response to IL-1β 127
4.2.2.3 Response to TGFβ1 130
4.2.2.4 Pathway analysis of cytokine response in cell lines 133
4.3 Characterisation of Primary Stromal Lines 135
4.3.1 Generation of Primary Stromal Cell lines 135
4.3.1.1 Processing of Patient samples 135
4.3.1.2 Indication of differential gene expression response between Tumour and Normal adjacent 136
4.3.2 Quality Control of RNA sequencing data 137
4.3.3 Clustering of the Primary Stromal cells 140
4.3.3.1 Overall clustering of samples 140
4.3.3.2 Comparing unstimulated primary stromal cells with cell lines 142
4.3.3.3 Comparing gene expression and pathways between Primary stromal cell clusters 145
4.3.3.4 Defining the identity of Primary stromal cell clusters 147
4.4 Primary Stromal Cell response to Cytokines 150
4.4.1 Comparing Unstimulated Tumour and Normal Adjacent Primary Stromal cells 150

4.4.2 Comparing Tumour and Normal Adjacent Stromal cells response to TGFβ 152
4.4.3 Comparing Tumour and Normal Adjacent Stromal cells response to IL-1β..... 162
4.5 The cytokine response in the Primary Stromal lines by Cluster and comparing Tumour and Normal adjacent 167
4.5.1 The cytokine response in primary stromal cells by Cluster 167
4.5.1.1 Clustering to Primary Stromal cells under all stimulation conditions 167
4.5.1.2 IL-1β stimulation of Primary Stromal cells 170
4.5.1.3 TGFβ1 stimulation of Primary Stromal cells 174
4.5.2 Comparing Cluster 1 Primary Stromal cells by Tumour Site 178
4.5.3 Responses in Individual Cell Lines 182
4.6 Mapping to Pelka scRNAseq 184
4.6.1 Manual correlation of DEGs with cytokine stimulation and Pelka scRNAseq cluster marker genes 184
4.6.2 Multi-subject Single-cell Deconvolution (MuSiC) 185
4.6.2.1 Mapping of all samples using MuSiC 185
4.6.2.2 Mapping of the 12 primary stromal lines in Cluster 1 to the scRNAseq using MuSiC 186
4.6.2.3 Identity of primary stromal cells and cytokine response 190
4.6.3 Mapping with the SingleR algorithm 192
4.7 Conclusions 194
4.8 Limitations and Future Work 196
5 Comparison of the TGFβ isoforms on PBMCs and fibroblasts 200
5.1 Introduction to TGFβ Isoforms 200
5.2 Isoform expression and CMS groupings 202
5.3 The effects of the 3 TGFβ isoforms on PBMCs 205
5.3.1 Proliferation of PBMCs 205
5.3.2 Activation of PBMCs 207
5.3.3 Checkpoint expression on T-cells 210
5.4 Comparison of the TGFβ isoforms on CCD18-Co fibroblasts. 211
5.5 Comparison of the Isoforms in Primary CRC cell lines 215
5.6 Conclusions from TGFβ Isoforms Comparison 219
5.7 Limitations and Future Work 220
6 Discussion 221
6.1 Summary 221
6.2 CD56+ CAFs 222
6.3 iCAFs 225
6.4 GREM1+ CAFs 226
6.5 myCAF-like stromal cells 227

6.6 Further Work 229
6.7 Implications for Treatment 230
6.8 Conclusions 230
7 References 231
8 Appendices 241
8.1 Appendix 1 241
8.2 Appendix 2 242

List of Abbreviations

ACTA2: Actin alpha 2
ACK: Ammonium-chloride-potassium
APC: Adenomatous Polyposis coli
APC: Antigen Presenting Cell
αSMA: Alpha smooth muscle actin
BCA: Bicinchoninic acid
BMP: Bone morphogenetic protein
BSA: Bovine serum albumin
BRAF: B-raf proto oncogene
CAF: Cancer associated fibroblast
CCL: CC chemokine ligand
CD: Cluster of Differentiation
CIMP: CpG island methylator phenotype
CIN: Chromosomal instability
CK: Cytokeratin
CMS: Consensus Molecular Subtype
COL: Collagen
CpG: 5'-C-phosphate-G-3'
CRC: Colorectal cancer
CRIS: CRC intrinsic subtype
CTLA4: Cytotoxic T-lymphocyte associated protein 4
CXCL: CXC chemokine ligand
DCC: Deleted in colorectal cancer
DEG: Differentially expressed gene
DFS: Disease free survival
DMSO: Dimethyl sulfoxide
DTT: Dithiothreitol
ECM: Extracellular Matrix
EDTA: Ethylendiaminetetraacetic acid
EGFR: Epidermal growth factor receptor
ELISA: Enzyme-linked immunosorbent assay
EMT: Epithelial to Mesenchymal Transition
EPCAM: Epithelial cellular adhesion molecule
FAP: Fibroblast activation protein alpha
FACS: Fluorescence activated cell sorting
FBS: Fetal bovine serum
FC: Fold change
FFPE: Formalin-fixed paraffin-embedded
FSP-1: Fibroblast-specific protein 1
GI tract: Gastrointestinal tract
GO: Gene Ontology
GREM1: Gremlin 1
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HER2: Human epidermal growth factor receptor 2
H&E: Haematoxylin & Eosin
HLA-DR: Human Leucocyte antigen – DR isotype
HRP: Horseradish peroxidase
IBD: Inflammatory bowel disease
iCAF: Inflammatory cancer associated fibroblast
ICMS: Intrinsic Consensus Molecular Subtype
IFN: Interferon
IGF2: Insulin-like growth factor 2
IHC: Immunohistochemistry
IL: Interleukin
KRAS: Kirsten rat sarcoma virus
MDSC: Myeloid Derived Suppressor Cell
MEMα: Minimum Essential Medium alpha
MHC: Major Histocompatibility Complex
MLH1: MutL protein homolog 1
MMP: Matrix metalloproteinase
MMR: Mismatch repair
MSH: MutS homolog
MSI: Microsatellite instability
MTT: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MuSiC: Multi-subject Single-cell Deconvolution
MUTYH: MutY DNA glycosylase
Myc: Myelocytomatosis oncogene
myCAF: Myofibroblastic cancer associated fibroblast
N.adj: Normal adjacent
NCAM: Neural cell adhesion molecule
NK: Natural Killer
NTHL-1: NTH like DNA glycosylase 1
NTRK: Neurotrophic tyrosine receptor kinase
PBMC: Peripheral blood mononuclear cell
PBS: Phosphate buffered saline
PCA: Principal component analysis
PD-1: Programmed cell death protein 1
PDGFR: Platelet derived growth factor receptor
PD-L: Programmed cell death ligand
PDPN: Podoplanin
PDS: Pathway-derived subtypes
PMS2: Postmeiotic segregation increased 2
POLD1: DNA polymerase delta
POLE: DNA polymerase epsilon
P/S: Penicillin/streptomycin
PTEN: Phosphatase and TENsin homolog deleted on chromosome 10
RGS5: Regulator of G-protein signalling 5
RIPA buffer: Radioimmunoprecipitation assay buffer
RNAseq: RNA sequencing

RORγt: RAR-related orphan receptor gamma
RPMI: Roswell Park Memorial Institute media
RSPO3: R-spondin 3
RT: Room temperature
RT-qPCR: Real Time quantitative Polymerase Chain Reaction
S:CORT: Stratification in Colorectal Cancer
scRNAseq: Single cell RNA sequencing
SMAD: Portmanteau of Caenorhabditis elegans SMA ("small" worm phenotype) and MAD family ("Mothers Against Decapentaplegic") of genes in Drosophil
TBET: T-box transcription factor
TBRS: TGFβ response signature
TCR: T-cell receptor
TGFβ: Transforming growth factor beta
Th: Helper T-cell
TME: Tumour Microenvironment
TNM: Tumour-Node-Metastases
Wnt: Portmanteau of gene Wingless and Int-1

1 Introduction

1.1 Colorectal Cancer Overview

1.1.1 Incidence and Risk Factors

Colorectal cancer (CRC) is currently the third most commonly diagnosed cancer globally and the second leading cause of cancer death[1] and therefore is a major public health concern. Incidence is increasing in younger populations under the age of 50, and there is an associated increase in CRC mortality in this age group[2]. There has been a decline in incidence and mortality in older age groups, primarily due to the introduction of screening programmes[3]. Development of CRC is due to a combination of genetics and environmental risk factors although most cases are sporadic. The rising incidence in younger people may be due to the many lifestyle factors that increase the risk of developing this disease at any age; increased consumption of red or processed meat or processed sugar, poor fibre intake, smoking, obesity and lack of exercise[2]. Colorectal cancers also occur in the context of chronic inflammation; those with ulcerative colitis have a 2.4 fold increased risk of developing CRC[4].

1.1.2 The healthy GI tract

The gastrointestinal (GI) tract is a continuous tube running from the mouth, to the oesophagus, stomach, small intestine, colon, rectum and anus. In the colon there is a single layer of epithelial cells lining the lumen and it is these epithelial cells that become neoplastic in CRC. In health, the epithelial layer has many indentations or crypts, at the base of these are stem cells, that continuously renew the layer. The stem cells give rise to transit amplifying cells that produce the intestinal epithelial cells that gradually move up to the surface where they are continuously shed. The intestinal epithelial cells comprise both the absorptive enterocytes and a range of other cell types; Paneth cells, goblets cell, enteroendocrine cells and tuft cells, that secrete mucus and antimicrobial peptides, help to maintain the stem cell niche and are involved in immune and metabolic regulation. The stem cells in the crypt continuously replicate due to the

production of Wnt (a Portmanteau of gene Wingless and Int-1) and epidermal growth factor (EGF), by the fibroblasts near the crypt. There is a gradient of decreasing Wnt going further up the crypt and increasing bone morphogenetic protein (BMP), which encourages differentiation and prevents further replication of the epithelial cells. The fibroblasts in the crypt also produce R-spondin and Gremlin 1 to inhibit BMP signalling there[5].

1.1.3 CRC Staging

The intestinal epithelial cells lie on a basement membrane, below which is the lamina propria. This layer contains blood vessels, lymphatics, nerves and tertiary lymphoid structures. There are therefore many different cells types present: immune cells, endothelial cells, pericytes, neurones, glial cells and fibroblasts. Below this is a thin muscle layer called the muscularis mucosa, then the submucosa, and then a thicker layer of muscles termed the muscularis propria, and then the serosa, outside of which is the peritoneal cavity.

CRC staging is based on the TNM (tumour-node-metastases) system. The T (tumour) staging of colorectal cancer is based on depth of invasion of these layers, with a Tis remaining within the mucosa, a T1 tumour entering the submucosa, a T2 tumour invading the muscularis propria, a T3 invading the serosa and a T4 reaching other organs or structures **(Fig 1.1.1.3)**. The N (node) staging describes the lymph node involvement; cancers are described as N0 when no lymph nodes are involved, N1 when 1-3 lymph nodes are affected, and N2 when 4 or more lymph nodes contain neoplastic cells. The M (metastasis) stage is M0 if there are no distant metastases and M1 if metastases are present[6]. Occasionally staging is given by the older, and simpler, Duke's staging system, where a Duke's A denotes tumours that have not reached the muscularis propria, Duke's B those that have invaded the muscle, Duke's C if lymph nodes are involved and sometimes the term Duke's D is used if distant metastases are present[7].

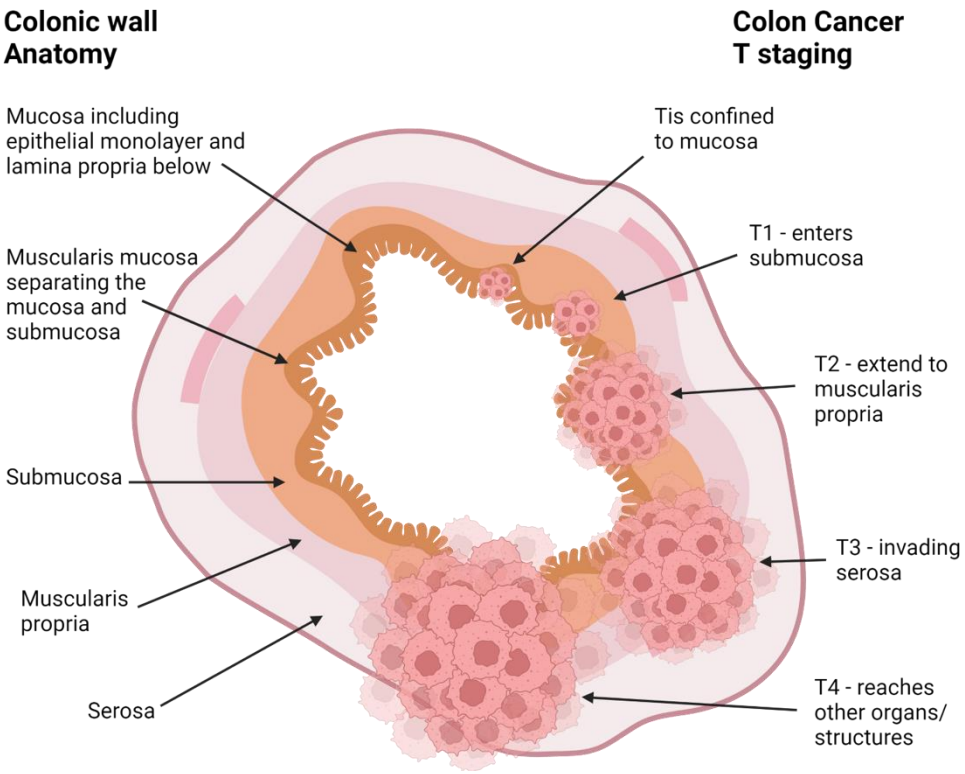


Fig 1.1.1.3 The layers of the colonic wall and colorectal cancer staging
The mucosa (dark orange) including the epithelial monolayer sitting on the basement membrane and the lamina propria. Below this is a thin layer of muscle, the muscularis mucosa, separating the mucosa from the submucosa, in light orange. The muscularis propria layer is shown in purple, with the serosa below in pale purple.
T staging of colorectal cancer is demonstrated, with Tis, a carcinoma in situ being confined to the mucosa, T1 entering the submucosa, T2 invading the muscularis propria, T3 reaching the serosa and T4 invading adjacent organs and tissues.

1.1.4 Genetics of CRC

1.1.4.1 Pathways to CRC development

Colorectal cancer has long been described as caused by a stepwise series of mutations in the neoplastic cell, where each step gives a selective advantage to the progeny of that cell, leading to clonal expansion and gradual development from an adenoma through to an invasive carcinoma. Often the earliest change leading to polyp formation is loss of the adenomatous polyposis coli (APC) gene, and then later changes include mutations or deletions of other tumour suppressor genes such as TP53 and DCC, and activating mutations or amplifications of

oncogenes such as Kirsten rat sarcoma virus (KRAS) and c-Myelocytomatosis oncogene (c-Myc)[8].

.

About 20% of CRCs follow an alternative pathway of development, with lesions growing from serrated polyps, rather than the classical adenoma[9, 10]. KRAS and B-raf proto oncogene (BRAF) mutations, Microsatellite instability (MSI) and the 5'-C-phosphate-G-3' (CpG) island methylator phenotype, contribute to tumorigenesis by this pathway. TP53 mutations are often found earlier in the course of mutagenesis, and APC loss later, than in the classical pathway. Serrated polyps have serrated epithelium, clear and abundant cytoplasm, and an absence of necrosis. They include hyperplastic polyps, sessile serrated adenoma and traditional serrated adenoma.

Contrary to the series of stepwise mutations described in the adenoma to carcinoma sequence, recent studies have postulated a theory of neutral evolution, where all driver mutations are present in the first malignant cell, and intra- tumour heterogeneity simply represents passenger mutations that are not altering disease course[11, 12]

1.1.4.2 Common sporadic mutations in CRC

APC is one of the most frequently mutated tumour suppressor genes in sporadic CRC as well as being the gene affected in the inherited condition familial adenomatous polyposis (FAP). APCs normal role is to sequester B-catenin in the cytoplasm, preventing it entering the nucleus. In health, Wnt signalling in the crypts leads to degradation of APC, allowing B-catenin to enter the nucleus and bind to the transcription factor TCF. This leads to transcription of genes triggering replication of the epithelial stem cells. Where APC is lost the Wnt pathway is constitutively active and uncontrolled replication occurs[13].

RAS and RAF are both in the MAPK/ERK pathways and BRAF and KRAS mutations are seen to be mutually exclusive. In health, EGF binding to the Epidermal growth factor receptor (EGFR) leads

to activation of RAS, which activates RAF, which is a kinase that phosphorylates MEK and finally mitogen-activated proteins kinase (MAPK)/extracellular signal-regulated kinases(ERK) is activated. These then activate transcription factors such as c-Myc and cyclin dependant kinases to promote progression into the S-phase of the cell cycle[14]. Therefore, constitutive activation of KRAS or BRAF, and therefore the MAPK pathway, leads to cell survival, proliferation and invasion and metastases.

Myc levels are also increased through other pathways in addition to MAPK/ERK, such as Wnt, PI3K/mammalian target of rapamycin (mTOR) and Shh signalling[15]. The PI3K/mTOR pathway is also activated in multiple cancer types and many signals such as EGF, Shh, IGF-1 and insulin can lead to PI3K signalling. Once PI3K is active it phosphorylates and activates AKT, which activates the cAMP response element (CREB) and mTOR and inhibits p27, which promote cell growth and replication. PTEN inhibits this pathway, and therefore loss of PTEN can lead to uncontrolled replication.

TP53 is a tumour suppressor which is mutated in over half of human cancers. It encodes p53, which prevents the cell entering either the S-phase or the mitosis phase of the cell cycle if DNA damage is present, and can trigger apoptosis if damage cannot be repaired[16].

Deleted in colorectal carcinoma (DCC) acts as a tumour suppressor, as when it is not bound to its ligand, netrin-1, it has a role in promoting apoptosis. In the healthy villi, netrin-1 is present at the crypts, preventing cell death, but there is a gradient of reducing netrin-1 going up the villi, such that at the tips unbound DCC triggers apoptosis and shedding. When DCC is not present this does not occur and cells can accumulate[17].

1.1.4.3 Germline mutations leading to CRC

Most CRCs are sporadic, although in patients under fifty, 26% have a family history of CRC in a first degree relative, even when a monogenic cause cannot be identified[18], suggesting common, low-risk genetic changes are important in CRC development. The proportion of colorectal cancer patients with a high-risk germline mutation is predicted to be 3-5%[19]. The most common germline mutations increasing CRC risk are in the mismatch repair (MMR) genes: MutL protein homolog 1 (MLH1), MutS homolog 2 (MSH2), MutS homolog 6 (MSH6) or Postmeiotic segregation increased 2 (PMS2). Germline mutations in these genes are known as Lynch syndrome or Hereditary Non-polyposis Colorectal cancer.

Colonic polyposis syndromes, where large numbers of polyps are seen usually from a young age, are rarer and caused by germline mutations in APC (known as familial adenomatous polyposis); MutY DNA glycosylase (MUTYH); BMPR1A; SMAD4 (SMAD is a Portmanteau of Caenorhabditis elegans SMA ("small" worm phenotype) and MAD family ("Mothers Against Decapentaplegic") of genes in Drosophil; Phosphatase and TENsin homolog deleted on chromosome 10 (PTEN); or STK11[19].

1.1.5 Stratification of CRC and CMS subtypes

Outcomes for patients with CRC vary enormously, even amongst those with the same stage disease. There have therefore been many attempts to stratify CRC, both to provide prognostic information, but also to guide treatment choices.

1.1.5.1 Stratification using bulk RNA sequencing data

One of the most commonly described methods of classification in research is consensus molecular subtype (CMS)[20]. This was originally based on combining subgroups from several, separate bulk RNA-sequencing studies. It has 4 categories: CMS1 (MSI-H or Immune) is associated with large amounts of immune infiltrate, is often seen in right sided disease, and is

associated with BRAF mutations and MSI. They are more common in females, and have undifferentiated histopathological grading and poor prognosis after relapse. CMS2 (Canonical) is associated with canonical APC mutations, Wnt signalling and Myc pathway activation. These tumours are often 'immune deserts' with very few immune cells in the tumour. They are found more often in the left colon and rectum and have a better prognosis. CMS3 (Metabolic) is associated with metabolic changes and these cancers often harbour KRAS mutations. KRAS increases glucose uptake, glycolysis and the pentose phosphate pathway, which allows rapid proliferation and the ability to survive relatively hypoxic environments. It also upregulates synthesis of some amino acids and fatty acids, protecting neoplastic cells against metabolic stress[21]. CMS4 (Mesenchymal) disease is associated with large amounts of stroma, particularly fibroblast cells, and high transforming growth factor beta (TGFβ) levels, increased angiogenesis and complement activation. These tumours have worse outcomes than the other three CMS subtypes, with higher rates of recurrence after treatment of localised disease and poorer survival. These tumours often tend to be immune excluded, there are immune cells present around the periphery of the tumour, but they are either unable to infiltrate or have become anergic to the tumour. Patients with these tumours therefore represent those with great clinical need and are a group where there is the possibility of harnessing an immune response against the tumour.

Another study suggested that in all tumours the neoplastic cells were either an underlying CMS2 or an underlying CMS 1/3, referred to as iCMS2 and iCMS3 respectively, and either of these 2 groups could become stroma high[22]. This is supported by the fact that there do not seem to be specific genetic changes associated with CMS4.

In reality many tumours do not fall neatly into one of the CMS subtypes, with intra-tumour heterogeneity meaning that different regions of the same tumour may resemble different CMS subgroups. One study found that 55% of tumours may overlap between CMS groups[23] and that

tumours with poor alignment to a specific CMS group have poorer prognosis than those clearly belonging to an individual group. Khaliq et al suggested that all CRC tumours lie on a transcriptomic continuum, rather than falling into discrete subgroups[24]. Metastatic disease may not always resemble the CMS subgroup of the primary, with metastases more often being classified as CMS 2 or 4[25]. This is likely to be an effect of the complex relationship between the tumour cells and the microenvironment. Attribution of a tumour to a specific CMS group may also be affected by tissue sampling, the invasive front of a tumour is more likely to be called CMS4 where more of the surrounding stroma is inadvertently included in the analysis[26].

More recently, CRC has also been stratified by performing pathway analysis on bulk transcriptional data to identify three pathway-derived subtypes (PDS). PDS2 overlaps with CMS1 and 4 and has increased immune and stromal infiltrate and increased expression of pathways involved in inflammation, immune signalling, epithelial to mesenchymal transition (EMT) and TGFβ signalling. PDS1 and PDS3 both overlap with CMS2 and CMS3. PDS1 is characterised by having highly proliferative stem-like epithelial cells. PDS3 was associated with low Ki67, well differentiated epithelial cells and expressed few pathways considered to be cancer hallmarks. These PDS3 tumours had poorer outcomes, more similar to that of CMS4 tumours[27].

1.1.5.2 Stratification using bulk RNA sequencing of the neoplastic cells

Since the original CMS classification other methods have been used to assign tumours to the same 4 groups. The intrinsic (ICMS) aims to classify tumours to the original four CMS groups based purely on the signatures coming from the neoplastic cells themselves, therefore allowing profiling of tumour cell lines, organoids and pre-clinical xenograph models[28].

Similar to the ICMS, another classification system that utilises the transcriptional profile of only the neoplastic cells is the CRC intrinsic subtype (CRIS) classification[29]. It was generated using patient derived xenographs, where human cancer cells were implanted into animals, so that only the neoplastic cells would be human and any infiltrating stromal or immune cells from the

host. Human-specific primers were then employed for bulk RNA-sequencing. CRIS A and B were associated with right sided tumours, more mucinous lesions, and CpG island methylation. CRIS A particularly had more MSI high disease and KRAS and BRAF mutations, whereas CRIS B was associated with high TGFβ levels, EMT and poorest prognosis. CRIS C-E were associated with Chromosomal instability (CIN) with heavy copy number alteration and Wnt activation. Within these, subgroups C to E were associated with Myc amplification, Insulin-like growth factor 2 (IGF2) amplification and TP53 mutations respectively. CRIS A seems to most closely resemble CMS subtypes 1 and 3, CRIS B overlaps with CMS4, and CRIS C-E with CMS2.

Another way of classifying colorectal cancer is based on the nature of the underlying genetic and epigenetic changes in the neoplastic cells. These can be divided into 3 groups; chromosomal instability (CIN) with many copy number alterations, micro-satellite instability (MSI) with many mutations throughout the genome, and CpG island methylator phenotype (CIMP) leading to widespread epigenetic changes and alterations in gene expression[30]. MSI and CIMP are often seen together in 13-16% of CMS1-like tumours and the remaining 84% are associated with CIN[31].

Table 1.1.5.2: **Correspondence between CRC classification systems:** Comparing CMS, CRIS and PDS and the underlying genetic landscape

<table>
<tr><td>CMS</td><td>CMS4</td><td colspan="2">CMS1</td><td colspan="2">CMS3</td><td colspan="3">CMS2</td></tr>
<tr><td>CRIS</td><td>CRIS B</td><td colspan="3">CRIS A</td><td>CRIS C</td><td>CRIS D</td><td>CRIS E</td></tr>
<tr><td rowspan="2">Genetic background</td><td></td><td colspan="3"></td><td colspan="3">CIN
APC mutations and Wnt activation</td></tr>
<tr><td></td><td colspan="2">MSI
CIMP
BRAF mutations</td><td>KRAS mutations</td><td>Myc amplification
EGFR signalling</td><td>IGF2 amplification</td><td>TP53 mutations
Paneth cell phenotype</td></tr>
<tr><td>Clinical features</td><td></td><td colspan="3">Right sided
Sessile/serrated polyps
Mucinous lesions</td><td colspan="3">Left sided
Classical polyps</td></tr>
<tr><td>Immune infiltrate</td><td>Immune excluded</td><td colspan="2">Immune infiltration</td><td></td><td colspan="3">Immune desert</td></tr>
<tr><td>Other features</td><td>High TGFβ
High stromal content
EMT
Angiogenesis
Complement activation</td><td colspan="2"></td><td>Metabolic dysregulation</td><td colspan="3"></td></tr>
<tr><td>Prognosis</td><td>Poor prognosis</td><td colspan="2">Good prognosis initially.
Poor prognosis after relapse.</td><td></td><td colspan="3">Good prognosis</td></tr>
<tr><td>PDS</td><td colspan="3">PDS2</td><td>PDS1</td><td>PDS3</td><td colspan="2">PDS1</td><td>PDS3</td></tr>
<tr><td></td><td colspan="3">Pathways associated with immunity, inflammation EMT and TGFβ signalling</td><td colspan="2">PDS1 – highly proliferative stem-like epithelial cells.
Good prognosis.</td><td colspan="3">PDS3 – well differentiated, low replication epithelial cells.
Poor prognosis</td></tr>
</table>

1.1.5.3 Histological stratification of CRC

CRC can also be classified based on histology, with conventional adenocarcinomas featuring irregularly-shaped glandular structures with desmoplastic stroma around the outside and necrotic debris inside. Variants to the conventional adenocarcinoma include mucinous adenocarcinoma in about 15% of cases where >50% of the tumour volume is extracellular mucin, signet-ring adenocarcinoma where an intracellular vacuole pushed the nucleus to the periphery and medullary carcinoma to name a few[32].

Efforts have been made to attribute histological subtypes to CMS subgroups, to allow quicker stratification of patients, though alignment has not been perfect, largely due to intra-tumour heterogeneity. One study performed transcriptomic profiling by micro-array on distinct tumour regions belonging to one of six morphological groups: classical complex tubular, desmoplastic, mucinous, papillary, serrated and solid/trabecular. To use the CMS description in the paper by Joanito *et al*, desmoplastic was a high-stromal iCMS2 and mucinous a high-stromal iCMS3. Solid/trabecular most overlapped with CMS1. The classical complex tubular and papillary morphotypes overlapped with a CMS2 molecular profile. Finally, the serrated tumours were a mixture of CMS 2 and 3[33].

The image-based CMS utilised Haematoxylin & Eosin (H&E) sections and machine learning to classify tumours to a CMS group[34]. This method allows classification of routine cases in the clinic with no additional procedures or testing required.

CRC can also be stratified by the Immunoscore[35]. This measures the intensity of staining for Cluster of Differentiation 3 (CD3) and CD8 on immunohistochemistry (IHC), and therefore gives an estimate of the level of total T-cell and cytotoxic T-cell infiltration. Higher Immunoscores are associated with better outcomes. Similarly, the presence of tertiary lymphoid structures at the tumour invasive front has been associated with better outcomes[36].

1.1.6 CRC Treatment

The management of most colorectal cancer has not changed significantly in recent years and there are few personalised treatment options. Management is based on the stage of the disease at diagnosis. For localised disease surgical resection is the mainstay of treatment, with chemoradiotherapy used for rectal cancers, followed by surgery if a complete response is not seen. Surgery is *en bloc* with the associated mesentery to allow analysis of lymph node

involvement. For very early disease, pT1 in a pedunculated tumour or Tis, endoscopic resection is possible[37]. For advanced local disease; those with lymph node involvement or other factors giving a high risk of recurrence such as T4 primary tumours or perforation, adjuvant or neo-adjuvant treatment is given. This is with 5-fluorouracil based chemotherapy with or without oxaliplatin[38, 39]. For metastatic disease palliative chemotherapy is the mainstay of treatment, though where possible oligometastatic disease is treated with curative intent[40].

The two currently used personalised treatment options for CRC are EGFR inhibitors and checkpoint inhibitors. EGFR inhibitors are combined with chemotherapy where tumours are wild type for the KRAS and BRAF genes; as RAS and RAF signalling are downstream of EGFR activation, EGFR inhibition has no benefit when they are constitutively active[40].

Checkpoint inhibitors have proved beneficial only where there is high tumour mutational burden, most often caused by microsatellite instability (MSI) due to MMR deficiency or deletions in Epithelial cellular adhesion molecule (EPCAM). MMR deficiency is either caused by mutations in the MMR genes or epigenetic inactivation of the MLH1 gene, often found in conjunction with BRAF mutation. Rarer causes of a high tumour mutational burden are DNA polymerase delta (POLD1) or DNA polymerase epsilon (POLE) mutations leading to errors in DNA replication, or mutations in the NTH like DNA glycosylase 1 (NTHL-1) gene, affecting the DNA base-excision repair mechanism[41]. The high tumour mutational burden results in production of many tumour neo-antigens which allow the immune system to recognise the neoplastic cells and results in greater responses to immunotherapy[42]. The mechanism of action of checkpoint inhibitors will be discussed in section 1.2.4.2.

Other personalised treatment options showing some promise in clinical trial include BRAF inhibitors in BRAF V600E mutant colorectal cancer, anti-HER2 therapy in human epidermal

growth factor receptor 2 (HER2) amplified CRC and topomysin receptor kinase inhibitors for Neurotrophic tyrosine receptor kinase (NTRK) fusion tumours[43]. BRAF V600E mutations occur in 10% of those with metastatic disease; single-agent BRAF inhibitors had minimal response in these tumours due to upregulations of EGFR signalling. However, in trials combinations of the BRAF and EGFR inhibitors gave response rates of 20%, which could be increased further by the addition of MEK1/2 inhibitors, which act downstream of BRAF.[44]

1.1.7 The TME in CRC

Colorectal cancer outcomes are not purely dependent upon the mutations in the neoplastic cells; many cell types in the tumour microenvironment (TME) impact on outcomes, including a multitude of immune and stromal cells alongside the supporting extracellular matrix (ECM). Crosstalk between these components is critical for immunoediting; leading to either clearance of the tumour at an early stage, a prolonged equilibrium where the rate of tumour growth matches destruction, or tumour progression and metastases[45]. Colorectal cancers develop in the context of the complex balance in the healthy colon between the immune system allowing tolerance to food antigens and the commensal microbiota, while being able to generate a rapid protective response against potential pathogens.

1.2 The Immune System in Colorectal Cancer

1.2.1 The conflicting role of the Immune System in Colorectal cancer

While the immune system can produce a beneficial anti-tumour response in cancer, it can also generate pro-tumorigenic inflammation, which is considered one of the enabling characteristics of cancer[45]. Chronic inflammation due to bacterial or viral infection can increase cancer risk, for example Helicobacter pylori with gastric cancer, Hepatitis B or C with hepatocellular carcinoma or human papilloma virus with cervical cancer to name a few; autoimmune conditions such as Inflammatory bowel disease (IBD) increase cancer risk, and exposure to irritants such as asbestos can also generate cancer promoting inflammation[46, 47]. Inflammation resulting in non-specific release of cytokines and growth factors and can stimulate proliferation and epithelial-to-mesenchymal transition in the neoplastic cells directly. Inflammatory cells can also produce reactive oxygen species that are mutagenic. Inflammation leads to release of extracellular matrix remodelling enzymes, allowing invasion by neoplastic cells, and release of pro-angiogenic factors, new vessel formation further facilitating tumour growth. The non-specific inflammation can alter the phenotype of many cells types in the TME, including immune cells and stromal cells, such that they perpetuate this tumour-promoting environment[46].

1.2.2 The Innate Immune system in Colorectal cancer

If there is a breach of the epithelial barrier the first line of defence is the innate immune cells, consisting of myeloid cells (mainly neutrophils, macrophages and dendritic cells), the innate lymphoid cells (ILCs) and Natural killer (NK)-cells. Myeloid cells have pattern recognition receptors to identify infection by pathogen-associated molecular patterns or host cell components released on injury or cell death, termed damage-associated molecular patterns. They can respond through phagocytosis of the pathogens, production of antimicrobial peptides,

upregulation of the machinery for antigen presentation to the adaptive immune system, and release of cytokines and chemokines to recruit other immune cells. Antigen presenting cells (APCs), in particular dendritic cells, function by uptake and processing of pathogen components to display as antigens for recognition by cells of the adaptive immune system on migration to the lymph nodes. NK-cells kill transformed cells through recognising cells where Major Histocompatibility Complex (MHC) class 1 has been downregulated preventing eradication by the adaptive immune system, or by responding to ligands expressed in times of cellular stress such as DNA damage[48].

In cancer many of these innate immune cells are altered to become permissive of the tumour or even promote metastases. Many tumours have infiltration of Myeloid Derived Suppressor Cells (MDSCs), which are neutrophils and monocytes that supress T-cells and NK-cell responses. It is unclear if MDSCs are immature neutrophils and monocytes, or if they are a separate mature cell phenotype[48]. Similarly, macrophages change from an anti-tumour 'M1' phenotype to having tumour permissive 'M2' characteristics, where they produce more immunosuppressive cytokines such as Interleukin 10 (IL-10) and IL-4[49]. Dendritic cells also enter a more immature state where expression of MHC Class 2 molecules is downregulated, preventing effective antigen presentation[50]. These cells that suppress immune responses are associated with poor prognosis. Many of these phenotypic changes in myeloid cells are triggered by TGFβ1[51] and in mouse models of cancer where myeloid cells lacked TGFβ receptors fewer metastases were seen[49, 52].

In cancer NK-cells become less able to target the neoplastic cells. TGFβ1 reduces the ability of NK cells to lyse target cancer cell lines[53] through inhibiting interferon gamma (IFNg) and T-box transcription factor (TBET) expression[54] and reducing localisation of perforin to the immune synapse[53]. TGFβ1 also downregulates expression of the activating receptors NKG2D[55, 56] and NKp30[55], and reduces surface localisation of NKp44[53].

1.2.3 The adaptive immune system

1.2.3.1 B and T Lymphocytes

The second line of defence against microbes is through the adaptive immune system, consisting of B and T lymphocytes. On each cell, the B or T-cell receptors that recognise the antigen are clonal, and the enormous diversity between cells is generated by random recombination of the gene segments encoding these receptors. The T-cells are further divided into those expressing CD4 and CD8, which recognise MHC class 2 and class 1 respectively.

MHC class 1 is expressed on all nucleated cells and is used to present intracellular proteins, therefore recognition of the antigen by a CD8 T-cell occurs when there is an intracellular infection, often by a virus. The CD8+ or cytotoxic T-cells kill the cell displaying its cognate antigen, though secretion of perforin and granzymes. In cancer TGFβ1 in the TME reduces the cytotoxicity of CD8+ T-cells by reducing activation, proliferation, and transcription of IFNg and Granzyme B[57].

CD4+ T-cells, or T helper cells, recognise MHC class 2 on antigen presenting cells and, depending on the cytokines present in the microenvironment, can orchestrate a variety of immune responses, though do not kill pathogens or infected cells directly. The function of different subsets of T helper cells is discussed in section 1.2.4.3.

1.2.3.2 Co-stimulatory and co-inhibitory receptors

In addition to the interaction between the MHC molecules and the T-cell receptor, other co-stimulatory or inhibitory ligands and receptor pairs interact at the same time to determine whether the T-cell produces an activating response or becomes anergic, where the cell will not respond to subsequent stimulation. APCs express CD80 and CD86, which can bind with either CD28 or Cytotoxic T-lymphocyte associated protein 4 (CTLA4) on a CD4+ T-cell. CD28 is a co-

stimulatory receptor, when it binds its ligand it promotes activation of the T-cell, whereas CTLA-4 is coinhibitory, resulting in anergy or tolerance to the antigen being presented. CD28 is expressed on naïve T-cells, whereas CTLA-4 is only upregulated 24 hours after activation, but binds to CD80/86 with higher affinity, and therefore probably exists to limit immune responses and prevent autoimmunity[58]. Similarly, T-cells and B-cells express the Programmed cell death-1 (PD-1) inhibitory receptor, which binds the ligands programmed cell death ligand 1 (PD-L1) and PD-L2 resulting in anergy. PD-L2 is expressed mainly on APCs, whereas PD-L1 is expressed by most non-lymphoid cell types. Many cancer cells upregulate PD-L1, to prevent destruction by CD8+ T-cells. Therefore, immunotherapy drugs, or checkpoint inhibitors, have been developed for cancer treatment to target these inhibitory interactions to allow the immune system to target cancer cells[59, 60]. Checkpoint inhibitors currently in clinical practice include those directed against CTLA-4, PD-1 and PD-L1, though immunotherapies targeting other checkpoints are in development or trials[61] **(Fig 1.2.4.2)**.

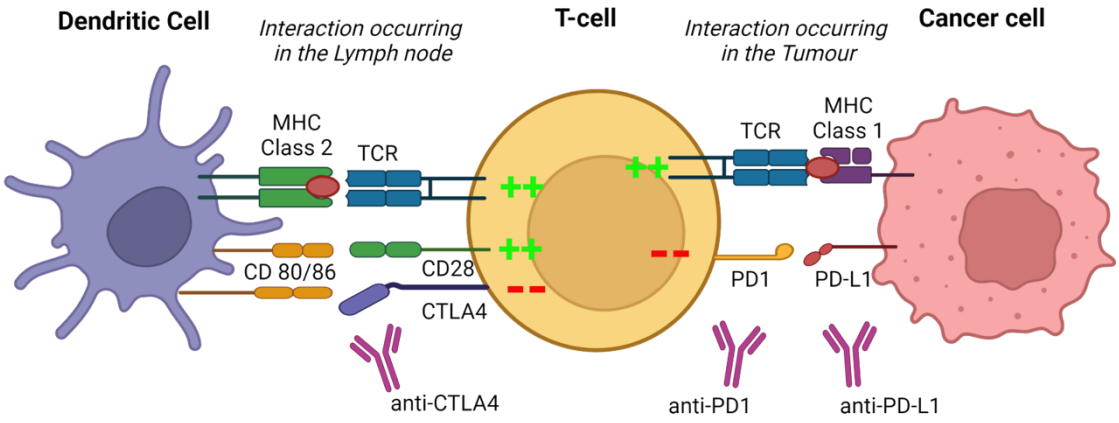


Fig 1.2.4.2 The mechanism of action of immune checkpoint inhibitors
In the lymph node dendritic cells (DCs) present antigen on MHC Class 2 to the TCR on the T-cell. CD80 or CD86 on the DC can either bind co-stimulatory CD28 or co-inhibitory CTLA4 on the surface of the T-cell. Binding of CD28 promotes activation of the T-cell, whereas binding to CTLA4 promotes anergy. Anti-CTLA4 antibodies are used to prevent this inhibitory signal and increase T-cell activation.
Activated T-cells then migrate to the tumour, where they encounter cancer cells displaying tumour antigens on MHC class 1. The cancer cells often upregulate PD-L1, which binds to PD1 on the T-cell, resulting in anergy and failure of the T-cell to lyse the cancer cell. Anti-PD1 and anti-PD-L1 antibodies disrupt this interaction, preventing the negative signal and enhancing T-cell killing of the tumour cells.

1.2.3.3 CD4+ T-cell subsets

If a T-cell receives the appropriate co-stimulatory signals in addition to T-cell receptor (TCR) binding it then enters the cell cycle giving expansion of that clone of effector cells. A third signal, from the cytokines present, enhances proliferation and determines the type of helper cell that CD4+ T-cells become[62]. IL-2 is the key cytokine in promoting T-cell proliferation, though the presence of TGFβ in the TME limits IL-2 production and therefore reduces T-cell expansion[63]. The presence of IL12 and IFNg causes a CD4 cells to express the transcription factor T-Bet and to become a Th1 type cell, producing IFNg and TNF. These lead to the expansion of CD8+ T-cells and the recruitment of NK-cells to combat intracellular pathogens, and the activation of macrophages for phagocytosis. If IL-4 is present this skews T-cell towards a Th2 subtype, that express GATA-3. These then produce IL-4, IL-5 and IL-13 which activate eosinophils and help combat helminth infection. The presence of TGFβ leads to upregulation of the transcription factor FoxP3 and the generation of Tregs which suppress immune responses and produce IL-10 and TGFβ. Combinations of IL-6, IL-23 and TGFβ lead to development of Th17 cells, which express RAR-related orphan receptor gamma (RORγt) and can combat extracellular pathogens. CD4+ T-cells can also differentiate into T follicular helper cells, Th9 and Th22 cells[62].

In cancer the increased TGFβ in the TME skews CD4+ T-cells towards either Treg (tumour permissive) or Th17 (tumour promoting) phenotypes, depending on the concentration of TGFβ and synergistic effects of other cytokines[64]. It drives cells away from Th1 phenotypes, which orchestrate the most useful anti-tumour responses, through silencing TBET and STAT4, which are needed for IFNg production[63].

1.3 Fibroblasts in CRC

1.3.1 CMS4 is a stromal signature

CMS4, associated with the worst prognosis CRCs, is defined by having a high proportion of stroma[20]. In cancer, the stroma is often used to describe all cell types apart from the neoplastic cells, untransformed epithelial cells and the immune cells. This includes fibroblasts, pericytes, vascular and lymphatic endothelial cells, smooth muscle cells, neuroendocrine cells and adipocytes. The fibroblasts in a tumour are often called Cancer associated fibroblasts (CAFs).

Soon after the publication of the CMS subtypes, studies began to sort cells based on expression of specific surface markers (FAP, CD45, CD31 and EPCAM) prior to performing bulk RNA-sequencing. This identified the poor-prognosis signature found in CMS4 as coming specifically from the FAP+ fibroblast component[65, 66]. Fibroblast-Activated-Protein (FAP) is a transmembrane serine protease expressed by fibroblasts exclusively in pathological conditions, including cancer and fibrosis, and is associated with poor prognosis in CRC[67]. These studies also found a linear association between expression of CAF genes and risk of recurrence, and between numbers of fibroblasts and poor outcomes[68] that persisted across CMS subgroups[65]. These data have highlighted fibroblasts as a potential target for CRC treatment.

Depleting activated fibroblasts by targeting FAP has reduced tumour growth in animal models[69-77], but trials in CRC patients have been disappointing[78-80] and loss of these vital multifunctional supporting cells can lead to serious adverse effects, such as anaemia and cachexia[81, 82]. Similarly, in some genetically engineered mouse models, depleting alpha smooth muscle actin (αSMA) positive fibroblasts[83] or preventing collagen production from myofibroblasts[84] have potentiated progression of pancreatic tumours, contrary to the aims of the authors.

These results suggest that fibroblasts in CRC are heterogeneous and they can function to be either detrimental or beneficial in this disease. Specific understanding of fibroblast heterogenous phenotypes is therefore needed before cancer treatment involving modifying their function can be successfully deployed.

1.3.2 Subsets of fibroblasts in health

Fibroblasts are not only heterogeneous in cancer, but are also heterogeneous in the healthy colon[85, 86]. Historically, they've been described as 'supporting cells', but their role is greater than to simply provide structural support to epithelial cells. Fibroblasts are resident cells that organise tissue structure and maintain discrete niches through communicating with other cells. They do this by responding to and secreting signalling molecules, including angiogenic factors, cytokines and chemokines. Markers of fibroblasts in health include fibroblast-specific protein 1 (FSP-1) and PDGFRα.

Within the healthy colonic mucosa three main fibroblast phenotypes have been identified **(Fig 1.3.2.1)**. At the crypt bases a subset of fibroblasts are seen that produce Wnt 2, Wnt 2b, R-spondin 3 (RSPO3) and Gremlin 1 (GREM1). These function to maintain the stem-cell phenotype of the epithelial cells of the crypts, promoting continued cell division. Progressing up from the crypts, but still in close proximity to the epithelial cells, are a group of fibroblasts referred to as telocytes, which are marked by expressing the transcription factor SOX6 and F3/CD142. These produce Wnt 5 and many BMP proteins, which encourage differentiation of the epithelial cells. Deeper in the lamina propria are a group of fibroblasts expressing CC chemokine ligand 8 (CCL8) and CCL11, as well as ABCA8, ADAMDEC1 and DCN. These probably recruit immune cells to the colon and interact with immune and perivascular cells[85, 86].

As well as maintaining the structure of the healthy colon, fibroblasts also play a role in response to injury and infection. When fibroblasts become activated after injury they secrete more extracellular matrix components and generate greater contractile forces, taking on a phenotype resembling a smooth muscle cell characterised by the upregulation of αSMA. These cells are critical for wound healing and are termed a myofibroblasts, but are also found inappropriately in fibrotic disease. In infection they can secrete cytokines and chemokines, recruiting more immune cells to the region[87].

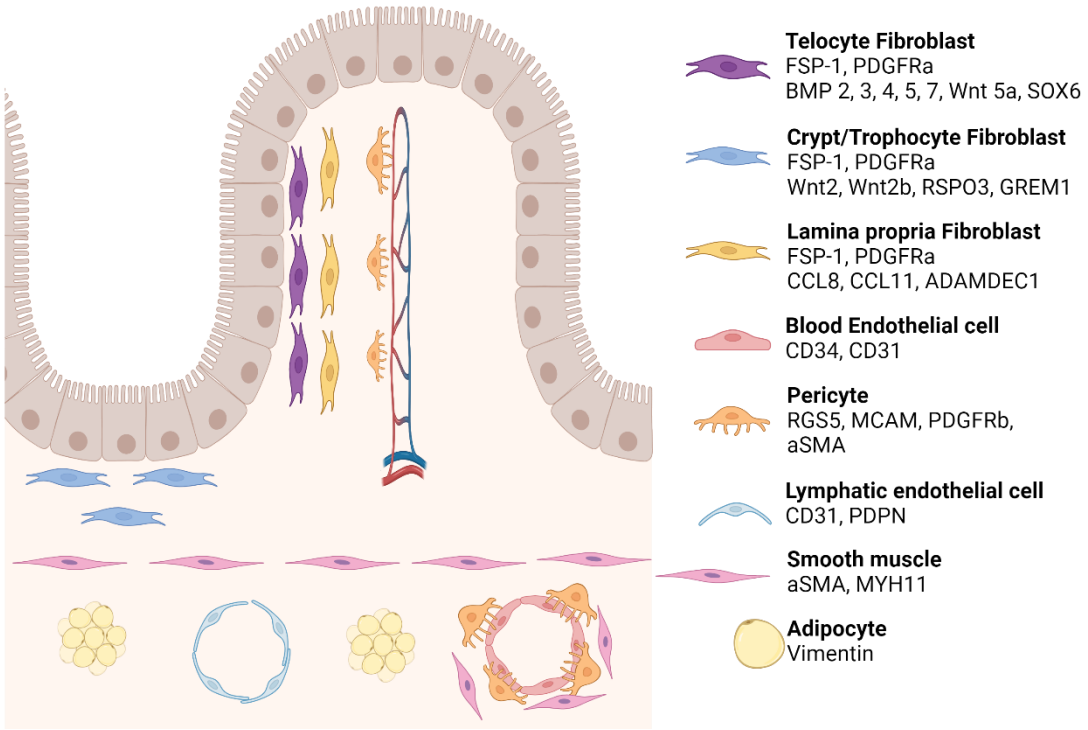


Fig 1.3.2.1 Stromal cells in the Healthy Colon

Schematic of the healthy colon showing the three main group of fibroblasts: Telocytes, Crypt fibroblasts and Lamina propria fibroblasts. Other stromal cell types; pericytes, endothelial cells, smooth muscle cells and adipocytes are also included.

1.3.3 Stromal subsets in health beyond fibroblasts

The stromal cell types seen in the colon beyond fibroblasts also have their own markers. Pericytes are identified by Regulator of G-protein signalling 5 (RGS5), PDGFRβ and MCAM/CD146; Smooth muscle cells by MYH11, CNN1 and actin alpha 2 (ACTA2); Vascular

endothelial cells by CD31/PECAM1, CD34, VE-cadherin and ACE/CD143; Lymphatic endothelium expresses lymphatic vessel endothelium hyaluronan 1 (LYVE1) and podoplanin (PDPN) in addition to CD34 and CD31; Schwann cells can be identified by expression of S100B.

1.3.4 Subsets of fibroblasts in disease

A cancer-associated fibroblast (CAF) refers to any fibroblasts found within a tumour, but not the normal adjacent (N.adj or N.adjacent) tissue. Activated fibrobroblasts are characterised by expression of FAP and PDPN in addition to the normal fibroblast markers. In recent years many single-cell RNA sequencing (scRNAseq) studies have identified differing groups of CAFs across a multitude of cancer types. Multiple fibroblast clusters have also been identified on scRNAseq of tissues from a variety of inflammatory diseases[88], including, most relevant to CRC, IBD[86, 89, 90], and across different species[91-94]. Each study has mapped its own clusters, but there are fibroblasts with considerable concordance in their gene expression profiles described across studies. Many have clusters that can be described as inflammatory CAFs (iCAFs), characterised by the expression of chemokines and immune cell interaction, and myofibroblastic CAFs (myCAFs), characterised by production of αSMA and other proteins involved in matrix deposition or remodelling[95-97]. This terminology seems a sensible functional paradigm to help understand the roles of these heterogenous cells in CRC.

The variety of clusters from different studies could be attributable to differences in sample size and depth (biopsy versus resection), number of cells captured, the sequencing technique and the algorithms and resolutions selected in the clustering process. This heterogeneity of fibroblasts subpopulations reflects the diversity of their function and is likely driven by their tissue location and the cells they interact with.

Prior to the emergence of scRNAseq and unsupervised clustering of cells, several studies had already identified discrete groups of fibroblasts in cancer. For example, in pancreatic cancer IHC has been used to locate αSMA high fibroblasts close to neoplastic cells and IL-6-positive fibroblasts further away[98]. In breast cancer flow cytometry identified a CD29+ FAP+ fibroblast subset associated with increased macrophages and Tregs and reduced CD8+ T-cells[99], and another breast cancer study found a group of fibroblasts characterised by expressing the complement receptor was associated with worse treatment response in patients and mouse models[100].

In the past 5 years there have been several comprehensive scRNAseq studies in CRC[68, 101-103] **(Table 1.3.4.1).** The earliest study by Li *et al.* in 2017 described a CAF-A group, characterised by expression of Matrix metalloproteinase 2 (MMP2), DCN and Collagen 1A2 (COL1A2), and a CAF-B group of typical myofibroblasts expressing ACTA2, TAGLN and PDGFRA[101]. Hae-Ock Lee *et al*.[102] in 2020, reported three normal and four myofibroblastic CAF clusters. The four myofibroblast clusters were involved in matrix synthesis, matrix degradation, a highly proliferative matrix degradation subset, and an RGS5 expressing group, which are probably pericytes. They did not describe an iCAF cluster, although one of the four 'myofibroblast' clusters expressed CXC chemokine ligand 1 (CXCL1) and 8 and MMP 1 and 3, which we interpreted as possibly equivalent to the iCAF clusters seen in other studies. Khaliq *et al.* in 2022 found the COL1A1 positive cells clustered into 'CAF-S1' fibroblasts, 'CAF-S4' vascular or pericyte like cells, and normal fibroblasts, basing their terminology on a previous scRNAseq study in breast cancer. Within CAF-S1 they were able to identify cells that matched the 3 myCAF clusters (ECM, wound and TGFβ) and 2 iCAF clusters (IL and detox) seen in the breast cancer study[95]. They found all 5 CAF clusters were enriched in CMS4 disease[24].

Wang *et al* compared matched CRC primary, liver metastases and normal adjacent for each tissue in six patients who had previously received neo-adjuvant treatment. They found an F3 positive fibroblast subset only in the colorectal tissue and absent from the liver, at higher levels in the primary tumours than normal adjacent[104]. The F3+ subset expressed C3, CXCL1, MMP2, MMP3, and NRG1. They found the NRG1 could promote proliferation of epithelial cells via binding to ERBB3 and that CRC patients with high expression of F3 and NRG1 had poorer overall survival.

Table 1.3.4.1: Fibroblast clusters in CRC scRNAseq publications

	Sample number	**Number of cells analysed**	**Number of fibroblast clusters**	**Clusters found more frequently in normal adjacent tissue**	**CAF clusters**
Li *et al*. in 2017[101]	11 CRC tumours & matched normal adjacent tissue	590	3	1 Normal cluster	CAF-A expressing MMP-2, DCN and COL1A2 CAF-B expressing ACTA2, TAGLN, and PDGFRA
Hae-Ock Lee *et al.* 2020 [102]	29 CRC tumours & matched normal Adj	91,103 cells from tumour tissues	13	3 groups more highly abundant in normal tissue than tumour. • Lipofibroblasts • PDGFRA/B high • Matrix fibroblasts	Study described 4 myofibroblast clusters: MF1 – matrix synthesis MF2 – matrix degradation, MF3 – RGS5+ expression MF4 – similar to MF2, but highly proliferative.
Pelka *et al.* 2021[103]	62 CRC tumours & 36 matched normal adjacent tissues	371,223	11	5 clusters primarily seen in n. adjacent	6 clusters, including those expressing myofibroblastic and inflammatory gene programmes
Khaliq *et al.* 2022[24, 68]	16 CRC tumours & 7 normal adjacent tissues	49,589	6	1 normal cluster	IL- and detox- iCAFs ECM-, Wound- and TGFβ- myCAFs
Wang *et al.,* 2023[104]	6 CRC tumours & liver metastases and matched normal adjacent for each tissue	41,892 CD45- nonimmune cells	8	CCL11 expressing cluster.	F3 expressing cluster, which on IHC formed a rim around epithelial cells, enriched in primary tumour MCAM cluster enriched in liver metastases

1.3.4.1 The Pelka et al CRC scRNAseq dataset

Of the published scRNAseq studies in primary CRC, one of the most comprehensive is that published by Pelka *et al* in 2021, identifying four times as many individual cells than the study with second highest numbers, and taken from 62 patients, over twice as many patients as the second largest published study. They identify multiple CAF and normal adjacent fibroblast clusters and their well-annotated results are also available for further analysis by other research groups, making this study ideal for use as a comparator for our results[103].

This study identified 32 stromal clusters, labelled cS1-32, including 11 fibroblast clusters, named cS21-31. While they identify 5 clusters from normal adjacent tissue, these form 3 discrete groups which match the normal colonic fibroblast phenotypes shown in **Fig 1.3.2.1**. cS21-22 are stem-cell like fibroblasts which match those normally seen at the crypts, cS23-24 are BMP-producing fibroblasts and are probably the telocytes found directly underlying and supporting the differentiated epithelial cells, and cS25 are CCL8+ fibroblasts which probably represent the subset normally seen in the lamina propria and involved in immune recruitment to the colon.

The 6 fibroblast clusters from the tumour tissue, named cS26-31, each cluster closely on UMAP to one of the three sets of normal fibroblasts, and indeed the differentially expressed genes (DEGs) for each cluster strongly overlap between those tumour and normal fibroblasts seen together on UMAP **(Fig 1.3.4.1).** cS30 CAF CCL8-fibro is overlapping with cS25 CCL8+ fibroblasts. cS31 CAF stem-niche lies close, but not overlapping with, cS21-22 stem-cell like fibroblasts. Clusters cS26-29 all lie close, but separate to cS23-24 telocytes. This group includes cells described as: cS26 myofibroblasts, cS27 CXCL14+ CAFs, cS28 GREM1+ CAFs and cS29 MMP3+ CAFs. Clusters 26-28 expressed a contractile gene program and clusters 28-29 expressed an inflammatory gene programme, suggesting a gradient from myCAF-like cS26 and 27, through to iCAF-like cS29, with sC28 showing features overlapping the two. Therefore, this

dataset suggests that both myCAF and iCAF like clusters derive from the telocyte cells that support mature epithelial cells. cS26 myofibroblasts additionally expressed a smooth muscle programme, and this cell cluster did sit slightly separately from cS27-29, near to both the smooth muscle cells (shown in purple on the upper row) as well as to fibroblast clusters 23,24,27,28 and 29 on UMAP.

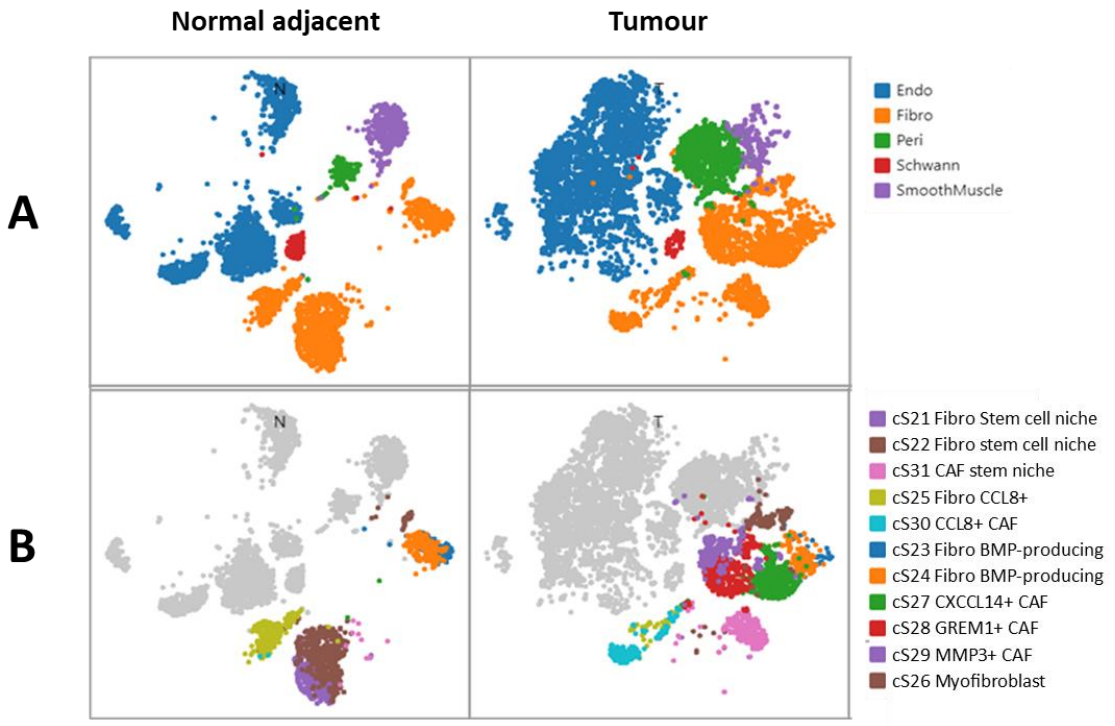


Fig 1.3.4.1 UMAP of stromal cell clusters from the Pelka *et al.* scRNAseq study
Row (A) shows mid-level clustering, of the 5 broad stromal cells types. Row (B) shows full-depth clustering of the 11 fibroblast clusters seen. These fall into three broad groups; the stem-cell niche normal fibroblast and CAFs, the CCL8+ normal fibroblasts and CAFs, and thirdly the BMP-producing normal fibroblasts alongside CXCL14+, GREM1+, MMP3+ and myofibroblast CAFs.

Fig 1.3.4.2 shows gene expression in these fibroblast clusters, with size of the dots indicating the proportion of cell expressing that marker and colour indicating the expression level, focussing on genes to identify fibroblasts generally, markers of the individual clusters and key genes expressed by them. Of note none of the widely-used markers of the fibroblasts are perfect: *PDPN* and *PDGFRA* are barely expressed in cS26, although otherwise they seem sensitive and specific, with *PDGFRA* expressed in all fibroblasts and *PDPN* in just CAFs. Whereas FAP is expressed only in a minority of CAF clusters, and *S100A4* (also known as FSP-1)

is expressed in pericytes and smooth muscle as well, reducing their utility as general markers of fibroblasts. Of the traditional pericyte markers, *PDGFRB* and *THY1* are both upregulated in CAFs, and while *RGS5* and *MCAM* are more specific they are also elevated in smooth muscle. Markers of the three normal fibroblast subsets (telocytes, crypt and lamina propria fibroblasts) are shown as well as the markers *CXCL14, GREM1* and *MMP3* of three of the CAF subsets. Of note, while MMP3 seems specific to its eponymous cluster, the other two are less specific.

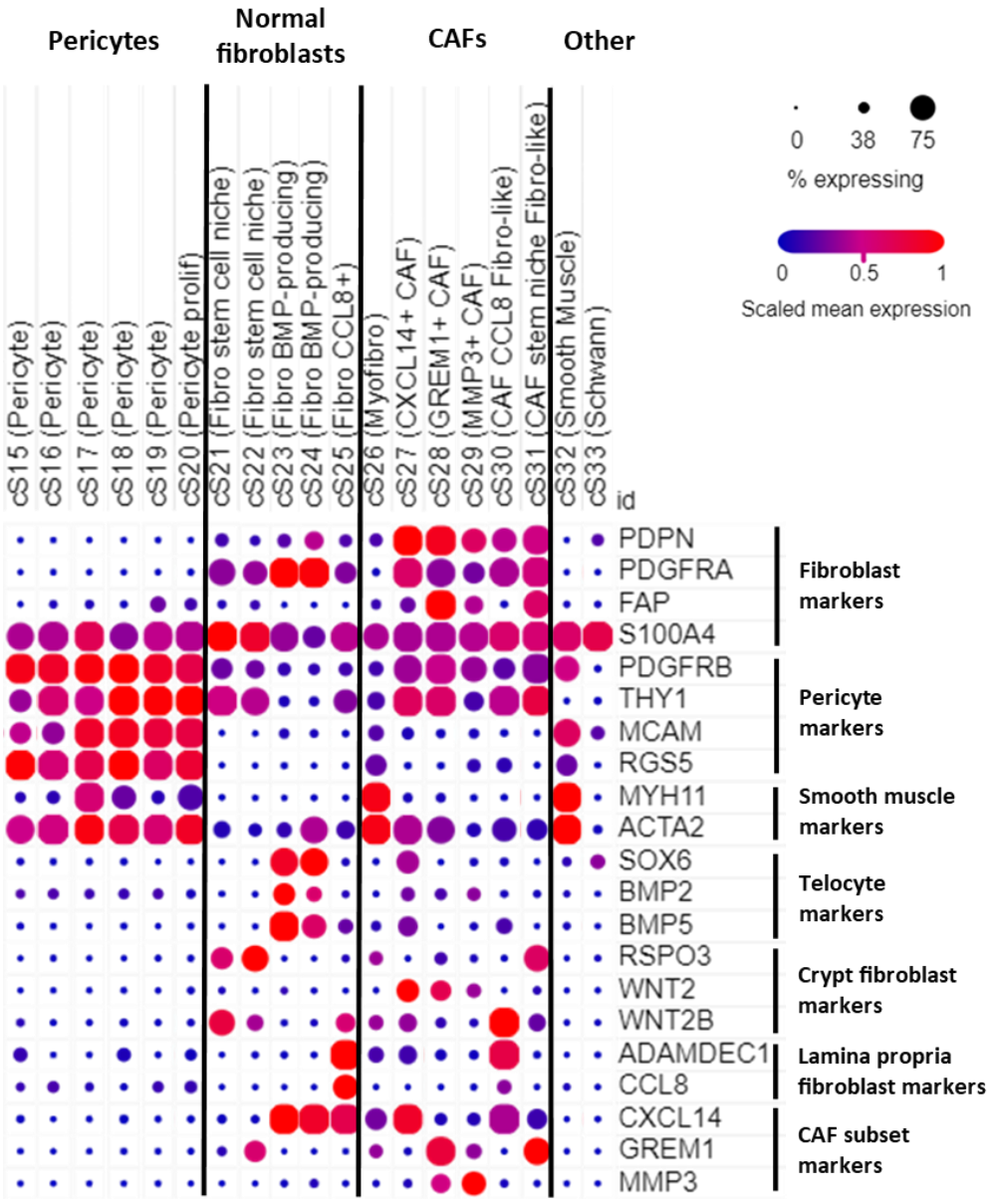


Fig 1.3.4.2 Gene expression in selected stromal cell clusters from the Pelka *et al.* scRNAseq study
Gene expression is examined in the pericyte, normal fibroblast, CAF, smooth muscle and Schwann cell clusters. Expression of general fibroblast, pericyte and smooth muscle markers are shown, as well as more specific markers of individual normal fibroblast types, and the markers for each of the three CAF subsets identified by Pelka *et al.*

Of particular note is cS26, the cluster described as a myofibroblast. Not only did this cluster sit as close to the smooth muscle cells on UMAP as the fibroblasts, it expresses none of the classic fibroblast or pericyte markers significantly (*PDPN, PDGFRA, PDGFRB, THY1, FAP, MCAM,* or *RGS5*) and does express the smooth muscle marker *MYH11*. Therefore, we consider if this is actually a cancer- associated smooth-muscle cell, as opposed to a fibroblast at all.

RNA *in situ* hybridization (ISH) and immunofluorescence were performed on selected cases within the study to visualise the fibroblast subsets *in situ*. This suggested CAFs (cS26-29) were found at different locations within the tumour; with CXCL14+ CAFs found in close association with the epithelial cells, MMP3+ CAFs around vessels close to the luminal surface, and the GREM1+ CAFs found within stromal tracts deep within the tumour. The location of the CXCL14+ CAFs close to the epithelial cells is where telocytes would be expected to be seen within normal tissue and supporting this of these four CAF groups the CXCL14+ CAFs did sit in greatest proximity to the normal telocyte cluster on UMAP, show greatest expression of BMPs, and greatest overlap in DEGs with the normal BMP-producing fibroblast clusters cS23-24.

Individual patients had differing proportions of cells in these four groups of CAFs and not every group was seen in each patient. MMP3+ CAFs were seen more frequently in MMR deficient disease and MMR proficient disease had more CXCL14+ CAFs[103].

1.3.5 Identifying fibroblasts at the protein level

While multiple studies have identified fibroblast clusters from RNA data; protein level data, to locate them *in situ* in the context of the complex TME or to sort them *ex vivo*, is emerging more slowly. Challenges include, firstly, lack of generic markers that are sensitive and specific for all fibroblasts[105]. Fibroblasts are often additionally identified as cells that are negative for other markers, such as epithelial, endothelial and leukocyte markers, and lacking the genetic mutations seen in neoplastic cells, to avoid malignant cells that have undergone EMT being inappropriately assigned as fibroblasts[105]. So-called fibroblast markers are often expressed by other cell types. For example, many common features exist between myCAFs and pericytes, which are both αSMA-high stromal cells. In some studies, pericytes are specifically excluded from fibroblast analyses through negative selection for the pericyte markers RGS5, with the risk that genuine fibroblasts are excluded, whereas in other studies pericytes are likely to contribute to the myCAF group[101, 106].

A cell type that shows considerable overlap with iCAFs are senescent fibroblasts. The senescence-associated secretory phenotype (SASP) may exist to limit ECM deposition and fibrosis during the completion of wound healing[107], but markers of this phenotype include SERPINE1[107], MMP3 and IL-6[108], all of which have been used as iCAF markers. However, iCAFs cannot completely overlap with this cell type, as they have been successfully cultured *ex vivo* maintaining their inflammatory phenotype[100], whereas a senescent cell has greatly reduced replicative potential. Fixed terminology, with terms applying to cells expressing certain markers, would improve concordance in findings across studies.

A second challenge is that scRNAseq clusters can be characterised by relatively small changes in expression of these markers[93], for example αSMA is more highly expressed in myCAFs than iCAFs though both express it to some extent, and a discriminatory marker found in one dataset doesn't

necessarily cross-reference perfectly into other scRNAseq datasets. Further challenges exist in a lack of protein-tools for some genes identified by RNA-sequencing.

Despite these caveats, fibroblast clusters have been spatially identified in a small proportion of scRNAseq papers; in CRC in the aforementioned Pelka *et al.*[103] study, and in other cancers, αSMA[96, 109] and POSTN[96] have identified myCAFs by multiplex IHC, and TEM8[95] by flow cytometry, whilst iCAFs have been identified on imaging by SERPINE-1[96], IL-6[98], GPR77[100] and CD10[100].

1.3.5.1 Interactions

The ability to identify the subsets *in situ* will provide information about which cells types they are interacting with, providing valuable information about the likely factors impacting CAF phenotype. Improved understanding of what drives fibroblast phenotypes will enable generation of different fibroblast flavours *in vitro/ex vivo,* allowing exploration of the mechanisms of these interactions. **Figure 1.3.5.1** shows the putative location of fibroblast phenotypes within CRC tumours and their interactions.

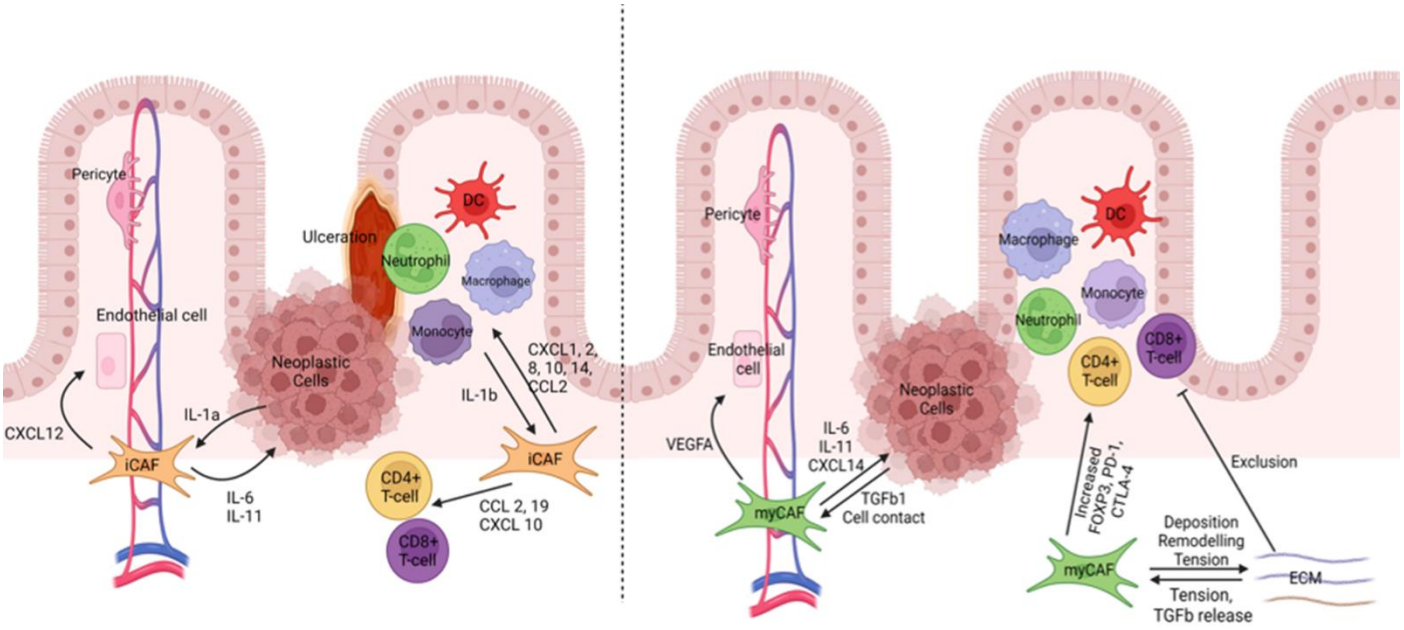


Fig 1.3.5.1 Hypothesised locations and interactions of myCAF and iCAF cell

Studies in pancreatic cancer found proximity to epithelial cells determines their phenotype[98]; with αSMA+ myCAFs seen close to the neoplastic cells, and additionally fibroblasts co-cultured with organoids developing myCAF features, whereas transwell systems led to iCAF properties[98]. This suggests that direct cell contact may drive generation of myCAF phenotypes, though how this occurs remains to be explored. Multiplex imaging in lung cancer has shown FAP+ αSMA+ CAFs surrounding tumour nests[109]. Another study found GREM1+ myCAF like cells to be further away from the tumour cells than a comparator group[110].

Proximity analysis of the fibroblast phenotypes to blood vessels on multiplex imaging has also resulted in conflicting results. Pelka *et al*.[103] saw the iCAF clusters specifically around the dilated blood vessels of the colonic luminal margin, whereas analysis of tissues from several inflammatory diseases by Korsunsky at al.[88] described a myCAF-like group as vascular fibroblasts closely associated with blood vessels. These may be due to differences in markers used to detect these phenotypes. Fibroblasts stimulate angiogenesis through production of several factors, such as CXCL12 and VEGFA, which may be differentially expressed between CAF subsets[111, 112].

1.3.6 Drivers of the fibroblast phenotypes

While many clusters of CAFs have been proposed, genes that define these clusters are often not specific for one cluster, but are rather expressed at different levels across several CAF subsets. The gradient of expression could represent different activation stages of the fibroblasts or location within the tissue.

Multiple studies have concluded that the key driver of the myCAF group is TGFβ and the iCAF group is driven largely by IL-1[96, 113]. These two cytokines both play a role in tumour progression, but have opposite effects on the immune response; IL-1 is an inflammatory cytokine produced in response to pathogens, whereas TGFβ is anti-inflammatory, dampening the effects of many immune cell types. However, the specific niche and cell types responsible for the expression of IL-1β and TGFβ in different CRC subsets have not been mapped yet. In the following sections we will describe the roles of IL-1β and then TGFβ in the TME, before describing other cytokines that may be impacting on fibroblast phenotype.

1.3.6.1 IL-1β production

IL-1 actually refers to two distinct, but similar cytokines: IL-1α and IL-1β. Studies differ in which of these family members they attribute to generating the iCAF phenotype. Computer algorithms predict both IL-1α and IL-1β as possible drivers of iCAFs[96] and studies across CRC[103, 114], pancreatic cancer[113, 115] and lung cancer[96] have shown either recombinant IL-1α or recombinant IL-1β increase inflammatory markers in primary tumour fibroblasts. Blockade of IL1α detected in PDAC tumour organoid supernatants reduced the iCAF phenotype, whilst IL-1β was undetectable by Enzyme-linked immunosorbent assay (ELISA)[113, 115]. This suggests that IL-1α is probably the more significant player in the effects of the neoplastic cells on CAFs. IL-1β, however, is largely produced by the innate immune cells, in response to the recognition of

pathogens[116]. IL-1β therefore may have a major role in CRC, given there is only a single layer of epithelial cells forming a barrier to the gut lumen which is often disrupted in cancer[117] and bacterial cells are known to be present within CRCs[118]. In line with this, iCAFs have been identified in areas of necrosis at the luminal border in CRC, where bacteria and neutrophils are expected to be present[103]. This concurs with work in IBD, where inflammatory fibroblasts have been seen in areas of ulceration in close association with IL-1β-producing neutrophils[89]. IL-1 is also released from necrotic cells[119], and therefore cancer treatment may lead to increased IL-1 levels driving the formation of iCAFs.

1.3.6.2 IL-1β signalling

The two isoforms of IL-1 both bind to the type 1 IL-1 receptor (IL-1R1), which forms a heterodimer with the IL-1 receptor accessory protein (IL-1RAcP)[120]. Ligand induced conformational changes allow recruitment of MYD88 and the serine-threonine kinase IRAK4, which together phosphorylate IRAK 1 and IRAK2, which then form a complex with TRAF6. This leads to formation of further signalling complexes, TAK1 and MEKK3, which then activate the NF-κB, JNK and p38 MAPK pathways. The signalling pathway from IL-1 binding to activation of transcription factors in an iCAF is summarised in **Fig 1.3.6.2**.

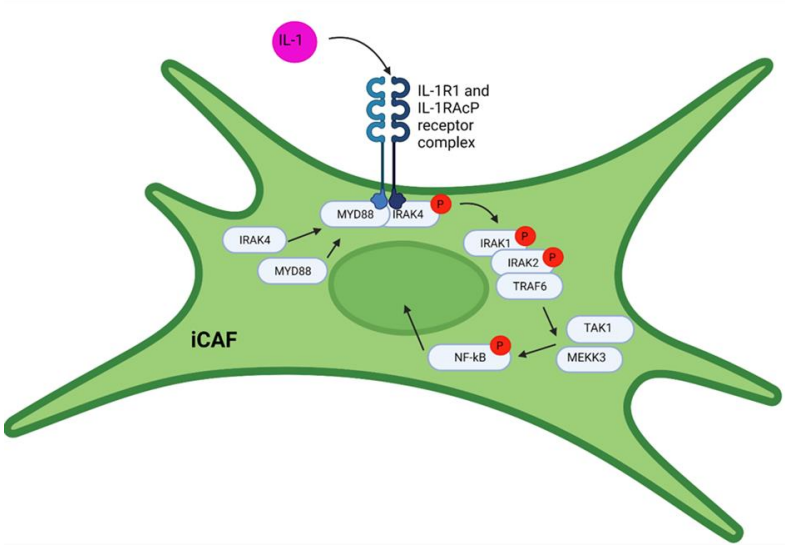


Fig 1.3.6.2 IL-1 signalling in the iCAF.
Ligand binding to the heterodimeric receptor complex recruits IRAK4 and Myd88, the former of which becomes phosphorylated. This leads to phosphorylation of IRAK1 and IRAK2, which then form a complex with TRAF6. This triggers formation of the TAK1 and MEKK3 complexes, which then have multiple actions, including phosphorylation and activation of the transcription factor NF-κB.

1.3.6.3 TGFβ expression in Colorectal Cancer

High TGFβ levels in CRC are correlated with worse outcomes irrespective of CMS subtype or stromal content. When localised tumours are resected, those with TGFβ expression levels in the bottom third, when measured on bulk RNA sequencing, almost never recur, whereas those in the top third recur in about half of cases[112]. CRC patients additionally have 2.5-3x higher circulating levels of TGFβ1 and 2 than healthy controls[121]. However, TGFβ is a pleiotropic cytokine that is produced by, and has effects upon, most cell types, and can also have protective roles. We will first describe the intracellular signalling pathway activated by TGFβ, and then its role in cancer.

1.3.6.4 TGFβ signalling

TGFβ is secreted from the cell in a latent form and requires activation in the ECM to exert its effects. TGFβ is synthesised as a pro-hormone, and cleaved into the cytokine and the latency-associated peptide (LAP), the LAP encircles the active TGFβ preventing its activity. Upon

secretion from the cell it binds latent TGFβ-binding proteins (LTBPs) to form the large latent complex (LLC). The LLC binds to fibrillin, resulting in the latent TGFβ becoming sequestered in the ECM. It can also bind to Glycoprotein A Repetitions Predominant Protein (GARP) on the surface of Tregs, endothelial cells and platelets, or Leucine-Rich Repeat-containing Protein 33 (LRRC33) on macrophages. Active TGFβ is released from the latent complex either through tension in the ECM, acting through the integrins aVb6 or aVb8, or by enzymes such as thrombospondin 1, plasmin, cathepsin D, MMP9 or MMP14. Once activated, TGFβ signals through a tetrameric transmembrane receptor, made up of a pair of each of TGFβ receptor 1 (TGBR1) and TGFβ receptor 2 (TGFBR2). TGFβ binding its receptor leads to canonical signalling with phosphorylation of SMAD2 and SMAD3, which then form complexes with SMAD4 to shuttle to the nucleus, where they interact with many other transcription factors to regulate gene expression. TGFβ also triggers non-canonical, SMAD-independent, signalling, mediated through the same receptor, including activating the MAPK and PI3K cascades[63]. The canonical signalling pathway is summarised in **Fig 1.3.6.4.**

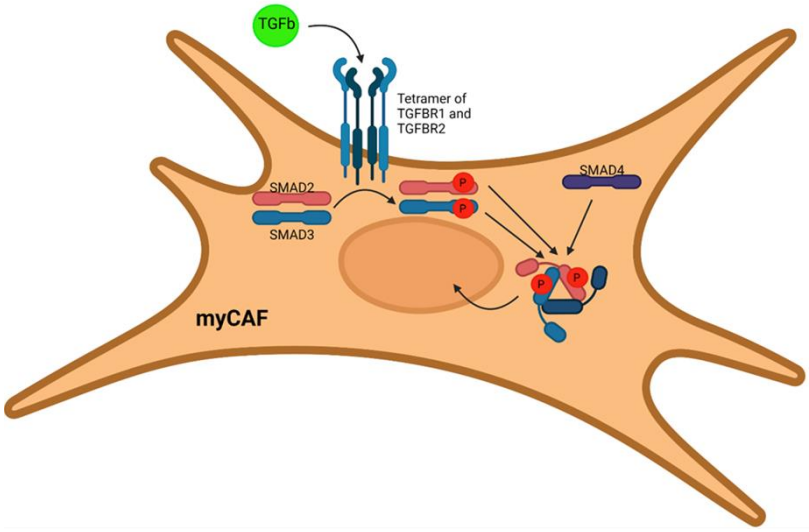


Fig 1.3.6.4 Canonical TGFβ signalling in the myCAF.
Active TGFβ binds to a tetramer of pairs of TGFBR1 and TGFBR2. This results in phosphorylation of SMAD2 and SMAD3, which form a complex with SMAD4. The SMAD complex shuttles to the nucleus, where it controls expression of many genes.

1.3.6.5 TGFβ effects in Cancer

In health, TGFβ has a tumour-supressing role in colonic epithelial cells, through increasing CDK inhibitors, repressing c-myc, and triggering apoptosis in pre-malignant RAS-mutant cells[63]. However, in over half of colorectal cancers the response to TGFβ by neoplastic cell is lost or altered due to mutations or deletions, giving it a tumour promoting effect. The most common alteration in the pathway is loss of SMAD4[122], which would only affect the canonical pathway, but in some tumours the receptor is affected. When neoplastic cells do retain some responsiveness to TGFβ, it may drive epithelial to mesenchymal transition (EMT)[123, 124], a process where epithelial cells lose their polarity and adhesion to other cells and become more migratory, which occurs in wound healing as well as cancer. Upregulation of transcription factors promoting EMT; ZEB, SNAIL and TWIST, would normally promote apoptosis in the cell, but in cancer neoplastic cells can undergo EMT without triggering apoptosis[63]. In neoplastic cells that still respond normally to TGFβ the relative cytostasis may make them more resistant to immune surveillance and chemotherapy[63].

Similar to in epithelial cells, TGFβ signalling to fibroblasts can prevent malignancy formation, but promote progression in established tumours. Mice that have a conditional loss of the TGFβ receptor in fibroblasts have increased formation of new tumours, which were thought to develop due to unrestrained inflammatory responses where fibroblasts did not respond to TGFβ[125]. However, the poor-prognostic signatures in stromal high, CMS4 CRCs are largely derived from genes that are upregulated in fibroblasts by TGFβ[65]. Co-injection of tumour cells in mice with both fibroblasts and TGFβ1 leads to synergistic increases in tumour formation, compared to either fibroblasts or TGFβ1 alone, suggesting the actions of TGFβ1 on fibroblasts greatly enhances initial tumour growth at new sites[65].

TGFβ also promotes tumour growth through acting on endothelial cells to promote proliferation and differentiation, leading to angiogenesis, and through driving more tolerogenic immune phenotypes described previously, in section 1.2.

1.3.6.6 Targeting TGFβ for cancer treatment

Due to the strong correlation between TGFβ and poor prognosis in CRC many drugs have been developed to target it; from direct inhibitors of the active cytokine, inhibitors of the receptor or prevention of TGFβ activation. The various strategies for blocking TGFβ activity are shown in **Fig 1.3.6.5.** As a single agent treatment[126, 127] or in combination with chemotherapy[128] targeting TGFβ has led to minimal response. In mice more promising results have been seen when TGFβ inhibitors are combined with immune checkpoint inhibitors or radiotherapy, with trials in human subjects ongoing[129-131]. These are logical treatment combinations, because TGFβ inhibition improves immune cell infiltration in tumours, but increases the expression of checkpoints, and radiotherapy leads to activation of latent TGFβ. Interestingly, TGFβ inhibition in mice has been much more effective in preventing sites of new metastatic disease than of restricting growth of the primary tumour[112, 132, 133]. This may be due to reversing systemic immunosuppression which may prevent the early eradication of new metastases, or by stopping TGFβ priming distant fibroblasts to provide the soil for circulating neoplastic cells to extravasate and grow. Alternatively, TGFβ could have local effects making it possible for the neoplastic cells to extravasate, which are halted by its inhibition.

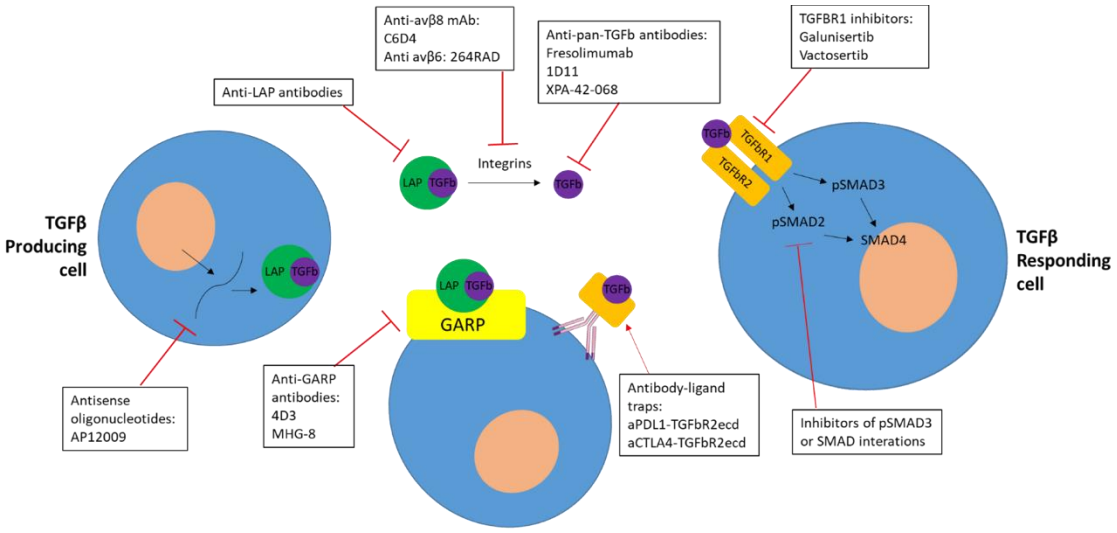


Fig 1.3.6.5 Mechanisms of targeting TGFβ for CRC treatment
Methods attempted to target TGFβ include (1) antisense-oligonucleotides to prevent TGFβ production, (2) preventing activation of the latent cytokine through anti-LAP, anti-GARP and anti-integrin antibodies, (3) directly targeting active TGFβ, (4) targeting the TGFβ receptor or (5) inhibiting SMAD or other downstream signalling molecules.

In the short term most TGFβ inhibitors are well tolerated, with several studies not reaching a maximum tolerated dose[127, 134, 135]. Formation of keratoacanthomas is frequently seen, but these resolve on cessation of treatment[126, 135, 136]. However, there have been concerns about cardiovascular side effects, as heart valve damage was seen with the first TGFβR1 inhibitors[137]. There are also concerns about longer term side-effects; given TGFβ has a tumour supressing role there is a risk its inhibition may increase occurrence of new, unrelated, primary tumours.

1.3.6.7 The three Isoforms of TGFβ

TGFβ exists as a family of three cytokines in humans; TGFβ1, TGFβ2 and TGFβ3. They are encoded by separate genes on different chromosomes, but with high sequence homology[138]. The isoforms are known to have differing roles in development, through studies selectively mutating individual genes. Where mice are heterozygous for any of the *TGFB* isoforms they are phenotypically normal and able to reproduce. When *TGFB1* heterozygote mice are crossed fewer than the expected 25% of offspring are homozygous mutant, suggesting lower survival of embryos lacking *TGFB1*, but the homozygous mutants identified appeared phenotypically

normal at birth. However, after about 3 weeks they rapidly deteriorated and died with multi-organ tissue inflammation, supporting TGFβ1 as having a key role in supressing immune responses. The timing of the widespread inflammation was approximately at the time of weaning, suggesting that prior to this point homozygous mutant mice may be receiving normal TGFβ1 through the placenta or in the mother's milk[139]. Mice homozygous for non-functional *TGFB2* die in the perinatal period, either in the days before, or minutes after birth due to cyanosis. They have multiple congenital abnormalities of the cardiac, lung craniofacial, limb, spine, eye, and urogenital systems, thought to be due to failure of migration of neural crest components[140]. Mice homozygous for non-functional *TGFB3* die within 4 to 24 hours of birth due to cyanosis and failure to suckle, on examination they are found to have cleft palate due to failure of adhesion of the epithelia, without any further craniofacial abnormalities[141].

Mechanisms of activation of TGFβ differ by isoform. The latent TGFβ binding proteins (LTBP) 1 and 3 bind to all isoforms of TGFβ, whereas LTBP4 only binds to TGFβ1. Only TGFβ1 and TGFβ3 bind to integrins, TGFβ2 does not have the required RGD binding motif[142]. Different isoforms of TGFβ also have differing affinity for members of the family of receptors, with TGFβ3 leading to greater phosphorylation of SMAD1 and 5 than the other isoforms, and TGFβ2 being more reliant on the presence of TGFBR3 to aid binding to TGFBR2 than the other isoforms[143]. Additionally, although the active forms of TGFβ generally exist as homodimers, heterodimers between TGFβ1 and TGFβ2 have been identified[144].

Studies of tissue distribution of the TGFβ isoforms are hampered by a lack of correlation between measurements of the TGFβ ligands at the mRNA and protein levels both across multiple normal tissues and organs[145] and in colorectal tumours, and a lack of specificity for the advertised isoform with some antibodies[121, 146] The three isoforms are differentially expressed in the normal colonic mucosa; with very low TGFβ1 expression, TGFβ2 staining in the columnar epithelium, but not in the crypt cells, and TGFβ3 expression throughout[121]. On analysis by IHC,

TGFβ1 and TGFβ2 are significantly overexpressed in CRC compared to normal colonic tissue, TGFβ3 expression is also increased in CRC, but non-significantly, largely due to the higher baseline expression [121].

As latent TGFβ binds to the surface of platelets, Tregs and macrophages, the site of activation of TGFβ may be far removed from that of production. Additionally, free TGFβ1 can be measured in the plasma of healthy individuals at a level of around 4ng/ml, whereas TGFβ2 and 3 are not thought to be detectable in plasma[147]. Plasma TGFβ1 and TGFβ2 levels are higher in CRC patients compared to healthy controls, although plasma levels of TGFβ3 were not investigated in this study due to lack of a specific ELISA for TGFβ3[121]. .

The vast majority of published work on TGFβ *in vitro* uses only TGFβ1. For TGFβ1, RNAseq has characterised TGFβ-response signatures (TBRS) for cultured CD4+ T-cells, fibroblasts, endothelial cells and macrophages[112], in addition to specific studies of TGFβ1 in numerous other cells types. CD4+ T-cells are the only one of these cell types for which the isoforms have been extensively compared[143], where induction of Th17 cells with TGFβ3 rather than TGFβ1 leads to a more pathogenic phenotype for autoimmune disease. There is little published work on the effects of TGFβ2 or 3 on other immune cells or fibroblasts. In epithelial cells TGFβ3 is thought to be the main contributor towards EMT[123, 124].

Studies of the contribution of each isoform to poor prognosis in CRC have led to differing conclusions, with some suggesting TGFβ1 and 3 are most strongly associated with worse outcome[112], and others that TGFβ2 is the main driver of metastatic disease[52]. Calon et al, compared recurrence rates in stage 1-3 colorectal cancers divided into those tumours in the lowest, middle, and highest thirds for bulk mRNA expression of each TGFβ isoform. This showed significantly greater recurrence rates in *TGFB1* high tumours compared to those middle or low

for *TGFB1*, and significantly higher recurrence rates with *TGFB3* high compared to *TGFB3* low or medium tumours and with *TGFB3* medium compared to *TGFB3* low tumours. There was, however, no differences in recurrence rates between *TGFB2* low, medium or high tumours[112]. Jackstadt et al. directly looked at *TGFB2* expression and recurrence, finding more recurrences in *TGFB2* high tumours, and also compared gene expression profiles of organoids that metastasised and resembled CMS4 tumours in mice with those that did not metastasise, and found significantly higher *TGFB2* in the organoids that metastasised, but no significant difference in the other isoforms[52].

Given potential side effects from cancer treatment targeting TGFβ generically, there has been some interest in using specific inhibitors that target some, but not all of the 3 isoforms of TGFβ. AVID200 is a receptor ectodomain trap that has been designed to specifically target TGFβ1 and TGFβ3 without binding TGFβ2. The reasoning is that TGFβ2 is known to be required for normal cardiovascular function[148] and haematopoiesis and that treatment with the pan-TGFβ inhibitor Galunisertib had led to cardiac toxicity[149]. Conversely, specific inhibition of TGFβ2 has been attempted in some tumour types though using oligo-mRNAs to this isoform[52, 150].

1.3.6.8 Fine-tuning the fibroblast phenotypes

Despite their important role, *in vitro* TGFβ1 and IL-1β stimulation of primary fibroblast from normal tissue cannot completely replicate the myCAF and iCAF signatures derived from the scRNAseq data[96]. Many other cytokines may play a role in fine-tuning fibroblast phenotypes, resulting in the multiple clusters seen from the scRNAseq data. TNFα is another cytokine that activates the NF-κB pathway and can also significantly upregulate iCAF marker genes, such as IL-6 and CXCL1[113]. Work in lung cancer suggests that autocrine IL-11 drives myCAF differentiation and autocrine IL-6 promotes iCAF differentiation[96]. This concurs with work in PDAC, which suggests autocrine IL-6 and LIF drive JAK/STAT3 signalling to maintain the iCAF phenotype[113]. Other cytokines that signal through STAT3, such as Oncostatin M (OSM), epidermal growth

factor and hepatocyte-derived growth factor, or other members of the IL-1 family, such as IL-18 and IL-33, may contribute to the many precise fibroblast phenotypes depicted on unsupervised clustering. Indeed, elevated OSM[151], IL18[152], IL-33[153] and HGF[154] have all been associated with more advanced stage and worse outcomes CRC, and EGFR inhibition is already widely used in CRC, due to its direct effects on the malignant cells. Potential mechanisms of targeting both myCAF and iCAF development or function are shown in **Figure 1.3.4.7.**

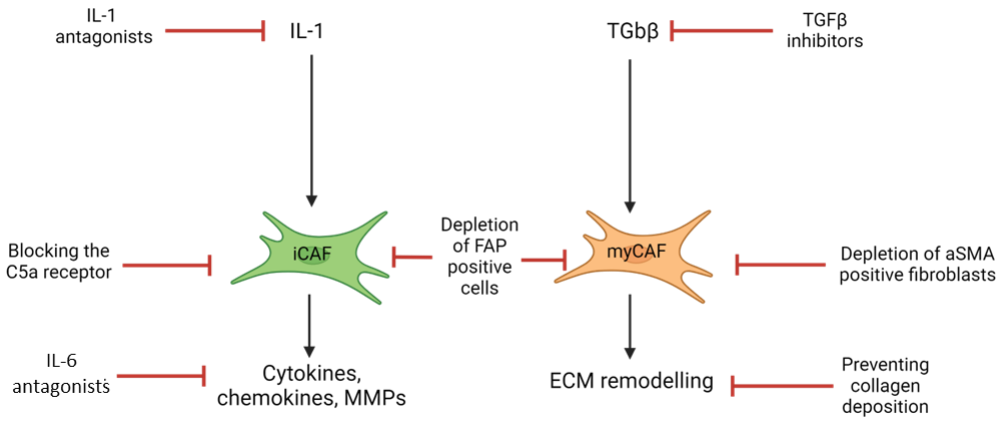


Fig 1.3.6.7 Mechanisms to target Fibroblast subsets for CRC treatment
Several trials in mice have tried to modulate the function of specific fibroblast subsets. As well as inhibition of the two driving ligands, attempts have been made to reduce the numbers of myCAFs by depleting αSMA positive fibroblasts[88], or to modulate their function through preventing collagen deposition[89]. Attempts have been made to block formation of the iCAF phenotype through targeting of the C5a receptor[50].

1.3.7 Fibroblasts across disease

1.3.7.1 Permanency of the Fibroblast subsets

Both scRNAseq and *in situ* imaging give information about a tissue at a snapshot in time. It is unclear if the fibroblast phenotypes found using these techniques are temporary responses to stimulation, stages along a differentiation pathway, or differentiated fibroblast subtypes. If they develop along separate differentiation pathways, the degree of permanence versus plasticity of these cell states is also debatable.

Pseudotime analysis in lung cancer suggests myCAFs and iCAFs are separate differentiation trajectories, but the subsets within these two groups are stages along the differentiation pathway towards those two cell states[96]. Another lung cancer study found an intermediate activation state CAF in early stage tumours, that was thought to be the pre-cursor to both iCAF-like T-cell permissive FAP+ αSMA- CAFs and myCAF-like T-cell excluding FAP+ αSMA+ CAFs seen in more advanced stage tumours[109]. However, others have suggested iCAFs could be an intermediary between 'normal' fibroblasts and myCAFs[101].

It has been possible to skew cells towards myCAF and iCAF phenotypes *in vitro* through growing cells in differing culture conditions, such as in monolayer to resemble myCAFs, and in 3D culture to represent quiescent cells or generate iCAFs[98]. The ability to alter fibroblast state by growth in new culture conditions suggests inherent plasticity[113], though whether these fibroblasts generated *in vitro* represent the full phenotype seen *in vivo* is not clear. A CRC study looking at chemokine and MMP expression in iCAFs, generated by treatment with organoid conditioned supernatant, found that the effects were largely reversed by treatment with regular medium or an IL-1 antagonist for 24 hours[114], suggesting an iCAF phenotype is a temporary response to stimulation. However, another study found that the IL-6 and IL-8 producing CAF phenotype was maintained *in vitro* for at least 14 days after isolation of these cells from human breast tumours, due to a feed-forward loop between autocrine C5a production and NF-κB phosphorylation[100].

Notably, TGFβ-stimulation of fibroblasts can lead to epigenetic changes which persist even after the initial cytokine stimulation is withdrawn[155, 156]. Kidney fibroblasts treated *in vitro* with 5-days of TGFβ1 have shown increased DNA methylation and persistent gene expression changes after TGFβ1 withdrawal, whereas the effects of 24 hours TGFβ1 stimulation on gene expression were reversible with no accompanying epigenetic changes[157]. Whether long-term IL-1 stimulation leads to epigenetic alterations in fibroblasts remains to be determined. While several studies

have reported epigenetic changes in CAFs[158-160], one study comparing CRC and normal adjacent fibroblasts found only a small proportion of differentially expressed genes between the two groups were associated with changes in promotor methylation[161]. Given a risk factor for CRC development is chronic inflammation, alterations in cell state developing only after longer-term environmental changes are likely to be more clinically-relevant than shorter term stimulations.

Another piece of evidence that fibroblasts from tumours may undergo longer term changes are the results of studies comparing the response of tumour and matched normal adjacent fibroblasts to cytokine stimulation. Investigations in lung cancer have shown hyper-responsiveness of CAFs to TGFβ1[162] and similar experiments in colorectal cancer have suggested *ex-vivo* CAFs are hyper-responsive to IL-1β[163]. These alterations in response could indicate changes in expression of receptor molecules, altered expression of downstream signalling cascade proteins, or epigenetic changes to affected genes.

A further line of evidence that CAFs undergo permanent changes is the finding that, in CRC, CAFs have many more chromosomal copy number changes than fibroblasts from matched normal adjacent tissue[164]. The mechanism behind this was not fully explored, but possibilities include the generation of reactive oxygen species in the tumour resulting in genetic damage.

To define a group of cells as a discrete subset, rather than a temporary activation state, often key transcription factors are identified; for example, RORγt specifies Th17 cells and FOXP3 expression is indicative of a Treg. While we know that myCAF differentiation is triggered by TGFβ1 leading to nuclear localisation of the SMAD transcription factor, and phosphorylation and nuclear localisation of the transcription factor NFkB occurs in response to IL-1 on the way to iCAF formation, these transcription factors are ubiquitously expressed. RUNX1, RUNX2 and CEPB are transcription factors with increased expression and chromatic accessibility in CRC CAFs

compared to normal colonic fibroblasts[165], but as yet no transcription factors with expression specific to one fibroblast phenotype have been identified.

1.3.7.2 Fibroblast subsets across disease stage

The majority of research into the role of fibroblasts in cancer has looked at established primary tumours from patients having resections. In addition, work is needed to better understand the role of the CAF phenotypes in pre-cancerous lesions and metastatic disease. A recent study found a pre-CAF population in tubular adenomas, that were distinct from the CAFs seen in established CRC, but had more similarity to CAFs than that of normal fibroblasts on gene expression scores[165]. Whether the emergence of pre-CAFs and later CAFs is a cause or consequence of CRC tumour progression remains to be explored, though another study of pre-cancerous changes in the breast found that fibroblasts generated inflammation that drove development of neoplastic cells[166].

Study of metastatic lesions is an important requisite for cancer treatment, given that it is metastases that lead to fatality. Che *et al*. found a lower frequency of CAFs in liver metastases than matched primary tumours[167] and Giguelay *et al.* found CAFs in the liver metastases were associated with a better prognosis desmoplastic invasion pattern and lower Ki67 in the tumour cells[168]. scRNAseq revealed an enrichment in myCAFs in the metastatic lymph nodes of patients with primary head and neck cancer as compared to their primary tumours[169], but there has been little fibroblast phenotyping in CRC metastases. Wang *et al*. found greater number of MCAM positive mesenchymal cells in liver metastases than in the primary, which interact with a CXCL13+ CD8 T-cell subset, and that the F3+ fibroblast subset seen surrounding tumour nests was completely absent in the metastatic deposits[104]. Brouwer *et al* found greater numbers of fibroblasts, as well as macrophages and Tregs, in metastatic tumour deposits within the fat surrounding the bowel than in primary tumours or lymph nodes, and a corresponding increase

in tumour deposits being classified as imCMS4[25]. However, this study did not investigate variation within fibroblasts.

We do not know if fibroblasts present in metastases come from the primary, either migrating in association with the cancer cells or earlier to form a niche, or were local to the site of metastases prior to cancer developing. Giguelay *et al.* found CAFs from CRC liver metastases mapped closely with a local fibroblast subset, both identified through scRNAseq[168]. A recent investigation of lung cancer metastases from breast cancer models in mice found IL-1β acting on fibroblasts prior to the formation of metastases stimulated prostaglandin production which had immunosuppressive effects on myeloid cells, and blockade of the prostaglandin receptors reduced metastatic burden[170]. It therefore seems most likely that CAFs in metastatic deposits originate from the site of metastasis. Study of metastatic disease is complicated by the fact that the nature of the metastasis may differ by site; in CRC lung metastases had greater immune infiltration, whereas peritoneal metastases had more fibroblasts and were associated with worse outcomes[171].

1.3.7.3 Fibroblast subsets and Prognosis

The relative contribution of these subtypes to poor prognosis in breast and pancreatic tumours has been studied, with conflicting results. Work in pancreatic cancer has suggested that iCAFs promote tumour-suppressive environments and poor prognosis, whereas myCAFs are protective, possibly through forming a defensive barrier around the tumour preventing migration of neoplastic cells[83]. Whereas studies in breast cancer have suggested the opposite, that iCAFs attempt to generate an anti-tumour response which is impeded by myCAFs forming a physical barrier to immune cells around the tumour. In breast cancer myCAF numbers negatively correlated with CD8+ T-cell infiltration, whereas the opposite was true for iCAFs[95]. In the same study, response to immunotherapy in metastatic melanoma and non-small cell lung

cancer patients, was found to be negatively correlated with a myCAF gene expression signature in publicly available datasets.

In CRC some studies have suggested that high IL-1R1 levels on CAFs are associated with worse prognosis[163] and that more iCAFs are associated with poorer responses to radiotherapy in rectal cancer[114]. However, there has also been work suggesting myCAFs are associated with poor prognosis. Additionally, while CMS4 disease is associated with greater total numbers of fibroblasts and poor prognosis, it is unclear if the increased fibroblast numbers in CMS4 is due to expansion of a specific fibroblast phenotype or all subsets equally. If a particular fibroblast subset is expanded in CMS4, other CMS groups may have a different fibroblast phenotype predominating.

In addition to understanding how the fibroblast subsets impact on prognosis, it will be important to understand how these change in response to cancer treatment. Che *et al*. found that, comparing resections of primary and metastatic CRC from individual that had and had not undergone neo-adjuvant chemotherapy, there were more ECM-CAFs expressing chemokines, MMPs and collagens in the treatment-naïve groups and more contractile-CAFs expressing ACTA2, TAGLN and RGS5 in the post-chemotherapy samples[167].

These apparent contradictions on the impact of iCAFs and myCAFs on prognosis may be due to differences in the underlying biology of heterogenous tumours, resulting in the prognostic implications of the CAF subsets depending on factors such as underlying mutations, the tumour stage, what treatments are being given, or other alteration in the TME. A full understanding of the role of fibroblasts in variable tumours will allow appropriate patient selection for any new therapeutics. Any targeting of fibroblast subsets conferring poor-prognosis will need to return

these cells to a genuinely quiescent state, rather than converting them to a different phenotype also driving progressive disease.

1.4 Research hypotheses and experimental aims

Fibroblasts have long been considered a target for CRC treatment given their association with poor prognosis, but generic targeting has not led to treatments entering the clinic. Fibroblasts cannot be considered in isolation, their phenotype both reflects and shapes the underlying tumour microenvironment, which is heterogeneous within tumours and between individuals. A better understanding of the heterogeneity of fibroblasts in CRC and how they interact with the multitude of other cells in the TME to impact disease course is needed for effective targeting of these cell types for cancer treatment.

Our first aim is therefore to identify CAF subsets associated with different, well defined, functions: myCAF and iCAF cells *in situ* in multiple human colorectal cancer sections. By studying these fibroblast subsets *in situ* we can assess which other cell types or structures within the tissue they locate with, giving information about their likely interactions and function. We also aim to correlate frequency of these fibroblast phenotypes with other characteristics of individual tumours, such as molecular subtype, mutations, responses to treatment and survival.

Our second aim is to define the intrinsic and external factors leading to CAF transformation and how these compare to normal fibroblasts. To this aim, we will study primary stromal cells extracted from resection samples from localised human CRC and the adjacent normal tissue. To study the phenotype of these cells both unstimulated and after stimulation with two hypothetical driver cytokines of the iCAF and myCAF phenotypes; IL-1β and TGFβ respectively. This is to see if the fibroblasts from the tumour respond differently to these key cytokines, we hypothesise there will be a feed-forward loop where cells already exposed to TGFβ or IL-1β will become more responsive to repeat stimulation, resulting in the generation of more divergent phenotypes.

TGFβ is also known to be associated with poor prognosis CRC and TGFβ inhibitors are under investigation in clinical trials, but understanding of the relative contributions of the three isoforms of TGFβ to CRC development and progression is limited. Elucidating the effects of TGFβ2 and TGFβ3 on key cell types in TME, to see if they lead to the same pro-tumorigenic phenotypes as TGFβ1 signalling, would be helpful in considering the future role of more specific TGFβ targeting for CRC treatment. Our third aim is therefore to compare the outcomes of stimulating fibroblasts and peripheral blood mononuclear cells (PBMCs) with each of the TGFβ isoforms, to identify any differential effects.

2 Materials and Methods

2.1 Ethical approval

Patients eligible to donate CRC resection specimens were identified through surgical screening programmes at Oxford University Hospitals. All patients gave informed consent. Storage of CRC tumours was under a human tissue authority (HTA) licenced tissue bank, and ethical approval was obtained from the UK National Research Ethics Service (NRES), Rec Nos. 16/YH/0247. Cones from healthy blood donors were obtained and stored under NRES, Rec Number 11/H0711/7. Slides from the Grampian arm of the Stratification in Colorectal Cancer (S:CORT) study were held under REC 15/EE/024 from the East of England – Cambridge South Research Ethics Committee.

2.2 Cell Isolation

2.2.1 Digestion and expansion of fibroblasts from resection tissue

External muscle and adipose tissue were removed, the rest of the tissue containing the mucosal and submucosal fraction were washed with Dulbecco's Phosphate buffered saline (PBS) containing antibiotics: 1% Penicillin/Streptomycin (SIGMA, P0781, Penicillin 10,000U/ml, Streptomycin 10mg/ml), Gentamicin 50ug/ml (SIGMA, G1397, stock 50mg/ml), Amphotericin B 2ug/ml (SIGMA, A294, stock at 20mg/ml) and ciprofloxacin 1ug/ml (SIGMA 17850 5GF 10mg/ml in 0.1N HCl). For the first wash 1mM Dithiothreitol (DTT) was added to prevent disulphide bonds forming between proteins and cross-linking. In the following two washes 5mM ethylendiaminetetraacetic acid (EDTA, Sigma-Aldrich) was added to the Dulbecco's PBS with antibiotics to remove the epithelial cells. Tissue was then minced with surgical scissors and underwent multiple rounds of digestion with RPMI containing 5% fetal bovine serum (FBS), 5mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 1% Penicillin/Streptomycin , collagenase A (stock 3 U/ml, used at 0.3 U/ml) and 1:200 DNase 1 (stock 2mg/ml) (all Sigma-Aldrich). After each 30-minute digestion at 37°C the liquid was

filtered and the single cell filtrate was spun down and resuspended in PBS with 5mM EDTA, to chelate calcium and prevent further digestion by collagenase A, and 0.1% Bovine serum albumin (BSA). Remaining cells were topped up with fresh digestion medium until no more cells were liberated. Stromal cells were resupended in Minimum Essential Medium alpha (MEMα) + 20%FCS + Penicillin/streptomycin (P/S) + HEPES, and expanded by plating the single cell suspension, at approx. 0.5 million cells/cm2, onto plastic cell culture vessels, the adherent cells were then retained.

2.2.2 Extraction of PBMCs from cones

Peripheral blood mononuclear cells (PBMCs) were extracted using a lymphoprep (Stem Cell Technologies, 07861) density gradient and remaining erythrocytes lysed with ammonium-chloride-potassium (ACK) lysing buffer. PBMCs were stored in 10% Dimethyl sulfoxide (DMSO):90% foetal calf serum (FCS) (Atlanta biologicals) in liquid nitrogen.

2.3 Cell line culture

2.3.1 Cytokines and pharmacological inhibitors

Recombinant Human (rh) TGFβ1, rhTGFβ2 and rhTGFβ3 (Thermofisher, PHG9214, PHG9114 and PHG9305 respectively) were reconstituted with 40mM acetic acid (TGFβ1) or distilled water (TGFβ2) and then prepared into working aliquots of 10µg/ml in 0.1% BSA (all isoforms). rhIL-1β (Peprotech 200-01B) was reconstituted in distilled water to a concentration of 0.1mg/ml and then prepared into aliquots of 10ug/ml in 0.1% BSA. IL-2 (Thermofisher, PHC0026) was reconstituted in 100mM acetic acid, then diluted in 0.1% BSA. Vactosertib (Stratech, S7530-SEL) was reconstituted in DMSO to a stock concentration of 10mmol/L.

2.3.2 Maintenance and stimulation of PBMCs

PBMCs were thawed and washed in RPMI 1640 media (Gibco), supplemented with 5% FBS and 1% P/S (Sigma). Cells were suspended in PBS at 1x106 cells/ml and labelled with 1µM Cell Trace

Violet (CTV) for 20 minutes at room temperature prior to washing and then seeded at 500,000 cells in 500uL RPMI 1640 media (Gibco), supplemented with 10% serum knockout replacement (Thermofisher, A3181502) and 1% P/S, in a 48-well plate. Cells were stimulated with either anti-CD3 (2µg/ml) (OKT3, Biolegend) and anti-CD28 (1µg/ml) (CD28.8, Biolegend), or anti-CD3 (1µg/ml) and rhIL-2 (100 units/ml), and rhTGFβ (0-10ng/ml), for 3-7 days.

2.3.3 Maintenance and stimulation of fibroblasts and primary stromal cells

Ccd-18Co fibroblasts (ATCC) or Primary stromal cells were maintained in MEMα (Gibco) supplemented with 10% FCS, 1% P/S and 5mM HEPES. For collection of mRNA cells were plated in 12-well plates at 5,000 cells/cm2 (for TGFβ comparison experiment in Chapter 5) or 10,000 cells/cm2 (in RNAseq experiments in Chapter 4) and for collection of protein at 90,000 cells/cm2 (to measure pSMAD2 in Chapter 5) or 25,000 cells/cm2 (for fibroblast/pericyte markers in chapter 3). Once at confluence the cells were starved of FCS for 24 hours prior to cytokine stimulation.

2.3.4 Maintenance and stimulation of hPL-PC pericytes

hPC-PL pericytes (PromoCell) were maintained in Pericyte Growth Medium 2 (Merck, C-28041) and the included Supplement Mix. For collection of mRNA cells were plated in 12-well plates at 10,000 cells/cm2 and for protein at 25,000 cells/cm2. Once at confluence the cells were starved for 24 hours prior to cytokine stimulation.

2.3.5 MTT test

2.5×10^3 Ccd-18Co cells per well, were seeded in 96-well plates and allowed to adhere for 1 hour before stimulation with 0-5ng/ml TGFβ alone or in combination, or 1µmol/ml Vactosertib, for 48, 72 or 96 hours. Number of viable cells was estimated by adding 100µg of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to each well 2 hours prior to the end of the incubation. Supernatants were aspirated and formazan particles solubilized with

DMSO. Absorbance was measured at 570nm on a BMG SPECTROstar NANO Microplate Reader (BMG Labtech).

2.4 Gene expression analysis

2.4.1 RNA extraction and cDNA synthesis

Cells were lysed with 350uL per well of RLT buffer with 7ul of 2mM DTT. 350uL of 70% ethanol was added to lysates and RNA was extracted using the ZymoBIOMICS DNA/RNA Miniprep kit (Zymo Research) for experiments in Chapter 5 and with the Qiagen RNeasy Micro kit for experiments in Chapter 3. In both on-column DNase digestion was performed. Samples were eluted from columns in RNase free, DNase free H20. cDNA was synthesised using the High Capacity cDNA kit (applied biosystems) according to the manufacturer's protocol using a C1000 Touch Thermal Cycler (Bio-Rad). Quantities of reagents per sample are shown in **Table 2.4.1.1.** Synthesised cDNA was diluted with ddH_20 and used for PCR analysis.

Table 2.4.1.1: Reagents for synthesis of cDNA

Reagent	**Volume per sample (uL)**
Eluted RNA	6
10x RT buffer	1
25x dNTP mix (110mM)	0.4
10x RT random primers	1
MultiScribe TM reverse transcriptase	0.5
RNase inhibitor	0.5
H20	0.6
Total Volume	**10**

2.4.2 RT-qPCR

Real-time quantitative PCR (RT-qPCR) was performed using Taqman probes (life Technologies) **(Table 2.4.2.1)** and Precision Fast 2xqPCR Mastermix with LR (long range) in 384-well plates (FrameStar), prepared in the quantities shown in **Table 2.4.2.2**. Plates were sealed with MicroAmp Optical Adhesive Film (Life Technologies) and run on a ViiA7 Real-Time PCR system (Applied Biosystems) using the fast run method **(Table 2.4.2.3)**. Data was analysed using $2^{-\Delta CT}$,

where $\Delta CT = C_{T(\text{target gene})} - C_{T\ (\text{endogenous control})}$. Large ribosomal protein (RPLPO) was used as the endogenous reference gene for all qPCR analyses.

Table 2.4.2.1: Taqman primers used for RT-qPCR

RPLPO	Hs99999902_m1
ACTA2	Hs00909449_m1
COL1A1	Hs00164004_m1
CXCL1	HS00605382_gh
CXCL12	Hs03676656_mh
IL-6	Hs00985639_m1
IL-11	Hs01055413_g1
MCAM	HS00174838_m1
MKI67	Hs04260396_g1
MMP10	HS00233987_m1
PDGFRA	HS00998026_m1
PDPN	HS00366766_m1
RGS5	HS00186212_m1
TIMP3	Hs00165951_g1
TGFβ1	Hs00998133_m1
TGFβ2	Hs00234244_m1
TGFβ3	Hs01086000_m1
THY1	HS00174816_m1
VEGFA	Hs00900055_m1

Table 2.4.2.2: RT-qPCR Reaction Mix per well

Reagent	**Volume for rxn (uL)**
Diluted cDNA	2.9
Precision fast 2xqPCR Mastermix	3
TaqMan Real Time PCR Assay	0.1
Total Volume	**6**

Table 2.4.2.3: RT-qPCR thermal profile, fast run method

Stage	**Step**	**Ramp rate**	**Temperature**	**Time**
Hold stage	1	1.9°C/s	95°C	2 minutes
PCS stage	1	1.9°C/s	95°C	5 seconds
Number of cycles: 40	2	1.6°C/s	60°C	20 seconds

2.4.3 RNA sequencing

To prepare sequencing libraries RNA quality and concentration were assessed using the Agilent Tapestation and equal quantities of RNA were loaded per sample. RINe was 9.4 or above for the majority of samples.

Sequencing was performed by the Oxford Genomics centre. Libraries were prepared using the Ultra-low SMARTseq2 protocol with 1ng of RNA per sample and were sequenced on the NovaSeq 150bp, PE platform.

Raw data was processed using a pipeline by Dr Nicholas Ilott. Quality control was performed with fastQC (v.0.11.9). Gene mapping and counts were produced by kallisto (v.0.46.0).

Subsequent analysis was performed using R software. Analysis included Principal Component Analysis (PCA) of transformed read counts (log2 counts per million), differential gene expression analyses (DESeq2) and gene ontology (GO) enrichment analysis (fgsea).

2.5 Protein expression analysis

2.5.1 BCA Assay

Protein quantification was performed with the Pierce Bicinchoninic acid (BCA) Protein assay (ThermoFisher) according to manufacturer's instructions and analysed at 562nm on a SpectroStar Nano.

2.5.2 Western blotting

Cells were lysed in Radioimmunoprecipitation assay (RIPA) buffer (Sigma-Aldrich) with 1:100 HALT protease/phosphatase inhibitor (Thermo-Scientific). Equal amounts of protein, determined by BCA assay, were reduced with 0.05M DTT, denatured at 99◦C, and supernatants were loaded into precast NuPAGE Novex 4-12% Bis-Tris Gels (Life Technologies), separated by SDS-PAGE, and transferred onto Immobilon-FL PVDF membranes (Millipore). Membranes were blocked (Odyssey Blocking Buffer, Li-COR), incubated with primary antibodies **(Table 2.5.2.1)**

overnight at 4°C, and fluorescently conjugated secondary antibodies for 1 hour at room temperature, before detection using the Odyssey CLx detection system (Li-COR).

Table 2.5.2.1: Antibodies used for Western blotting

Detection stage	Reactivity	Dilution – change to actual concentration	Clone	Manufacturer
Primary	Rabbit anti-p-SMAD2	1:1,000	138D4	Cell signalling, 3108S
Primary	Mouse anti- β-actin	1:10,000	AC-15	Sigma, A1978
Primary	Goat anti-PDGFRα	1:2,000	AF-307-NA	RnDSystems
Primary	Rabbit anti-PDGFRβ	1:1,000	28E1	CST, 3169
Secondary	Goat anti-rabbit-800CW	1:15,000		LiCOR 925-32211
Secondary	Donkey anti-goat-800CW	1:15,000		LiCOR 926-32214
Secondary	Donkey anti-mouse-680RD	1:15,000		LiCOR 925-68070

2.5.3 ELISA

IL-11 was quantified in supernatants using Human IL-11 DuoSet ELISA (R&D Systems, DY218) according to the manufacturer's instructions. Capture antibody was allowed to adsorb to the plate overnight, plates were then washed and blocked before 50uL of sample in triplicate, or standards in duplicate, were incubated for 2 hours at room temperature. After washing detection antibodies were incubated for 2 hours, and following further washes Streptavidin-Horseradish peroxidase (HRP) was added for 20 minutes, followed by further washes. Substrate solution was applied and after 20 minutes stop solution added and absorbance was measured at 450nm by a spectrophotometric plate reader (SPECTROstar Nano).

2.5.4 Flow cytometry

PBMCs were washed with Fluorescence activated cell sorting (FACS) buffer (PBS + 5mM EDTA + 0.1% BSA) and spun (1500rpm for 4minutes), stained for 30 minutes at 4°C **(Tables 2.5.4.1-2)** and washed again before fixation in Fix/Lyse solution (BD). Samples were run on an LSR (BD) and analysed using FlowJo version 10 (Tree Star).

Table 2.5.4.1: Panel for flow cytometry analysis of immune checkpoint expression

Reactivity	Fluorophore	Manufacturer	Clone	Dilution
Live/dead	APC-Cy7	eBiosciences	N/A	1:1000
CD3	APC	Biolegend	OKT3	1:100
CD4	PE-Cy7	Biolegend	OKT4	1:100
CD8	AF700	Biolegend	SK1	1:100
CD25	BV510	Biolegend	M-A251	1:100
CD39	BV421	Biolegend	A1	1:100
CD45RA	BV785	Biolegend	HI100	1:100
CTLA4	PE	Biolegend	BNI3	1:100
PD1	BV711	Biolegend	EH12.2H7	1:100
TIGIT	BV605	Biolegend	A15153G	1:100

Table 2.5.4.2: Panel for flow cytometry analysis for proliferation and activation markers

Reactivity	Fluorophore	Manufacturer	Clone	Concentration (ug/ml)	Dilution
Cell Trace Violet	BV421/510	Invitrogen			1uM
Live/dead	APC-Cy7	eBiosciences	N/A		1:1000
CD3	APC	Biolegend	OKT3	0.5	1:100
CD4	PE-Cy7	Biolegend	OKT4	1.5	1:100
CD8	AF700	Biolegend	SK1	1	1:100
CD56	FITC	Biolegend	HCD56	2	1:100
CD25	BV605	Biolegend	BC96	1	1:100
CD69	BV785	Biolegend	FN50		1:100
CD16	BV711	Biolegend	3G8	1.4	1:100
NKG2D	PE	BDBiosciences	1D11		1:100

2.6 In situ analysis

2.6.1 IHC on FFPE tissue

Sections of tumour and normal adjacent were fixed in 4% formaldehyde (Fisher scientific UK) overnight and then stored in 70% ethanol the next day. Samples were embedded in paraffin (Kennedy Histology Facility) and 5µm sections cut using a Leica RM2235 microtome. Sections were baked at 60°C for 60 minutes prior to use. Deparafinisation and rehydration was performed with the 'Tissue-Tek DRS 2000' (Sakura), with 2x 3 minutes Xylene, 2x 1 minute 100% ethanol, and then slides were placed in water. Slides were permeabilised with 0.3% TRITON 100x (Sigma) for 10 minutes. Antigen retrieval was for 20 minutes with each of pH6 citrate (DAKO) and pH9 TRIS-EDTA target retrieval solutions using a pressure cooker. Blocking was performed with Bloxall endogenous peroxidase, Avidin, and Biotin (when the secondary used was biotinylated). Sections were washed with TBS between each blocking step and all steps were

performed at room temperature (RT). A final blocking step with 10% serum (Sigma) of the species secondary antibodies were raised in in Tris Buffered saline (TBS, Insight Biotechnology) was not followed by washing.

Primary antibodies **(Table 2.6.1.1)** were incubated overnight at 4°C in TBS with serum. Slides were washed and, for biotinylated primaries, secondary amplification was at room temperature for 2 hours, followed by StrepAB/HRP complex for 30 minutes. Chromogenic visualisation with Vector laboratories DAB substrate was for 10 minutes, and counterstaining with Mayer's Haematoxylin for 3 minutes. Slides were dehydrated to xylene using the 'Tissue-Tek DRS 2000' , with 3x 1 minute in 100% ethanol and 2 x 1 minute in xylene, and mounted with DPX mountant (Sigma-Aldrich) and coverslips placed using the Tissue-Tek Glass. Slides were visualised on an Olympus BX51 microscope.

2.6.2 Immunofluorescence

Initial steps for immunofluorescence staining on FFPE sections were identical to those mentioned in 2.6.1. However, no avidin or biotin blocking was required. Primary antibodies were either directly conjugated to a fluorophore or conjugated secondary antibodies were incubated for 1 hour at RT. Nuclei were stained with Hoechst 28332 (Life technologies) prior to washing, mounting and visualisation on a fluorescent microscope.

Table 2.6.1.1: Antibodies for IHC

Detection stage	Reactivity	Clone	Conjugation	Species and Isoform	Concentration	Manufacturer
Primary	Podoplanin	NZ-1.3	-	Rat	2.5ug/ml	eBioscience
Primary	IL-6	ab9324	-	Mouse IgG2a	5ug/ml	Abcam
Primary	pSMAD2	138D4	-	Rabbit	0.5ug/ml	Cell signalling technology
Primary	CD45	2D1	AF647	Mouse	5ug/ml	BDBiosciences
Secondary	Anti-mouse		AF555	Donkey	1:500	
Secondary	Anti-rat		AF488	Donkey	1:500	
Secondary	Anti-rabbit		AF555	Donkey	1:500	

2.6.3 Immunofluorescence on Vectar Polaris

Imaging using a 6-plex panel, opal detection and scanning on the Vectar Polaris was kindly performed by the Translational Histopathology lab. Antibodies used are shown in **Table 2.6.2.1** for the fibroblast subsets panel and **Table 2.6.2.2** for the immune panel.

Table 2.6.2.1: Fibroblast subsets panel for immunofluorescence

Reactivity	Clone	Manufacturer	Concentration	Opal
IL-6	10C12	Leica	1:200	520
Phospho-Smad2	138D4	CST	1:400	690
CD34	EP373Y	Abcam	1:3,000	480
CD45	2B11 + PD7/26	Dako	1:200	570
Podoplanin	D2-40	GeneTex	1:200	620
Cytokeratin	AE1/AE3	Dako	1:1,500	780

Table 2.6.2.2: Immune panel for immunofluorescence

Reactivity	Clone	Manufacturer	Concentration	Opal
CD66b	G10F5	Biotechne	1:1,000	520
HLA-DR	TAL 1B5	Santa Cruz	1:10,000	620
CD3	LN10	Leica	1:500	570
CD45	2B11 + PD7/26	Dako	1:200	480
CD56	CD564	Leica	1:75	690
CD68	PG-M1	Dako	1:400	780

2.7 Statistical analysis

Due to the nature of the study, no statistical methods were used to predetermine sample size. Data are presented as individual points, mean and +/- 95% confidence intervals. Statistical analysis was performed in GraphPad Prism, version 8, using ANOVA with Turkey correction for multiple comparisons, or in R. Statistical tests were 2-sided, and p values less than 0.05 were considered significant.

3 Identifying stromal cell location within CRC

3.1 Introduction

Many studies have used scRNAseq to identify different clusters of fibroblasts in multiple types of cancer[95, 97, 103, 106, 109, 169, 172-176]. Similar gene signatures are seen repeatedly in many of these datasets, and two broad functional programmes often detected are myCAFs and iCAFs[68, 95-97, 101, 103]. In a few studies these fibroblast phenotypes have been identified through other methods, such as by IHC[98, 103] or flow cytometry[99], but only in small numbers of samples. To date, these cell types have not been shown *in situ* across CRC specimens from multiple patients. Therefore, conclusions drawn from these studies about the location and abundance of myCAFs and iCAFs are mostly speculative. Moreover, findings about myCAF and iCAF location in tissue have been generalised across all samples, rather than looked at in the context of patient subsets such as the one provided by the CMS classification. Given the huge heterogeneity between CRCs from different patients, and known variation in total fibroblast abundance between CMS subgroups, it is likely that relative contributions of fibroblast subsets to tumour development will also be heterogenous. Relative abundance of these two fibroblast subsets have not yet been conclusively linked to prognosis, or other features of the tumour such as immune cells landscape or mutational status. Those studies that have aimed to locate myCAFs and iCAFs within the TME have led to conflicting results on where the subsets are found relative to ulceration, neoplastic cell and vessels. For example, Pelka *et al* found iCAFs around dilated blood vessels of the colonic luminal margin[103], whereas analysis of tissues from several inflammatory diseases by Korsunsky *et al.*[88] described a myCAF-like group as vascular fibroblasts closely associated with blood vessels. These may be due to differences in markers used to detect these phenotypes.

My aim in this chapter is to identify myCAF and iCAF cells histologically across multiple CRC sections from human tumours for which molecular and histological classification has been determined. I aim to study their location within the tumour microenvironment and identify the

neighbouring cells of these CAF subsets by looking at where these cells are relative to the neoplastic cells, tumour vasculature and specific immune cells. Understanding with which cells types in the TME the fibroblasts interact will provide information about how they arise and their role in disease progression. I hypothesise that these cells will be found in different niches within the TME, suggesting interactions with different cell types and that their relative proportion will vary between CMS subgroups alongside immune cells subsets such as neutrophils, T cells and macrophages. I will also correlate the relative abundance of myCAFs and iCAFs to mutation status and patient outcomes. I have validated antibodies to stain for specific iCAF and myCAF markers which will be used alongside other pan immune cell and stromal cell markers by multiplex imaging on the Vectra polaris.

3.2 Stromal subsets panel on the Grampian dataset

3.2.1 Demographics of the 60 Grampian cases

To identify myCAFs and iCAFs, 60 cases from the Grampian arm of the S:CORT (Stratification in Colorectal Cancer) study were selected. The specimens were all biopsies from patients with locally advanced rectal cancer, who were subsequently treated with chemoradiotherapy with capecitabine. As part of the S:CORT dataset, demographic data and information about staging, treatment response, cancer recurrence and survival had been collected for these patients, shown in **Table 3.2.1.1.** The median age of 67.5 and slight male predominance, 36 cases versus 24, is typical of CRC. Nearly half of cases went on to have a complete response to chemoradiotherapy and there were very few recurrences among this selection of cases, limiting the utility of Disease-free survival (DFS) outcomes. As well as the availability of demographic data, as part of the S:CORT study, the biopsies had previously been used to perform micro-array, from which conventional CMS, Intrinsic CMS, CRIS and PDS grouping had been predicted, as well as TGFβ response signatures (TBRS). TBRS are a series of gene signatures derived from bulk RNAseq of cultured fibroblasts, endothelial cells, macrophages and T-cells stimulated with TGFβ1 at 5ng/ml for 24 hours[112]. The conventional CMS, Intrinsic CMS and CRIS allocations were not calculated in exactly the same way as in the original papers where these methods were developed. All three came from micro-array data rather than bulk sequencing, and from the entire human tumour including the stromal and immune components, whereas CRIS was derived from patient-derived xenographs using human-specific primers so that only the tumour cells were studied[29] and Intrinsic CMS was from a combination of cell lines, organoids and patient-derived xenographs[28]. Histological CMS classification had also been performed on the S:CORT cases, from H&E stains of the biopsies. **Table 3.2.1.2** describes the different classification methods used for the tumour in our study. The dataset also included mutation data, copy number analysis and methylation status for selected genes.

Table 3.2.1.1 Demographic data from the 60 selected cases from the Grampian study

Characteristic	**Number of Patients (n=60)**
Age	Median 67.5 (range 43-91)
Gender	Males (36), Females (24)
Dukes stage	Duke's A (12), Duke's B (5), Duke's C (15), NA (28)
TNM Stage	T1N0 (1), T2N0 (10), T3N0 (24), T4N0 (1), T2N1 (3), T3N1 (16), T3N2 (4), T4N2 (1)
Response to Treatment	Minimal (2), Partial (14), Good partial (16), Complete (28)
Recurrence	Recurred (7), Disease free (53)
Histological CMS	imCMS1 (15), imCMS2 (26), imCMS3 (14), imCMS4 (5)
Conventional CMS	CMS1 (10), CMS2 (16), CMS3 (14), CMS4 (3), Unclassified (17)
Intrinsic CMS	CMS1 (9), CMS2 (11), CMS3 (12), CMS4 (20), Unclassified (8)
CRIS	CRIS-A (15), CRIS-B (3), CRIS-C (15), CRIS-D (6), CRIS-E (12), Unclassified (9)
PDS	PDS1 (17), PDS2 (19), PDS3 (24)
MSI status	MSI (2), MSS (58)
KRAS mutation	Mutant (33), Wild-type (27)
BRAF mutation	Mutant (1), Wild-type (59)
APC mutation	Mutant (51), Wild-type (9)
SMAD3 mutations	Mutant (2), Wild-type (58)
SMAD4 loss	Homozygous loss (5), Heterozygous loss (42), No loss (7), NA (6)

Table 3.2.1.2 Classification methods used in this thesis. Publications where they were originally described, and the methods used in the original publication and in the S:CORT dataset.

Description used in this thesis	Reference for original publication	Original method	Modified method used in S:CORT
Conventional CMS	Guinney *et al*, Nature Med, 2015[20]	Bulk RNAseq from whole tumour	Micro-array from bulk tumour
Intrinsic CMS	Eide *et al*, Sci Rep, 2017[28]	Bulk RNAseq from organoids, cell lines and xenographs	Micro-array from bulk tumour
Histological CMS	Sirinukunwattana *et al*, Gut, 2021[34]	H&E sections of patient tumours	H&E sections of patient tumours
CRIS	Isella *et al*, Nature Commun, 2017[29]	Bulk RNAseq from patient derived xenographs with human-specific primers	Micro-array from bulk tumour

When CMS was measured by three different methods there were very different numbers of cases within each category, with histological CMS identifying large numbers of CMS2 and few CMS4, conventional CMS finding very few CMS4, but similar numbers of CMS 1-3, and intrinsic CMS identifying many cases of CMS4. It is unexpected that intrinsic CMS identified so many

cases of CMS4, given it is designed to look specifically at the neoplastic cells, rather than the TME. Of note with both conventional CMS and intrinsic CMS many cases were unclassified, whereas histological CMS gives every case a positive classification. Looking at CRIS classification of the 60 tumours, few cases of CRIS-B, the stromal grouping, were seen. We therefore looked at the overlap between the different classification methods in identifying the same CMS group **(Fig 3.2.1.1 A)** and there did seem to be fairly good overlap between the CMS groupings. We then looked at concordance between CRIS grouping the CMS grouping by all 3 methods **(Fig 3.2.1.1 B).** For this CMS 1 and 3 were combined, as they should both correspond to CRIS-A, and CRIS C, D and E were combined, as these should all overlap with CMS2. CMS4 should overlap with CRIS-B. We found reasonable concordance between CMS1+3 and CRIS-A, and CMS2 and CRIS-C-E, with all CMS types, but no CMS4 tumours by any classification were described as being CRIS-B. This may reflect that one of the analysis methods is not working well in this dataset, given the differences between this analysis and the original methods used. We would expect greatest concordant results between CRIS and intrinsic CMS, as both classification methods are meant to study the transcriptomic profile coming from only the tumour cells themselves, but all three methods of determining CMS gave similar concordance with expected CRIS grouping.

We studied the relationship between CMS groupings and response to treatment and recurrence. Fewer complete responses were seen in histological CMS4 disease than the other groups, which is consistent with the known poorer outcomes in stromal high disease. Histological CMS and Intrinsic CMS showed more complete responses in the CMS1 immune, whereas with conventional CMS there was no correlation with response **(Fig 3.2.1.1 C).** None of the few recurrences identified were seen in histological or conventional CMS4 disease, though a few were in intrinsic CMS4 **(Fig 3.2.1.1.D)**.

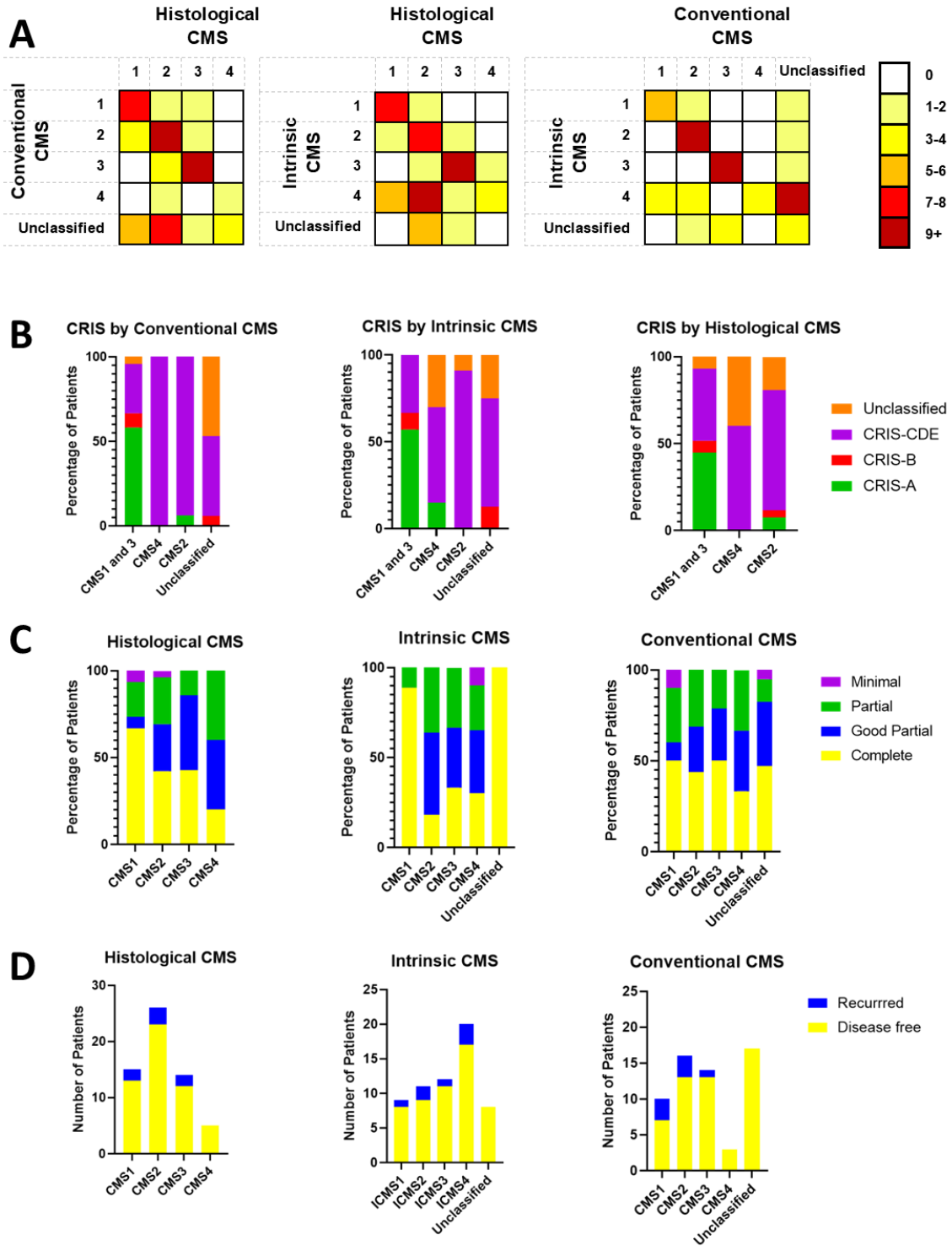


Fig 3.2.1.1 CMS and CRIS subtyping of the 60 selected cases from the Grampian study

(A) Concordance between CMS grouping when performed by Histological CMS, Intrinsic CMS or Conventional CMS. (B) Correlation between CRIS groups with CMS by each of the three methods. CMS 1 and 3 care combined as these should overlap with CRIS-A, and CRIS C-E are combined as these should overlap with CMS2, and CMS 4 should overlap with CRIS-B. (CD) Treatment Response by CMS subgroup, (D) Recurrence status by CMS group.

3.2.2 Markers to Identify myCAFs and iCAFs

Two tumour sections were stained for each of the 60 patients, with a stromal panel and an immune panel **(Table 3.2.2.1).** This was kindly performed by the Translational Histology lab and imaged using the Vectra polaris microscope. This uses an opal system, allowing for a 6-plex panel of markers with primary antibodies from the same species. **Fig 3.2.2.1 A** shows the mechanism of opal staining, where the HRP conjugated to the secondary antibody activates a tyramide-fluorophore, which then binds to tyrosine residues on the tissue. When the tissue is heat-treated the primary and secondary antibodies are removed, but the tyramide-fluorophores remain in situ, allowing repeated staining with difference fluorophores.

Table 3.2.2.1 Markers chosen for each of the panels and the main cell type they identify

Fibroblast Panel		**Immune Panel**	
Marker	**Cell type**	**Marker**	**Cell type**
Pan-CK	Epithelial cells	CD45	Immune cells
CD45	Immune cells	CD66b	Neutrophils
CD34	Endothelial cells	CD3	T-cells
PDPN	Fibroblasts	CD56	NK-cells
pSMAD2	myCAFs	CD68	Macrophages
IL-6	iCAFs	HLA-DR	NK-cells

For the stromal 6-plex panel, PDPN was chosen as a generic CAF marker, pan-cytokeratin (CK) to identify epithelial cells, CD45 for haematopoietic cells and CD34 for endothelial cells. CD34 is generally considered to be an endothelial cells marker, though has been used as a marker of the adventitial fibroblast subset in lung cancer[177]. However, in the Pelka *et al.* scRNAseq dataset *CD34* mRNA was only expressed at low levels in the stem-niche fibroblast subsets (both CAF and normal) and not expressed in any of our CAF subsets of interest. An alternative endothelial cell marker would be CD31 (gene name *PECAM1*), but this can also be expressed on granulocytes, lymphocytes and colon cancer cells. PDPN was chosen as a marker of cancer associated fibroblasts based on its expression in the myCAF and iCAF clusters (clusters 27-29) from Pelka *et*

al., but lack of expression in most of the fibroblast clusters seen in adjacent normal tissues from the dataset **(Fig 3.2.2.1 B)**. Another marker that was considered was PDGFRα, but this was expressed at a lower level than PDPN in the main clusters of interest (sC27-29) and was more highly expressed in the normal fibroblast clusters cS23-24. FAP was an alternative potential fibroblast marker, but this was only highly expressed in clusters 28 and 31. **(Fig 3.2.2.1 B).** Therefore, a single marker for each of the myCAF and iCAF subsets was required to distinguish these subsets amongst the PDPN+ CAFs.

IL-6 was chosen as an iCAF marker based on its specificity for iCAFs (cluster 29, MMP3+ CAFs) in the scRNAseq study by Pelka *et al* **(Fig 3.2.2.1 B)** and its previous use as a marker for histological identification of iCAFs by Ohlund *et al*. IL-6 staining with the Leica antibody was first validated by immunohistochemistry using chromogenic DAB staining **(Fig 3.2.2.1 C),** and then by immunofluorescence alongside other markers to detect fibroblasts (PDPN) and leukocytes (CD45) and imaged on the CellDive fluorescent microscope scanner **(Fig 3.2.2.1 D)**. The cytoplasmic IL-6 staining in the cells with elongated nuclei that also had cytoplasmic PDPN staining suggested these were iCAFs.

Nuclear pSMAD2 was chosen as the marker for myCAFs, as this indicates TGFβ signalling, the hypothetical driver of this fibroblast phenotype, and the lack of specifically upregulated genes in the myofibroblast cluster from the Pelka *et al* dataset. Other potential myCAF markers were considered: while αSMA (gene name *ACTA2*) has often been previously used as a myCAF marker, it is also expressed on other CAF subsets, albeit at a lower level, and is a marker of smooth muscle cells and pericytes present in homeostasis **(Fig 3.2.2.1 B)**. pSMAD2 was also validated by DAB staining and then by fluorescence staining **(Fig 3.2.2.1, E+F)** and was seen in the nuclei of cells staining positive for PDPN, suggesting these are fibroblasts responding to TGFβ. Positive nuclear staining for pSMAD2 was also seen in CD45+ immune cells and the tumour cells.

Immune cells would be expected to respond to TGFβ and, while TGFβ response is lost in roughly half of CRCs, this is usually due to loss of SMAD4 and so phosphorylation of SMAD2 would still be expected.

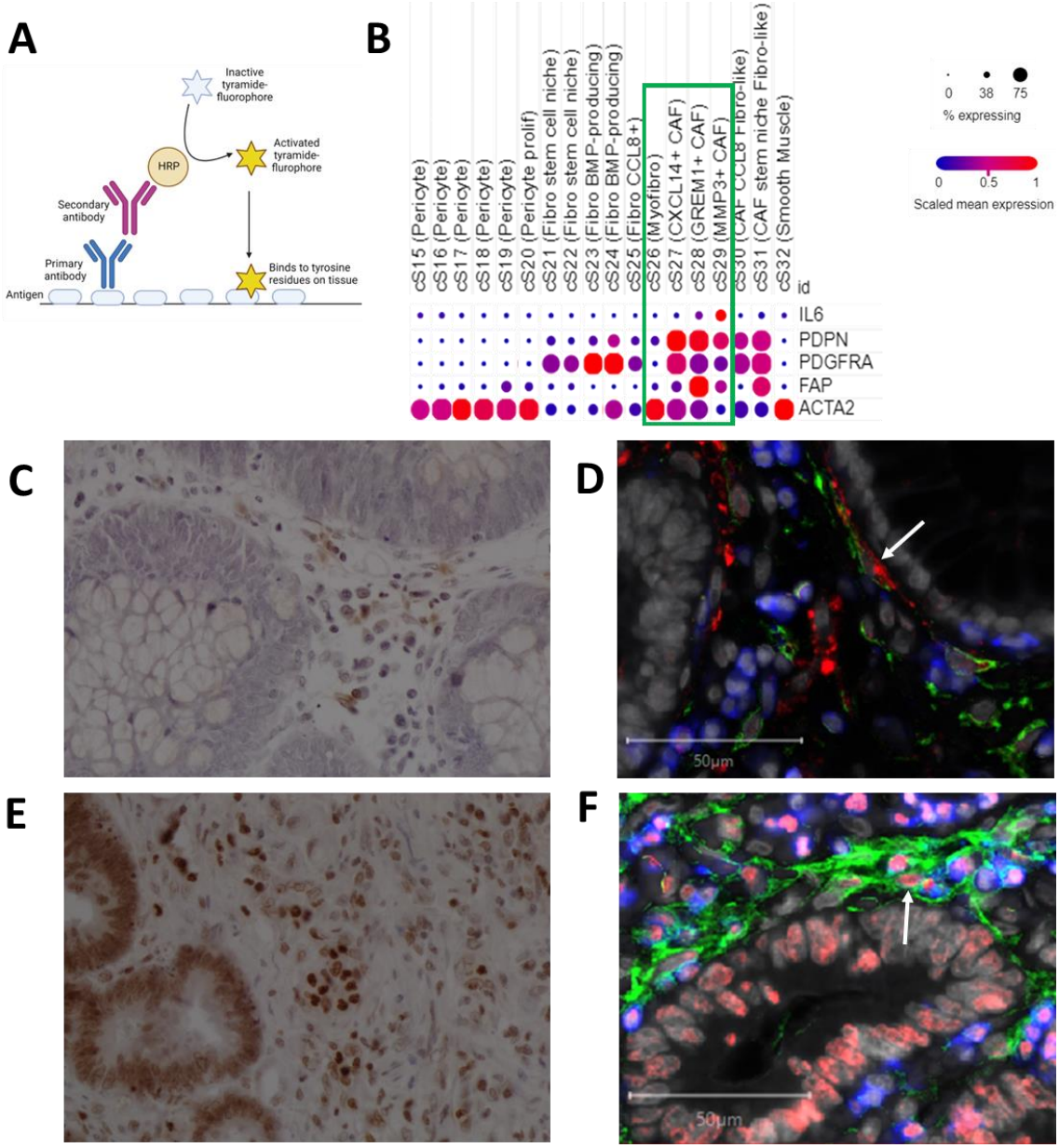


Fig. 3.2.2.1 Validation of markers for 6-plex imaging panel to locate fibroblast subsets.
(A) Diagram of staining with the opal system on the Vectra-Polaris. **(B)** Expression of potential fibroblast markers within each cluster from Pelka *et al.* scRNAseq data. Clusters of interest highlighted in green. **(C-F)** IHC of CRC sections. **(C)** DAB staining for IL-6. **(D)** Imaging using CellDive: IL-6 red, PDPN green, CD45 blue, DAPI grey. White arrow highlights iCAF cell. **(E)** DAB staining for pSMAD2. **(F)** Imaging using CellDive: pSMAD2 red, PDPN green, CD45 blue, DAPI grey. White arrow highlights myCAF cell.

Fig 3.2.2.2 gives examples of the individual stains on the Vectra-Polaris, and the combination together for the fibroblast subsets panel. The staining for DAPI, CD45, CD34, PDPN, pSMAD2 and IL-6 were all clear and specific. The pan-CK staining did not work well in identifying areas of tumour epithelium. Larger tumour nests can be recognised by morphology and therefore this did not affect their identification. However, clearer staining of the tumour epithelium may have helped identify isolated clusters of tumour cells budding into the surrounding TME.

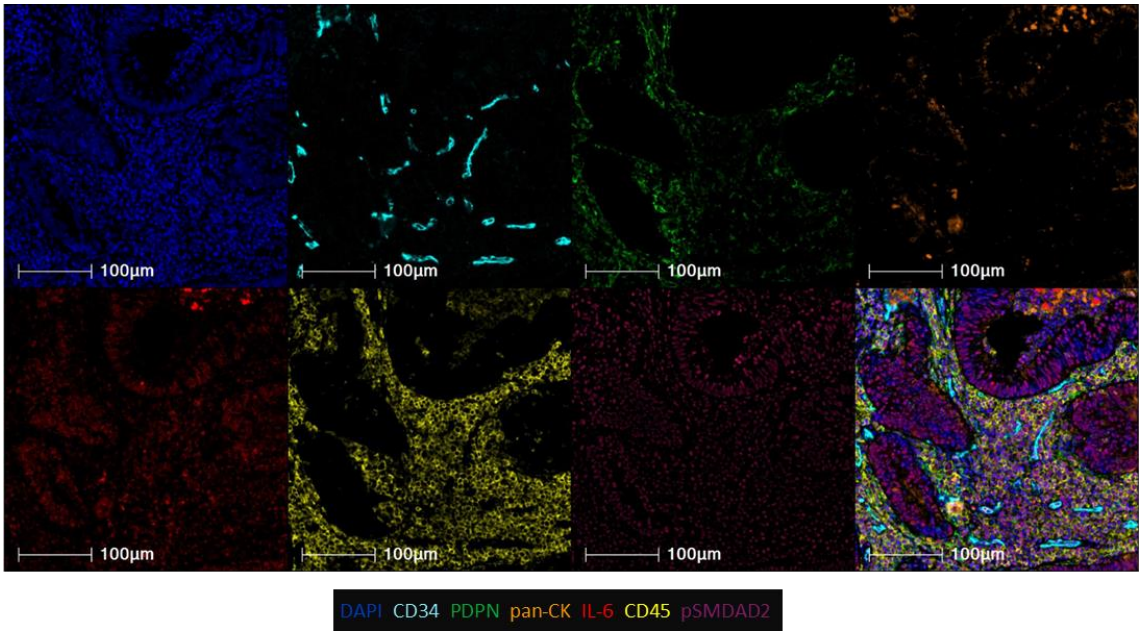


Fig 3.2.2.2 Staining the Fibroblast subsets panel
Staining with the 6-plex fibroblast subsets panel on the Vectra-Polaris. Individual stains with DAPI, CD34, PDPN, pan-CK, IL-6, CD45 and pSMAD2 are shown, followed by the combination, for the same section of tumour.

3.2.3 Image Analysis of the Fibroblast panel by Cell Segmentation

The stromal panel was analysed using HALO v 3.5.3577.140. First a region of tumour was annotated for further analysis, as these were biopsy specimens, there were often regions of adjacent normal epithelium or muscle tissue without tumour infiltration on the slide, which were not included in the analysis **(Fig 3.2.3.1 A)**. 59 out of 60 tumour sections were annotated, one had no clear area of tumour on the biopsies and so could not undergo further analysis. The second step was to use the Mininet (HALO AI) classifier to define the entire annotated region as

one of 3 classes: tumour epithelium, the surrounding tissue microenvironment (TME), or glass **(Fig 3.2.3.1 B)**. This was done by drawing multiple example regions across all the slides and then using machine learning, classification of small regions, and providing further example drawings until performance of the classifier was satisfactory to perform across the annotated region. This gave total area of tumour epithelium and TME for the areas annotated as tumour for each of the 59 slides. Adding these together, the total area of tissue analysed for each slide ranged from 0.2 - 38mm^2, with a median area analysed of 2.51mm^2 and a mean of 5.26mm^2. On some slides several, small regions were annotated, meaning these figures do not always represent one continuous area. TME area was divided by tumour epithelium area to give a ratio for each slide. TME:Epithelium ratio for the 59 cases was analysed, divided by histological CMS, conventional CMS, intrinsic CMS and CRIS subgroups **(Fig 3.2.3.1)**. This shows slightly higher average proportions of stroma for the CMS4 subgroups, which is consistent with other studies in the literature, but this was not significant.

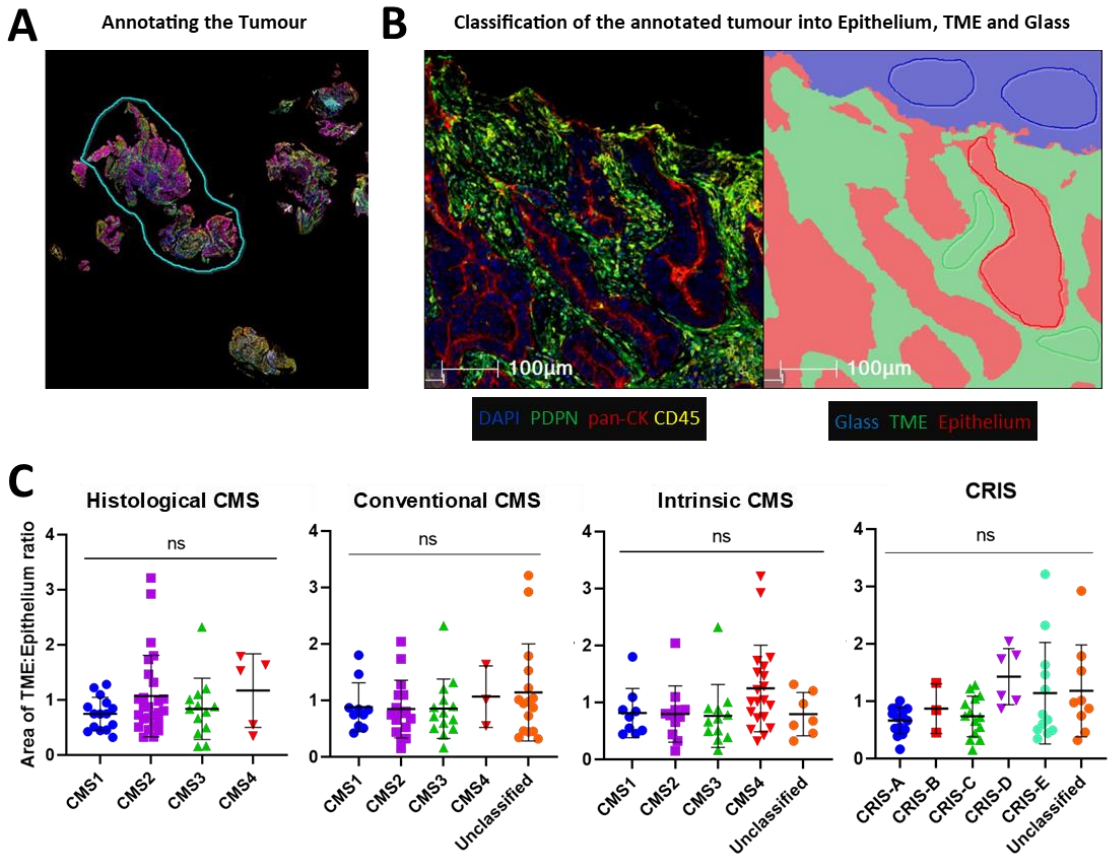


Fig 3.2.3.1 TME : Epithelial area ratio by CMS and CRIS subgroup
(A) Tumour regions for analysis were annotated. (B) Examples were drawn for machine learning to classify the area of interest into tumour epithelium (red), TME (green) and glass (blue). (C) The area classified as TME divided by the area classified as epithelium for each slide shown by histological CMS, conventional CMS, intrinsic CMS and CRIS groups. Data analysed by one-way ANOVA (ns=non-significant).

Having identified the tumour microenvironment region the following step was segmentation of the cells in that region. For this the HALO HighPlex FL v.4.2.3 was used where nuclei are identified and then cell borders are drawn around these up to a maximum diameter of 1.5um. A small maximum diameter was chosen to prevent areas outside of an elongated, narrow cell being attributed incorrectly to a cell, which would lead to greater inaccuracy than missing some of the cell. Once cell segmentation was performed thresholds were set for each of the markers defining the intensity level in the nucleus or cytoplasm for which a cell would be classified as positive for that marker **(see Appendix 1)**. Phenotypes are then set depending on positivity or negativity for certain markers. **Fig 3.2.3.2 A-C** shows examples of cells phenotyped as immune (CD45+ CD34-), endothelial (CD34+) or fibroblasts (PDPN+ CD34- CD45-). While CD34 was used

as a negative marker for immune cells, CD45 was not used as a negative marker for endothelial cells, so that immune cells inside a vessel would not necessarily be identified separately.

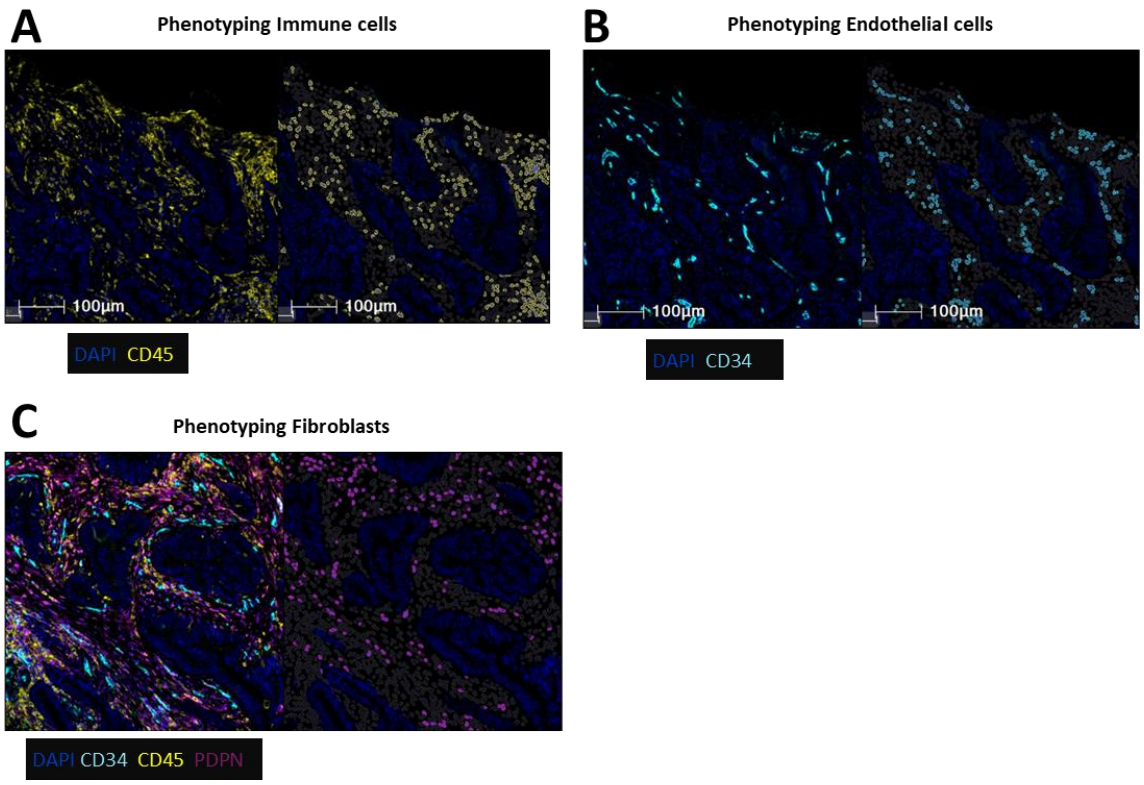


Fig 3.2.3.2 Cell segmentation and phenotyping in HALO
(A-C) Cell segmentation in the regions classified as TME and phenotyping of these cells as: (A) Immune cells (CD45+, CD34-), (B) Endothelial cells (CD34+), and (C) Fibroblasts (PDPN+ CD45– CD34-).

We then looked at the proportion of each cell type including CAF, immune and endothelial cells across CMS and CRIS classification **(Fig 3.2.3.3 A).** While we expected to see an enrichment of immune cells in CMS1, there was no clear patterns to the data. **Fig 3.2.3.3 B** show the percentage of cells in the region classified as TME that were classified as fibroblasts by CMS subgroups and CRIS subgroups. It shows no correlation between percentage of fibroblasts in the TME and these, which is surprising as we would expect to see a higher proportion of fibroblasts in CMS4 and CRIS-B disease.

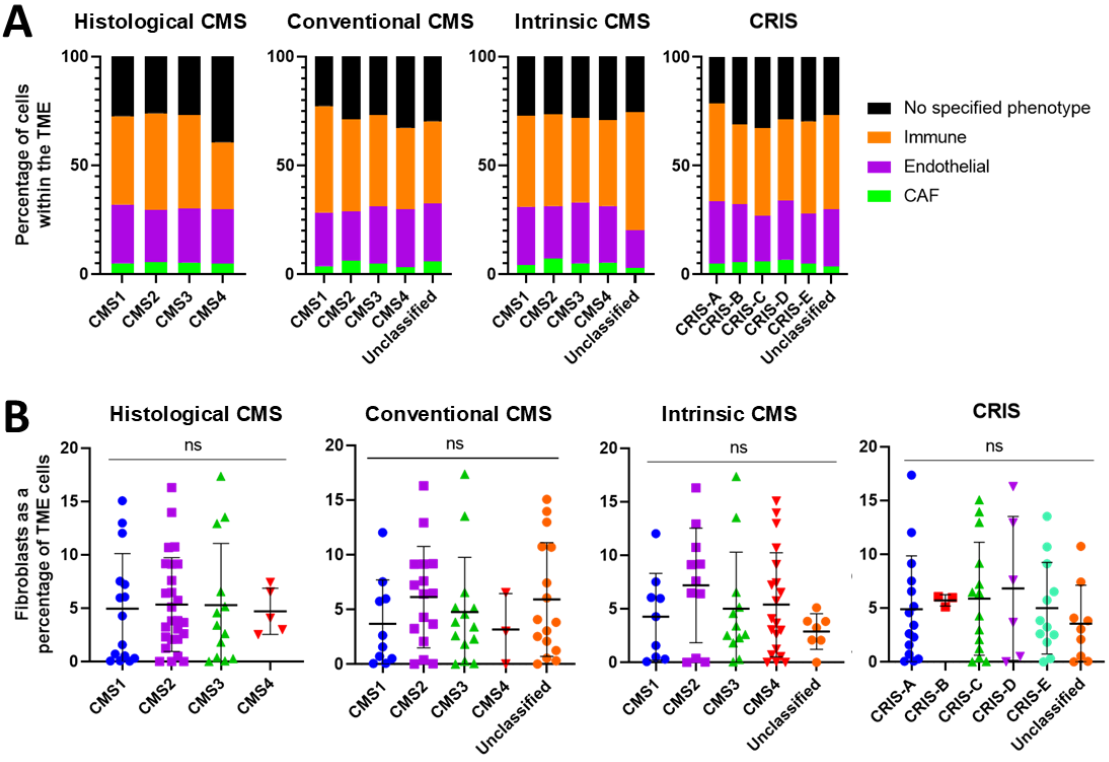


Fig 3.2.3.3 Cell types within the TME by imCMS, CMS, ICMS and CRIS subgroups

(A) CAFs (PDPN+ CD34- CD45-), immune cells (CD45+ CD34– PDPN-) endothelial cells (CD34+) and cells not otherwise specified as a percentage of cells in the TME by histological, conventional and intrinsic CMS, and CRIS subgroups. (B) The percentage of cells in the area classified as TME that were defined as fibroblasts (PDPN+ CD45– CD34–) by Histological CMS, Conventional CMS, Intrinsic CMS and CRIS groups. Data analysed by one-way ANOVA (ns=non-significant).

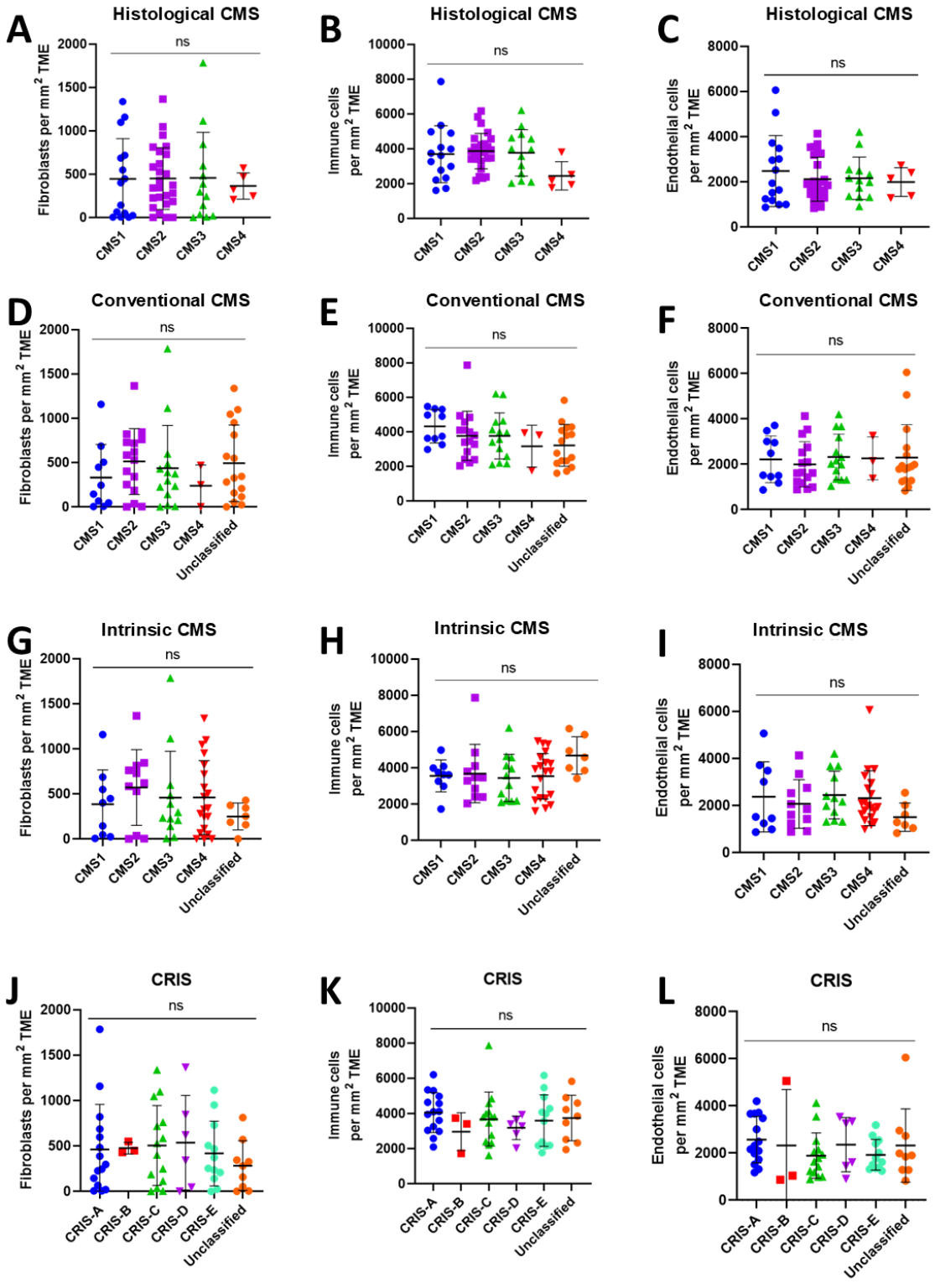


Fig 3.2.3.4 Fibroblasts, Immune and Endothelial cells per mm^2 of the TME by CMS and CRIS groups CAFs (PDPN+ CD45– CD34–) **(A,D,G,J),** Immune cells (CD45+ CD34-) **(B,E,H,K)** and Endothelial cells (CD34+) **(C,F,I,L)** per mm2 of the TME by **(A-C)** imCMS, **(D-F)** CMS **(G-I)** ICMS and **(J-L)** CRIS groups. Data analysed by one-way ANOVA (ns=non-significant).

Fig 3.2.3.4 shows CAFs, immune and endothelial cells per mm^2 of TME by CMS groups, and again no significant trends are seen. **Fig 3.2.3.5** shows associations between the number of fibroblasts per mm^2 of TME and the associated metadata, including mutations, response and survival. The only association reaching significance was fewer fibroblasts were seen in FBXW7 mutation (p=0.015), a known tumour suppressor in many human cancers and low FBXW7 expression is associated with poorer outcomes in CRC[178].

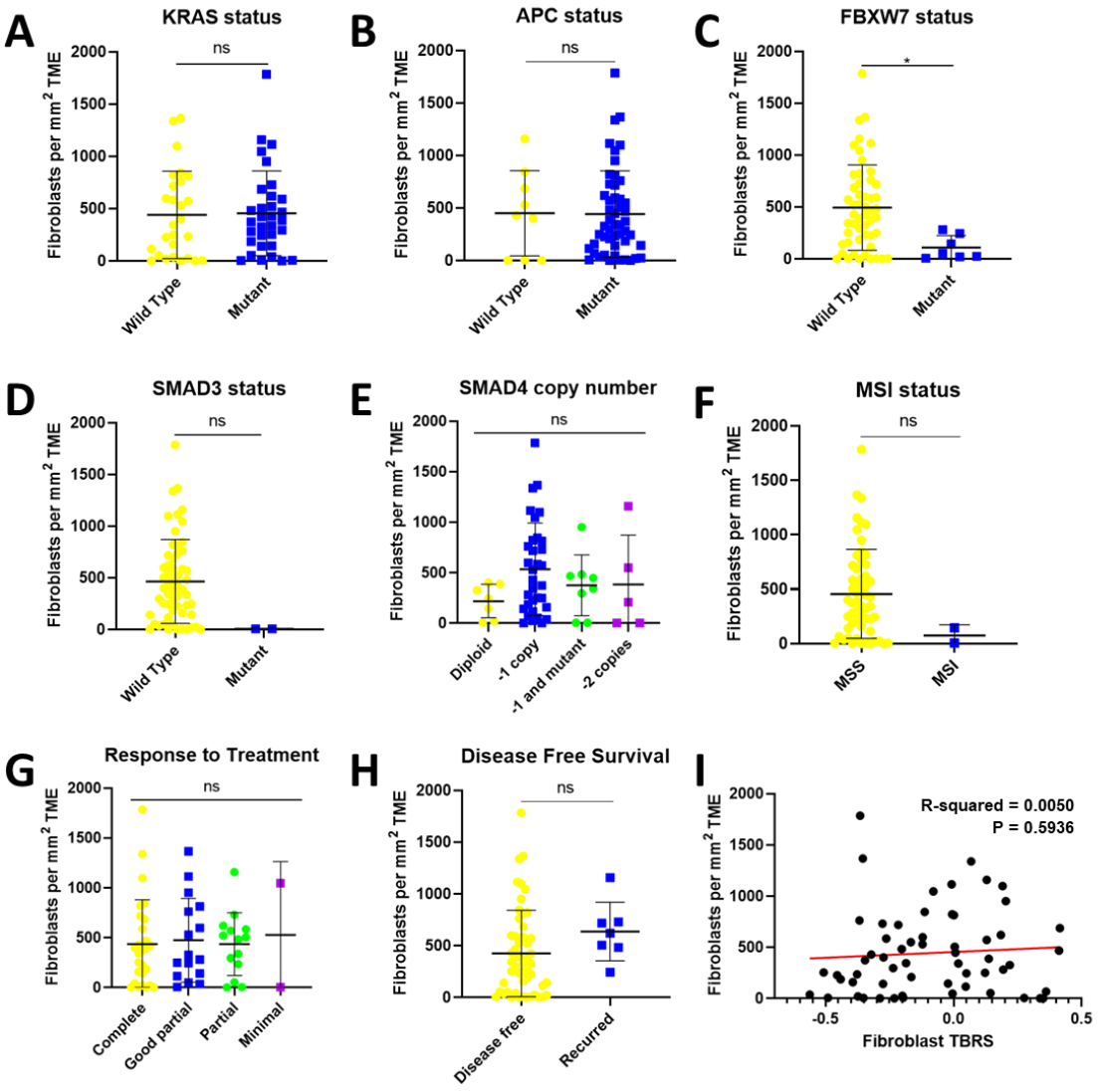


Fig 3.2.3.5 Fibroblast per mm^2 of TME and correlations with the metadata
Fibroblast numbers per mm^2 of TME and **(A)** KRAS mutation, **(B)** FBXW7 mutation, **(C)** SMAD 3 mutation, **(D)** SMAD4 deletion, **(E)** response to treatment and **(F)** disease free survival. Data were analysed by T-test **(A-D, F, H),** one-way Anova **(E+G)** and simple linear regression **(I)** (*p<0.05, ns=not significant).

The cells classified as fibroblasts (PDPN+ CD34- CD45-), were further subdivided into myCAFs (also pSMAD2+ IL-6), iCAFs (also IL-6+ pSMAD2-) or other CAFs (either pSMAD2+ IL-6+ or pSMAD2- IL-6-). **Fig 3.2.3.6 A+B** shows examples of these. No correlation was seen between the percentage of cells in the TME that were positive for pSMAD2 and IL-6, meaning that across all cell types in the TME these markers are independently expressed from each other **(Fig 3.2.3.6 C).** We then looked at phenotyping of all the cells in the region classified as TME, by histological CMS subgroup **(Fig 3.2.3.6 D)**. 25.1% of cells were endothelial and 41.9% immune cells overall. We found that only a very small percentage of the cells in the TME were classed as fibroblasts, an average of 5.2% across all tumours, with only 2.0% of all cells in the TME being defined as myCAFs and 0.2% being defined as iCAFs. Numbers of iCAFs in each slide ranged from 0 to 473, in over half of the tumours analysed less than 10 iCAFs were counted, the median was 8. Numbers of myCAFs ranged from 0 to 2322, with a median of 134. There were large numbers of cells that were not phenotyped as any of fibroblasts, immune cells or endothelial cells, 27.7% overall, and their proportion seemed higher in histological CMS4, suggesting that stromal cells within this CMS could be a different origin than PDPN+ fibroblasts. The majority of these were negative for all three of PDPN, CD45 and CD34, though a very small number of cells were positive for more than one marker. The cells negative for all 3 markers could include pericytes, smooth muscle cells and neuroendocrine cells. Cells positive for CD34 and PDPN could be lymphatic endothelial cells, PDPN positive immune cells have also previously been described.

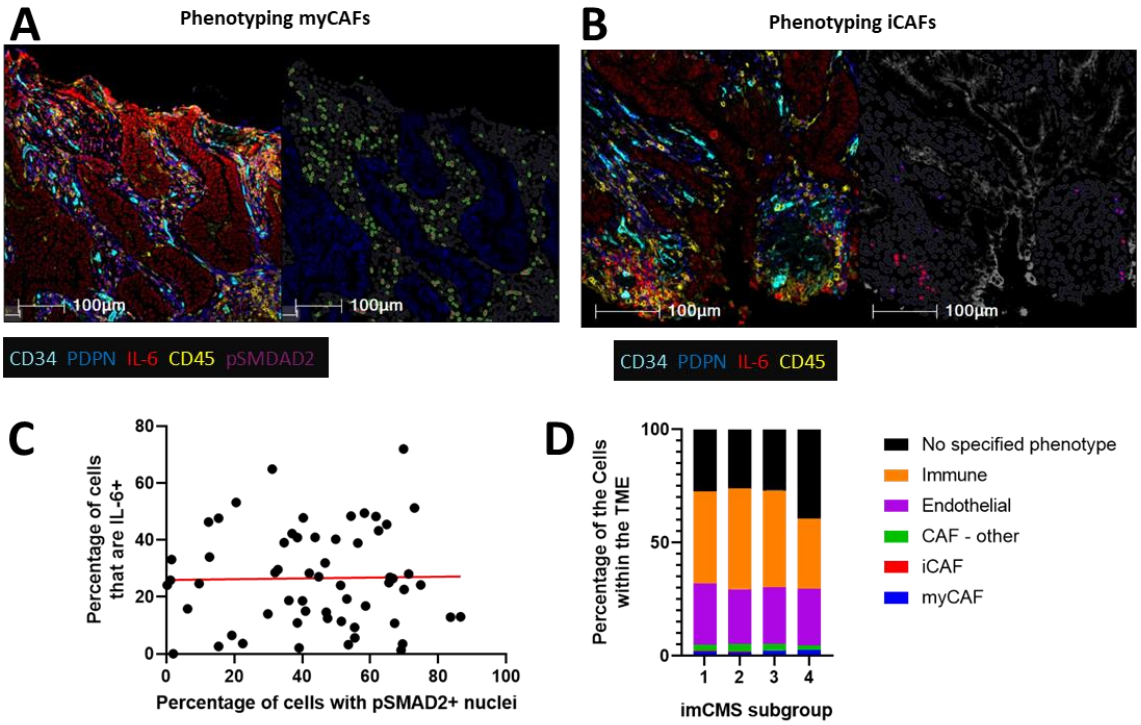


Fig 3.2.3.6 Cell segmentation and phenotyping of myCAFs and iCAFs

(A-B) Cell segmentation in the regions classified as TME and phenotyping of these cells as: **(A)** myCAFs (pSMAD2+ PDPN+ IL-6-CD45-, CD34-) and **(B)** iCAFs (IL-6+ PDPN+ pSMAD2- CD45– CD34-). **(C)** For each slide the proportion of cells in the region classified as TME that had pSMAD2 positive nuclei are plotted against the proportion of these cells that were positive for IL-6. **(D)** For each imCMS group how the cells in the TME have been classified.

To further interrogate the relative proportion of myCAF and iCAF amongst total CAF, we looked at their relative abundance in each CMS and CRIS subgroups **(Fig 3.2.3.7 A).** There was a trend towards a greater proportion of myCAFs in histological and intrinsic CMS4 disease, consistent with the known higher TGFβ levels in CMS4 disease, but this did not reach statistical significance. A higher proportion of CAFs were classified as iCAFs in CMS3 disease than the other CMS types combined; on T-test this had a p-value of 0.0259 for histological CMS, 0.0159 for conventional CMS and 0.0982 for intrinsic CMS **(Fig 3.2.3.7 B)**.

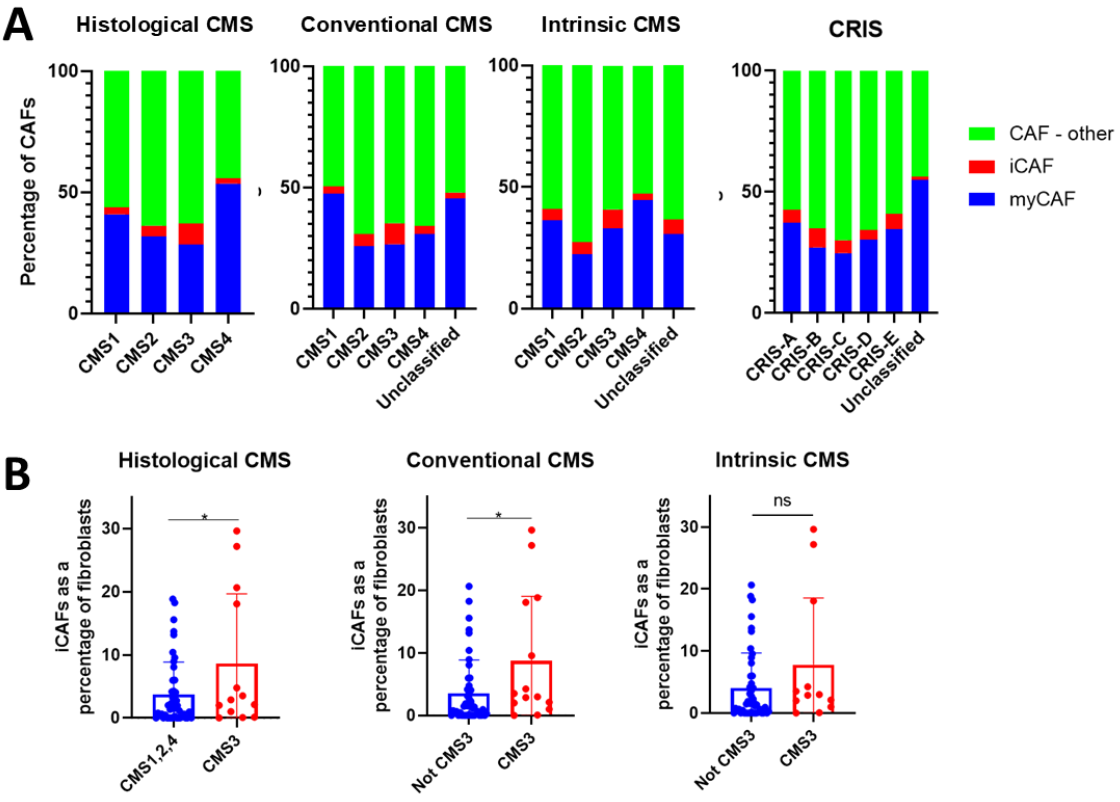


Fig 3.2.3.7 A higher proportion of fibroblasts are iCAFs in CMS3 CRC
(A) CAFs (PDPN+ CD34- CD45-) are divided into myCAFs (also nuclear pSMAD2+ IL-6-), iCAFs (also IL-6+ nuclear pSMAD2-) or CAF– other (also either nuclear pSMAD2+ IL-6+ or nuclear pSMAD2– IL-6-) and percentages of each are shown by imCMS, CMS, CMS and CRIS subgroups. **(B)** The percentage of CAFs defined as iCAFs in CMS3 versus the other CMS groups. Data analysed by T-test (* $p<0.05$, ns = non-significant).

We then looked at how the percentage of iCAFs and myCAFs correlated with the presence of immune and endothelial cells **(Fig 3.2.3.8).** This showed significant positive correlation between the percentage of iCAFs and immune cells (adjusted R-squared = 0.09588, p=0.017), consistent with the known inflammatory associated phenotype of iCAFs. The other correlations were not significant.

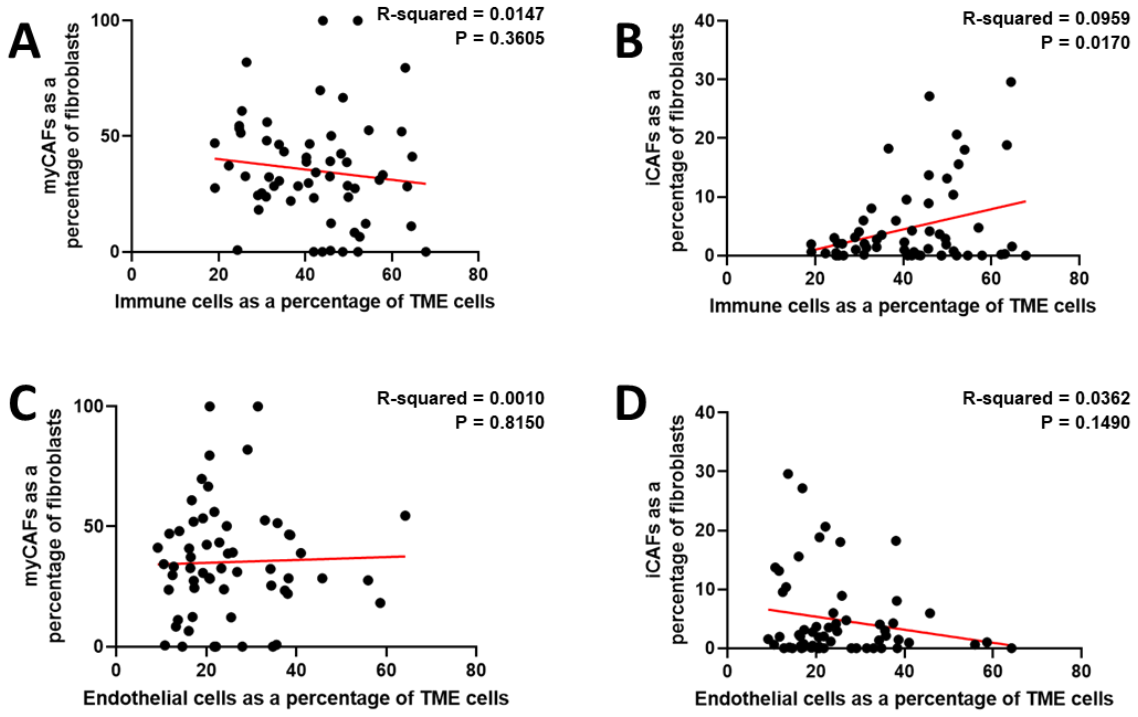


Fig 3.2.3.8 Percentage of iCAFs correlates with percentage of immune cells in the TME
myCAFs (PDPN+ CD34- CD45– pSMAD2+ IL-6-) **(A, C)** and iCAFs (PDPN+ CD34- CD45– pSMAD2+ IL-6-) **(B, D),** as a percentage of CAFs, correlated with the percentage of cells in the TME defined as immune cells **(A+B)** and endothelial cells **(C+D) .** There was positive correlation between percentage of CAFs that were iCAFs and percentage of immune cells in the TME (adjusted R-squared = 0.09585, p=0.017), all other graphs showed no significant correlation. Data analysed by simple linear regression, (significant p<0.05).

3.2.4 Repeat analysis by cell segmentation, removing PDPN

Many of the slides analysed showed very little PDPN staining and therefore few fibroblasts were being detected. Of note, looking in the Pelka *et al.* dataset PDPN is not universally expressed by CAFs, while it is expressed in most of CXCL14+ and GREM1+ CAFs, it is only found in half of cells in the MMP3+ CAF cluster, and even fewer of the myofibroblasts **(Fig 3.2.2.1 B)**. This phenotyping algorithm was probably excluding most of our cells of interest from analysis. We therefore repeated the analysis defining any CD45- CD34- pan-CK- cell in the TME as a stromal cell regardless of PDPN status. Within this group of stromal cells, we defined pSMAD2+ IL-6- cells as myCAF-like stromal cells and IL-6+ pSMAD2- cells as iCAF-like stromal cells. **Table 3.2.4.1** shows the new phenotyping method. **Fig 3.2.4.1 A+B** show the percentage of cells in the TME defined as stromal cells, and stromal cells per mm2, by CMS/CRIS group. This shows a trend towards greater numbers of stromal cells in CMS4 disease, defined by histological and conventional CMS, which is consistent with published literature. The fact that trends were seen between CMS4 disease and percentage of cells in the TME defined as stromal cells by this method, but not by the previous method of identifying fibroblasts based on PDPN expression (Fig 3.2.3.2 and Fig 3.2.3.4) helps validate this new method for defining stromal cell landscape in CRC. Intrinsic CMS4 and CRIS-B showed no enrichment for stromal cells, which suggests that this stratification methodology, designed to look at the epithelial component only, are not adapted to capture TME features in CRC.

Table 3.2.4.1 Algorithm to phenotype immune cells

Phenotype	Positive markers	Negative markers
myCAF-like stromal cell	pSMAD2	IL-6, CD45, CD34, panCK
iCAF-like stromal cell	IL-6	pSMAD2, CD45, CD34, panCK
Stromal cell – other	pSMAD2 and IL-6 (either both positive or both negative)	CD45, CD34, panCK
Immune cells	CD45	CD34
Endothelial cells	CD34	

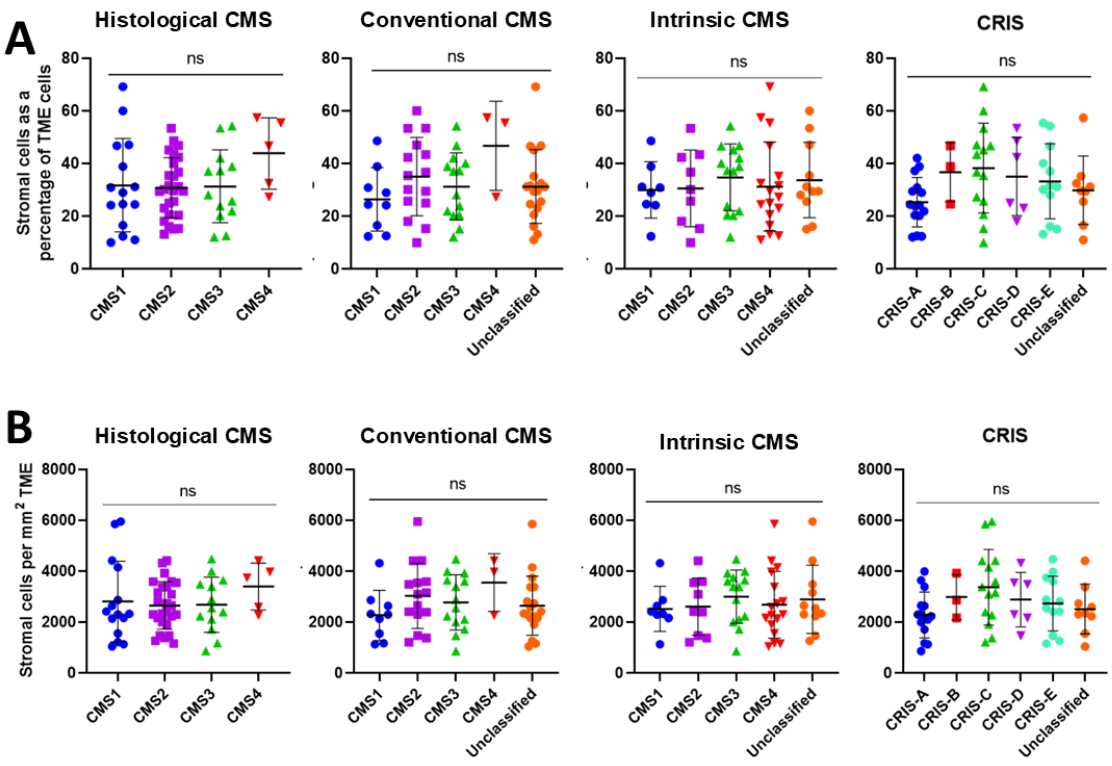


Fig 3.2.4.1 Higher numbers of stromal cells are associated with imCMS4
(A) The percentage of cells in the area classified as TME that were defined as stromal cells (CD45– CD34– panCK-) and (B) stromal cells per mm2 of TME by CMS groupings. Data analysed by T-test (ns = non-significant).

The overall numbers of stromal cells per mm^2 of tumour microenvironment for each case was correlated with the metadata available. No associations were found between stromal cell numbers per mm^2 and MSI status, or mutations in APC, AMER1, ARID1A, ATM, ATR, BRAF, CREBBP, CTNNB1, ELF3, EP300, ERBB2, ERBB3, KRAS, MBD6, NRAS, PIK3CA, PTEN, SMAD3, SMAD4, SMARCA4 SOX9, TCF7L2 or ZFP36L2. Significantly fewer stromal cells were seen in FBXW7 mutated tumours, p=0.0166 **(Fig 3.2.4.2 A-C)**. There was also a trend towards fewer stromal cells when SMAD4 was diploid in the epithelial cells, but this did not reach statistical significance **(Fig 3.2.4.2 D).** No association was seen with response to treatment or disease-free survival **(Fig 3.2.4.2 E+F)**. Negative correlations were seen between stromal cell numbers and immune cell numbers per mm^2 TME **(Fig 3.2.4.2 G)**, R-squared 0.3024, p<0.0001, and stromal cell and endothelial cell numbers **(Fig 3.2.4.2 H),** R-squared 0.2405, p<0.0001, both given per

mm^2 of tumour microenvironment. No correlation was seen between numbers of immune and endothelial cells per mm^2.

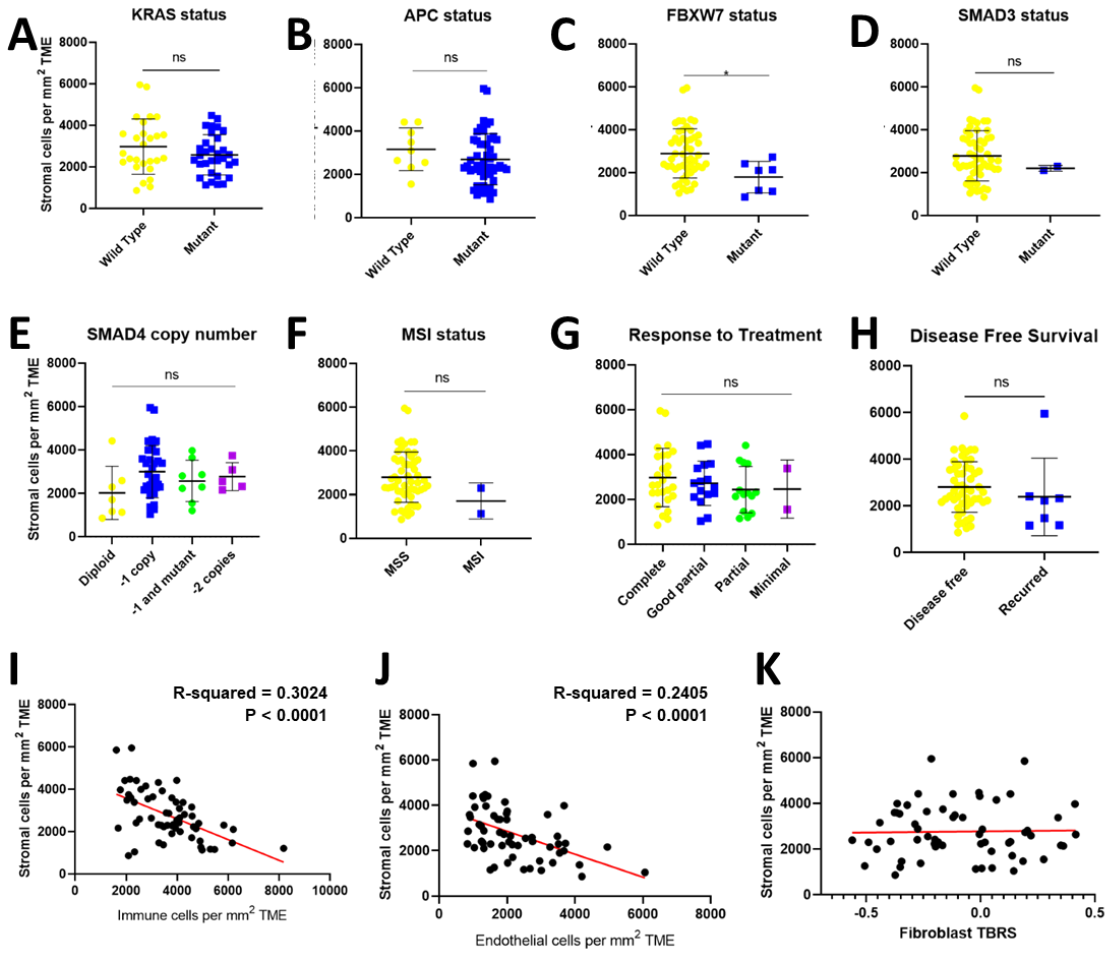


Fig 3.2.4.2 Stromal cell numbers within the TME and correlations with other data

Stromal cell numbers per mm2 of TME and **(A)** KRAS mutation, **(B)** FBXW7 mutation, **(C)** SMAD 3 mutation, **(D)** SMAD4 deletion, **(E)** response to treatment, **(F)** disease free survival., **(G)** immune cells per mm2 TME and **(H)** endothelial cells per mm2 of TME. Data analysed by T-test **(A-D, F, H),** one-way Anova **(E+G)** and simple linear regression **(I-K)** (*p<0.05, ns = non-significant).

Since tumours are heterogenous even within one CMS subset, we looked at the relative abundance between the cells composing the TME across patients in each histological CMS group **(Figure 3.2.4.3 A).** There is enormous variability, with some slides such as SC01283A having no myCAF or iCAF-like stromal cells, and others such as SC00598A being nearly a third myCAF-like stromal cells. No slides showed very large numbers of iCAF-like stromal cells. **Fig 3.2.4.3 B** shows the combined relative abundance of TME cell types per histological and intrinsic CMS, in the histological CMS4 subgroup there were a higher proportion of total stromal cells and within these specifically of myCAF-like stromal cells than other subgroups, which is consistent with the known features of CMS4 in having more fibroblasts and greater TGFβ levels. Where the proportion of TME cells phenotyped as myCAF-like stromal cells are compared between histological CMS1-3 combined versus CMS4 (**Fig 3.2.4.3 D**), significantly more myCAF-like cells are seen in CMS4 (T-test, $p = 0.0066$). This trend was not seen when intrinsic CMS was used **(Fig 3.2.4.3 C+E)**. There also appear to be less iCAF-like stromal cells in CMS4 disease, although the difference was not statistically significant. The increased proportion of iCAFs seen in CMS3 disease when PDPN was used as a marker was not seen by this analysis method.

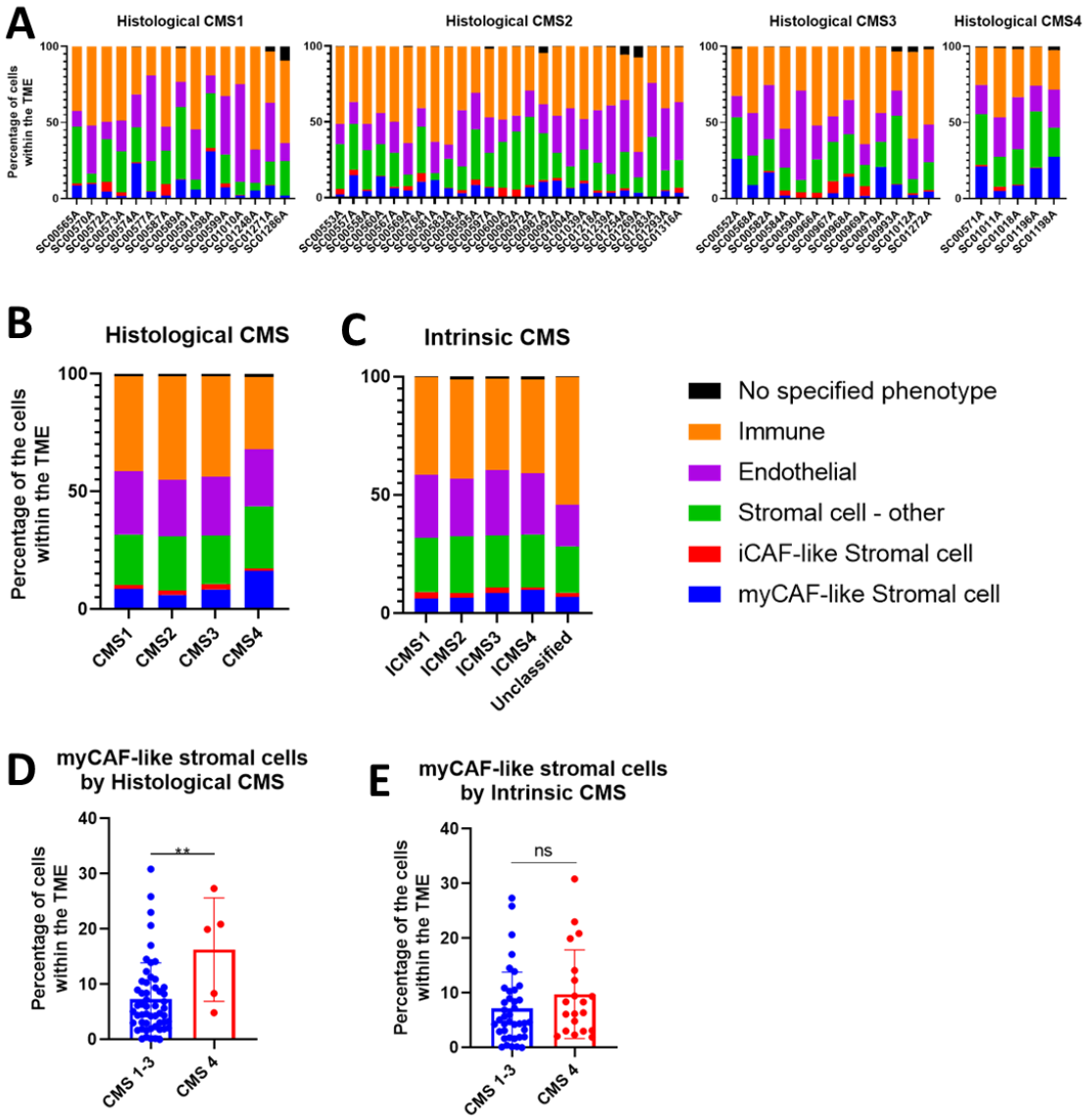


Fig 3.2.4.3 Phenotyping of Cells in the Tumour Microenvironment: more myCAF-like stromal cells in CMS4 disease

Cells within the TME were phenotyped as either myCAF-like stromal cells (pSMAD2+ IL6- CD45- CD34- panCK-), iCAF-like stromal cells (pSMAD2- IL6+ CD45- CD34- panCK-), stromal cell – other (CD45- CD34- panCK- and either pSMAD2+ IL6+ or pSMAD2- IL6-), Endothelial (CD34+ panCK- PDPN-), Immune (CD45+ CD34- panCK- PDPN-) or No specified phenotype. **(A)** shows the distribution of cells in each individual tumour by histological CMS group. **(B+C)** shows the tumours within each histological CMS group **(B)** and intrinsic CMS group **(C)** combined. **(D+E)** show the proportion of TME cells classified as myCAF-like stromal cells for CMS1-3 combined versus CMS4 using histological **(D)** and intrinsic **(E)** CMS. Data analysed by T-test (**p<0.01, ns=non-significant).

Having looked at how overall numbers of stromal cells associated with clinical features and molecular signatures of CRC, the next stage was to look at numbers of iCAF-like and myCAF-like stromal cells specifically and how they relate to other measures of the tumours. Negative correlation was seen between the number of stromal cells classified as iCAF-like and myCAF-like per mm^2 TME for each of the tumour, Adjusted R-squared = 0.1021, p= 0.0136 **(Fig 3.2.4.4 A)**. There was positive correlation between myCAF-like stromal cells per mm^2 TME and the fibroblast TBRS (F-TBRS) **(Fig 3.2.4.4 B)** and negative correlation between iCAF-like stromal cells per mm^2 TME and F-TBRS **(Fig 3.2.4.4 C)** although neither of these reached statistical significance, with p values of 0.0634 and 0.0575 respectively. There was no association between the individual stromal cell subsets and tumour mutations **(Fig 3.2.4.4 D-G)** or MSI status **(Fig 3.2.4.4 H)**. While no trend was seen between stromal cell subsets and treatment response **(Fig 3.2.4.4 I)**, in tumours where the disease recurred there were less iCAF-like stromal cells per mm2 TME, hinting towards a protective role, although this was not statistically significant **(Fig 3.2.4.4 J).**

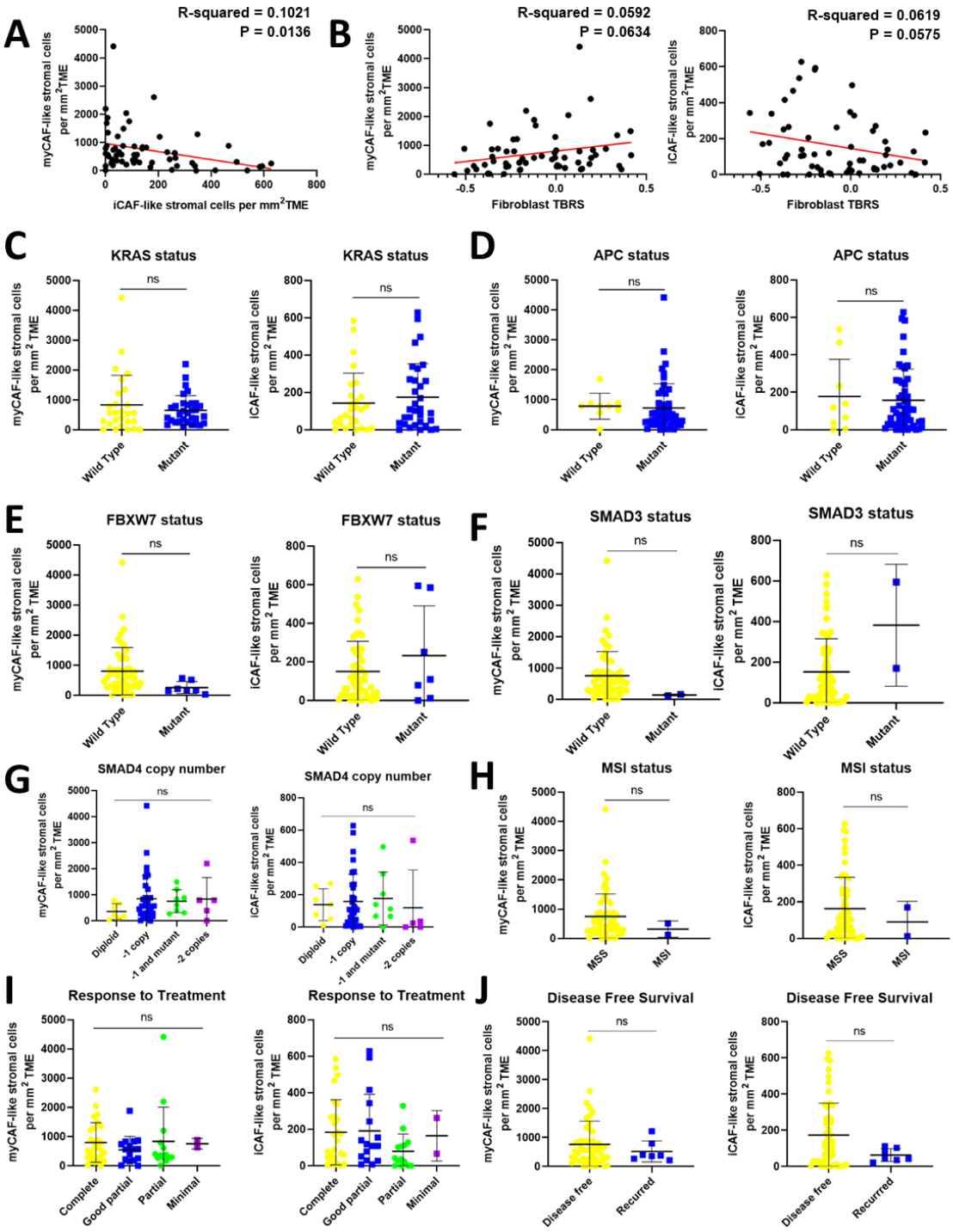


Fig 3.2.4.4 The Fibroblast-TBRS shows positive correlation with the proportion of fibroblasts that are myCAFs

Correlation between numbers of iCAFs and myCAFs per mm^2 of TME **(A)** and between each of these and the Fibroblast TBRS**(B+C) .** Correlations between mutations in KAS (**D-E)** and APC **(F-G)** and myCAFs per mm2 TME **(D+C)** and iCAFs per mm2 TME **(E+G).** Correlations between MSI status **(H+I),** Response to Treatment **(J+K)** and Disease Free Survival **(L+M)** and myCAFs per mm^2 TME **(H,J,L)** and iCAFs per mm^2 TME **(I,K,M).** Data analysed by T-test **(C-F, H, J),** one-way ANOVA **(G,I)** and simple linear regression **(A,B)** (ns=non-significant).

The proportion of stromal cells defined as myCAF and iCAF-like was next correlated with the proportion of cells in the TME defined as immune cells and endothelial cells **(Fig 3.2.4.5)**. There was no correlation between either stromal cell subset and endothelial cells. There was no correlation between myCAF-like stromal cells and immune cells, whereas there was significant positive correlation between the proportion IL6+ stromal cells and immune cells consistent with previous results using PDPN+/IL-6+ markers to identify iCAFs (R=0.0797, p= 0.0303). On this slide we were only able to compare iCAFs with total numbers of immune cells. The next step is to be able to see how iCAF numbers compare with specific immune subsets, and to do this an additional panel was used.

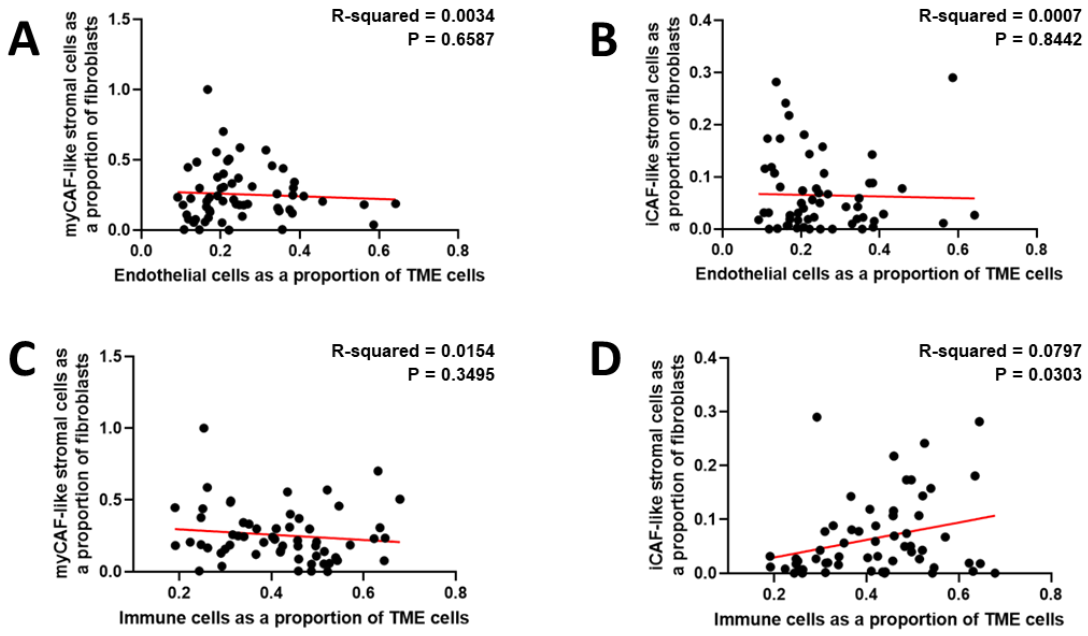


Fig 3.2.4.5 Higher proportion of iCAFs are associated with more immune cells
The proportion of fibroblasts that were defined as myCAFs (A,C) and iCAFS (B,D) were correlated with endothelial cells (A,B) and immune cells (C,D) as a proportion of all cells in the TME. Data analysed by simple linear regression, (significant, p<0.05).

The next aim was to identified the niche within the TME in which myCAFs and iCAFs are located. Specifically, their locations relative to the epithelial tumour nests, the vessels and the immune cells was studied. Only slides with at least 50 myCAF-like and 50 iCAF-like stromal cells were included for analysis, due to the likelihood of individual cells skewing the results when smaller numbers were present. This led to the exclusion of 15 cases due low numbers of iCAF-like stromal cells, 3 cases due to low numbers of myCAF-like stromal cells and 1 due to low numbers of both. With 1 case with no area of tumour for analysis, this resulted in 40 cases being suitable for proximity analysis.

The nuclei of the cells phenotyped as myCAF-like or iCAF-like stromal cells were drawn as dots onto spatial plots in HALO, in addition to either the borders of the tumour epithelium **(Fig 3.2.4.6 A)**, the immune cells **(Fig 3.2.4.6 B)** or the endothelial cells **(Fig 3.2.4.6 C)**. Numbers of iCAF-like and myCAF-like stromal cells in a series of bands from areas of interest was studied. No differences were seen in distribution of iCAF-like and myCAF-like stromal cells when distances greater than 10um from the region/cell type of interest were used. A significantly higher proportion of iCAF-like stromal cells were found within 10um of the tumour epithelium than myCAF-like stromal cells (32.8% versus 24.9%), p=0.0004 on paired T-test **(Fig 3.2.4.6 D)**. A significantly higher proportion of iCAF-like stromal cells were within 10um of an immune cell than myCAF-like stromal cells (45.9% versus 39.5%, on paired T-test p <0.0001) **(Fig 3.2.4.6 E)**. A significantly higher proportion of iCAF-like stromal cells were found within 10um of tumour vessels (endothelial cells) than myCAF-like stromal cells (23.8% versus 20.9%, p= 0.0205 on paired t-test) **(Fig 3.2.4.6 F)**. These last two findings are likely to be linked, with immune cells entering a tumour via the vasculature and therefore being more concentrated around the tumour vessels.

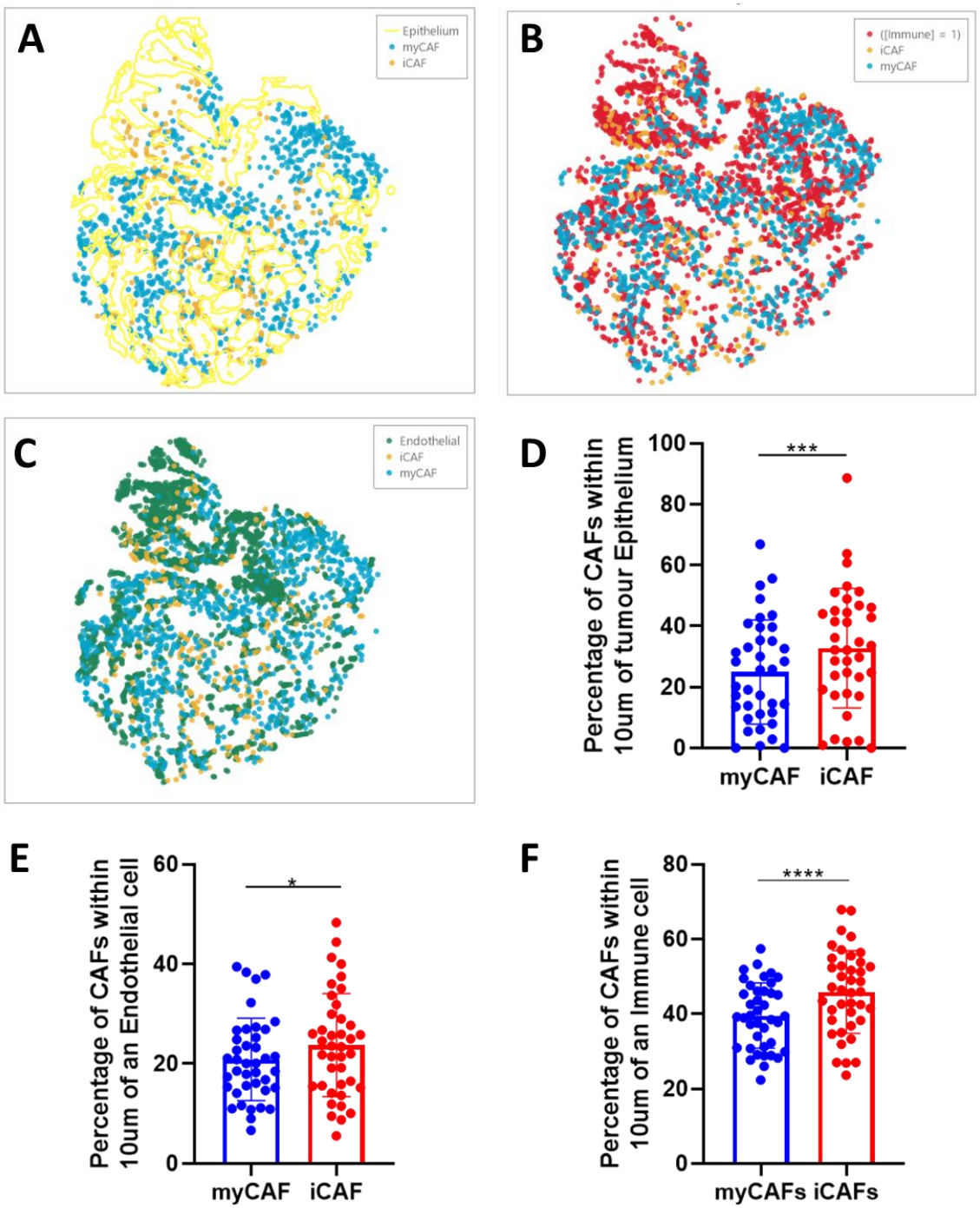


Fig 3.2.4.6 iCAFs are more often located close to immune cells and endothelial cells than myCAFs **(A-C)** Spatial plots of one slide with myCAFs represented by blue dots and iCAFs by yellow dots, with the border of the tumour epithelium **(A)**, the immune cells in red **(B)** and the endothelial cells in green **(C)**. **(D)** The density of fibroblasts within each band radiating out from the tumour epithelial cells, compared to the overall fibroblast density for the TME in that slide. **(E)** The percentage of myCAFs and percentage of iCAFs located within 10um of an endothelial cell. **(F)** The percentage of myCAFs and percentage of iCAFs located within 10um of an immune cell. Data analysed by paired T-test (*p<0.05, ***p<0.001, ****p<0.0001).

To summarise the analyses so far, IL-6 positive stromal cells are associated with immune cells regardless of whether PDPN is used as a marker for the cells of interest. When PDPN is required as a marker, iCAFs are associated with CMS3 disease, and when we analyse stromal cells generally without using PDPN as a marker, myCAF-like cells are associated with CMS4 disease. This suggests that the cells often described as myCAFs, seen in CMS4 disease and associated with poor prognosis, may represent smooth muscle cells enriched at the tumour front as it invades deeper into underlying tissue.

3.2.5 Image Analysis of the Stromal panel by Area Quantification

There are problems with using cell segmentation followed by phenotyping of individual cells. The cell segmentation algorithms are imperfect, as they merely attribute cytoplasm to the nearest nucleus. While this may be effective for rounded cells, such as haematopoietic cells, it can perform less well for elongated cells, which includes fibroblasts**. Fig 3.2.5.1** demonstrates how cell segmentation may not be ideal for cell for this shape. We tried to mitigate the problems with incorrect drawing of cell borders by cell segmentation by keeping the maximum cytoplasm radius low at 1.5um **(see Appendix 1)**, to avoid incorrectly attributing a marker present on another cell to one where the nucleus is closer, but it still may not perform perfectly. Therefore, an alternative second analysis was performed, where the HALO package Area quantification FL v.2.3.3. was used to work out the percentage of tissue that is over a pre-set intensity level for IL-6 and pSMAD2, without drawing boundaries for individual cells.

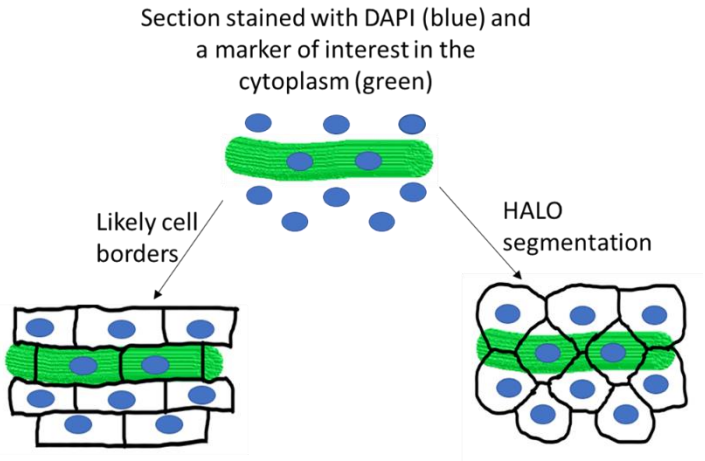


Fig 5.2.5.1 Challenges with cell segmentation
Due to the HALO cell segmentation algorithm drawing cell boundaries at an equal distance between nuclei up to a maximum distance, this can poorly represent cell shape of elongated cells, where the linear distribution of staining may better represent cell boundaries.

To perform this new, area quantification-based analysis, first a new classifier had to be developed, to ensure that immune cells and endothelial cells were not included in the quantification analysis, as they would be present in the region classified as TME before. The TME class from the original classifier was split into three: immune, endothelial and stroma. Therefore the 5-classes in the new classifier were immune, vessels, glass, epithelium and the stroma **(Fig 3.2.5.2 A),** with the stroma being the region of interest containing the fibroblasts in which to quantify the percentage positivity for pSMAD2 and IL-6. Developing a new classifier brought further challenges, as the minimum object size recognised by the machine learning algorithm had to be reduced greatly to allow the classifier to recognise regions as small as individual haematopoietic cells. However, by reducing the minimum object size to this extent it became less effective at successfully delineating epithelial tumour nests, as initially many 'gaps' of glass were found, and this also occurred in some stromal areas. Once the classifier was optimised, the next step was to set thresholds of intensity at which an area would be defined as positive for markers of interest, **Fig 3.2.5.2 A** shows the new classifier used for IL-6 and pSMAD2 area quantification within the stroma.

Firstly, the location of pSMAD2 positive and IL-6 positive stromal regions within the TME was studied, with aim of identifying the likely niche for myCAF and iCAF cells. The location of these areas relative to both the neoplastic cells and the tumour vessels was investigated. First the percentage positivity for these markers relative to distance from the neoplastic cells was analysed. The regions classified as tumour epithelium were converted into annotations. Negative regions within the tumour nests and very small epithelial regions <300um2 were excluded. Then a series of 10 concentric bands were drawn around the regions classified as tumour at 10um intervals **(Fig 3.2.5.2 B).** Within each band the percentage of tissue classified as stroma that was positive for IL-6 and pSMAD2 was calculated. Due to some of these small tumour biopsies having very little additional tissue surrounding the tumour nests themselves, where there was < 0.01µm2 of stromal tissue within a band results were not analysed, as the

results became erratic when small regions were analysed. This band analysis was performed on a representative selection of 5 cases from each histological CMS subgroup, due to the large amounts of processing time required for each case precluding use on the entire dataset **(Fig 3.2.5.3)**. For IL-6, 7 of the 20 cases show higher levels close to the epithelium, falling with distance from the tumour. In one case there are low levels next to the tumour, which increase with distance, and in 12 of the cases there is no clear trend. For pSMAD2, in 8 cases levels increased away from the epithelium, in 1 case it dropped with distance from the epithelium, and in 11 cases there was no clear trend. This finding that in more tumours IL-6 positivity within the stroma reduced with distance from the epithelium and pSMAD2 levels increased is the opposite to the findings by Ohlund *et al.* in pancreatic cancer, who identified myCAF cells close to the epithelium and iCAF further away. Of note, in this study, larger resection specimens were used, and the iCAFs were seen approximately 300um from the tumour, in these CRC biopsy specimens there was very little tissue further than 100um away from the tumour nests. In the majority of these biopsy specimens there was simply no correlation in levels of these markers in the stroma and distance from the tumour epithelium.

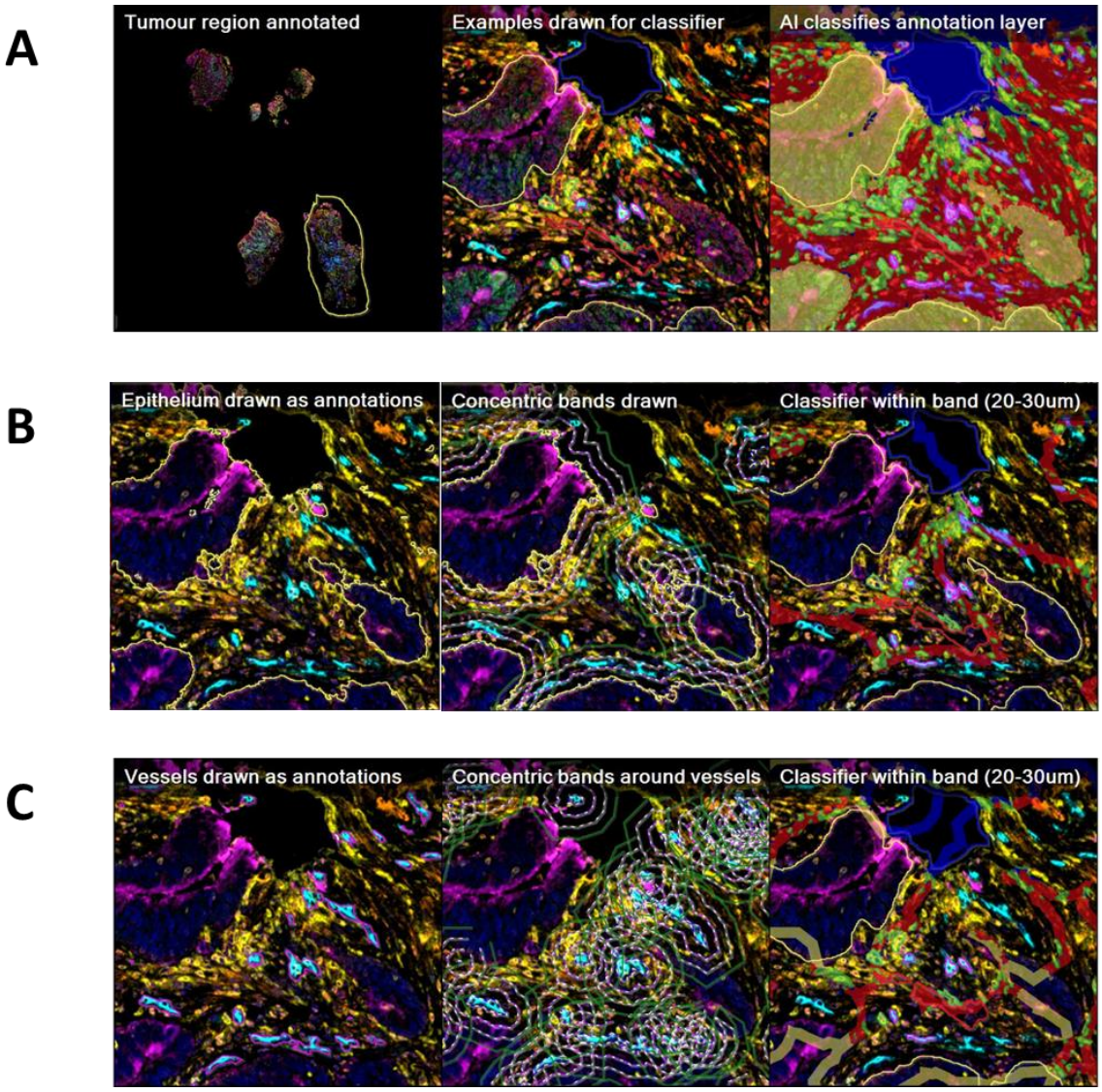


Fig 3.2.5.2 Method for analysing IL-6 and pSMAD2 positivity by Area Quantification

(A) The region to be analysed is annotated as tumour. Within this teaching regions are drawn for machine learning so that AI can classify the entire tumour region as Epithelium (yellow), Immune (green), Vessels (purple), Stroma (red) or Glass (blue). The percentage of the Stroma (red) that is positive for IL-6 and pSMAD2 can then be calculated. (B) The area classified as Epithelium (yellow) in A is converted to annotations. Concentric bands are drawn around these at 10um intervals. Within each band the classifier shown in A can be performed to find regions of Stroma (red) a set distance from epithelium, within which the percentage positive for IL-6 and pSMAD2 can be calculated. (C) The area classified as Vessels (purple) in A is converted to annotations. Concentric bands are drawn around these at 10um intervals. Within each band the classifier shown in A can be performed to find regions of Stroma (red) a set distance from vessels, within which the percentage positive for IL-6 and pSMAD2 can be calculated.

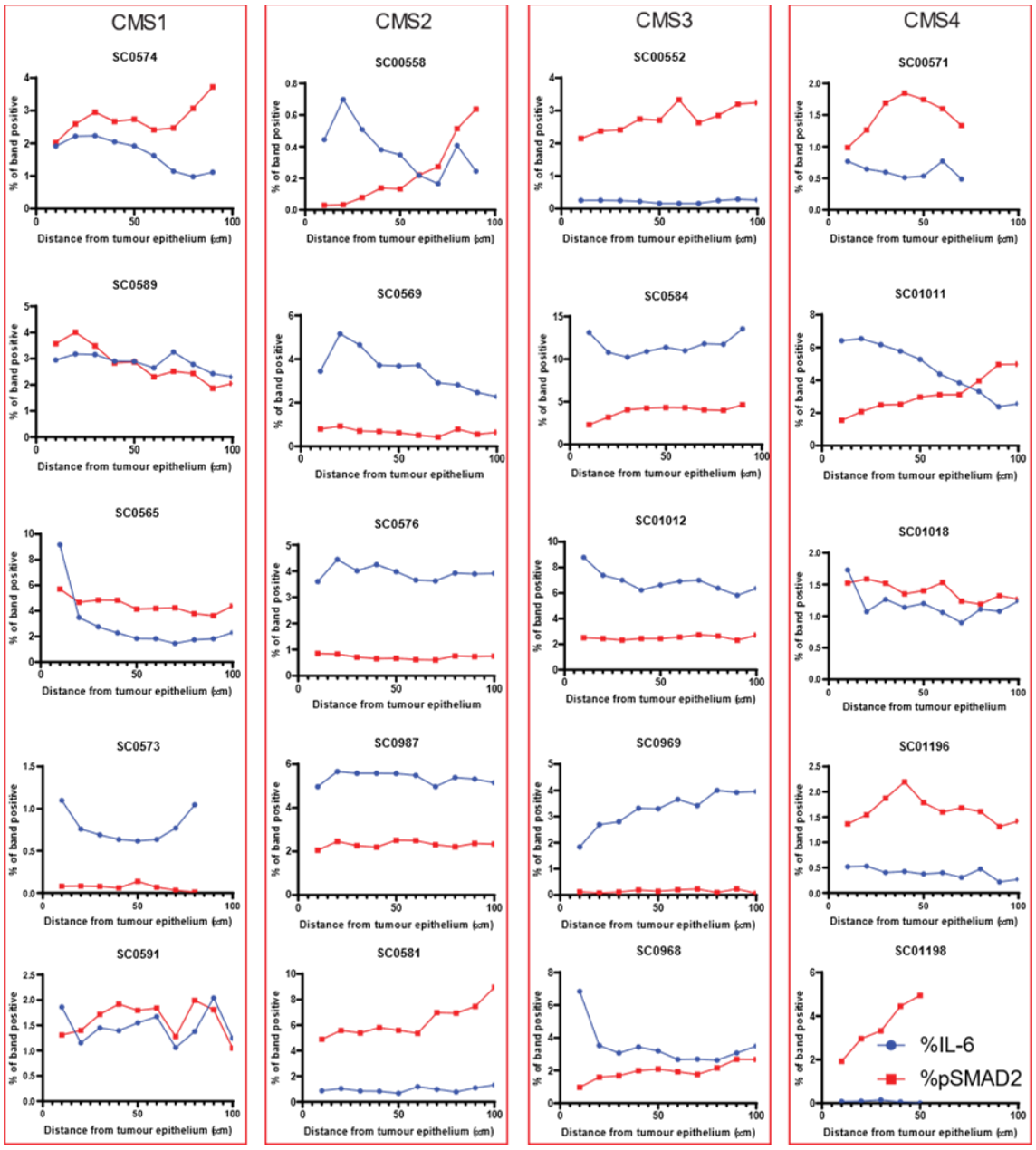


Fig 3.2.5.3 No correlation between IL-6 or pSMAD2 positivity in the stroma and distance from the tumour epithelium

Five representative cases were chosen from each imCMS subgroup. Around the epithelium a series of 10 concentric bands of 10um diameter were drawn. Within each band the region classified as stroma was assessed for percentage IL-6 and pSMAD2. positivity.

Area quantification of the percentage of stroma positive for pSMAD2 and IL-6 was also performed in bands radiating from the vessels. The same classifier was used to turn regions classified as vessels into annotations, and again negative regions or those of a very small area, less than 15um2 in this case, were excluded. Up to 8 bands of 10um diameter were drawn, but for many tumours fewer bands than this were analysed, due to their being very little tissue as for as 80um away from a vessel. As with the bands around the epithelium, once bands with less than 0.01um2 of stroma within them were reached values were not analysed. **Fig 3.2.5.2 C** shows the steps in proximity analysis to vessels. **Fig 3.2.5.4** shows graphs with percentage positivity for IL-6 and pSMAD2 by distance from the vessels in the 5 representative cases from each imCMS subgroup. For IL-6 there were 10 cases where percentage positivity fell with distance from the vessels, 2 cases where it appeared to increase, and 7 where no trend was seen. For pSMAD2 in the majority of cases no trend was seen between percentage positivity and distance from the vessels.

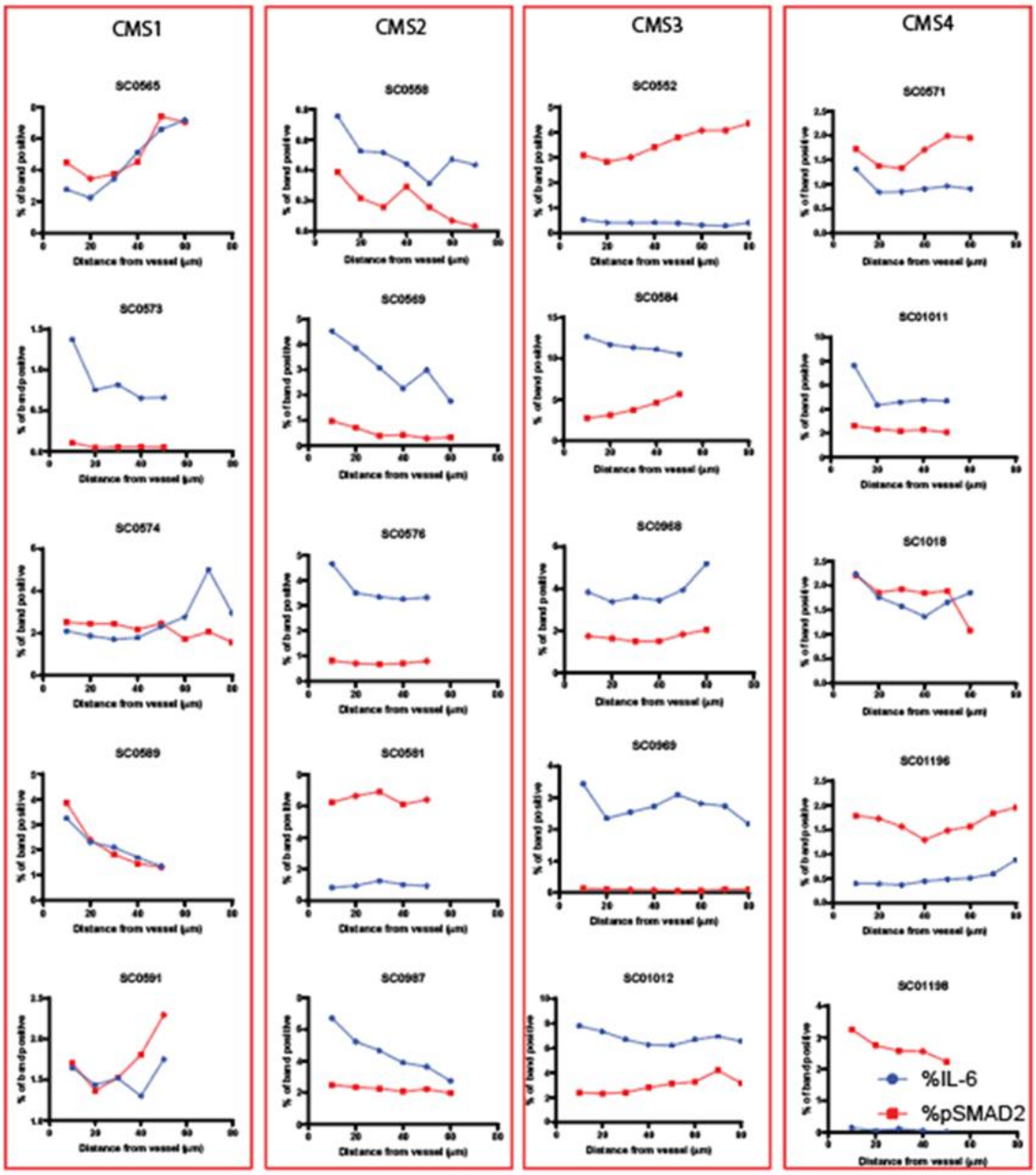


Fig 3.2.5.4 Higher stromal IL-6 seen closer to vessels in approximately half of tumours.
Five representative cases were chosen from each imCMS subgroup. Within each 10um band around the vessels the region classified as stroma was assessed for percentage IL-6 and pSMAD2 positivity, where bands contained less than 0.01um2 of stroma analysis was not performed.

3.3 The Immune panel on the Grampian dataset

3.3.1 Image Analysis of the Immune Panel

A broad 6-plex immune panel was developed for use on the Vectar-Polaris, which aimed to identify neutrophils, macrophages, T-cells, antigen presenting cells (APCs) and NK cells. **Figure 3.3.1.1** shows examples of staining for each of the following markers: CD45, CD66b, CD68, CD3, Human Leucocyte antigen – DR isotype (HLA-DR) and CD56.

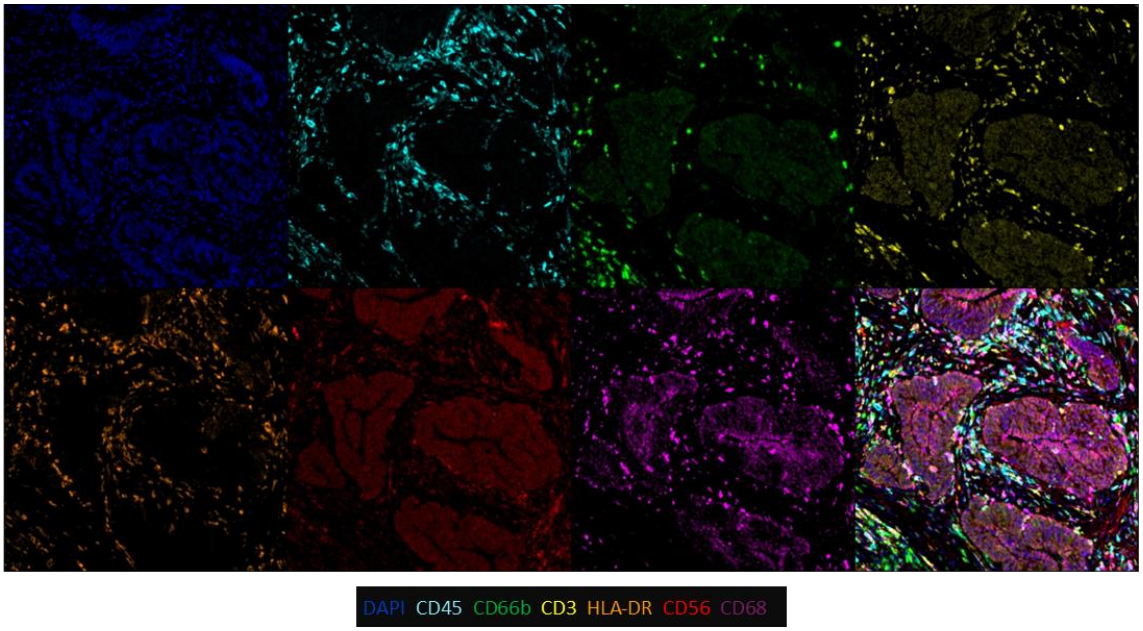


Fig 3.3.1.1 Staining of the Immune panel
Example staining of each marker individually and all markers together (bottom right).

Fig 3.3.1.2 shows the stages in analysis of the tumours on the HALO imaging platform. After annotating the tumour region, we developed a classifier to recognise tumour epithelium, tumour microenvironment and glass, and using machine learning with HALO AI MiniNet to classify the entire annotation region **(Fig 3.3.1.2 A)**. The next step was identification of the nuclei and cell segmentation within the TME, followed by setting thresholds for intensity for each of the markers at which a cell would be defined as positive for that marker. **Appendix 2** shows the settings used in HALO HighPlex F v.4.2.3 to identify the cells. This then allows all the cells to be phenotyped according to their positivity or negativity for a given marker or set of markers, shown in **Table 3.3.1.1** and **Fig 3.3.1.2 B-F**. As the size of the tumours vary greatly between

slides, all results are given as the number of cells per mm^2 classified as tumour microenvironment.

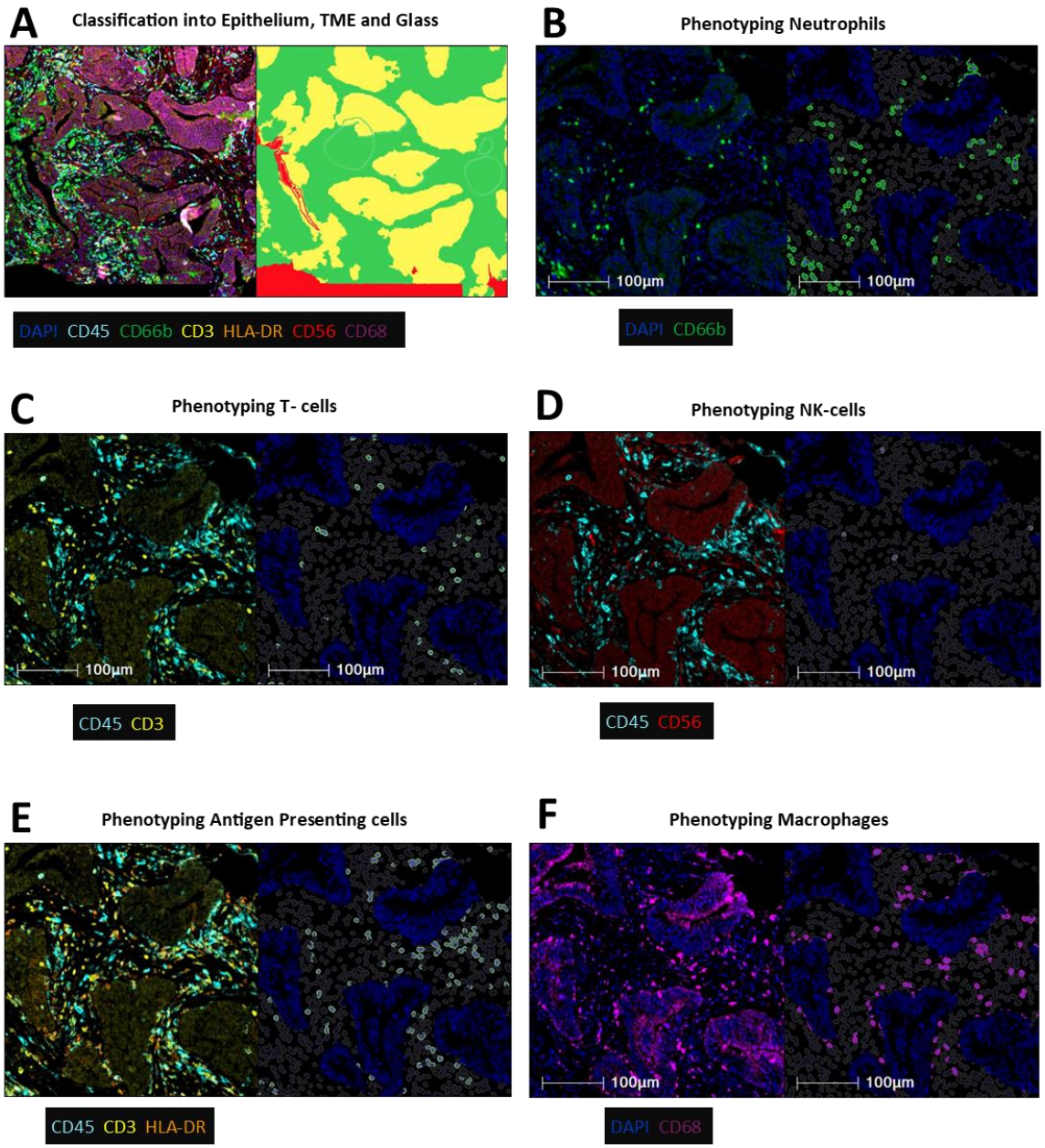


Fig 3.3.1.2 Analysis of the Immune panel

(A) Machine learning was used to classify the area of interest into tumour epithelium (yellow), TME (green) and glass (red). **(B-F)** Cell segmentation in the regions classified as TME and phenotyping of these cells as **(B)** Neutrophils (CD66b+), **(C)** T-cells (CD45+ CD3+), **(D)** NK-cells (CD45+ CD56+), **(E)** Antigen presenting cell (CD45+ CD3- HLA-DR+) and **(F)** Macrophages (CD68+).

Table 3.3.1.1 Strategy to phenotype immune cells

Phenotype	Positive markers	Negative markers
Neutrophils	CD66b	
T-cells	CD3, CD45	
Macrophages	CD68	
NK-cells	CD56, CD45	
APCs	HLA-DR, CD45	CD3

3.3.2 Correlation of Immune panel with metadata

The numbers of each immune cell type per mm^2 of TME were correlated with the metadata for these 60 cases. There were more T-cell seen in CMS1 disease when conventional CMS was used (CMS1 v CMS2 p=0.0101, CMS1 v CMS4 p=0.0403, CMS1 v Unclassified p=0.0124), and this trend was present with the other methods to determine CMS although did not reach significance **(Fig 3.3.2.1 A)**. There were also higher numbers of T-cells seen in CMS3 disease compared to CMS2, 4 and unclassified across the CMS classification methods, but this did not reach significance. There were more neutrophils in CMS1 disease when measured by histological CMS (CMS1 v CMS2 p=0.0082, CMS1 v CMS3 p=0.0053), and this trend was seen with other methods of classification **(Fig 3.3.2.1 B)**. There was also a trend toward CMS4 having the second highest number of neutrophils, after CMS1. Both these findings are consistent with existing literature, more T-cells in CMS1 disease was seen in the original CMS classification[20], and the association between neutrophils and CMS 1 and 4 was seen by Germann *et al.* who predicted a neutrophil score from bulk RNAseq across two datasets[179]. There were no associations between other immune cell types and CMS grouping. **Fig 3.3.2.1 C** shows the classification of all the immune cells in the TME, divided by histological CMS subgroup.

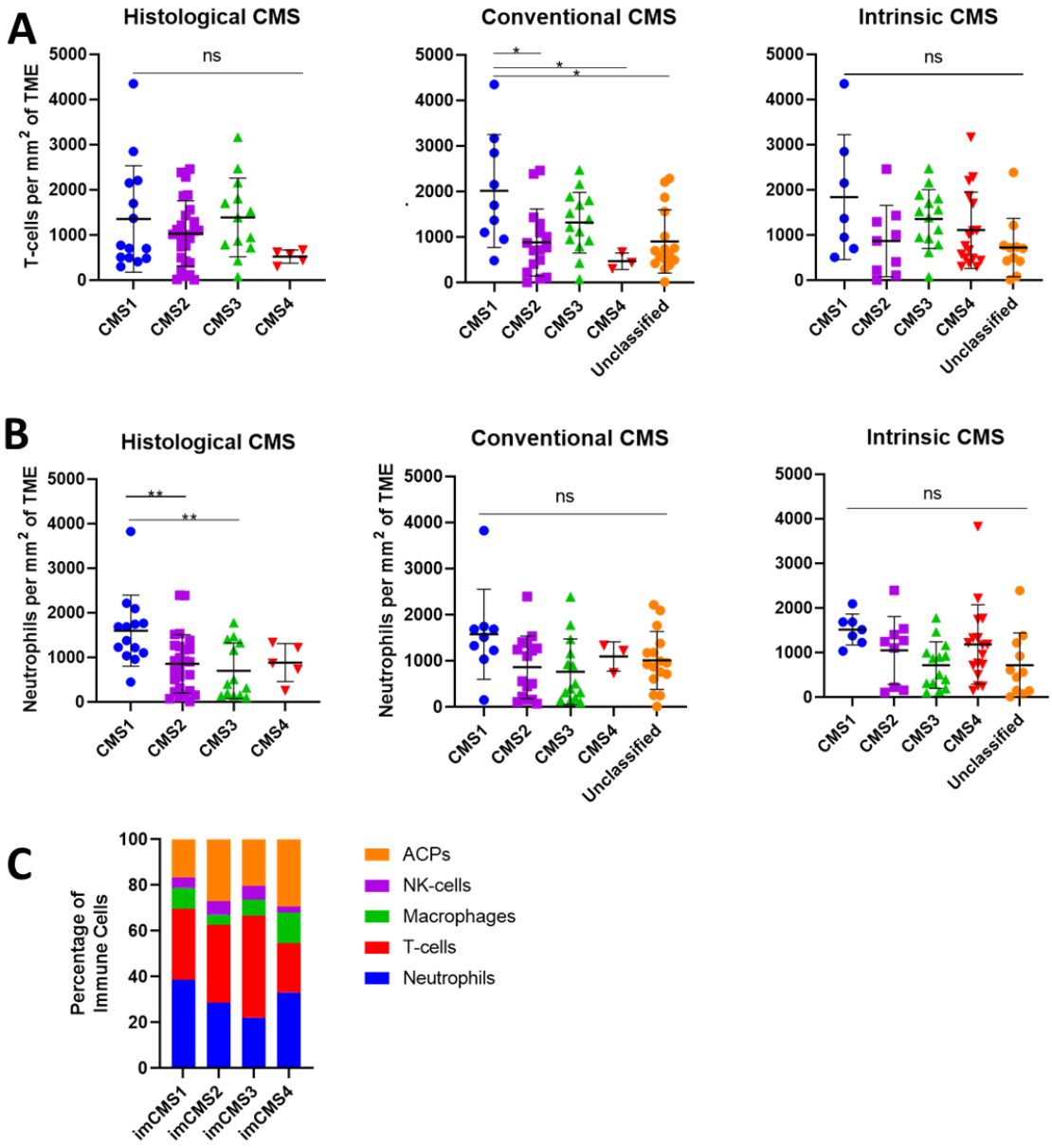


Fig 3.3.2.1 CMS1 is associated with greater numbers of T-cells and neutrophils

Number of T-cells **(A) and** neutrophils **(B)** per mm2 of TME by CMS. **(C)** Classification of all the immune cells in the TME, by histological CMS subgroup. Data analysed by T-test (* p<0.05, **p<0.01, ns = non-significant).

There was a trend towards MSI high tumours having more T-cells and neutrophils per mm^2 TME, but this did not reach statistical significance given there were only 2 MSI high tumours in the dataset **(Fig 3.3.2.2 A)**. There was also a trend towards greater numbers of T-cell in KRAS mutant tumours **(Fig 3.3.2.2 B)**. There were no associations between other immune cell groups and the mutations tested for in the S:CORT dataset. There was no association of numbers of specific immune cell subsets per mm^2 TME tissue with response to treatment or DFS.

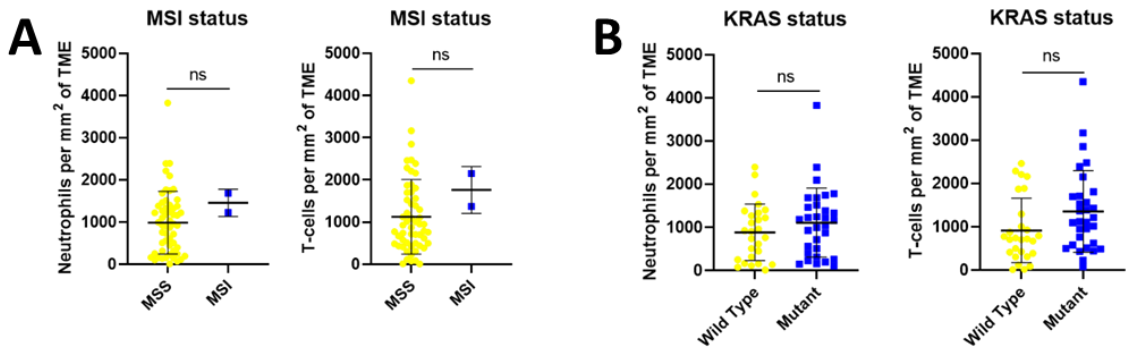


Fig 3.3.2.2 Associations between Immune cell numbers and the S:CORT metadata
T-cells and neutrophils per mm^2 by (A) MSI status and (B) KRAS status. Data analysed by T-test (ns=non-significant).

3.4 Correlations between the Fibroblast and Immune Panels

3.4.1 Correlations across the entire slide

The next step is to look for any correlation between the fibroblast subsets and specific immune cell subsets, through combining results from the two panels. Initially this was done using the cell segmentation method of identifying stromal cells and correlating cell numbers per mm^2 tissue microenvironment across the slide between the two sections. The density of total stromal cells, myCAF-like and iCAF-like stromal cells were each correlated with numbers of neutrophils, T-cells, macrophages, APCs and NK-cells. The only statistically significant correlation seen was between T-cells and total stromal cells, where there was negative correlation, Adjusted R-squared = 0.0839, p = 0.0274 **(Fig 3.4.1.1 A)**. There were no correlations between numbers of any immune subset and either iCAF-like or myCAF-like stromal cells specifically, although there was a trend towards iCAF-like stromal cells correlating positively with T-cells and myCAF-like stromal cells negatively **(Fig 3.4.1.1. B+C)**. Next, the area quantification method, using percentage of the stroma positive for pSMAD2 or IL-6 was correlated with each of the immune cell subsets. Positive correlation was seen between T-cells and IL-6 positivity in the stroma (Adjusted R-squared = 0.0996, p-value = 0.01582), but no other statistically significant correlations were identified **(Fig 3.4.1.1 D+E)**. Next the proportion of cells in the TME phenotyped as myCAF-like stromal cells was compared for tumours in the lower and upper half for neutrophils as a proportion of immune cells and T-cells as a proportion of immune cells. There were more myCAF-like stromal cells in the T-cell low tumour (p=0.0481 on T-test), but no difference depending on whether they were neutrophil high or low **(Fig 3.4.1.1 F)**. There was also no difference when the same comparison was studied for iCAF-like stromal cells.

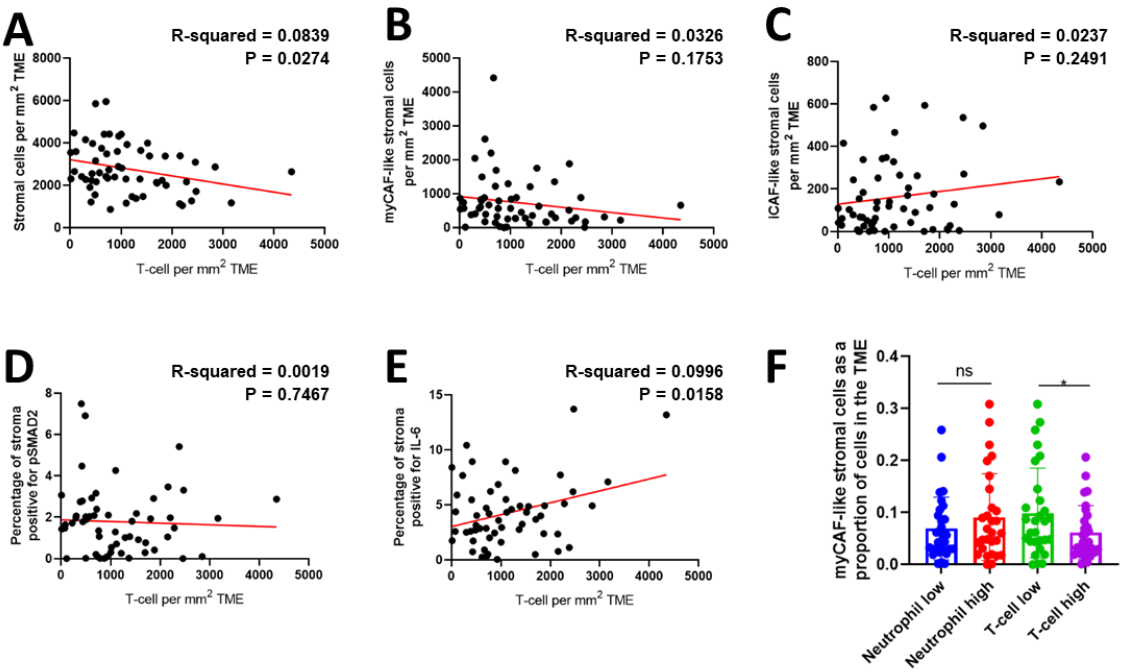


Fig 3.4.1.1. T-cells correlated negatively with stromal cells and positively with IL-6
Total stromal cells **(A),** myCAF-like **(B)** and iCAF-like **(C)** stromal cells versus T-cells per mm_2 of tissue microenvironment. Percentage of the stroma positive for pSMAD2 **(D)** and IL-6 **(E)** and plotted against T-cells. **(F)** myCAF-like stromal cells by neutrophil low and high tumour and T-cell low and high tumours. Data analysed by simple linear regression **(A-E)** and T-test **(F)** (* $p<0.05$, ns = non-significant).

3.4.2 Proximity analysis between immune cell subsets and fibroblast subsets

Next specific proximity between myCAF-like and iCAF-like stromal cells on the fibroblast subsets panel and Neutrophils and T-cells on the immune subsets panel was investigated. The immune and fibroblast subsets panels were performed on adjacent sections from the same block of tumour, allowing alignment. Of the 40 cases where at least 50 iCAFs and myCAFs were phenotyped (see section 3.2.2.4), successful alignment was possible in 27, in the rest architectural changes between the sections meant that it was not possible to be confident of alignment to a precision of 10um needed to identify cells likely to be directly interacting. Of these 27, in 2 cases there were less than 50 neutrophils and T-cells, meaning 25 cases were analysed. The nuclei of the immune cells and fibroblasts from the aligned slides were drawn on

a dotplot to allow distances between cells to be measured **(Fig 3.4.2.1 A)**. A significantly higher proportion of iCAF-like stromal cells were located within 10um of a neutrophil than myCAF-like stromal cells, 25.4% versus 15.7% respectively (P < 0.0001 on paired T-test). There was also a higher proportion of iCAF-like than myCAF-like stromal cells within 10um of a T-cell, 16.4% versus 13.1% (p = 0.0206 on paired T-test) **(Fig 3.4.2.1 B)**. The finding of more iCAF-like stromal cells within close proximity of neutrophils is consistent with the fact that neutrophils are a major producer of IL-1β , which promotes the iCAF phenotype.

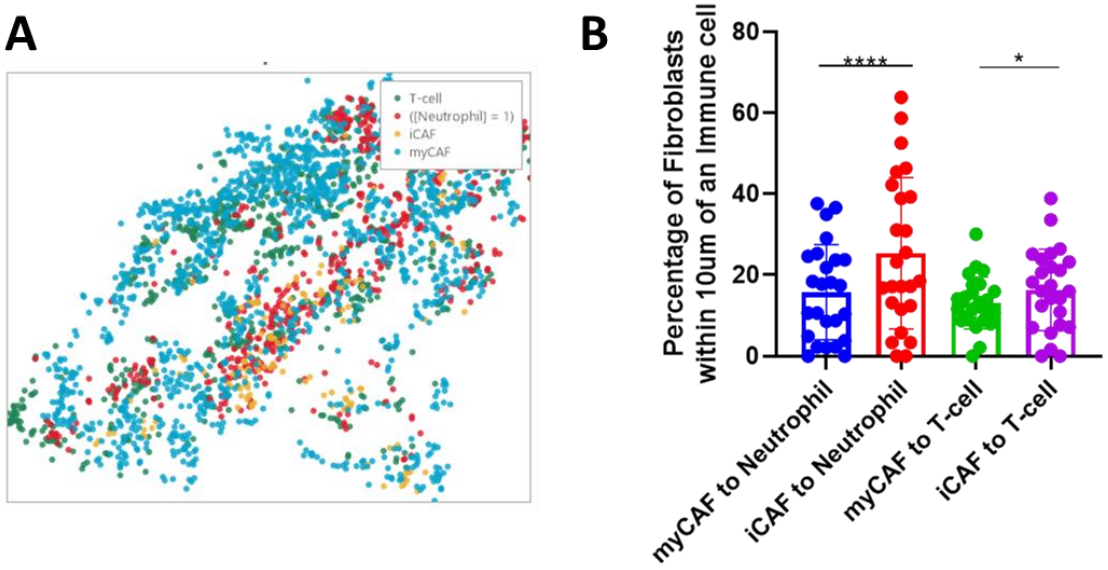


Fig 3.4.2.1 A higher proportion of iCAFs than myCAFs are seen within 10um of a neutrophil
(A) Dotplot showing locations of T-cells (green) and neutrophils (red) from the immune subsets slide and the iCAFs (yellow) and myCAFS (blue) from the adjacent fibroblasts subsets slide. (B) Percentage of myCAFs and iCAFs that are within 10um of a nuetrophil or T-cell. Data analysed by paired T-test (* p<0.05, **** p<0.0001).

3.4.3 CD56 staining around the Tumour Nests

It was noted that in many of the tumours, CD56 stained in a narrow band around the edge of the neoplastic cells **(Fig 3.4.3.1. A)**. This staining did not overlap with CD45 staining, and so was not thought to represent NK-cells, of which CD56 is generally used as a marker. Instead the elongated nature of the cells staining for CD56 suggested that they may be fibroblasts. On studying this in each slide it was recorded whether the rim of CD56 was present, absent or found around some, but not all tumour nests. Interestingly, this band was seen in in a high proportion of imCMS 2-3 tumours, but was absent in about half of imCMS 1 tumours **(Fig 3.4.3.1 D)**. Of note, looking at the scRNAseq data from Pelka *et. al*, within the stromal cells, CD56 (known by gene name *NCAM1*) was found very specifically in the CXCL14+ CAF cluster and Schwann cells **(Fig 3.4.3.1 E)** and, using the CCD18-Co fibroblast cell line, Neural cell adhesion molecule 1 (NCAM1) was downregulated by both TGFβ1 and IL-1β **(Fig 3.4.3.1 F)**. The area quantification method was used to look in 10um bands around the epithelial tumour cells to see what percentage of the TME in those bands were positive for CD56 staining, regardless of cell boundaries. Performing this in 5 cases from each imCMS subgroup showed a trend towards more CD56 staining near epithelial cells in histological CMS 2 and 3 tumours **(Fig 3.4.3.1 G)**. Using the same 10um bands of TME around epithelial cells, we performed cell segmentation and looked at the number of CD56+ CD45- cells per mm2 of that band. **Fig 3.4.3.1 B** shows CD56+ CD45- cells in red where all the cells in the TME have been segmented, and **Fig 3.4.3.1 C** shows a spatial plot with the edges of the tumour epithelium marked in yellow and the CD56+ CD45- cells as blue dots. Again, there was a trend towards more CD56+ CD45- cells next to the epithelium in histological CMS 2 and 3 CRCs than in histological CMS 1 and 4 **(Fig 3.4.3.1 H)**. However, there was no correlation between PDS groups and CD56+ cells.

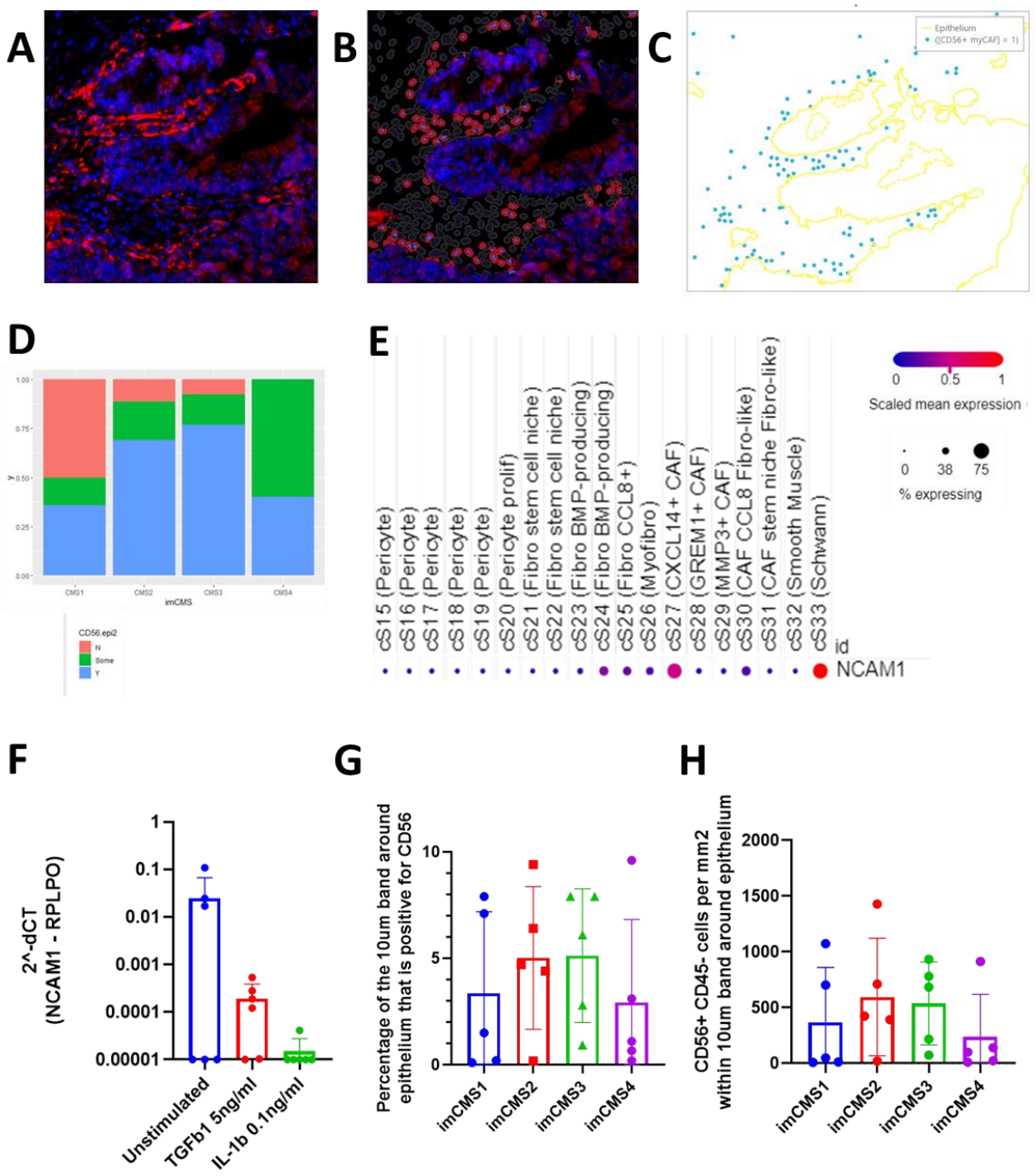


Fig 3.4.3.1 Most CRC tumours have a rim of CD56+ stromal cells surrounding the tumour nests, particularly in CMS2-3

(A). Example CD56 staining around the tumour nests (CD56 red, DAPI blue). **(B)** The same section, with cells in the TME segmented and CD56+ CD45– cells shown in red. **(C)** Spatial plot of this section, showing the outline of the tumour in yellow and CD56+ CD45– cells in blue. **(D)** The proportion of tumours that have a clear band of CD56 staining around the neoplastic cells (blue), do not have this band (pink) or the CD56 staining is present in some tumour nests but not all (green), by imCMS subgroup. **(E)** NCAM1 expression in the stromal clusters in the scRNAseq dataset by Pelka *et al*. **(F)** NCAM1 expression by RT-qPCR of CCD18-Co fibroblasts either unstimulated or stimulated with one of TGFβ1 or IL-1β. **(G)** Percentage of a 10um band around the tumour epithelium that stains positive for CD56, in 5 representative cases from each imCMS subgroup. **(H)** Number of CD56+ CD45– cells per mm2 within a band of 10um around the tumour epithelium, in 5 representative cases from each imCMS subgroup.

Given that I saw a trend towards more CD56+ stromal cells in CMS 2-3, and NCAM1 was seen in the CXCL14 cluster in the data from Pelka *et al*, we hypothesised that within the Pelka *et al* dataset more CXCL14+ CAFs would be seen in CMS2-3 disease than CMS1 and 4. The Pelka *et al* dataset had been used as a validation cohort in a study by Joanito *et al*, and for this each of the 62 tumour had been defined as MSS iCMS2, MSS iCMS3, MSI iCMS3 (which is the group that would be called CMS1 under the conventional CMS), MSI iCMS2 (again, more likely to be CMS1) or intermediate. We also had data on number of cells in each cluster by patient from the Pelka *et al* interactive portal, thus allowing me to calculate the proportion of cells in a given cluster out of total number of cells in clusters 15-31, the fibroblasts and pericytes **(Fig 3.4.3.2)**. This showed that CXCL14+ fibroblasts were most frequently seen in iCMS2 disease, consistent with my finding of this cell type in CMS2 and 3. There was a trend towards more GREM1+ CAFs and myofibroblasts being seen in MSI high disease.

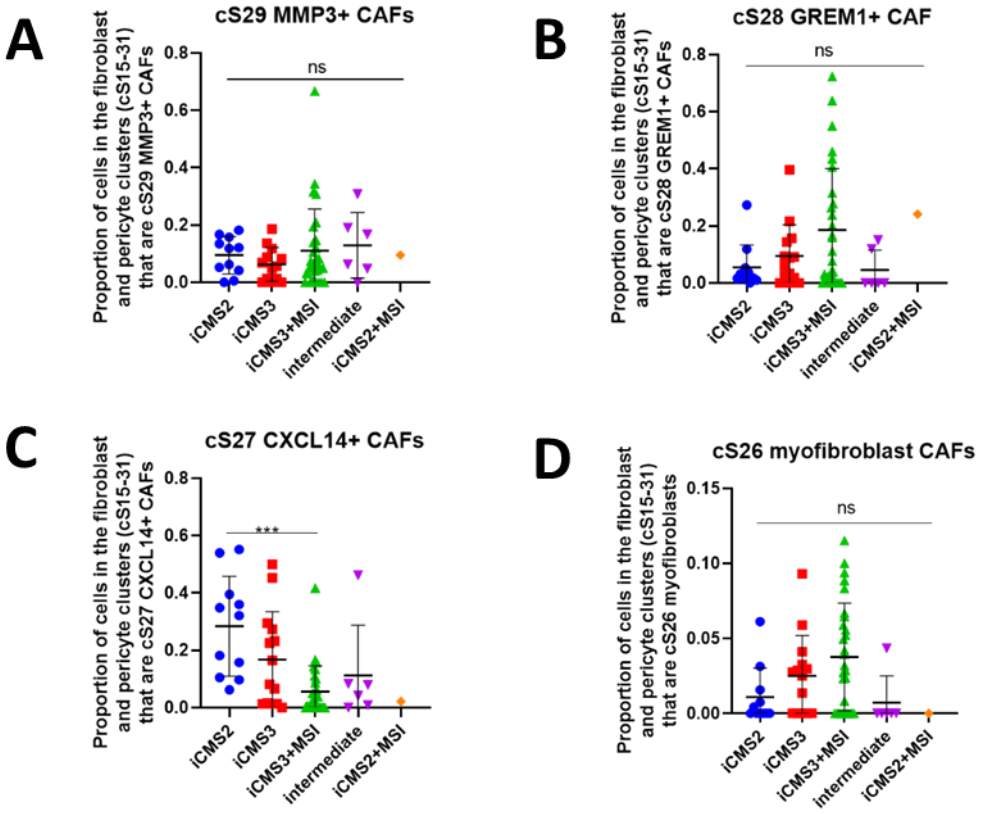


Fig 3.4.3.2 CXCL14+ CAFs make up a higher proportion of stromal cells in CMS2 disease
The cases from the scRNAseq dataset from Pelka *et al*. were divided into intrinsic CMS2 and 3, with and without MSI, by Joanito *et al*. Looking at the proportion of all fibroblast and pericyte clusters in each patient that were of **(A)** cS29 MMP3+, **(B)** cS28 GREM1+, **(C)** cS27 CXCL14+ and **(D)** cS26 myofibroblasts, significantly more cS27 CXCL14+ CAFs were seen in iCMS2 than iCMS3+MSI disease. Data analysed by one-way ANOVA (*** p<0.001, ns=non-significant).

3.5 Conclusions from Grampian Chapter

Our staining of 60 colorectal cancer cases with panels to identify fibroblast and immune subsets has successfully replicated known findings from the literature. Firstly, we saw trends towards higher numbers of stromal cells in CMS4 disease[20]. Secondly, the known association between T-cell and CMS1 and neutrophils and CMS1 and 4[179]. These known outcomes suggest that our dataset and our methods of analysis using HALO are likely to be generating reliable results.

We have also made some new findings through identification of myCAF and iCAF cells specifically. We have seen that there are significantly more PDPN+ iCAFs in CMS3 disease and more PDPN+/- pSMAD2+ myCAF-like stromal cells in CMS4 disease than the other subsets. There was positive correlation between numbers of PDPN+ iCAFs and numbers of immune cells, consistent with the hypothesis that iCAFs differentiate in response to IL-1 from immune cells and then produce cytokine and chemokines to recruit further immune cells. There was a trend towards increased numbers of myCAF-like stromal cells being associated with higher fibroblast TBRS. These new findings concur with the known higher TGFβ levels in CMS4 disease[20] and TGFβ1 as the predicted driver of the myCAF phenotype[96, 113].

We have also shown that iCAF-like stromal cells are significantly more likely to be found in close proximity, within 10um, of neutrophils than myCAF-like stromal cells. The distance of 10um means that, as fibroblasts are 10-12um across, these cells are likely to be next to each other and interacting directly. This finding again concurs with the consensus opinion that the iCAF phenotype is promoted by IL-1 signalling[96, 113], and the fact that neutrophils are a major producer of IL-1 in the TME. It is consistent with work from our lab in the IBD field, showing on multiplex IHC that an iCAF equivalent seen in IBD was found in areas of ulceration where large numbers of neutrophils are seen[85]. This is however the first time this co-localisation has been shown across multiple human CRC sections.

Another new finding is the increased likelihood of iCAFs being seen close to tumour vessels than myCAFs. Whether this is a direct effect, or an artefact of there being greater numbers of immune cells around vessels, remains to be investigated. Other studies have found both neutrophils[103] and tumour vessels[180] to be more frequently closer to the luminal border of primary CRCs than the deeper, invasive front. We also saw more iCAFs close to the tumour epithelium, this contradicts work in pancreatic cancer, that has located myCAFs close to the epithelium. Though it does concur with a recent study in colorectal cancer that showed a rim of fibroblasts staining positive for F3 around the epithelial cells[104], on scRNAseq these F3-positive cells were found to express CXCL1, MMP3 and CXCL14, markers of two clusters used by Pelka *et al,* including the most iCAF-like cluster identified by MMP3+. The finding that myCAF-like stromal cells are further away from other structures in the TME than iCAF-like stromal cells may suggest that the myCAFs are locating together in dense groups, whereas the iCAFs may be more spread out amongst other cells. This is supported by the fact that stimulating fibroblasts in culture with TGFβ1 leads to greater *TGFB1* production by the fibroblasts, meaning that the myCAF-like cells themselves may be driving the myCAF phenotype in adjacent cells. However, given that no specific fibroblasts markers were used to identify these cells as CAFs, and the myofibroblast cluster seen in Pelka also expressed no fibroblast markers, it is possible these are smooth muscle cells and not fibroblasts at all.

An interesting observation if that of CD56+ stromal cells forming a rim around the tumour nests and the fact that it is seen less frequently in CMS1 and CMS4 disease than other subsets. These cells do not seem to fall neatly into either a myCAF or iCAF phenotype. In the Pelka *et al.* dataset shows that NCAM is most highly expressed in the CXCL14+ cluster[103], but the pattern of staining we have observed with CD56 looks very similar to that seen by Wang *et al.* with F3[104], and F3+ fibroblasts were found in their dataset to express both CXCL14 and MMP3, markers of two of Pelka *et al* clusters.

3.6 Limitations and Future work

One limitation of this study is that all sections were taken were biopsies from locally advanced rectal cancers. The fact that these were biopsies meant that these were small areas that we were often unable to orientate as to whether they came from the luminal surface, centre or deep margin of the tumour. Much of the variation in proportions of fibroblast subsets seen and CMS group allocated probably reflected this rather than differences in the cancer as a whole, resulting in the lack of correlation with outcomes or mutations. We had tried to correlate proportions of neutrophils, T-cells and total fibroblasts with those estimated from the micro-array data using CIBERSORT and the MCP counter, but no significant correlation was seen. Being small areas also meant we were analysing TME in very close proximity to the neoplastic cells themselves, whereas some studies have primarily seen iCAFs at much greater distances[98].

Using rectal cancers rather than tumour from throughout the colon may have restricted the heterogeneity seen between individual samples, for example, MSI high and BRAF mutated tumours are more often found in the right side of the colon[20], and so were underrepresented in our study, although molecular alterations seen in rectal cancer are largely the same as those typically found in left-sided colon cancer[31]. By selecting cases that were found when the cancers were localised and suitable for treatment with curative intent may have reduced the chance of analysing the most aggressive CRCs.

We have correlated many of our findings with CMS subtype. There are many ways of measuring CMS subtype, while the original study utilised bulkRNAseq[20], the methods available to us from these cases were CMS predicted from microarray and image-based CMS from analysis of H&E sections[34]. The CMS status allocated did not always concur between these methods, it is known that there is considerable intra-tumour heterogeneity and biopsies from different parts of the same individual's tumour can result in different CMS subtyping. For many of the findings

analysed we got more significant results looking at histological CMS rather than conventional CMS from microarray. Two factors that may be contributing to this difference are that, firstly, the histological CMS was derived from H&E slides cut from the same block as our slides and therefore is from tissue that was physically closer to the tissue we analysed than that used for CMS from microarray, potentially reducing the impact of intra-tumour heterogeneity. Secondly, histological CMS has allocated every tumour a CMS subgroup, whereas CMS from microarray has several tumours allocated to an unclassified category; dividing the cases between a 5th subgroup will have reduced the power of any findings.

To be able to perform staining across 60 cases we had to use a high-throughput method, utilising the Vectra-polaris. The disadvantage of this is that we were limited to 6 stains on each slide and therefore to performing some analyses across two adjacent slides. To confirm our findings, it would be beneficial to have a much larger panel performed on a representative selection of cases, possibly using the CellDive system to stain up to 40 markers on a single slide. It would be helpful to have several markers to confirm the identity of cells as fibroblasts, for example other positive markers such as PDGFRα in addition to PDPN, and exclusion markers for pericytes, such as MCAM, PDGFRβ and RGS5, and smooth muscle cells, with MYH11, CNN1 or TAGLN. It would also be ideal to have more than one marker to identify each of the myCAF and iCAF subsets, for example adding αSMA for myCAFs and MMP3 for iCAFs. It would also be interesting to attempt to identify the CXCL14+ and GREM1+ clusters from Pelka *et al* using these markers. To have these on the same slide as an immune panel would allow us to confirm more confidently the association between iCAFs and neutrophils. To be able to confirm the status of the CD56 positive cells around the tumour nests it will be vital to perform CD56 staining on the same slide as the general fibroblast, myCAF and iCAF markers, to show positivity in the same cell. A further limitation of using multiplex imaging comes from the ability of the HALO software to accurately segment stromal cells.

If confirmed to come from fibroblasts, we would like to perform mechanistic studies to explain the CD56 staining around the tumour cells. We know that this is not an immediate, direct effect of TGFβ1 or IL-1β on transcription of *NCAM1* from our PCR data. Therefore, possibilities include that it is an indirect effect of cytokine stimulation on the fibroblasts, which would take more than 24 hours to be realised, or more likely that CD56 is induced by the neoplastic cells themselves. This could be explored through co-culture experiments between fibroblasts and tumour cells, and it would be fascinating to understand how the CMS subgroup leads to variation in the presence of these cells and how this impacts tumour progression.

The finding that iCAFs are more often seen in close proximity to vessels than myCAFs needs to be explained mechanistically, it is possible that it is an effect of there being more neutrophils nearer to the vessels. Experiments using co-culture, or transfer of conditioned supernatants from endothelial to fibroblasts, could be performed to see how this alters fibroblast phenotype. The proximity between the neoplastic cells and iCAFs also needs exploring mechanistically.

4 Cytokine response in Tumour and Normal Adjacent Stromal cells

4.1 Introduction to Chapter

It is hypothesised that the main cytokines driving the myCAF and iCAF phenotypes are TGFβ and IL-1β respectively. However, it is not known exactly which cell types the iCAF and myCAFs derive from, and whether they are both truly fibroblasts or if one or both subsets are pericytes that have adapted to the tumour microenvironment. Pericytes are contractile, mural cells of the vasculature, that have many similarities to both fibroblasts and smooth muscle cells. It is also currently unknown to what extent the iCAF and myCAF phenotypes are temporary responses to stimulation or longer-term alterations in cell state, likely due to epigenetic changes. If epigenetic changes are occurring, whether these alter the response to key cytokines, in particular enhancing the response to these signals to further cement the myofibroblastic or inflammatory state.

Previous work has suggested that TGFβ stimulation of fibroblasts leads them to secrete more TGFβ1 themselves, driving a positive feedback loop. A study in lung cancer, using 4 pairs of cancer and matched normal lung fibroblasts, showed that those stromal cells from the tumour were hyper-responsive to TGFβ1 in term of their production of extracellular matrix proteins[162]. A study in CRC suggested tumour fibroblasts were hyperresponsive to IL-1β compared to matched normal colonic fibroblasts in terms of the number of genes differentially expressed on RNAseq[163]. Stimulating the same primary fibroblasts with each of TGFβ1 and IL-1β has not been investigated, so it is unknown whether all tumour fibroblasts are hyper-responsive to both, or if there are distinct subsets of fibroblasts that are hyperresponsive to one of these cytokines only. While being hyper-responsive to one cytokine may consolidate a myCAF-like or iCAF-like

phenotype, hyper-responsiveness to both cytokines may lead to greater fluctuations in cell function over time.

The aim of this chapter is to study the activity of TGFβ isoforms and IL-1β on primary stromal cells derived from normal and tumour colorectal tissue. We will study the transcriptomic profile induced by these cytokines, look for hyper-responsiveness to one or both cytokines, both across all pairs of normal and tumour stromal cells and confirm their capacity to mimic the gene signature associated with CAF subsets. Comparing response of tumour and normal adjacent fibroblasts to both cytokines in the same pairs of cells will give clues as to the likely longevity of these phenotypes and the mechanisms promoting the divergent cell states.

Since primary stromal cells generated from plating out cell digest from normal and tumour are heterogenous, we will use two defined cell lines, one for fibroblasts and one for pericytes, as a reference to identify how different stromal cells respond to these cytokines and to identify the nature of the primary stromal cells grown for each patient. The fibroblast line used are CCD18-Co cells, these are non-malignant colonic fibroblasts isolated from a 2-month old girl. The pericyte line is hPC-PL derived from the placenta, as no colonic pericytes were available for culture. The advantage of utilising cell lines is that they grow readily in culture, though with repeated passage their gene expression profile may have altered from cells immediately *ex vivo.*

4.2 Characterisation of the Fibroblast and Pericyte cell lines

4.2.1 Characterisation of the Unstimulated cell lines

Prior to investigating the cytokine response in primary stromal cells, it was important to characterise them in their unstimulated state to confirm their cell type, as once the tissue is digested and a single cell suspension put into culture any adherent cell type can grow; fibroblasts, pericytes, smooth muscle cells, endothelial cells or those of the enteric nervous system. To identify the phenotype of the primary stromal cells we compared them to the CCD18-Co fibroblast and the hPC-PL pericyte lines. Therefore, we firstly studied potential markers of these cell lines, to choose the ones most likely to differentiate between them to perform in all the primary stromal lines. We also compared with expression in the fibroblast and pericyte clusters in the Pelka *et al.* single cell sequencing dataset, shown in **Fig 1.5.1.2.**

PDPN was seen at a low level in the normal fibroblasts, was not present in the pericytes, and was highly expressed in the CAFs, in the Pelka *et al* scRNAseq dataset. Therefore, the small, but non-significant increase in *PDPN* we observed in the CCD18-Co fibroblasts compared to the hPC-PL pericytes is consistent with these being fibroblasts from the normal colon, and I would expect levels to be higher in my primary stromal cells from the tumour **(Fig 4.2.1.1 A)**. *PDGFRA* was seen in most fibroblast clusters, but not the pericyte clusters, from the single cell sequencing, and accordingly we found them to be expressed more highly by RT-qPCR in the CCD18-Co fibroblast line than the pericyte line, p=0.0007 on T-test **(Fig 4.2.1.1 B)**. *THY1* gave a mixed picture in the scRNAseq dataset, with it being present in several though not all fibroblast and pericyte clusters, but in our RT-qPCR data was significantly higher in the fibroblast line than the pericyte line, $p < 0.0001$ on T-test **(Fig 4.2.1.1. C)**. We also looked at the expression of 3 genes generally considered to be pericyte markers: *MCAM, PDGFRB* and *RGS5*. While *MCAM* was present in the pericyte, but not fibroblast, clusters in the scRNAseq dataset it was actually more

highly expressed in our fibroblast than pericyte cell lines by RT-qPCR, p=0.0329 on T-test **(Fig 4.2.1.1 D)**. Given this disparity in results it was decided not to use this marker further in phenotyping our primary cells. *PDGFRB* was highly expressed in several clusters that were both CAFs and pericytes in the scRNA-seq data and was also lowly expressed in a couple of normal fibroblast clusters. Consistent with this *PDGFRB* was expressed at the same level in both our cell lines by RT-qPCR **(Fig 4.2.1.1 E)**. *RGS5* was present in the pericyte, but not fibroblast, clusters on the scRNAseq study, but in our cells no significant difference between fibroblast and pericytes was observed **(Fig 4.2.1.1. F)**.

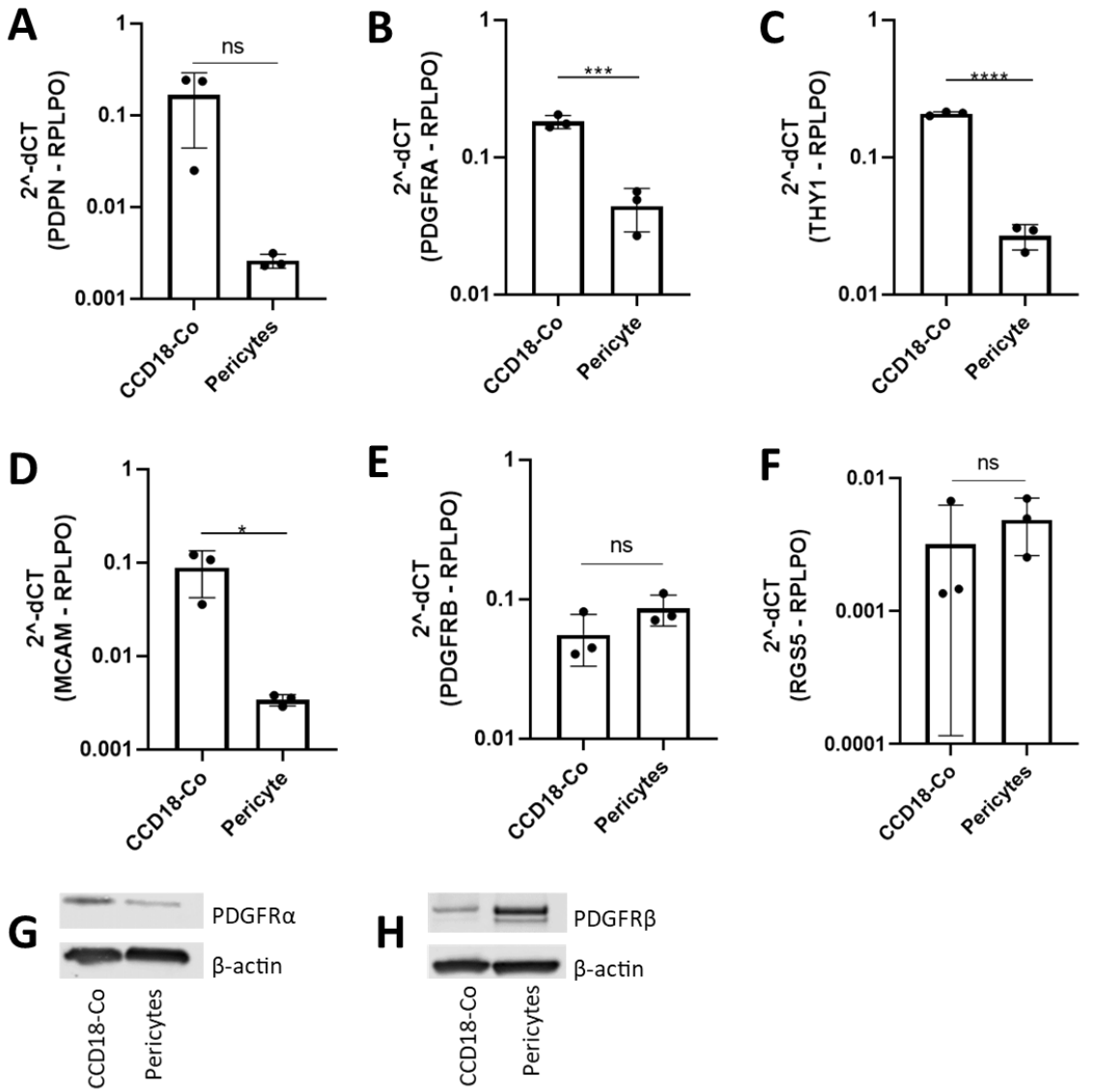
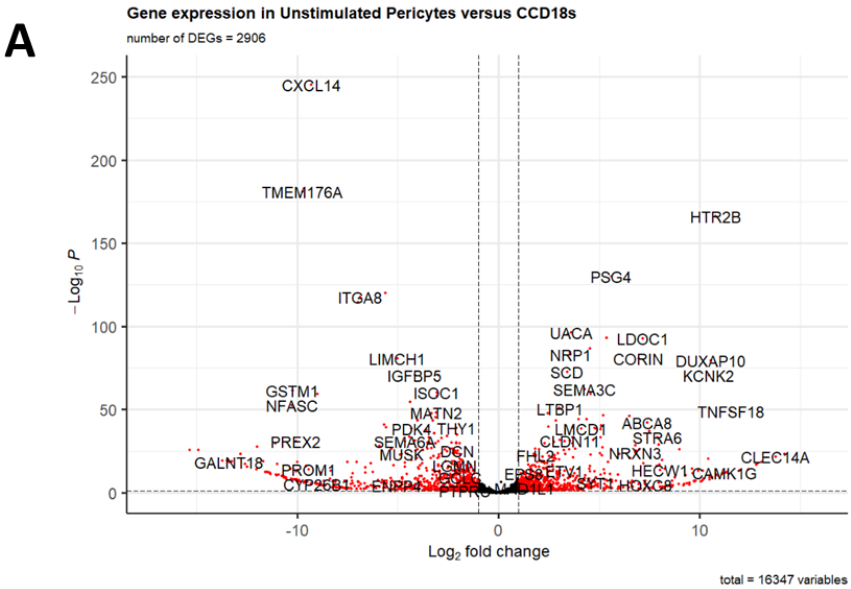


Fig 4.2.1.1 Markers of CCD18-Co Fibroblasts and hPC-PL Pericytes

RT-qPCR of CCD18-Co fibroblast and hPC-PL pericyte cell lines for (A) *PDPDN,* (B) *PDGFRA*, (C) *THY1*, (D) *MCAM* (E) *PDGFRB* and (F) *RGS5*. Each datapoint is an independent experiment, which is the average of two technical replicates, data were assessed by T-test (* p<0.05; *** p<0.001, **** p<0.0001, ns, no significance). (G-H) WB for (G) PDGFRα and (H) PDGFRβ.

We then compared protein expression of PDGFRα and PDGFRβ in the two cell lines by western blotting. PDGFRα was more highly expressed in the fibroblasts and PDGFRβ in the pericytes, despite there having been no difference in mRNA level **(Fig 4.2.1.1 G+H)**.

Next bulk RNAseq data was used to calculate differential gene expression between the unstimulated fibroblast and pericyte cell lines **(Fig 4.2.1.2 A)**. Looking at genes with a log fold change (FC) of at least +/- 1, those that were most significantly altered between the cell lines are shown in **Fig 4.2.1.2 B+C** and **Fig 4.2.1.2 D** shows a heatmap of these genes. Of note the gene most significantly upregulated in fibroblasts compared to pericyte is *CXCL14*, which is the key markers of the CXCL14+ CAF cluster identified by Pelka *et al.* *PDPN* was also very highly expressed compared to the pericytes in the sequencing data. Surprisingly, *FAP* and *GREM1* were some of the genes most highly expressed in the pericyte line compared to fibroblasts; *FAP* is generally considered to be a fibroblast marker and *GREM1* was the key marker for the GREM1+ CAF cluster identified by Pelka *et al.*



B **Genes significantly upregulated in Fibroblast compared to Pericyte cell lines**

Gene	log2FC	padj
CXCL14	-9.30	7.2E-246
TMEM176A	-9.73	1.3E-181
CADM1	-5.62	6.8E-121
ITGA8	-6.86	2.1E-118
CDH6	-6.97	6.3E-115
LIMCH1	-5.01	1.6E-81
IGFBP5	-4.16	1.7E-71
GPX3	-3.00	2.06E-71
GSTM1	-10.21	5.38E-62
ISOC1	-3.07	5.59E-61
SLCO2B1	-9.00	2.64E-60
SLC2A1	-3.06	4.48E-59
P3H2	-4.39	2.4E-55
NFASC	-10.26	1.73E-53
MATN2	-3.09	4.25E-49
NABP1	-3.31	1.94E-48
TRPA1	-3.11	3.54E-46
ACTG2	-4.00	1.52E-44
PDPN	-3.24	3.07E-44
ALDH1A1	-4.01	5.03E-44

C **Genes significantly upregulated in Pericyte compared to Fibroblast cell lines**

Gene	log2FC	padj
HTR2B	10.81	8.6E-167
PSG4	5.62	1.2E-130
UACA	3.62	6.56E-97
CYB561A3	3.47	2.39E-96
EBF1	5.38	8.35E-94
LDOC1	7.17	3.1E-93
ISLR	4.54	1.5E-87
CTHRC1	3.55	3.78E-85
NRP1	3.59	6.03E-84
CORIN	6.96	2.86E-81
DUXAP10	10.55	2.93E-80
SCD	3.44	1.47E-73
KCNK2	10.48	1.58E-71
SEMA3C	4.28	6.78E-63
LOXL4	4.58	5.6E-60
GREM1	3.01	1.48E-51
LTBP1	3.05	1.75E-51
TNFSF18	11.57	7.72E-50
FAP	2.44	1.39E-48
SFRP1	5.21	4.06E-47

D

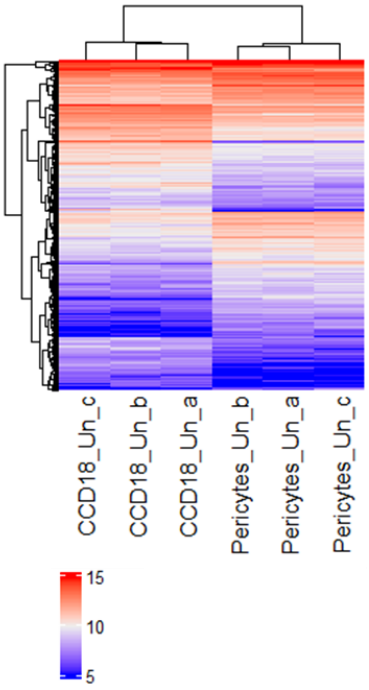
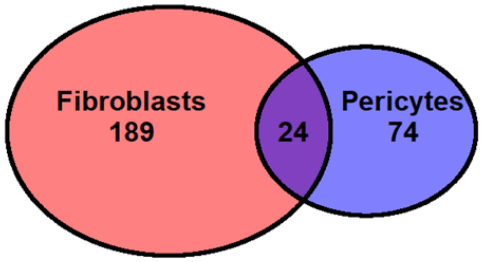


Fig 4.2.1.2 Comparing CCD18-Co Fibroblast and hPC-PL Pericyte cell lines using bulk RNAseq
(A) Volcano plot showing gene expression in unstimulated fibroblast versus pericyte cell lines, with log_2FC of the difference on the x axis and $-Log_{10}$p-value on the y axis *(DEseq, ggplot).* **(B)** Table of the 20 genes with expression most significantly upregulated in unstimulated fibroblasts compared to pericytes. **(C)** Table of the 20 genes with expression most significantly upregulated in unstimulated pericytes compared to fibroblasts. **(D)** Heatmap of the genes differentially expressed between fibroblasts and pericytes.

Gene set enrichment analysis was used to identify pathways differentially regulated between the cell types. Firstly, the relative expression of the 50 HALLMARK pathways was studied and none were found to be significantly altered between the two cell types. Secondly, differential expression of the gene ontology (GO) processes were analysed. **Fig 4.2.1.3 A** shows the numbers of GO fgsea processes significantly upregulated in fibroblasts and pericytes, some pathways overlapped between the cell types as different genes within the same pathway were differentially expressed. Once only pathways where at least 20% of the genes are differentially expressed are included there is no overlap, with 51 pathways more active in fibroblasts and 11 more highly expressed in pericytes. The most significantly upregulated pathways in each cell type are shown: pericytes have higher expression of pathways involved in angiogenesis and monocyte recruitment **(Fig 4.2.1.3 B),** whereas fibroblasts have upregulation of metabolic processing pathways **(Fig 4.2.1.3 C)**.

A



B **Pathways significantly upregulated in Pericyte compared to Fibroblast cell lines**

Pathway	padj	DEGs	Total genes in pathway	Ratio DEG:total
GOBP_POSITIVE_REGULATION_OF_LIPID_BIOSYNTHETIC_PROCESS	0.00102	16	68	0.24
GOBP_DEFENSE_RESPONSE_TO_GRAM_NEGATIVE_BACTERIUM	0.00398	10	35	0.29
GOBP_MONOCYTE_CHEMOTAXIS	0.00522	10	37	0.27
GOBP_HINDLIMB_MORPHOGENESIS	0.00691	8	25	0.32
GOBP_FORELIMB_MORPHOGENESIS	0.01663	8	30	0.27
GOBP_RETINA_VASCULATURE_DEVELOPMENT_IN_CAMERA_TYPE_EYE	0.01885	6	17	0.35
GOBP_CELL_MIGRATION_INVOLVED_IN_SPROUTING_ANGIOGENESIS	0.02815	10	50	0.20
GOBP_POSITIVE_REGULATION_OF_STEROID_METABOLIC_PROCESS	0.02979	6	19	0.32
GOBP_RENAL_FILTRATION	0.03615	7	27	0.26
GOBP_CELLULAR_RESPONSE_TO_LIPOPROTEIN_PARTICLE_STIMULUS	0.04156	7	28	0.25

C **Pathways significantly upregulated in Fibroblast compared to Pericyte cell lines**

Pathway	padj	DEGs	Total genes in pathway	Ratio DEG:total
GOBP_PRIMARY_ALCOHOL_METABOLIC_PROCESS	0.00021	19	78	0.24
GOBP_DIGESTIVE_SYSTEM_DEVELOPMENT	0.00026	22	108	0.20
GOBP_ICOSANOID_TRANSPORT	0.00092	11	34	0.32
GOBP_XENOBIOTIC_METABOLIC_PROCESS	0.00153	16	73	0.22
GOBP_NEGATIVE_CHEMOTAXIS	0.00202	11	38	0.29
GOBP_TERPENOID_METABOLIC_PROCESS	0.00219	15	68	0.22
GOBP_FATTY_ACID_TRANSPORT	0.00277	15	70	0.21
GOBP_PROSTAGLANDIN_TRANSPORT	0.00301	7	16	0.44
GOBP_MEMBRANE_DEPOLARIZATION	0.00733	12	53	0.23
GOBP_PROSTANOID_METABOLIC_PROCESS	0.00995	10	40	0.25

Fig 4.2.1.3 Processes differentially active between unstimulated fibroblast and pericyte lines
Gene ontology (GO) processes enriched in fibroblasts and pericytes (*fgsea*) (A) Venn diagram showing overlap between these pathways. (B) The 10 most significantly upregulated Pathways in pericytes compared to fibroblasts , of those where at least 20% of the genes in the pathway are upregulated. (C) The 10 most significantly upregulated pathways in fibroblasts compared to pericytes, of those where at least 20% of the genes in the pathway are upregulated.

4.2.2 Characterisation of the cytokine response in fibroblast and pericyte cell lines

4.2.2.1 Dose Response and Time-Course

A dose-response and time-course with TGFβ1-3 and IL-1β were performed using the CCD18-Co fibroblast cell line to find the ideal conditions to achieve maximal response. For the dose-response a series of 6 doses of cytokine were trialled for a 24-hour period and PCR quantification for *COL1A1* for TGFβ1-3, *CXCL1* and *IL-6* for IL-1β, and *MMP10* for all cytokines were performed. These showed that the dose-response plateaued at 5ng/ml for TGFβ1-3 and 0.1ng/ml for IL-1β and so these concentrations were used for all subsequent experiments, including those sent for RNAseq. Measuring *COL1A1*, comparing each isoform 5ng/ml with unstimulated, p=0.0096 for TGFβ1, p=0.0013 for TGFβ2 and p=0.0256 for TGFβ3, and comparing the same doses and examining *MMP10* expression, p=0.0001 for TGFβ1 and p<0.0001 for TGFβ2 and 3, on T-test. Comparing 0.1ng/ml IL-1β to unstimulated, p=0.022 for *CXCL1* and p<0.0001 for both *MMP10* and *IL-6* expression **(Fig 4.2.2.1 A-E)**. A time-course measured *COL1A1, IL-11, CXCL1* and *MMP10* by RT-qPCR after 3, 8 and 24 hours. For *COL1A1* and *CXCL1* a plateau had not been reached by the latest time-point, therefore 24 hours was used for further experiments and RNAseq. At the 24-hour timepoint, with TGFβ1-3, *COL1A1* expression had significantly increased with p values of 0.0008, 0.0021 and 0.013 respectively. *IL-11* also increased with TGFβ1-3 over time, with p-values of 0.031 for TGFβ1 and 0.0002 for each of TGFβ2 and 3. With 24hours of IL-1β *CXCL1* and *MMP10* both increased, with p-values of 0.0002 and 0.0304 respectively **(Fig 4.2.2.1 F-I)**.

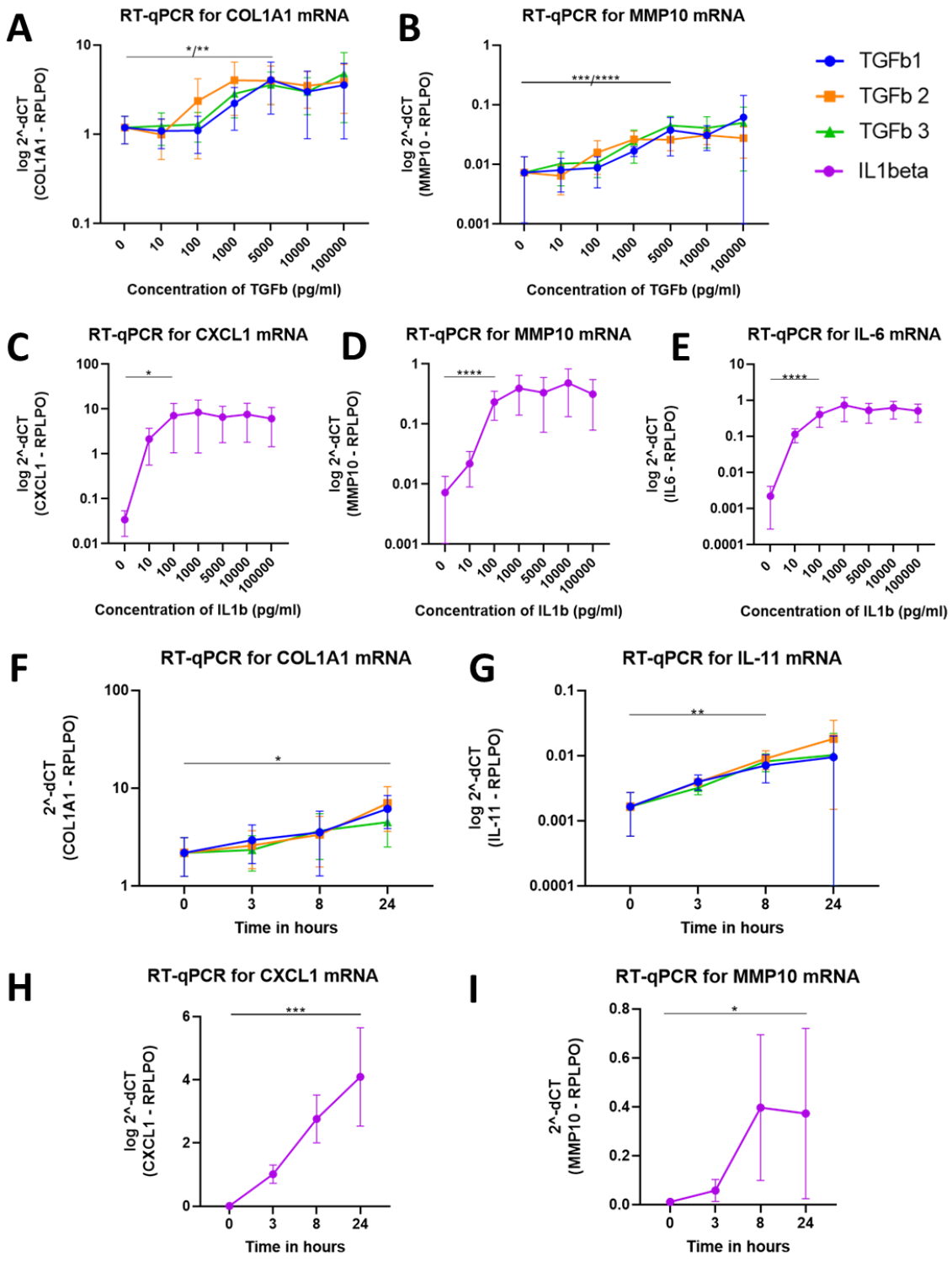


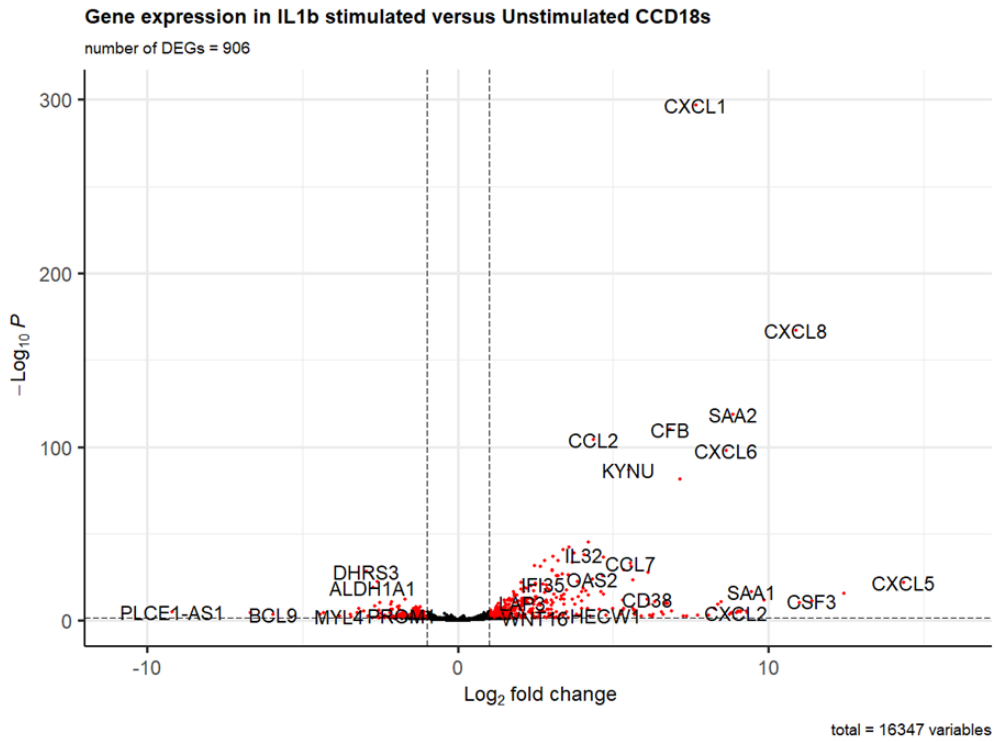
Fig 4.2.2.1 Dose-response and Time-course for TGFβ1-3 and IL-1β

(A-E) Dose-response at the 24 hour timepoint for TGFβ1-3 and IL-1β using RT-qPCR for **(A)** *COL1A1,* **(B)** *MMP10,* **(C)** *CXCL1,* **(D)** *MMP10* and **(E)** *IL-6.* **(F-I)** Time-course using either 5ng/ml TGFβ1-3 or 1ng/ml IL-1β using RT-qPCR for **(F)** *COL1A1,* **(G)** *IL-11,* **(H)** *CXCL1* and **(I)** *MMP10.* Data from 3 separate experiments, data was assessed by T-tests (* p<0.05; **p<0.01, *** p<0.001, **** p<0.0001).

4.2.2.2 Response to IL-1β

We compared the response to cytokine stimulation in CCD18-Co fibroblasts and hPC-PL Pericytes. DEseq was performed with 24-hours, 0.1ng/ml IL-1β versus unstimulated CCD18s to identify DEGs, shown in a volcano plot. The 20 most significantly up and down regulated genes, with log_2FC greater than 1 and less than -1 respectively, are given **(Fig 4.2.2.2 A-C)**. **Fig 4.2.2.3 A-C** shows the same comparison between IL-1β and unstimulated for pericytes and the overlap between the up and down regulated genes with IL-1β for the two cell types are shown **(Fig 4.2.2.3 D+E)**. Notably, there is much greater overlap in the upregulated genes than downregulated, with more upregulated genes in fibroblasts than pericytes, but more downregulated genes in pericytes than fibroblasts. Genes upregulated by IL-1β in both fibroblasts and pericytes include *CCL 2,5,7,8,11,12* and *20, CXCL 1,2,3,5,6,8,10* and *11, HLA B* and *F, ICAM1, IDO1, IL6, MMP1, MMP3, PDPN* and *RGS5.* Therefore IL-1β stimulation results in both these cell types becoming more active in terms of ECM degradation and immune cell recruitment. Genes upregulated in CCD18s only include *CD274* or PD-L1, suggesting this contributes to immune evasion, and in pericytes are *IL-1B* and *IL10*. There were only 23 genes significantly downregulated by IL-1β in both cell lines, including *MYH10* and *PDCD4*. Genes downregulated in pericytes only include *ACTA2* and *WNT2B.*

A



B **Genes significantly downregulated with IL1b in Fibroblast cell lines**

Gene	log2FC	padj
DHRS3	-2.91	7.9E-28
SELENOP	-2.57	2.04E-22
ALDH1A1	-2.71	2.71E-18
GPX3	-1.67	9.77E-12
SEMA6A	-2.11	1.63E-11
ZNF106	-2.48	5.01E-10
DLEU2	-2.12	5.31E-10
TOX	-1.86	2.47E-09
FGL2	-2.63	8.21E-09
ELL3	-1.92	1.48E-08
LAMA4	-1.17	3.19E-08
LACTB2-AS1	-3.17	4.34E-08
MATN2	-1.32	6.91E-08
APPL2	-1.40	1.24E-07
TTC3	-1.24	3.76E-07
GPAT3	-2.12	4.1E-07
KCND3	-1.85	9.35E-07
SCN3A	-2.06	1.48E-06
TPD52L1	-2.52	1.67E-06
LIMCH1	-1.19	2.62E-06

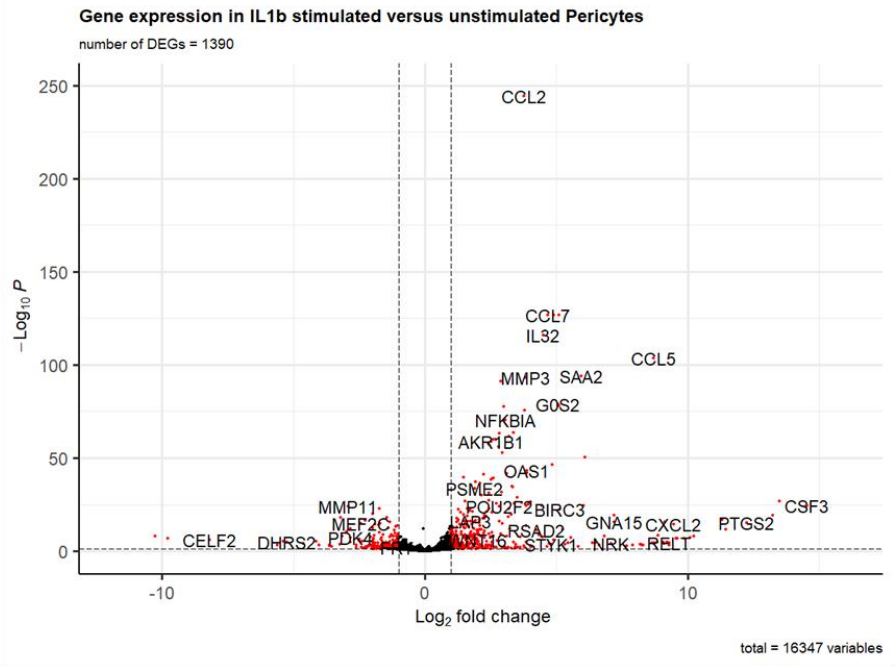
C **Genes significantly upregulated with IL1b in Fibroblast cell lines**

Gene	log2FC	padj
CXCL1	7.69	2.9E-291
CXCL8	10.90	5.2E-156
SAA2	8.89	1.9E-120
CFB	6.87	4.3E-106
CCL2	4.39	2.5E-105
CXCL6	8.66	6.32E-94
KYNU	5.54	2.57E-84
CXCL10	7.17	1.73E-77
CTSS	4.23	9.05E-45
MMP12	3.77	1.26E-41
NCOA7	3.42	1.77E-41
CA12	3.62	1.97E-41
IL32	4.10	6.91E-38
NFKBIA	3.09	2.96E-36
RGS5	4.71	1E-35
SLC39A8	3.27	2.46E-35
NAMPT	2.83	1.7E-34
CCL7	5.58	2.1E-34
ISG15	3.73	2.92E-32
VNN1	5.61	4.69E-32

Fig 4.2.2.2 IL-1β in the CCD18-Co Fibroblast cell line

(A) Volcano plot showing gene expression in IL-1β stimulated versus unstimulated fibroblasts, with log_2FC of the difference on the x axis and p-value on the y axis *(DEseq, ggplot)*. (B) Table of the 20 most significantly downregulated genes by p-value, of those genes with a log_2FC less than -1. (C) Table of the 20 most significantly upregulated genes by p-value, of those genes with a log_2FC greater than 1.

A



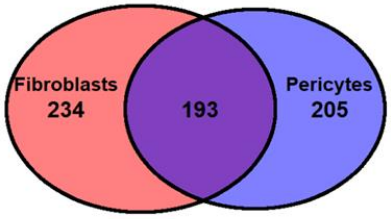
B **Genes significantly downregulated with IL1b in Pericyte cell lines**

Gene	log2FC	padj
MMP11	-2.91	6.04E-25
MFAP4	-1.73	1.45E-23
GPX3	-1.97	6.18E-21
SOX4	-1.46	6.93E-19
COL21A1	-3.2	1.37E-18
FRZB	-2.26	3.77E-17
SDC2	-1.32	2.94E-16
SLC40A1	-1.71	9.66E-16
MEF2C	-2.45	2.09E-15
TP53INP1	-1.89	3.64E-15
LBH	-1.77	6.86E-15
ACTA2	-1.03	2.18E-14
NR2F2	-1.11	7.37E-14
SELENOP	-1.4	3.34E-13
PPP1R14A	-1.99	3.61E-13
CRABP2	-2.82	1.14E-12
F2R	-1.17	8.00E-12
ADH1B	-2.98	5.52E-11
ASPN	-2.99	8.43E-11
ANGPTL4	-1.74	1.12E-10

C **Genes significantly upregulated with IL1b in Pericyte cell lines**

Gene	log2FC	padj
CCL2	3.78	5.3E-245
CFB	5.12	1.4E-127
CCL7	4.69	1.9E-127
SAA1	4.92	1.9E-127
IL32	4.50	1E-116
CCL5	8.71	2.9E-104
SAA2	5.95	1.14E-94
MMP3	3.82	7.21E-94
IFI27	2.88	9.89E-92
G0S2	5.07	2.04E-79
CTSS	5.09	7.3E-79
MMP1	3.00	3.3E-78
ALDH1A3	3.80	1.92E-76
NFKBIA	3.06	4.05E-71
TNFAIP6	3.36	3.79E-64
SERPINB2	2.83	4.75E-64
NAMPT	3.21	3.19E-62
IFI6	2.70	6.98E-61
PTGES	2.61	1.31E-60
AKR1B1	2.54	1.95E-59

D **Overlap in genes significantly upregulated with IL1b**



E **Overlap in genes significantly downregulated with IL1b**

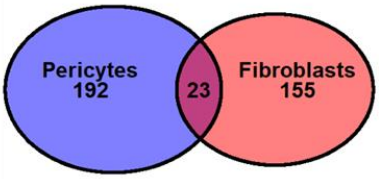


Fig 4.2.2.3 IL-1β in the hPC-PL Pericyte cell line

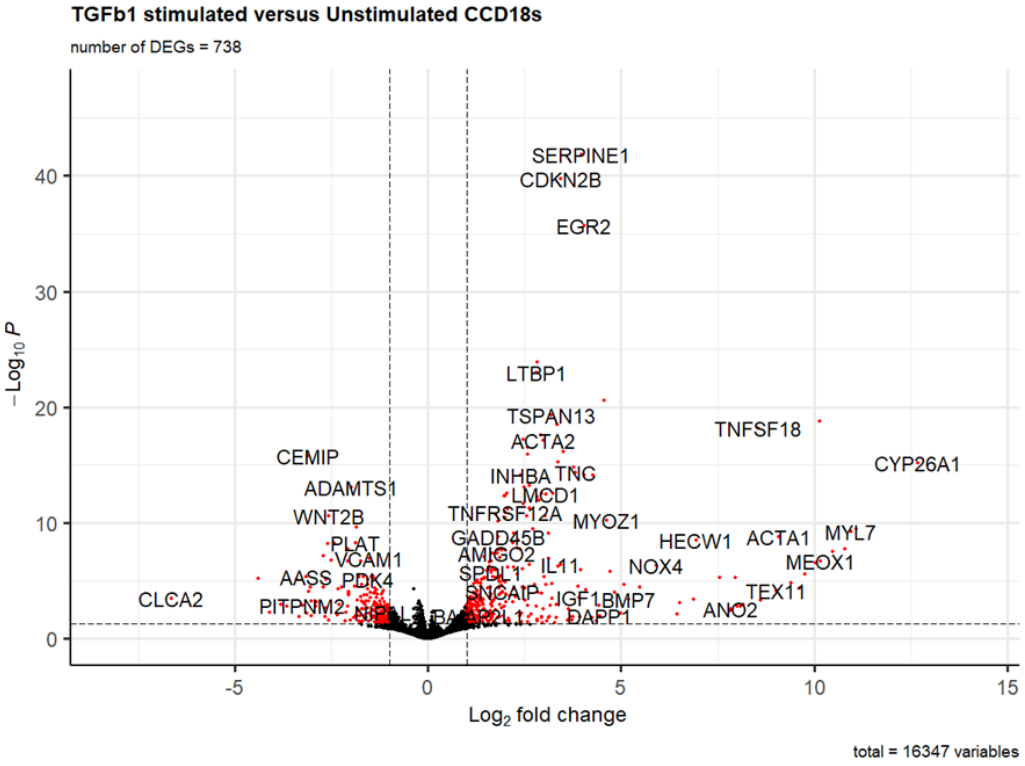
(A) Volcano plot showing gene expression in IL-1β stimulated versus unstimulated pericytes, with log_2FC of the difference on the x axis and p-value on the y axis*(DEseq, ggplot)*. **(B)** Table of the 20 most significantly downregulated genes by p-value, of those genes with a log_2FC less than -1. **(C)** Table of the 20 most significantly upregulated genes by p-value, of those genes with a log_2FC greater than 1. **(D)** Overlap in genes significantly upregulated ($Log_2FC > 1$ and adjusted p-value < 0.05) with IL-1β versus unstimulated samples, in fibroblast and pericyte cell lines. **(E)** Overlap in genes significantly downregulated ($Log_2FC < -1$ and adjusted p-value < 0.05) with IL-1β versus unstimulated samples in fibroblast and pericyte cell lines.

4.2.2.3 Response to TGFβ1

This was repeated for TGFβ1. **Fig 4.2.2.4** shows DEGs with 24-hours, 5ng/ml TGFβ1 versus unstimulated for CCD18s, as a volcano plot and the top 20 most significantly up and down regulated genes. **Fig 4.2.2.5 A-C** shows the information for pericytes and **Fig 4.2.2.5 D+E** show the overlap between the up and down regulated genes for the two cell types respectively. Genes that were upregulated in both fibroblasts and pericytes by TGFβ1 include the smooth muscle markers *CNN1, TAGLN* and *ACTA2, NOX4* and *PDPN*. Another smooth muscle marker, *MYH11,* was upregulated in fibroblasts, but not in pericytes. While *PDPN* was upregulated, *PDGFRA* remained unchanged. *CXCL14* was upregulated by TGFβ1 in pericytes, but not in fibroblasts, which is probably due to *CXCL14* being significantly higher in fibroblasts than pericytes at baseline **(Fig 4.2.1.2)**. *RGS5,* a pericyte marker, was upregulated by TGFβ1 in fibroblasts, but not in pericytes. *WNT2B* was downregulated in both cell types by TGFβ1, suggesting TGFβ1 reduces these cell's ability to support stem cells. In fibroblasts only, TGFβ1 led to downregulation of *IL1R1*, suggesting that once cells detect TGFβ1 they are less likely to respond to IL-1β and develop an iCAF phenotype.

Of note, looking at the upregulated genes, there is much greater overlap between fibroblast and pericyte response for IL-1β than TGFβ1. This may suggest that *in vitro* IL-1β could play a similar role in both cell types, whereas TGFβ1 leads to more disparate effects.

A



B **Genes significantly downregulated with TGFβ1 in Fibroblast cell lines**

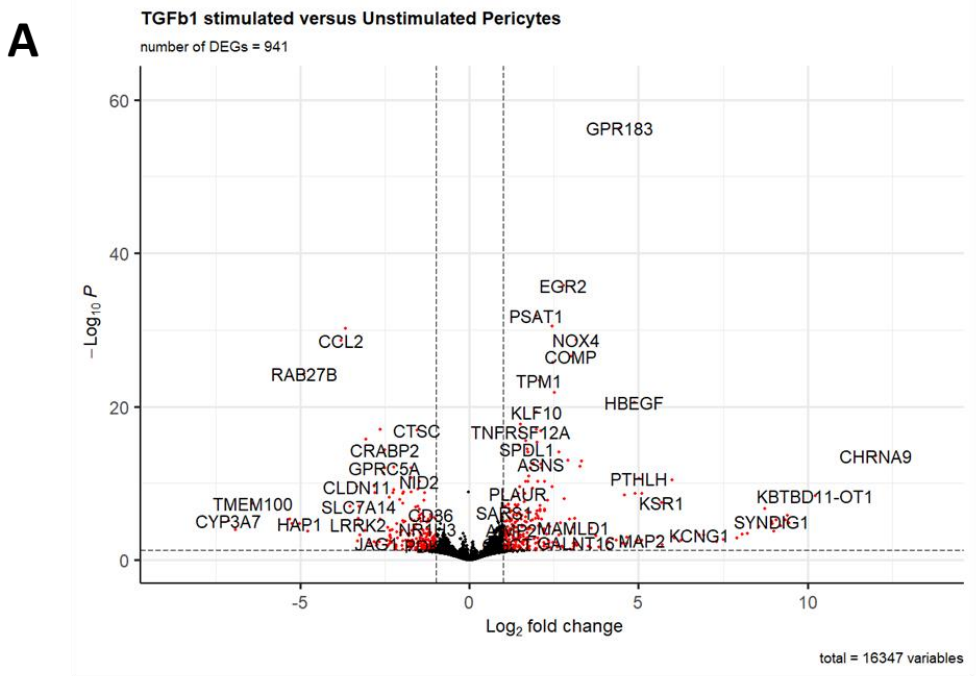
Gene	log2FC	padj
CEMIP	-3.09	7.44E-16
ADAMTS1	-1.96	1.86E-13
WNT2B	-2.55	5.46E-11
SELENOP	-1.82	7.86E-10
CCL2	-2.42	1.1E-08
RDH10	-2.59	1.41E-08
HS3ST3B1	-2.05	2.68E-08
PLAT	-1.83	6.81E-08
EFNA5	-2.68	9.6E-08
VCAM1	-1.49	1.21E-07
TNFRSF19	-2.04	2.55E-07
ADH1B	-2.47	4.44E-07
LIMCH1	-1.70	6.86E-07
LRRK1	-1.76	2.96E-06
TXNRD1	-1.42	6.25E-06
SYNE3	-3.36	8.62E-06
AASS	-3.14	8.87E-06
MCTP1	-4.36	0.000011
PXK	-1.40	0.000011
PDK4	-1.54	1.23E-05

C **Genes significantly upregulated with TGFβ1 in Fibroblast cell lines**

Gene	log2FC	padj
CDKN2B	3.46	5.44E-41
SERPINE1	3.99	1.93E-40
EGR2	4.07	1.29E-36
LTBP1	2.84	2.87E-24
SERPINE2	2.86	4.08E-23
HBD	4.58	3.27E-21
TSPAN13	3.22	6.48E-20
ACTC1	10.16	1.21E-19
TNFSF18	8.57	2.73E-19
MAP3K7CL	3.37	2.94E-18
UACA	2.49	7.08E-18
IGFBP3	2.94	2.81E-17
ARL4A	2.61	1.33E-16
ACTA2	2.99	1.59E-16
IL7R	3.81	2.13E-16
ALKAL1	3.52	2.13E-16
NGF	3.40	4.93E-16
CYP26A1	12.72	6.25E-16
CCN2	2.60	7.44E-16
GASAL1	4.07	1.15E-15

Fig 4.2.2.4 TGFβ1 in the CCD18-Co Fibroblast cell line

(A) Volcano plot showing gene expression in TGFβ1 stimulated versus unstimulated pericytes, with $\log_2$FC of the difference on the x axis and $-\log_{10}$p-value on the y axis *(DEseq, ggplot)*. (B) Table of the 20 most significantly downregulated genes by p-value, of those genes with a $\log_2$FC less than -1. (C) Table of the 20 most significantly upregulated genes by p-value, of those genes with a $\log_2$FC greater than 1.

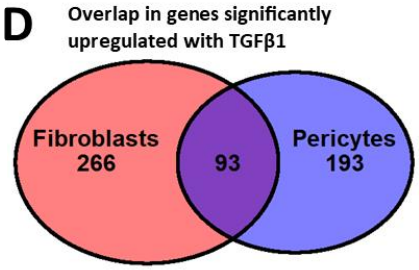


B **Genes significantly downregulated with TGFβ1 in Pericyte cell lines**

Gene	log2FC	padj
MTUS1	-3.66	6.25E-31
CCL2	-3.81	2.49E-29
RAB27B	-4.89	5.45E-25
IFIT1	-2.65	9E-18
CTSC	-1.55	1.34E-17
COL21A1	-3.06	1.71E-16
CRABP2	-2.50	4.37E-15
ABCA6	-2.25	8.19E-13
GPRC5A	-2.50	1.09E-12
PLK2	-1.73	1.09E-12
IFIT2	-2.56	5.25E-12
GJA1	-1.72	1.73E-11
NID2	-1.48	6.55E-11
TBC1D8	-2.82	2.35E-10
CLDN11	-3.34	2.71E-10
ABI3BP	-1.53	5.12E-10
SFRP1	-2.24	7.15E-10
ADAMTS5	-1.94	1.35E-09
DPP4	-1.72	1.35E-09
SLC40A1	-2.27	1.72E-09

C **Genes significantly upregulated with TGFβ1 in Fibroblast cell lines**

Gene	log2FC	padj
GPR183	4.45	5.24E-57
EGR2	2.78	1.56E-36
PSAT1	1.98	1.33E-32
CDK5RAP2	2.44	3.34E-31
NOX4	3.16	1.97E-29
COMP	3.01	2.68E-27
SPP1	2.46	1.89E-26
TPM1	2.07	4.43E-24
AHCYL1	2.50	1.38E-22
HBEGF	4.89	3.09E-21
KLF10	1.97	5.7E-20
NNMT	1.50	2.23E-18
SERPINE2	1.97	9.82E-18
TNFAIP6	2.12	1.51E-17
TNFRSF12A	1.52	1.8E-17
TP53I3	1.67	2.63E-16
LRRC32	2.01	4.11E-16
MAP3K7CL	1.47	1.17E-15
SPDL1	1.70	3.74E-15
TRIB3	1.74	8.16E-15



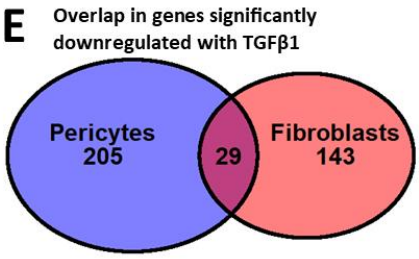


Fig 4.2.2.5. TGFβ1 in the hPC-PL Pericyte cell line

(A) Volcano plot showing gene expression in TGFβ1 stimulated versus unstimulated pericytes, with log_2FC of the difference on the x axis and p-value on the y axis *(DEseq, ggplot)*. **(B)** Table of the 20 most significantly downregulated genes by p-value, of those genes with a log_2FC less than -1. **(C)** Table of the 20 most significantly upregulated genes by p-value, of those genes with a log_2FC greater than 1. **(D)** Overlap in genes significantly upregulated (LogFC > 1 and adjusted p-value < 0.05) with TGFβ1 versus unstimulated samples, in fibroblast and pericyte cell lines. **(E)** Overlap in genes significantly downregulated (LogFC < -1 and adjusted p-value < 0.05) with TGFβ1 versus unstimulated samples in fibroblast and pericyte cell lines.

4.2.2.4 Pathway analysis of cytokine response in cell lines

We then studied the effects of TGFβ1 and IL-1β on expression of the HALLMARK pathways in fibroblast and pericyte cell lines. For IL-1β, with both cell lines, the 15 most significantly differentially regulated pathways are shown, with nearly all of these being upregulated pathways on normalised enrichment score. The pathways increased by IL-1β are largely the same in fibroblast and pericyte lines, with both including the inflammatory response, the interferon alpha and gamma response, TNFα signalling, JAK-STAT signalling and allograft rejection as their top pathways **(Fig 4.2.2.6 A+B)**. For TGFβ compared to unstimulated lines, only 5 pathways were significantly altered in fibroblasts and 4 in pericytes. As with the GO pathways analysis, there seemed to be slightly less overlap in cytokine response to TGFβ1 than IL-1β. In both lines the EMT, Myc and TNFα pathways were upregulated, then in fibroblasts apical junction and myogenesis pathways were also upregulated, and in pericytes the MTORC1 pathways was also significantly upregulated **(Fig 4.2.2.6 C+D)**.

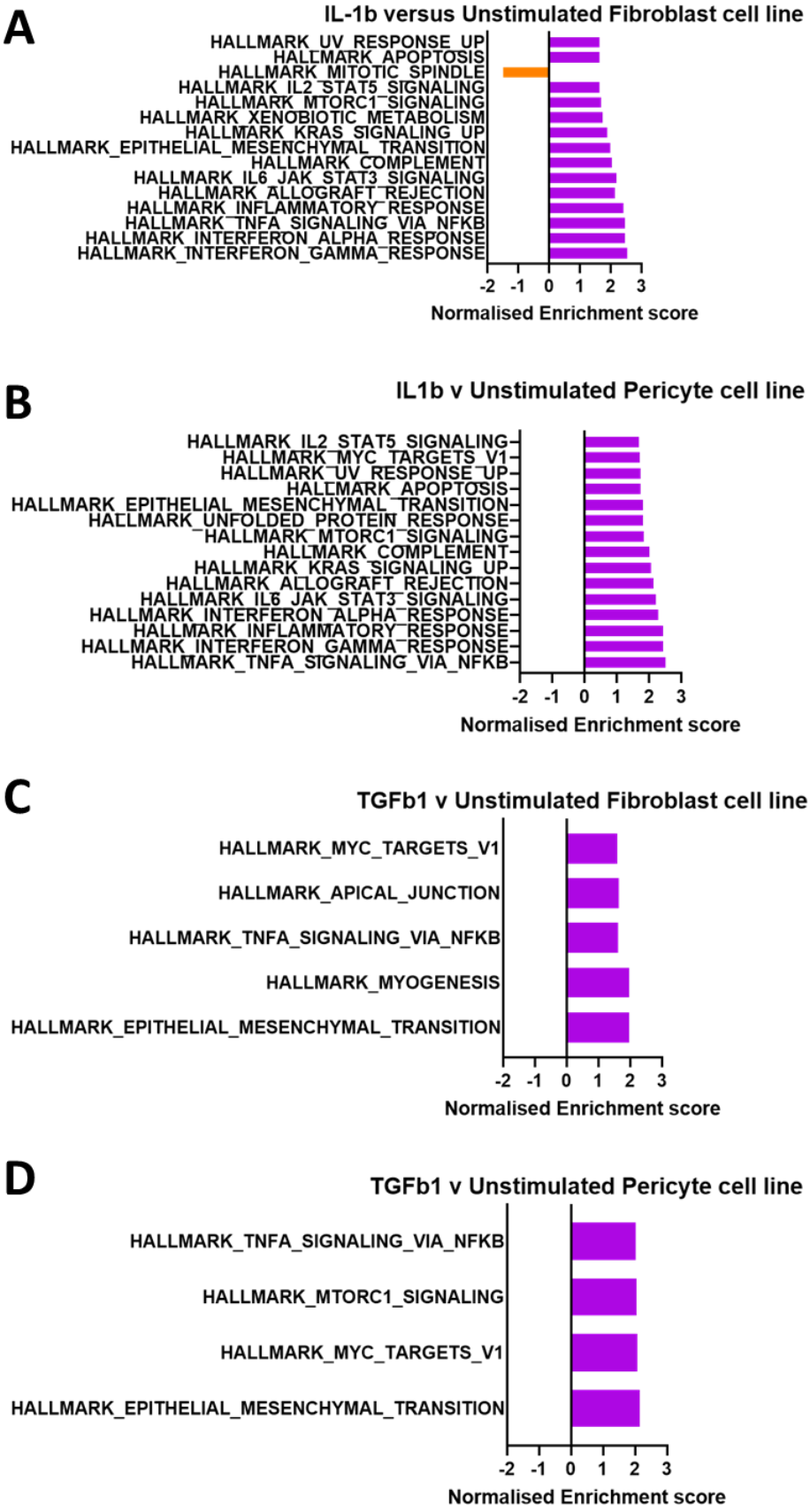


Fig 4.2.2.6. Hallmark Pathways altered in cell lines by IL-1β and TGFβ1

The 15 most significantly altered pathways with **(A+B)** IL-1β and **(C+D)** TGFβ1, in fibroblast **(A+C)** and pericyte **(B+D)** cell lines.

4.3 Characterisation of Primary Stromal Lines

4.3.1 Generation of Primary Stromal Cell lines

4.3.1.1 Processing of Patient samples

To compare tumour and normal adjacent fibroblasts, cells were extracted from resection samples from patients undergoing surgery for localised CRC. We received separate samples of tumour and adjacent normal tissue, the fat and muscle were resected from the normal colonic tissue so that only the mucosal layer remained. Both this and the tumour underwent EDTA washes to remove the epithelial cells, before being minced into smaller pieces and digested with Collagenase A and DNase 1. The digested tissue was filtered so that a single cell suspension remained. This was then plated for adherent cells to grow **(Fig 4.3.1.1).**

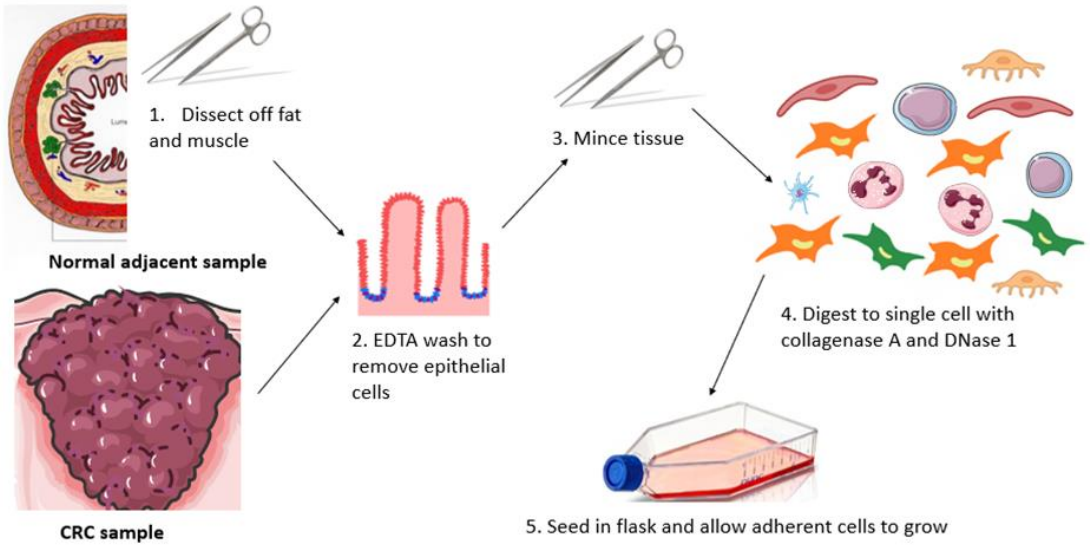


Fig 4.3.1.1 Schematic of processing Tumour and Normal adjacent samples to grow stromal cells

For some samples received either the tumour or normal adjacent failed to grow, therefore 8 pairs were available for analysis. Initially, within each sample, separate clusters of cells grew with varying morphology. After reaching confluence and undergoing passage the cells became more homogenous within each sample, however, significant morphological variation remained between samples, with some cells being spindle-shaped, others large with projections, and some small and rounder. As well as appearing morphologically heterogenous, the cells grew at

varying rates; the normal adjacent often grew faster than the tumour stromal cells, with a mean time to first passage of 17.5 versus 20.25 days respectively.

4.3.1.2 Indication of differential gene expression response between Tumour and Normal adjacent

The effects of cytokine stimulation with 24-hours of 5ng/ml TGFβ1-3 and 0.1ng/ml IL-1β were analysed by RT-qPCR of a large panel of genes. When all eight pairs were combined no difference was seen in the response of tumour and normal adjacent stromal cells **(Fig 4.3.1.2 A+B)**. However, on paired comparison some differences were noted between response of tumour and normal adjacent primary cells stromal to TGFβ2, although no significant differences were seen with TGFβ1, TGFβ3 or IL-1β **(Fig 4.3.1.2 C+D)**. *COL1A1* was more highly upregulated in the tumour stromal cells than the normal adjacent, with TGFβ2 leading to an average of 3.11 times higher expression than unstimulated in tumour stromal cells, compared to 1.95 times higher than unstimulated in normal adjacent cells (paired t-test, p=0.0089). However, *CXCL1* underwent greater downregulation with TGFβ2 in normal adjacent compared to tumour stromal cells, with an expression of 0.14 versus 0.30 of the unstimulated level (paired t-test, p=0.0415). While it was unexpected that on one measure TGFβ2 seemed to have a greater effect in tumour stromal cells and on another measure in normal adjacent stromal cells, these differences meant that further investigation of the differential response seemed worthwhile, despite the heterogeneity previously described in the primary lines. We therefore decided to perform bulk RNA-sequencing on the eight pairs of primary stromal cells, either unstimulated or stimulated with one of TGFβ1, 2 or 3 or IL-1β.

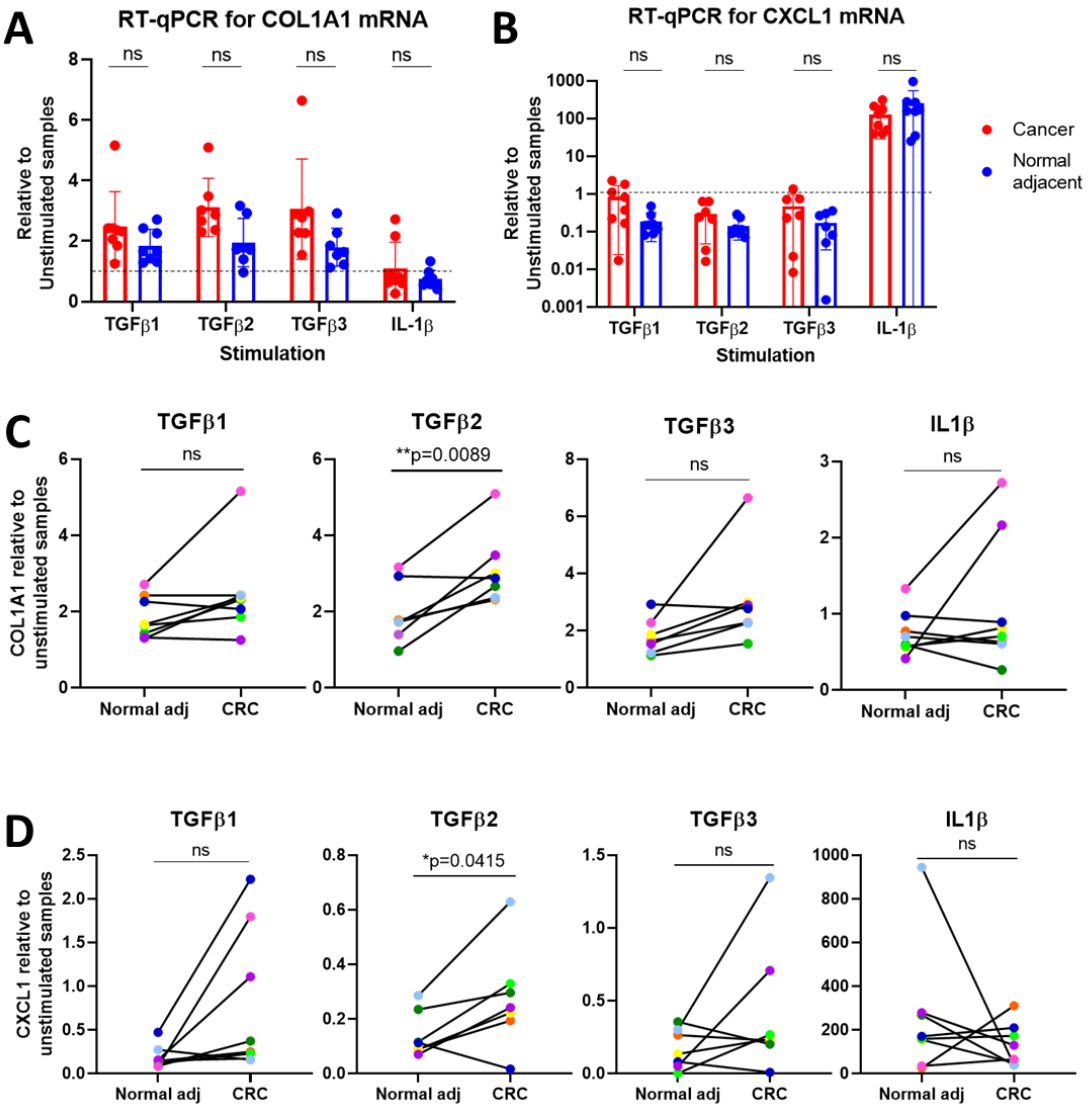


Fig 4.3.1.2 CRC samples are hyper-responsive to TGFβ2 compared to Normal adjacent
RT-qPCR for *COL1A1* **(A+C)** and *CXCL1* **(B+D)** , comparing response to stimulation with TGFβ1-3 or IL-1β to unstimulated samples, combing all cancer versus normal adjacent samples **(A+B)** or in paired cancer versus normal adjacent samples **(C+D).**

4.3.2 Quality Control of RNA sequencing data

As well as performing RNA sequencing on the primary stromal cells we decided to send the fibroblast and pericyte cell lines alongside these, stimulated with the same conditions; **Fig 4.3.2.1** shows a schematic of the planned experiment. We aimed for all conditions to be performed in triplicate before extraction of mRNA, but one primary stromal line, 3_Tumour, did not grow sufficiently to send all conditions in triplicate. Therefore, this line had only duplicate

samples that were either unstimulated or stimulated with TGFβ1 or IL-1β. For all other samples the experiments were performed as planned.

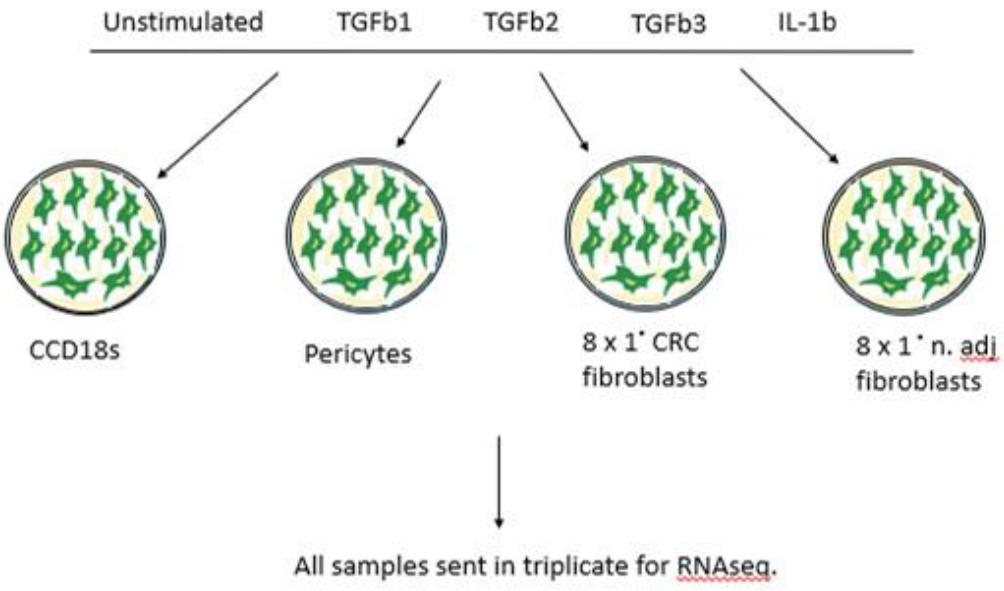


Fig 4.3.2.1 Schematic of cells types and stimulation conditions undergoing bulk RNA sequencing
CCD18-Co fibroblasts, hPC-PL pericytes and 8 pairs of matched primary tumour and normal adjacent stromal cells from resected CRCs were either left unstimulated or treated with TGFβ1, 2 or 3 or IL-1β for 24 hours. All experiments were performed in triplicate and mRNA extracted and sent for sequencing.

From the RNA sequencing adequate counts were seen in all samples and PCA plots and hierarchical clustering were used to check that the repeats yielded similar results. **Fig 4.3.2.2 A** shows a PCA using abundance data from all samples demonstrating that one TGFβ1 stimulated sample is an outlier. Performing hierarchical clustering on the samples from each of the five stimulation conditions showed 5_N.adj_T1_a, was an outlier with a height on the dendrogram of 1400 **(Fig 4.3.2.2 C)**, compared to a maximum of 1000-1050 for all other samples **(Fig 4.3.2.2 B, D-F)**. PCA of the TGFβ1 stimulated normal adjacent cells **(Fig 4.3.2.2 G)** again showed this sample as an outlier and so it was removed from all further analysis. Performing PCA analysis of each stimulation condition showed that all other repeats clustered together well.

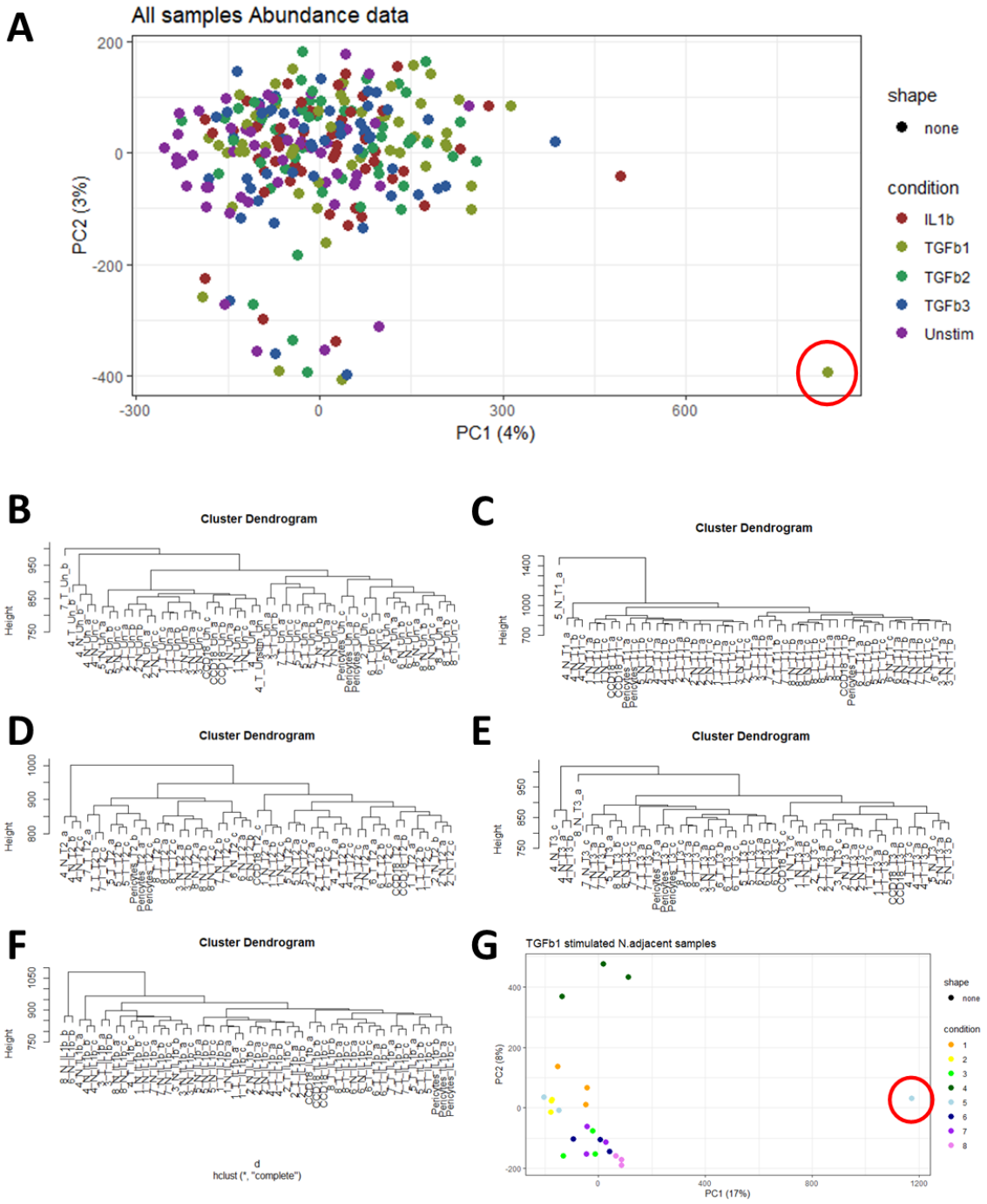


Fig 4.3.2.2 Identification of outlier samples

(A) PCA of all samples. **(B-F)** Hierarchical clustering of **(B)** Unstimulated or **(C)** TGFβ1, **(D)** TGFβ2, **(E)** TGFβ3 or **(F)** IL-1β stimulated samples. **(G)** PCA of TGFβ1 stimulated samples from N. adjacent primary cell lines.

4.3.3 Clustering of the Primary Stromal cells

4.3.3.1 Overall clustering of samples

The first question was how the samples clustered overall; whether by patient, the site either tumour or normal tissue, the cytokine stimulation, or the cell type either fibroblast or alternative stromal cell. Visualising the 2 main principal components on all samples showed that they primarily clustered by stimulation, with the unstimulated samples centrally, and the three isoforms of TGFβ and the IL-1β stimulated samples on opposite sides **(Fig 4.3.3.1 A)**. When showing the same PCA coloured by the individual patient and shaped by the site the sample came from, samples from the same patient often grouped together, whereas there was no clear grouping by tumour site **(Fig 4.3.3.1 B)**. When studying each stimulation condition separately, the samples from each primary line were most often seen clustering nearby the line from the other site (tumour or normal) from the same patient, rather than clustering with cells from the same site in different individuals. On PCA there was no suggestion of stromal cells from the tumour being more similar to each other than the normal stromal cells. This suggests that tumour stromal cells are not inherently different from normal adjacent stromal cells and they are not responding differently to cytokine stimulation. In addition to there being no clustering by site, when looking at just the TGFβ stimulated cells there was no clustering by which of the three TGFβ isoforms were used, suggesting they were having the same effect on these stromal cells **(Fig 4.3.3.1 C)**.

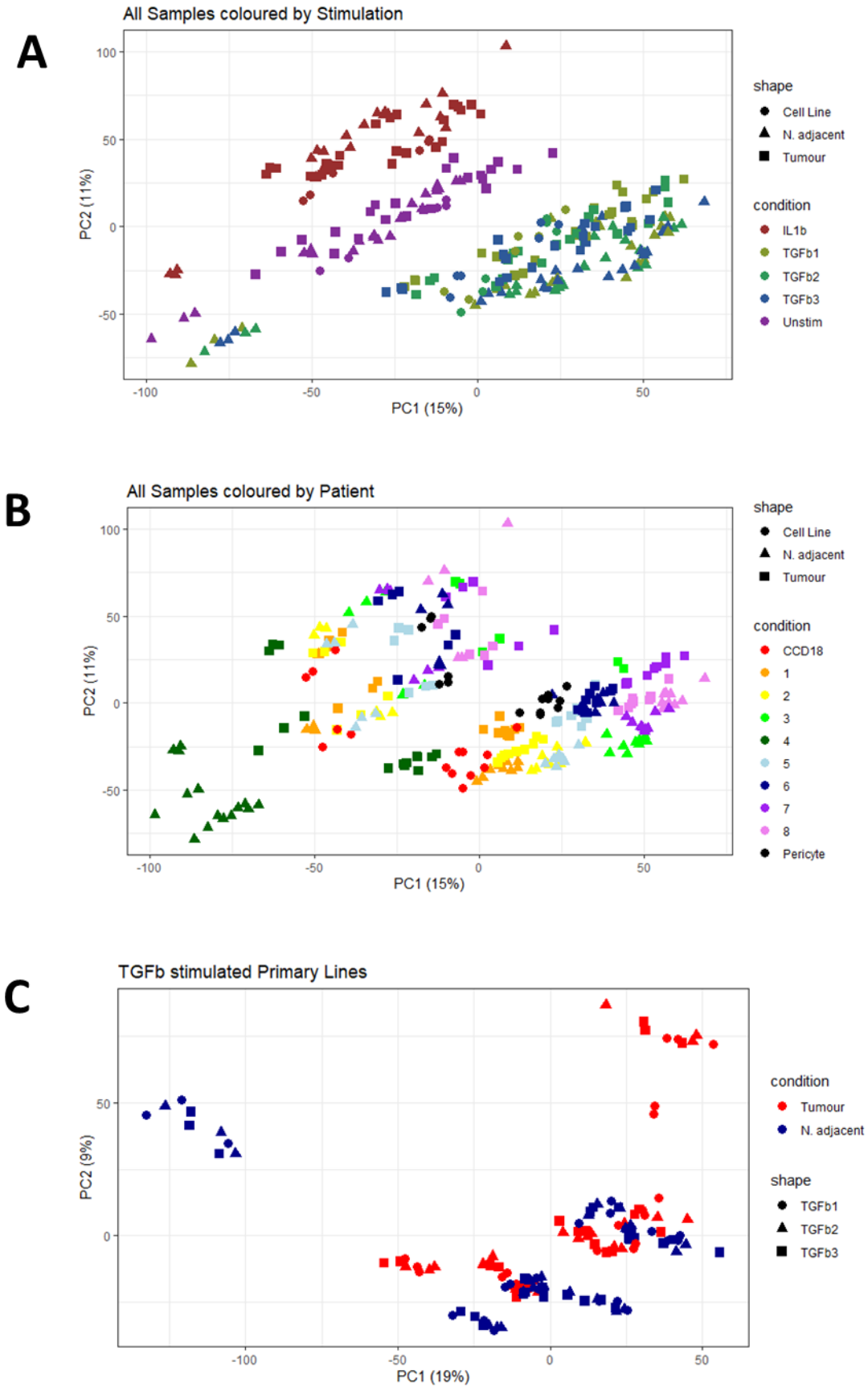


Fig 4.3.3.1 Samples cluster primarily by stimulation

(A) PCA of all samples coloured by stimulation conditions. (B) PCA of all samples coloured by the patient, shaped by the site the sample came from, Tumour or N. adjacent. (C) PCA of all primary cells stimulated with TGFβ, coloured by the site the sample came from, shaped by the TGFβ isoform.

4.3.3.2 Comparing unstimulated primary stromal cells with cell lines

The DEGs identified on comparing unstimulated CCD18s and pericytes were used for comparison in all unstimulated primary stromal cells and the results shown as a heatmap **(Fig 4.3.3.2)**. Two primary lines (3_Tumour and 7_Tumour) clearly clustered with the Pericyte cell line, whereas 6 samples (1_N.adj, 5_N.adj, and both tumour and normal from patients 2 and 4) clustered with the Fibroblast cell line. However, the remaining 8 primary samples did not clearly cluster with either.

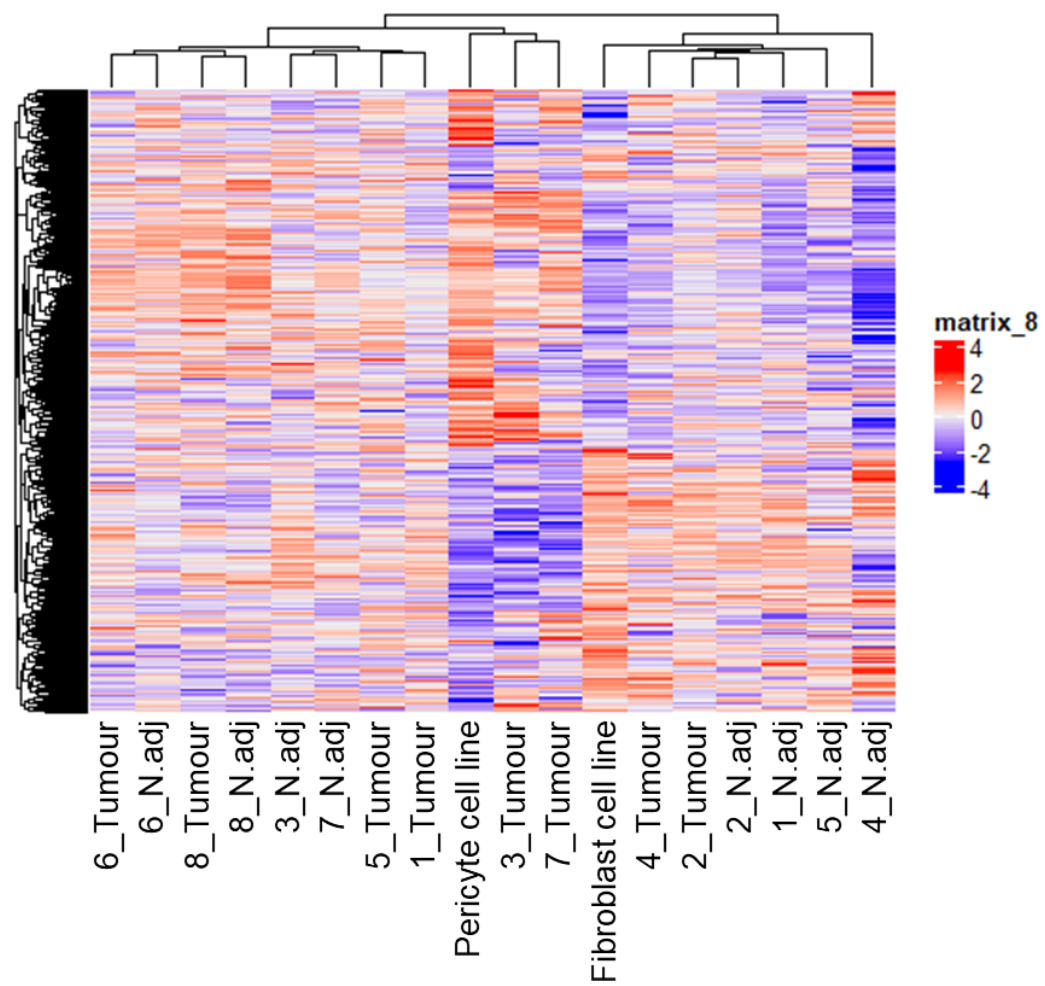


Fig 4.3.3.2 Most primary lines do not cluster with either fibroblast or pericyte cell lines

(A) Heatmap comparing gene expression in the unstimulated primary lines with gene expression in unstimulated CCD18-Co fibroblast and hPC-PL Pericyte cell lines, showing genes that are differentially expressed between the cell lines.

Performing a PCA of all the unstimulated cells revealed similar results, with the majority of the lines clustering close to the ccd18 fibroblast line, but some forming separate clusters further away **(Fig 4.3.3.3 A).** The main group of primary stromal cells found with the CCD18s were named "Cluster 1". The 3_Tumour and 7_Tumour primary stromal lines, which had clustered with the pericytes on the heatmap, did cluster together and were names "Cluster 2", but were separate from the pericyte line named "Cluster 4". 4_N.adj and to some extent 4_Tumour were also separate from the main group and were labelled as "Cluster 3". Of note 4_N.adj had clustered slightly separately from the main group in the PCA of all samples, shown in Fig 4.3.3.1 B. **Fig 4.3.3.3 B+C** show example expression of two of the main genes contributing to PC1, showing gradual changes in expression across the samples. For PC2 the pattern of expression was different, with some of the main contributing genes highly expressed in PC2 being seen in either Cluster 2 or Cluster 3, but not in both **(Fig 4.3.3.3 D+E).**

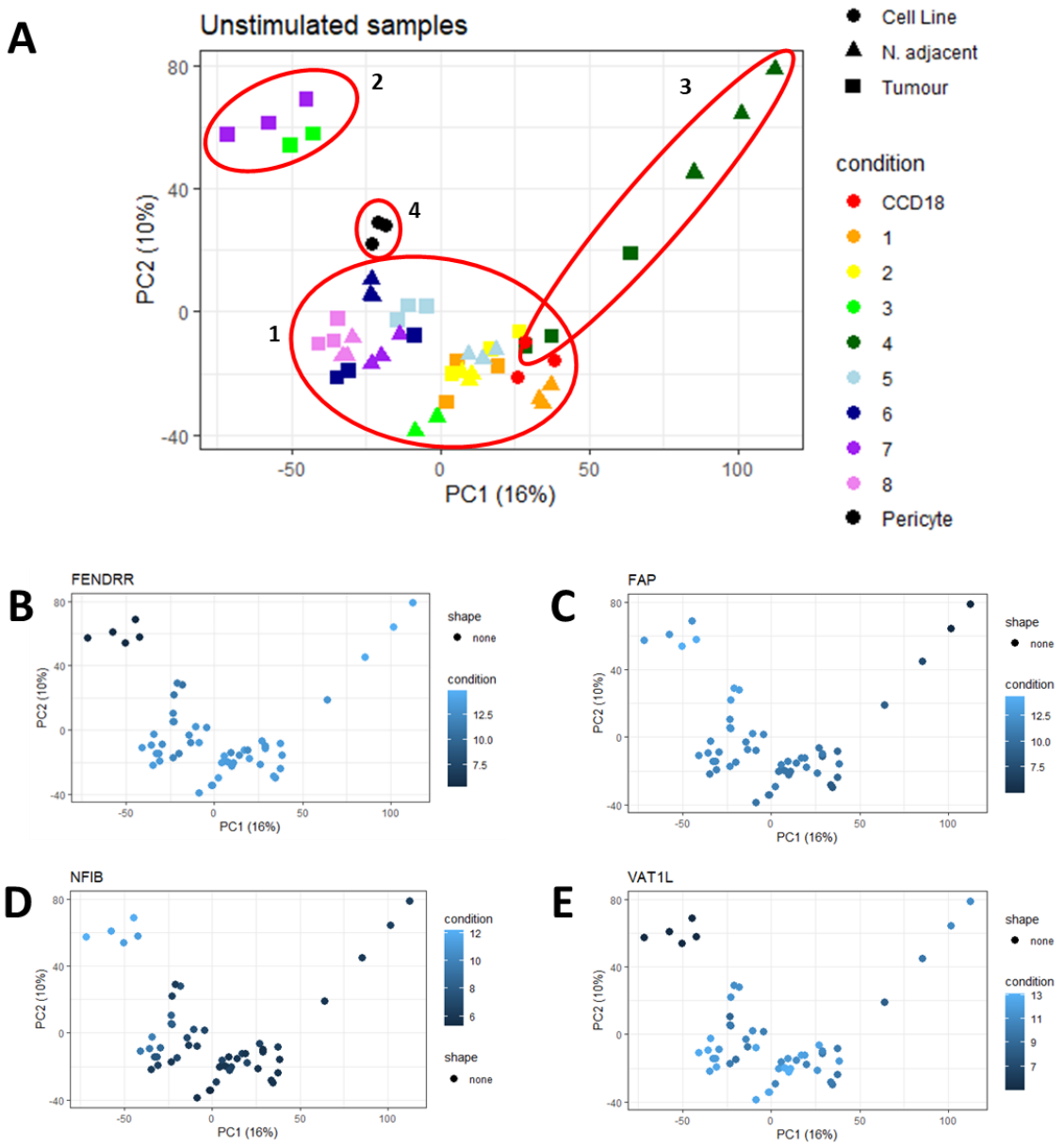
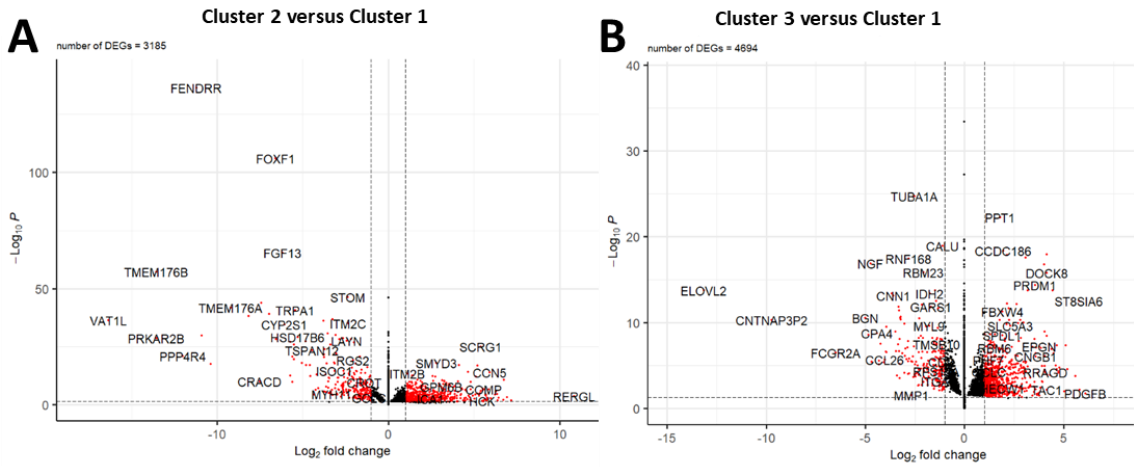


Fig 4.3.3.3 Unsupervised clustering of unstimulated samples revealed outlier primary stromal cell lines

(A) PCA of all unstimulated samples coloured by patient, shaped by site. **(B)** Example expression of a gene (*FENDRR*) with higher expression in PC1. **(C)** Example expression of a gene (*FAP*) with lower expression in PC1. **(D)** Example expression of a gene (*NFIB*) contributing to PC2, which is highly expressed only in Cluster 2. **(E)** Example expression of a gene (*VAT1L*), contributing to PC2, which is at lower expression only in Cluster 2.

4.3.3.3 Comparing gene expression and pathways between Primary stromal cell clusters

Fig 4.3.3.4 A+B compares each of Cluster 2 and Cluster 3 with the primary stromal cells within Cluster 1, with volcano plots highlighting DEGs. There were few significantly altered pathways on GO biological processes, although two of the four pathways upregulated in Cluster 2 compared to Cluster 1 were to do with TGFβ signalling. Those upregulated in Cluster 1 compared to Cluster 2 were to do with negative regulation of smooth muscle contraction. These two differences together suggest Cluster 2 may have a particularly contractile, myofibroblast-like phenotype **(Fig 4.3.3.4 C+D)**. Comparing clusters 1 and 3, there was only one pathway significantly altered, myelin assembly was greater in Cluster 1 **(Fig 4.3.3.4 E)**. Looking at the HALLMARK pathways, there were no differentially expressed pathways between cluster 1 and cluster 2. There were 3 pathways downregulated in cluster 3 compared to cluster 1: EMT, Myc targets and MTORC1 signalling.



C

Pathways significantly upregulated in Cluster 1 v 2	padj
GOBP_NEGATIVE_REGULATION_OF_MYOTUBE_DIFFERENTIATION	0.0069171
GOBP_HEPARIN_METABOLIC_PROCESS	0.0096775
GOBP_NEGATIVE_REGULATION_OF_MUSCLE_CONTRACTION	0.014482
GOBP_GROWTH_PLATE_CARTILAGE_DEVELOPMENT	0.015181
GOBP_CELL_PROLIFERATION_INVOLVED_IN_HEART_MORPHOGENESIS	0.015181
GOBP_HEPARIN_BIOSYNTHETIC_PROCESS	0.0287033
GOBP_CHONDROCYTE_DIFFERENTIATION_INVOLVED_IN_ENDOCHONDRAL_BONE_MORPHOGENESIS	0.0365762
GOBP_CYCLIC_NUCLEOTIDE_CATABOLIC_PROCESS	0.0365762
GOBP_NEGATIVE_REGULATION_OF_SMOOTH_MUSCLE_CONTRACTION	0.0365762
GOBP_TONGUE_DEVELOPMENT	0.0441555

D

Pathways significantly upregulated in Cluster 2 v 1	padj
GOBP_PEPTIDE_CROSS_LINKING	0.0040979
GOBP_TRANSFORMING_GROWTH_FACTOR_BETA_PRODUCTION	0.0321148
GOBP_TRANSFORMING_GROWTH_FACTOR_BETA_ACTIVATION	0.0461295
GOBP_LIPID_IMPORT_INTO_CELL	0.0461295

E

Pathways significantly upregulated in Cluster 1 v 3	padj
GOBP_MYELIN_ASSEMBLY	0.0442994

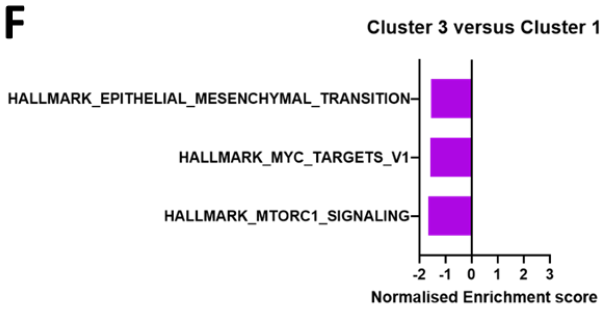


Fig 4.3.3.4 Comparison of primary stromal cell clusters

(A) PCA of all unstimulated samples, with clusters labelled. (B) Volcano plot showing DEGs between cluster 2 and cluster 1. (C) Pathways significantly upregulated in cluster 1 compared to cluster 2. (D) Pathways significantly upregulated in cluster in 2 compared to cluster 1. (E) Volcano plot showing DEGs between cluster 3 and cluster 1. (F) Pathways significantly upregulated in cluster 1 compared to cluster 3.

4.3.3.4 Defining the identity of Primary stromal cell clusters

We aimed to characterise the primary stromal cells from these different clusters by looking at their morphology, and the expression of fibroblast and pericyte markers by RT-qPCR and WB. The cluster 1 cells appeared to be more spindle-shaped, whereas those from Cluster 2 were smaller and more tightly packed and Cluster 3 smaller and round **(Fig 4.3.3.5 A-C)**. **Fig 4.3.3.5 D** gives a heatmap showing the relative expression by RT-qPCR of PDPN, PDGFRB and THY1 to the fibroblast and pericyte cell lines. All primary lines except 7_Tumour (from Cluster 2) showed high PDPN consistent with a fibroblast phenotype, all except 8_Tumour (Cluster 1) expressed PDGFRα at a level that could suggest they were fibroblasts, and only 4_N.adj (Cluster 3) and 8_Tumour (Cluster 1) had Thy1 levels more similar to that of pericytes than fibroblasts.

PDGFRα, a fibroblast marker, and PDGFRβ, which is expressed more highly on pericytes, were studied by western blot, **Fig 4.3.3.5 D** also shows a heatmap of these. This data comes from four separate western blots for each marker, each comparing two pairs of primary lines with CCD18-Co fibroblasts and pericytes, examples are shown in **Fig 4.3.3.5 E+F**. For each of the western blots the pixel density of the bands was measured and the ratio of PDGFRα to β-actin calculated. To make the four western blots comparable, for each one the PDGFRα/β-actin ratio in the CCD18-Co line and the PDGFRβ/B-actin ratio in the pericyte line were transformed to 1, and the log of these values are shown in the heatmap. Most samples express PDGFRα at a level similar to that seen in fibroblasts, whereas 1_Tumour, 3_N.adj and 8_Tumour (all cluster 1) are comparable to that of a pericyte. In most lines the levels of PDGFRβ are comparable to that seen in a fibroblast, however, for 2_Tumour (Cluster 1), 4_N.adj (Cluster 3) and 7_Tumour (Cluster 2) they are more comparable to pericytes.

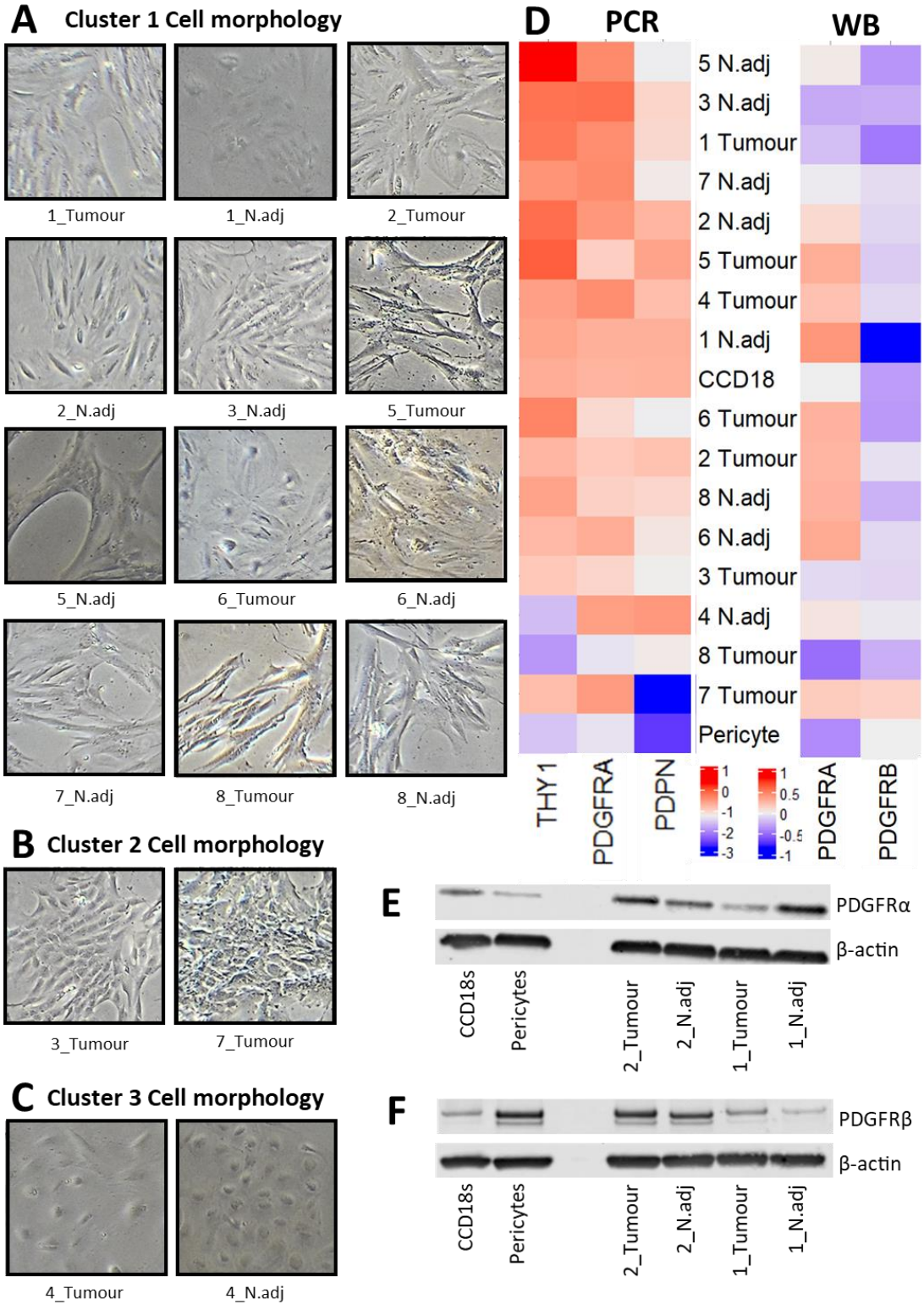


Fig 4.3.3.5 Phenotyping of the primary stromal cells

(A) Cell morphology of Cluster 1. (B) Cell morphology of cluster 2. (C) Cell morphology of cluster 3. (D) Heatmap showing expression of *THY1*, *PDGFRA* and *PDPN* by RT-qPCR, and the expression of PDGFRα and PDGFRβ by western blot, for each cell line. (E) Example western blot for PDGFRα. (F) Example western blot for PDGFRβ.

Overall it was therefore difficult to define each primary line as either a fibroblast or a pericyte, as a line may resemble a fibroblast cell line in one measurement and a pericyte cell line in another. This is unsurprising for a number of reasons. Firstly, the cell lines with which the primary lines were being compared will have been through multiple passages and are therefore unlikely to be 'normal' stromal cells. Secondly, it is possible that fibroblasts and pericytes are plastic, interchangeable cell types that have adapted to their tissue niche and therefore these adaptions will alter when put into 2D culture. However, on combining the results from sequencing, PCR, western blot and histology of the primary stromal cells it seems most likely that Cluster 1 represents fibroblast and Cluster 2 pericytes. Cluster 3 seems to have a telocyte fibroblast phenotype, with high expression of SOX6 **(Fig 4.3.3.6)** and BMP3. Given the lack of absolute certainty about cell phenotype, it was decided to initially investigate the effects of cytokine stimulation across all 8 pairs of primary cells, and then to investigate the effects of cytokine stimulation in just the 5 pairs where both tumour and normal adjacent stromal cells came from Cluster 1.

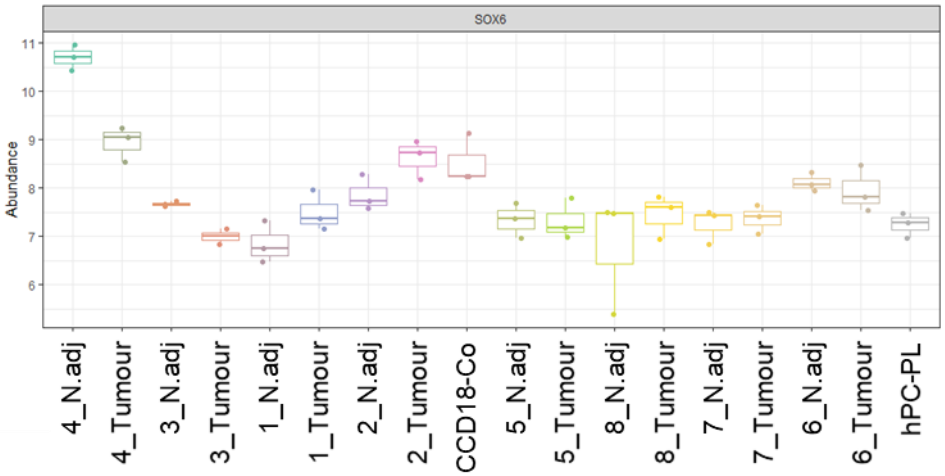


Fig 4.3.3.6 4_N.adjacent expressed high SOX6 suggestive of a telocyte identity
Boxplot showing SOX6 expression in each of the unstimulated cell lines.

4.4 Primary Stromal Cell response to Cytokines

4.4.1 Comparing Unstimulated Tumour and Normal Adjacent Primary Stromal cells

To compare the tumour and normal adjacent stromal cells first differential gene expression between the unstimulated cells was compared, shown as a volcano plot in **Fig 4.4.1.1 A**. This identified 591 DEGs, the most significantly altered of which was KRT16, which encodes keratin 16, a type 1 cytokeratin generally found in epithelial tissues. Drawing a boxplot of expression of this gene across the two groups of stromal cells showed it to be more highly expressed in just two of the tumour lines, 3_Tumour and 7_Tumour, specifically **(Fig 4.4.1.1 B)**. This may suggest that these two lines are epithelial in origin, rather than stromal cells; in the context of being tumour samples these could potentially be neoplastic cells that have undergone EMT. However, these lines also express fibroblast markers such as FAP **(Fig 4.4.1.1 C)**. Pathway analysis of the genes upregulated in tumour compared to normal stromal cells identified no significantly altered GO processes where at least 20% of genes in that process were altered, and looking at HALLMARK pathways, none were found to be significantly different between unstimulated tumour and normal adjacent.

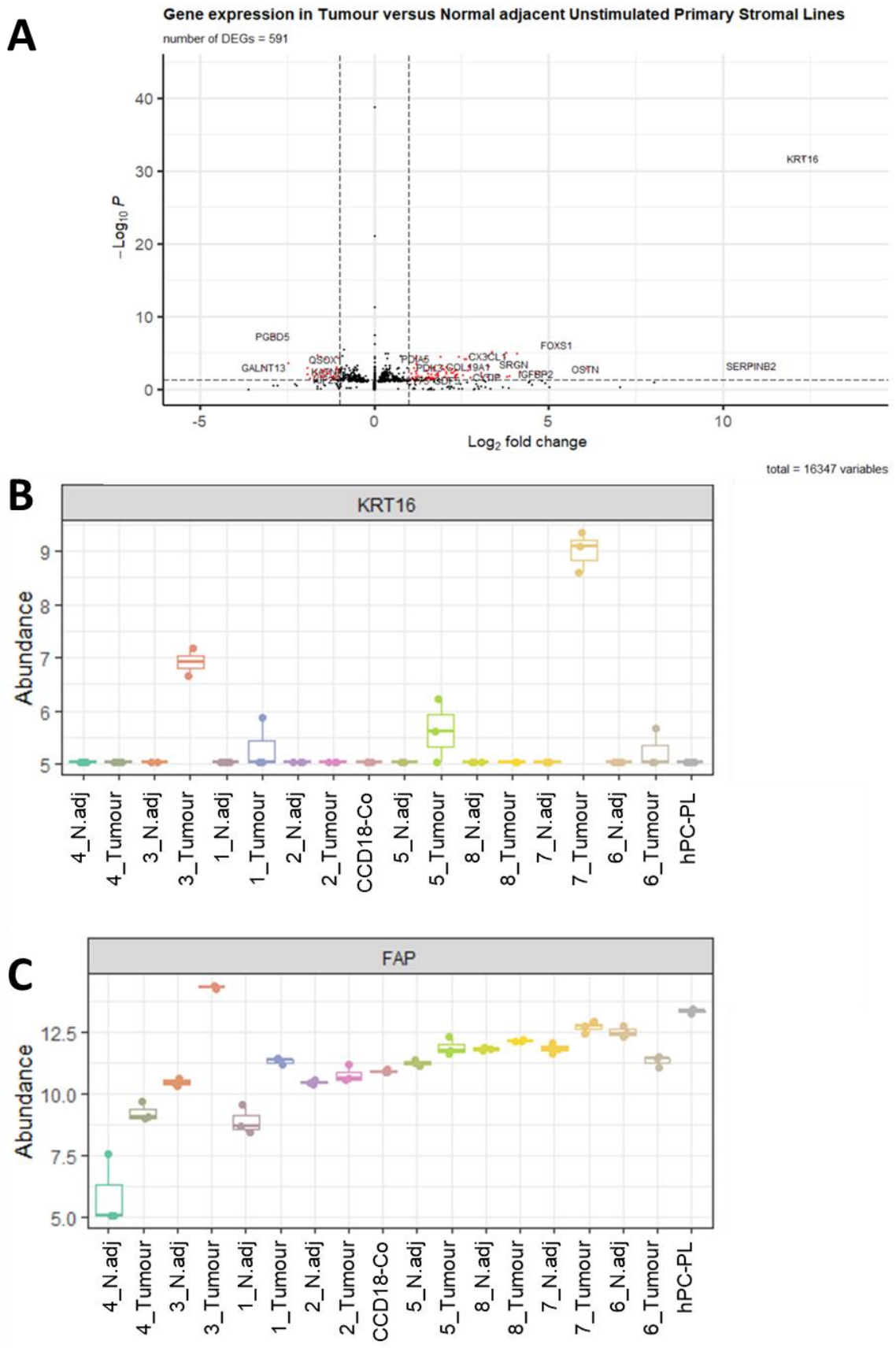


Fig 4.4.1.1 Differential gene expression between unstimulated Tumour and Normal adjacent primary stromal lines

(A) Volcano plot showing DEGs between unstimulated primary stromal lines from tumours and normal adjacent. (B) Boxplot showing KRT16 expression in each of the unstimulated cell lines (C) Boxplot showing FAP expression in each of the unstimulated cell lines

4.4.2 Comparing Tumour and Normal Adjacent Stromal cells response to TGFβ

Differential gene expression with 24-hours, 5ng/ml TGFβ1 versus unstimulated was calculated for the eight tumour lines combined and the eight normal adjacent lines combined and is shown in volcano plots **(Fig 4.4.2.1 A+B)**. **Fig 4.4.2.1 C** shows the fold change with TGFβ1 versus Unstimulated in the lines from the 2 different sites plotted against each other, each point represents one gene. Genes where the change is statistically significant are coloured; yellow where it is significant in cells from both sites, blue where the change is only significant in tumour cells and red where it is significantly changed in normal adjacent only. There is strong positive correlation, suggesting there is a similar response by the majority of genes to TGFβ whatever site the stromal cells were derived from. There were however some genes that were altered in one cell site, but not the other. The expression of the labelled genes that are significantly altered with TGFβ1 in cancer stromal cells with a logFC of at least +/- 4, but not significantly altered in stromal cells from normal adjacent tissue was visualised with boxplots **(Fig 4.4.2.2)**. These did not show any consistent changes across the 8 pairs in these genes. They were all genes of fairly low expression in the stromal cells with counts of below 10 except for the occasional individual line.

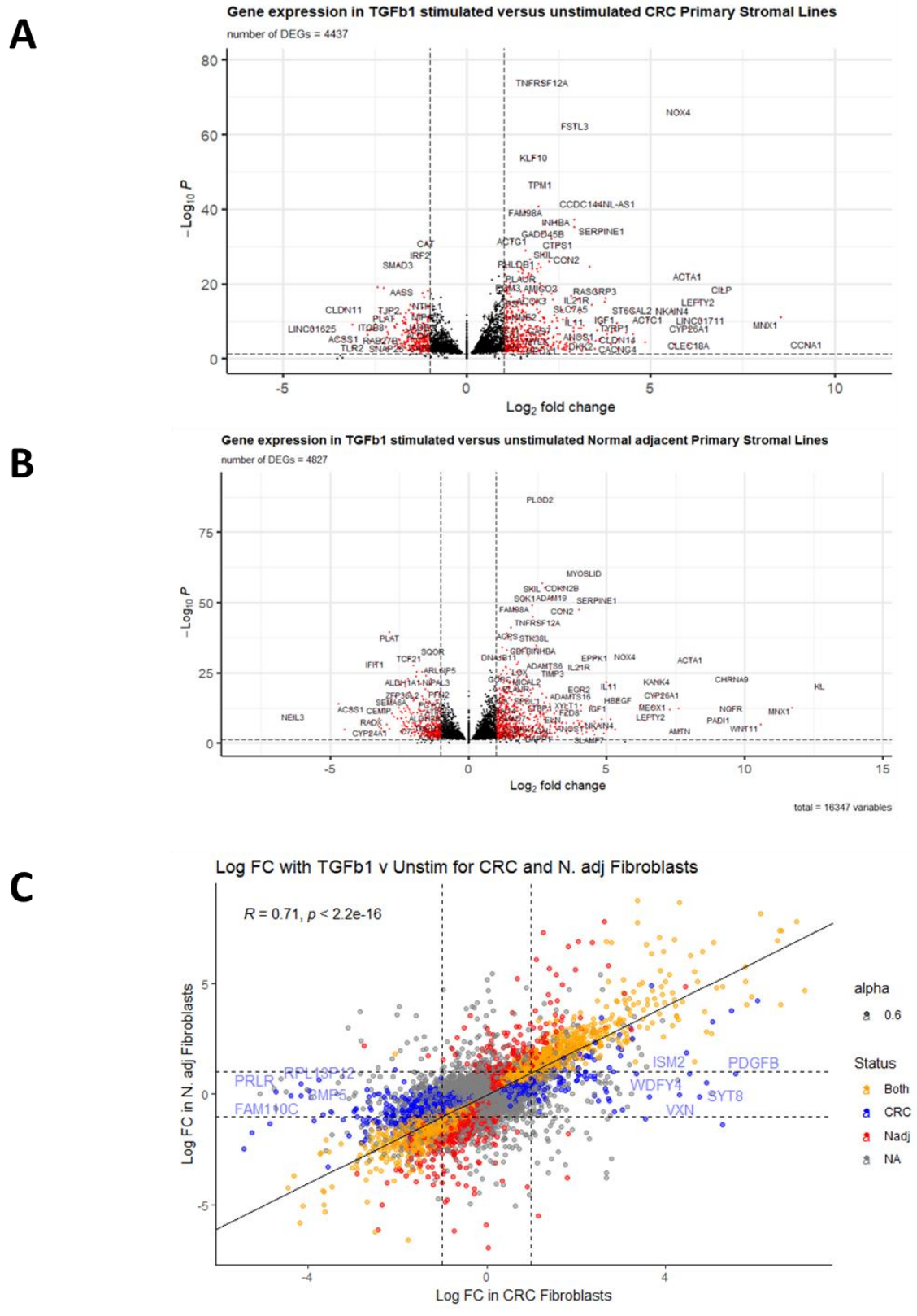


Fig 4.4.2.1 Differential gene expression with TGFβ1, comparing Tumour and Normal adjacent primary stromal line

(A) Volcano plot showing DEGs between TGFβ1 and unstimulated primary stromal lines from tumours. **(B)** Volcano plot showing DEGs between TGFβ1 and unstimulated primary stromal line from normal adjacent tissue. **(C)** Plot of the log fold change of every gene with TGFβ1 compared to unstimulated in tumour versus normal adjacent stromal cells.

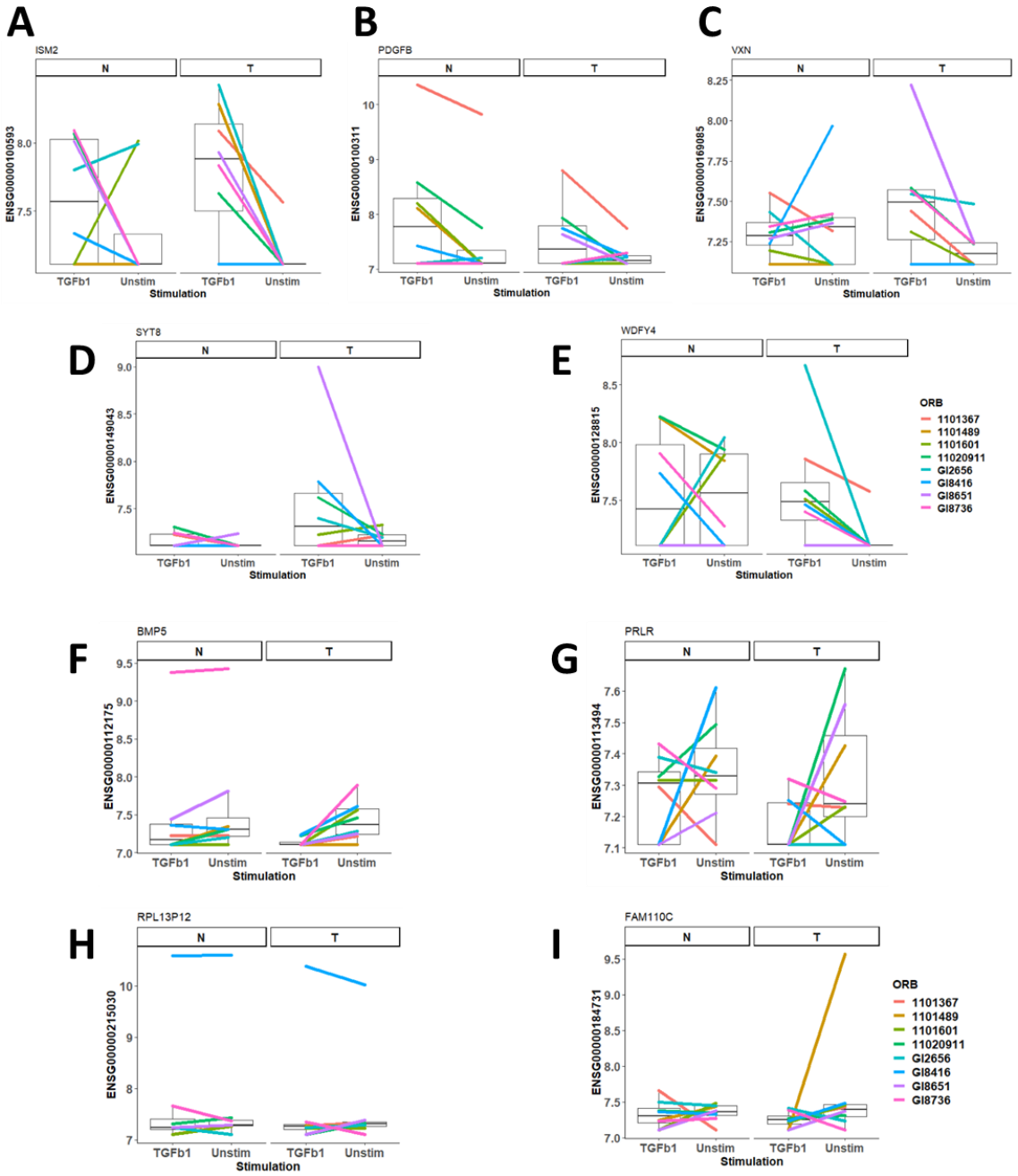


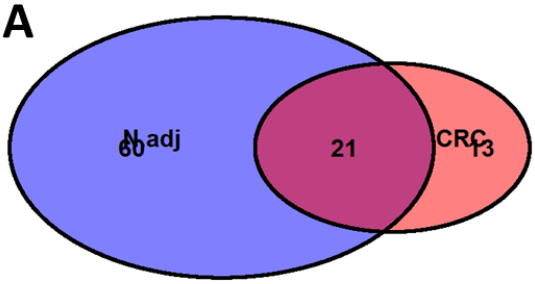
Fig 4.4.2.2 Boxplots of genes significantly altered in CRC, with a log Fold Change > +/- 4, but not in Normal adjacent fibroblasts

(A-E) Genes that are upregulated in CRC. **(A)** ISM2. **(B)** PDGFRB. **(C)** VXN. **(D)** SYT8. **(E)** WDFY4. **(F-I)** Genes that are downregulated in CRC. **(F)** BMP5. **(G)** PRLR **(H)** RPL13P12 **(I)** FAM110C.

Pathway analysis was performed to compare pathways where at least 20% of the genes in that pathway are upregulated by TGFβ1 in CRC and in N.adjacent lines. With TGFβ1 this identified 21 pathways that were upregulated in both, 60 just significant in N.adjacent stromal cells and 13 just significant in tumour stromal cells **(Fig 4.4.2.3)**. Pathways upregulated by TGFβ in the

normal adjacent cells only include 'TGFβ production', 'TGFβ activation', 'Extracellular matrix assembly', 'Positive regulation of extracellular matrix organisation', 'Positive regulation of EMT', 'Positive regulation of fibroblast migration', 'Positive regulation of fibroblast proliferation' and 'Angiogenesis involved in wound healing'. Pathways upregulated in the tumour stromal cells only by TGFβ include 'Collagen fibril organisation' and 'Positive regulation of cartilage development'. Studying changes in the HALLMARK pathways, all three pathways that were altered in N.adjacent were also altered in CRC; EMT and Myc targets were upregulated in both and interferon alpha response downregulated **(Fig 4.4.1.6)**. The three other pathways that were significantly upregulated in CRC stromal cells were also upregulated in N.adj stromal cells, but it did not reach statistical significance.

A



B

Pathway	padj	DEGs	Total genes in pathway	Ratio DEG:Total
GOBP_HOMOTYPIC_CELL_CELL_ADHESION	2.09E-05	16	75	0.2133333
GOBP_COLLAGEN_FIBRIL_ORGANIZATION	1.29E-04	13	59	0.220339
GOBP_REGULATION_OF_HOMOTYPIC_CELL_CELL_ADHESION	9.14E-03	7	29	0.2413793
GOBP_REGULATION_OF_CALCIUM_ION_IMPORT	1.43E-02	6	23	0.2608696
GOBP_EPIBOLY	1.60E-02	7	33	0.2121212
GOBP_CELL_SURFACE_RECEPTOR_SIGNALING_PATHWAY_INVOLVED_IN_HEART_DEVELOPMENT	1.98E-02	6	25	0.24
GOBP_REGULATION_OF_INSULIN_LIKE_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY	2.07E-02	5	17	0.2941176
GOBP_L_SERINE_METABOLIC_PROCESS	2.07E-02	4	10	0.4
GOBP_ELASTIC_FIBER_ASSEMBLY	2.07E-02	4	10	0.4
GOBP_POSITIVE_REGULATION_OF_CARTILAGE_DEVELOPMENT	3.41E-02	6	29	0.2068966

C

Pathway	padj	DEGs	Total genes in pathway	Ratio DEG:Total
GOBP_CARDIAC_CHAMBER_MORPHOGENESIS	3.90E-05	21	101	0.2079208
GOBP_CARDIAC_VENTRICLE_DEVELOPMENT	5.90E-05	21	104	0.2019231
GOBP_ARTERY_DEVELOPMENT	7.65E-05	19	89	0.2134831
GOBP_MUSCLE_ORGAN_MORPHOGENESIS	8.38E-05	15	58	0.2586207
GOBP_PATHWAY_RESTRICTED_SMAD_PROTEIN_PHOSPHORYLATION	1.37E-04	14	53	0.2641509
GOBP_MESENCHYME_MORPHOGENESIS	7.71E-04	13	54	0.2407407
GOBP_TRANSFORMING_GROWTH_FACTOR_BETA_PRODUCTION	1.21E-03	10	34	0.2941176
GOBP_VENTRICULAR_SEPTUM_DEVELOPMENT	1.39E-03	14	66	0.2121212
GOBP_EXTRACELLULAR_MATRIX_ASSEMBLY	2.04E-03	10	37	0.2702703
GOBP_STRIATED_MUSCLE_CELL_DEVELOPMENT	2.11E-03	12	53	0.2264151

Fig 4.4.2.3 Numbers of pathways upregulated in CRC and N.adj stromal cells with cytokine stimulation.

(A) Pathways upregulated with TGFβ1. (B) Pathways upregulated with TGFβ1 in CRC, but not N.adjacent primary stromal cells. (C) Pathways upregulated with TGFβ1 in N.adj, but not CRC, primary stromal cells.

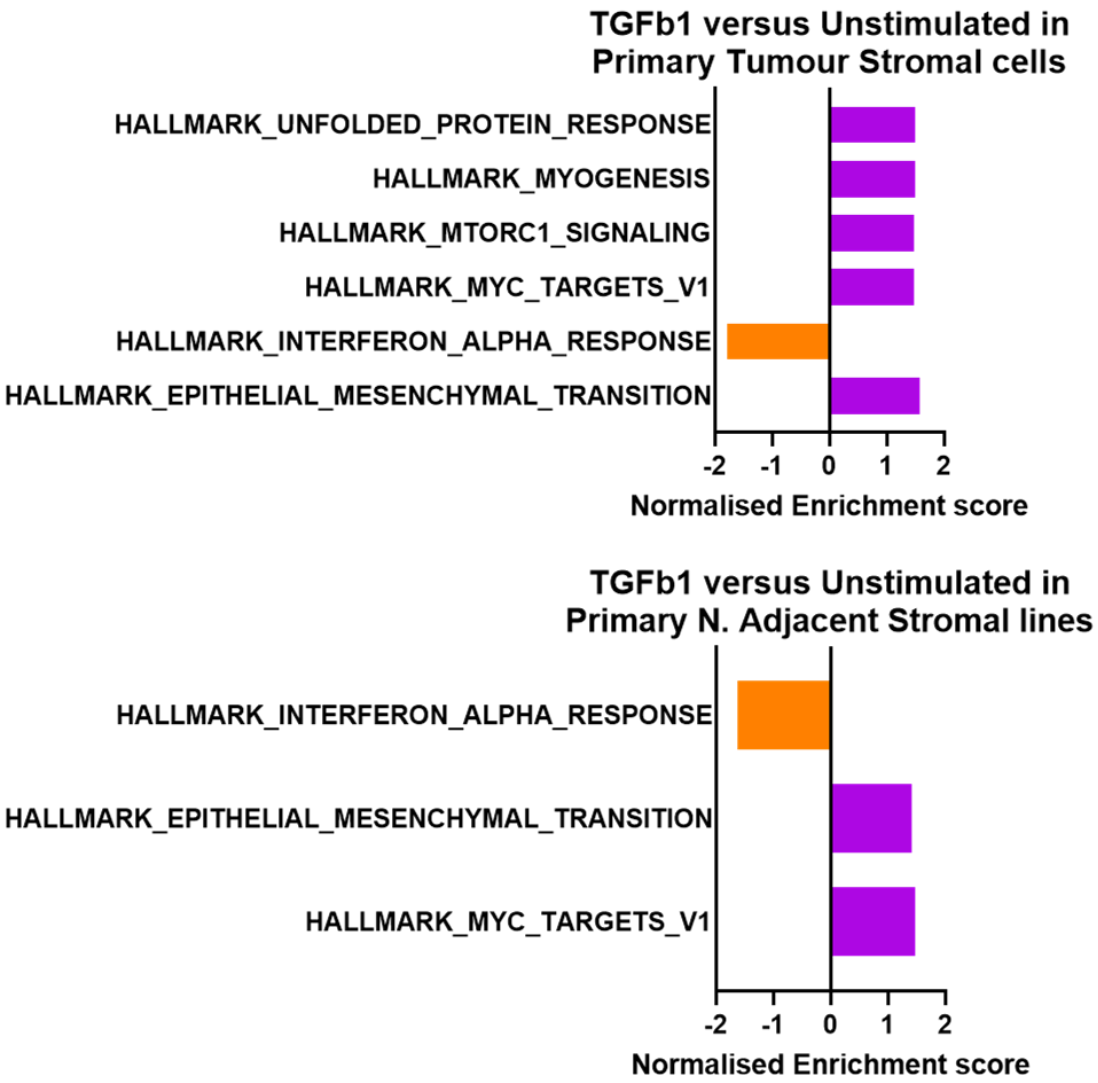


Fig 4.4.2.4 HALLMARK pathways altered by TGFβ1 in CRC and N.adj stromal cells
Pathways significantly up or down regulated with TGFβ1 in (A) tumour stromal cells and (B) normal adjacent stromal cells.

To compare, the response of *COL1A1* and *CXCL1* was studied for each of the TGFβ isoforms at 5ng/ml for 24 hours, as these had previously shown a differential response in tumour and normal adjacent to TGFβ2, but not to TGFβ1 or TGFβ3 in the PCR data in Fig 4.3.1.3 **(Fig 4.4.2.5)**. *COL1A1* and *CXCL1* are both expressed more highly than the genes previously identified, between 15 and 21 for *COL1A1* and between 7 and 15 for *CXCL1*, but while *COL1A1* was significantly increased with TGFβ and *CXCL1* decreased, there did not appear to be any differences between tumour and normal adjacent response or between the three isoforms.

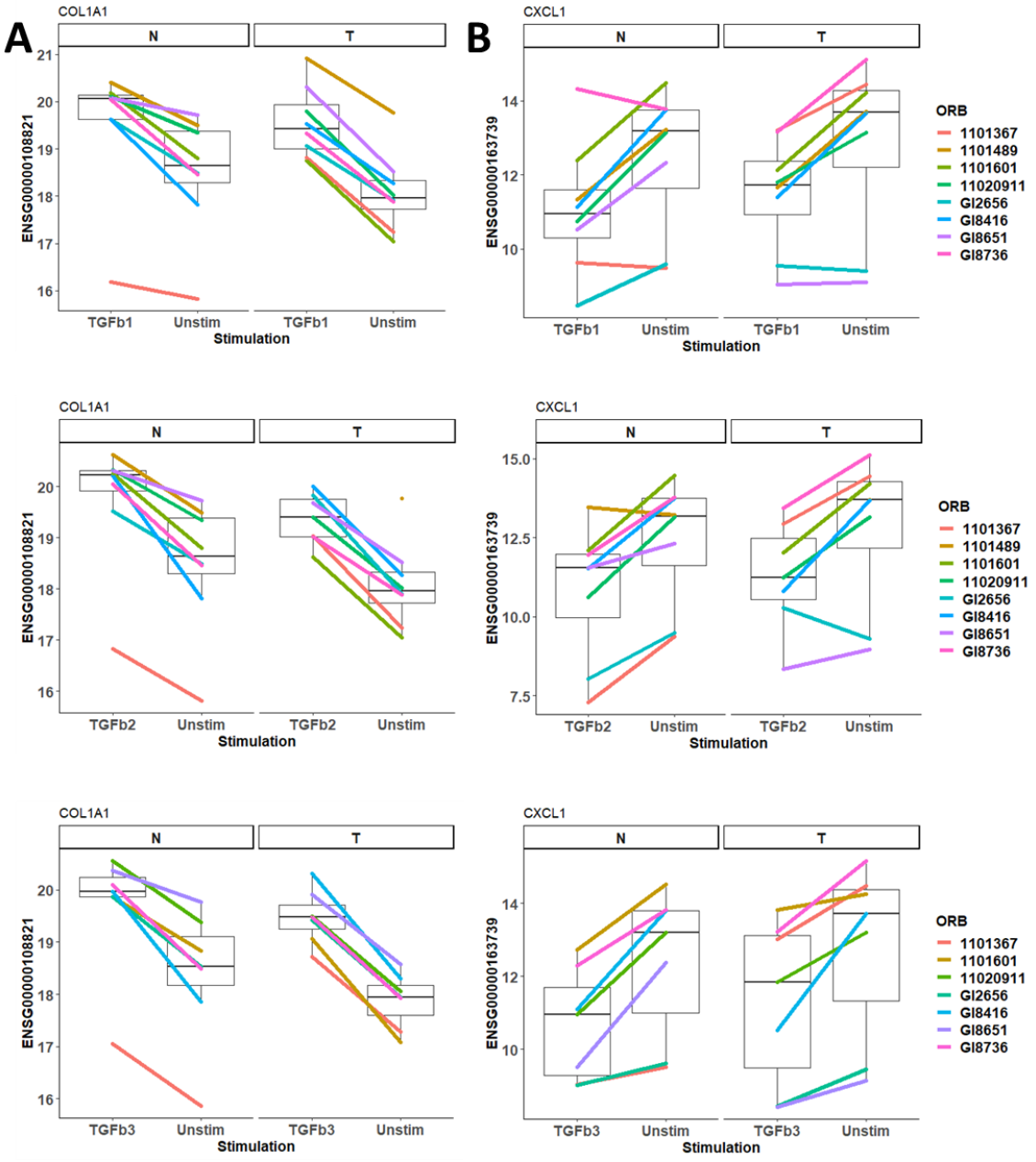


Fig 4.4.2.5 Boxplot s of genes shown by PCR to respond differentially to TGFβ2 stimulation in tumour and normal adjacent fibroblasts
(A) COL1A1. (B) CXCL1

This method of comparing tumour and normal adjacent did not account for the eight samples from each site being paired. Repeat analysis was performed utilising this information and it identified 4 genes that were significantly differentially altered by TGFβ1 in tumour and N.adj across the 8 pairs. The 4 genes identified were *IL36B, RAMP3, RGS7BP* and *IGLC3*. On drawing boxplots of expression of these genes differences were not seen across all 8 pairs in any of them, rather the results seemed very skewed by one or two individual lines **(Fig 4.4.2.6 A-D)**. Also of note, all 4 genes were lowly expressed and not normally considered to be expressed in stromal cells. Using the same method to compare TGFβ2 and unstimulated led to the identification of 3 genes: *BIMPA5, BIRC7* and *CAPN6* **(Fig 4.4.2.6 E-G)**. Comparing TGFβ3 with unstimulated identified 7 genes with differentially altered gene expression between the two sites: *CENPO, IGHA1, MGAM, SPINK13, SUPV3L1, YJEFN3* and *UGT1A8* **(Fig 4.4.2.7)**. As with the other isoforms, the majority of the genes were lowly expressed and changes were not consistent across the 8 pairs. However, *SUPV3L1* was more highly expressed, above 10 in all lines, and the upregulation in tumour and downregulation in normal adjacent did seem to be fairly consistent across most of the primary lines. *SUPV3L1* is Sup3 like RNA helicase found in the mitochondria of most tissues and cancers[181] and is known to be expressed in fibroblasts[182], but little else is known about it.

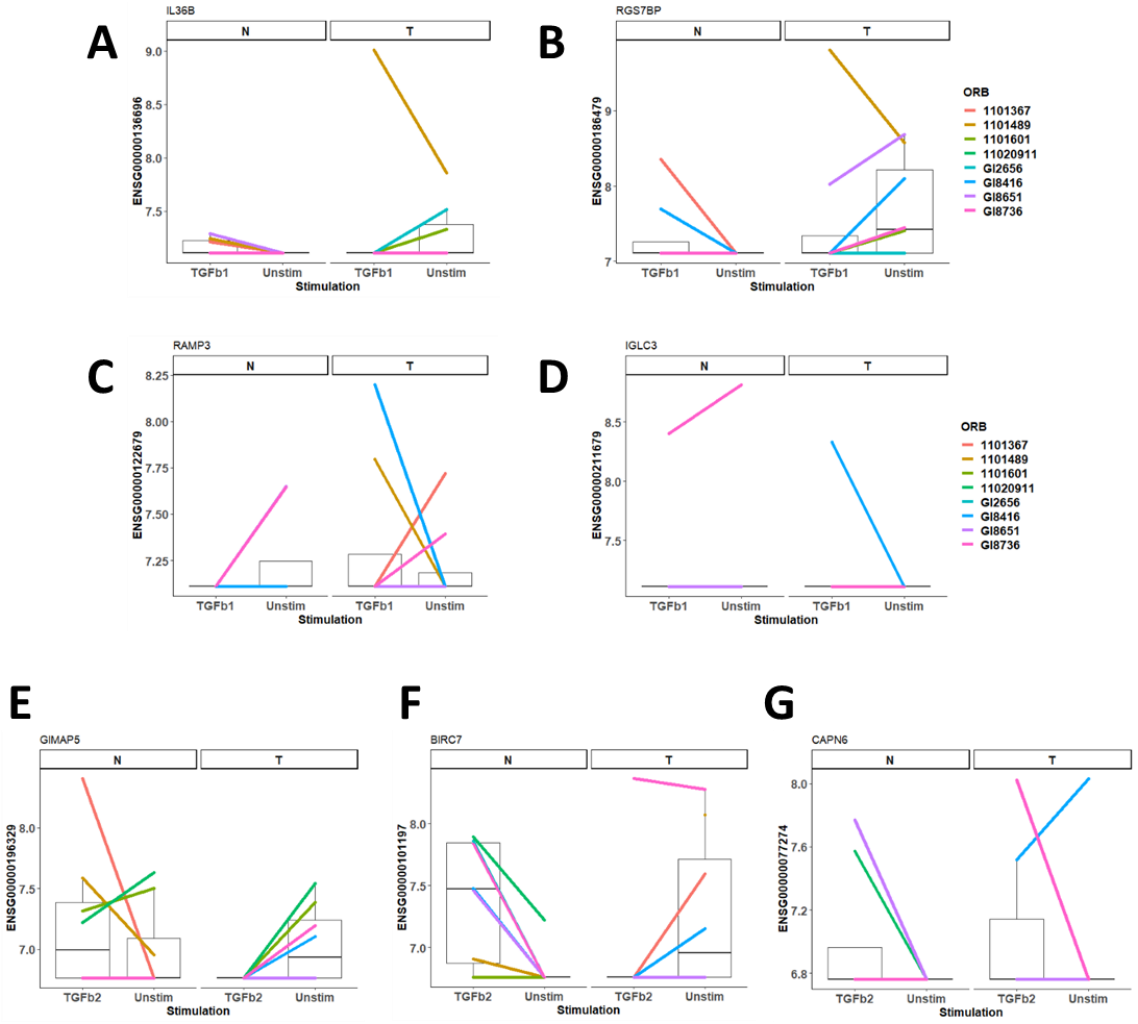


Fig 4.4.2.6 Boxplot s of genes significantly differentially altered by TGFβ1 and 2 in tumour and Normal adjacent stromal cells when the paired nature of the samples is accounted for.
(A-D) Genes differentially altered by TGFβ1: **(A)** IL36B, **(B)** RGS7BP, **(C)** RAMP3 and **(D)** IGLC3. **(E-G)** Genes differentially altered by TGFβ2: **(E)** GIMAP5, **(F)** BIRC7 and **(G)** CAPN6.

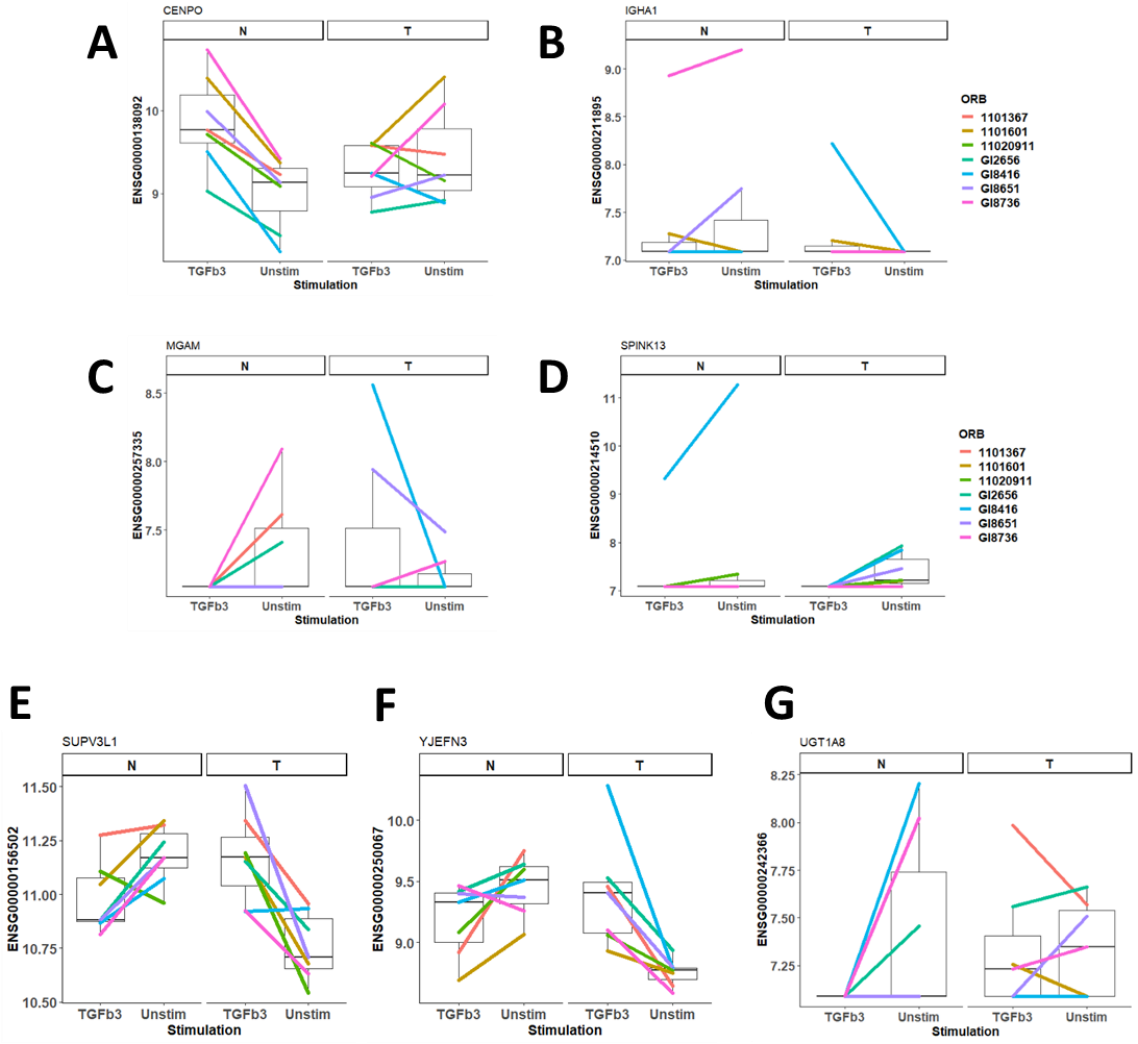


Fig 4.4.2.7 Boxplot s of genes significantly differentially altered by TGFβ3 in tumour and Normal adjacent stromal cells when the paired nature of the samples is accounted for.
(A) CENPO, (B) IGHA1, (C) MGAM, (D) SPINK13, (E) SUPV3L1, **(F)** YJEFN3 and **(G)** UGT1A8.

4.4.3 Comparing Tumour and Normal Adjacent Stromal cells response to IL-1β

Differential gene expression was calculated in 24-hours, 0.1ng/ml IL-1β stimulated versus unstimulated primary stromal lines from each of the tumour and Normal adjacent **(Fig 4.4.3.1 A+B)**. This enabled log FC of genes with IL-1β from each site to be compared and again showed strong positive correlation, suggesting tumour and normal adjacent fibroblasts are responding similarly to stimulation with this cytokine **(Fig 4.4.3.1 C)** and boxplots of expression of genes with statistically significant logFC of > +/- 3 in tumours, but no significant change in normal adjacent stromal cells, are shown **(Fig 4.4.3.2)**. Again, all 5 genes identified were lowly expressed.

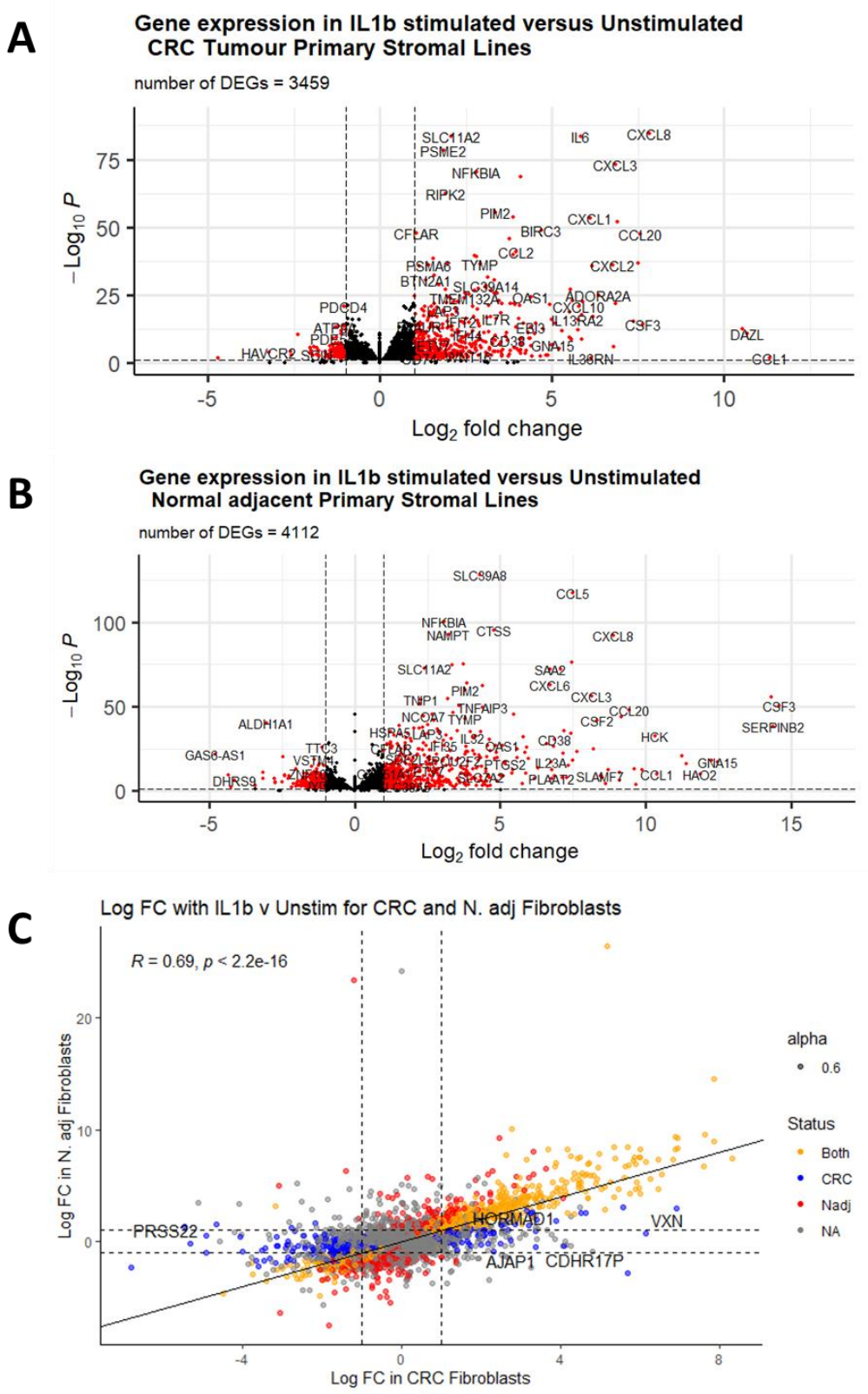


Fig 4.4.3.1 Differential gene expression with IL-1β, comparing Tumour and Normal adjacent primary stromal line

(A) Volcano plot showing DEGs between IL-1β and unstimulated primary stromal lines from tumours. **(B)** Volcano plot showing DEGs between IL-1β and unstimulated primary stromal lines from normal adjacent tissue. **(C)** Plot of the log fold change of every gene with IL-1β compared to unstimulated in tumour versus normal adjacent stromal cells.

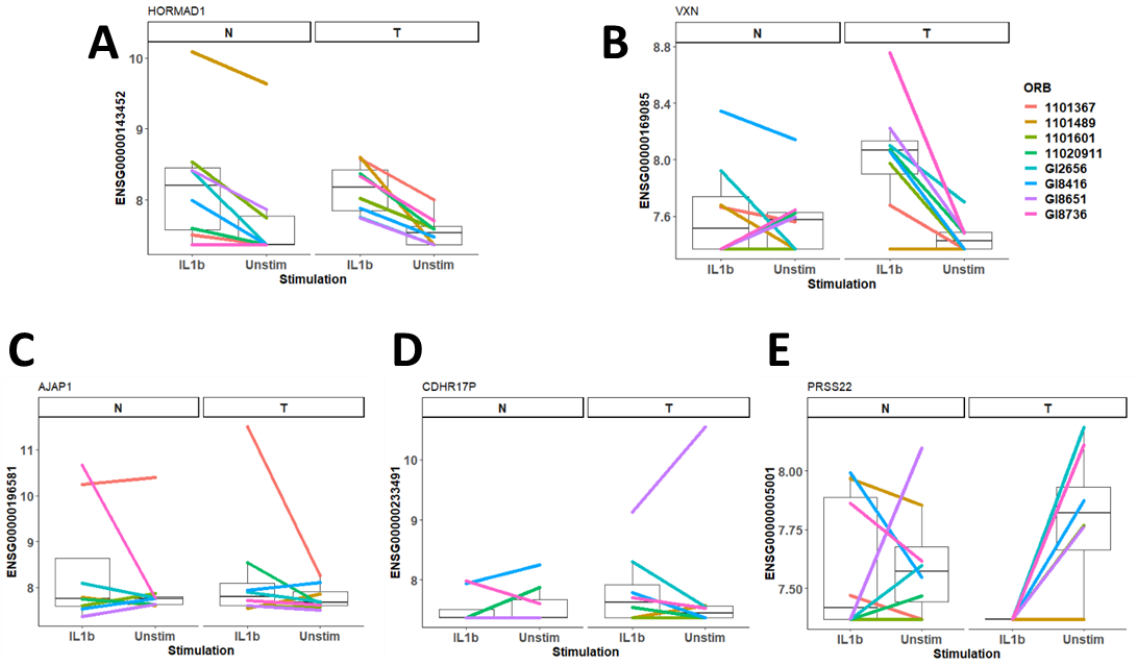


Fig 4.4.3.2 Boxplot s of genes significantly altered in CRC, with a log Fold Change > +/- 4, but not in Normal adjacent fibroblasts

(A-D) Genes that are upregulated in CRC. (A) HORMAD1. (B) VXN. (C) AJAP1 (D) CDHR17P. (E) PRSS22 is downregulated in CRC.

Pathway analysis for GO biological processes was performed to compare pathways where at least 20% of the genes in that pathway are upregulated by IL-1β in CRC and in N.adjacent lines. 253 pathways were upregulated by IL-1β in both groups of cells, 61 pathways were upregulated in just the N.adj and 52 in the tumour stromal cells only **(Fig 4.4.3.3 A)**. In both groups of primary stromal cells the most upregulated pathways were to do with cytokine signalling and leucocyte migration. Examining the HALLMARK pathways altered by IL-1β in primary stromal cells, 7 pathways were significantly upregulated in the tumour stromal cells and 11 in the normal adjacent stromal cells. All 7 HALLMARK pathways upregulated in tumour stromal cells were in the 11 upregulated in Normal adjacent stromal cells **(Fig 4.4.3.3 B+C)**. No HALLMARK pathways were significantly downregulated.

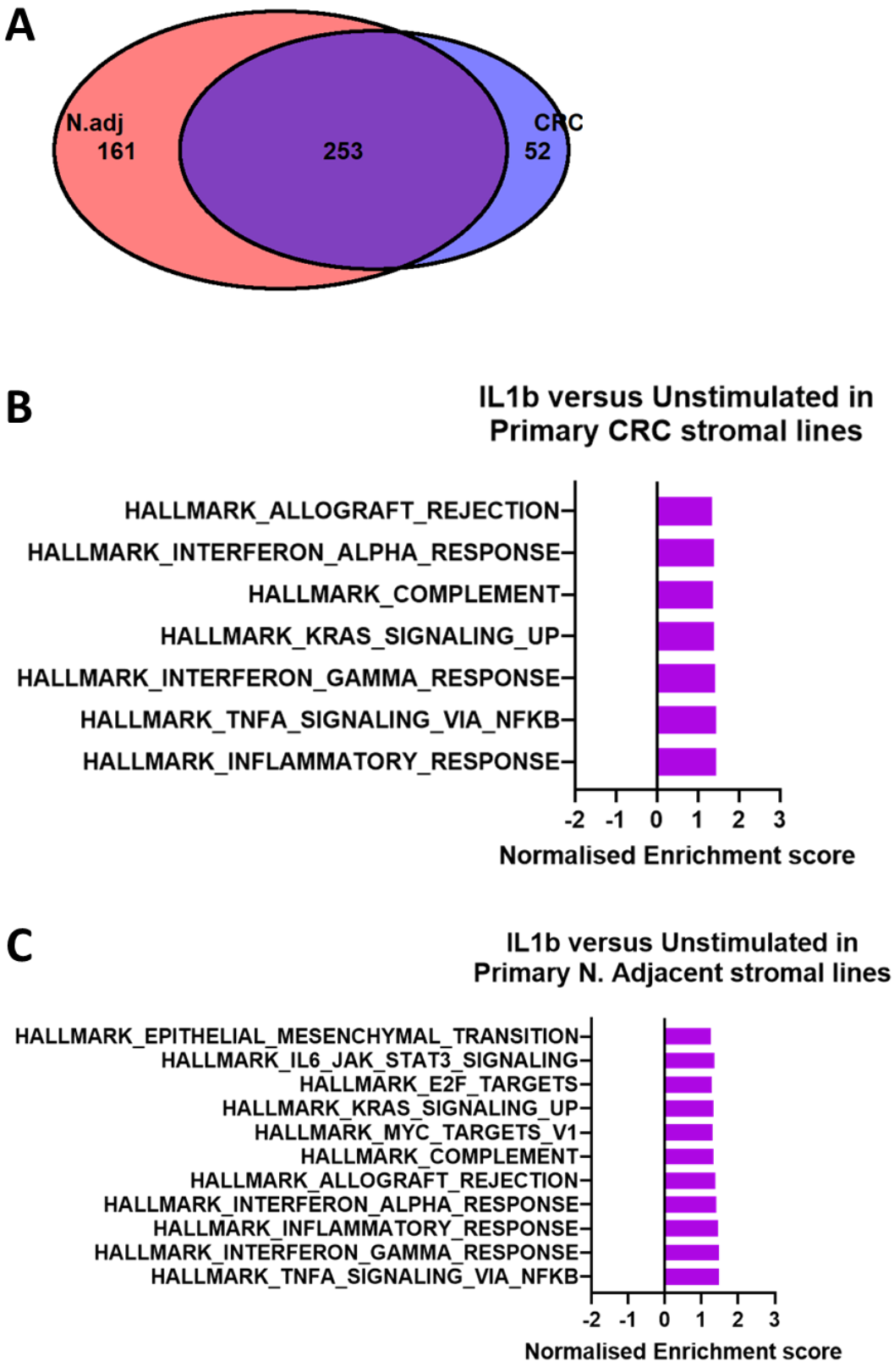


Fig 4.4.3.3 Numbers of pathways upregulated in CRC and N.adj stromal cells with cytokine stimulation.

(A) GO biological processes pathways upregulated with IL-1β. (B) HALLMARK pathways upregulated with IL-1β in primary tumour stromal cells. (C) HALLMARK pathways upregulated with IL-1β in primary normal adjacent stromal cells.

Comparing IL-1β response in tumour and normal adjacent accounting for the fact the lines were paired led to the identification of two significantly differentially altered genes; *CD70* and *SERPINB2* **(Fig 4.4.3.4)**. With both genes a slightly greater response was seen in N.adj than tumour, though increases were seen in cells from both locations. Both were generally very lowly expressed, <10, and CD70 is most commonly identified on immune cells. However, a recent paper found that CD70 was expressed on fibroblasts in 15% of CRCs, and in the 8% of CRCs where CD70 and POSTN were both present on fibroblasts, patients were more likely to have metastatic disease and had lower 5-year survival[183].

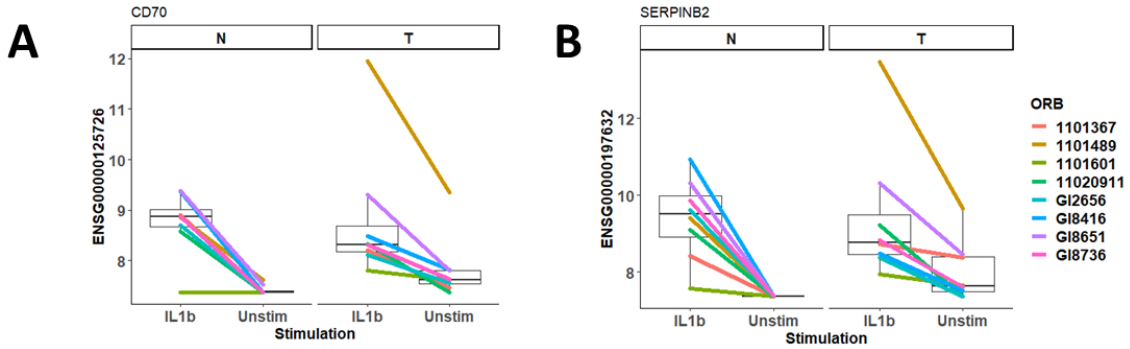


Fig 4.4.3.4 Boxplot s of genes significantly differentially altered by IL-1β in tumour and Normal adjacent stromal cells when the paired nature of the samples is taken into account
(A) CD70, (B) SERPINB2.

4.5 The cytokine response in the Primary Stromal lines by Cluster and comparing Tumour and Normal adjacent

4.5.1 The cytokine response in primary stromal cells by Cluster

4.5.1.1 Clustering to Primary Stromal cells under all stimulation conditions

The lack of a clear hyper-response to cytokines in tumour versus normal adjacent may indicate that tumour fibroblasts do not respond differently to stimulation than normal fibroblasts. Alternatively, there may be a difference that is being masked due to the huge heterogeneity already seen between the samples. To overcome this, it was decided to repeat the analysis using only the pairs from samples in Cluster 1 that were most likely to be fibroblasts. **Fig 4.5.1.1 A** shows a PCA of the 12 Cluster 1 primary stroma lines, with clustering occurring again by stimulation rather than cell type. Performing hierarchical clustering **(Fig 4.5.1.1 B)** of the primary stromal lines in all stimulation conditions showed that the primary stromal lines in Cluster 2, 3_Tumour and 7_tumour, and 4_N.adj in Cluster 3, clustered by primary line regardless of stimulation conditions, whereas the other lines clustered by stimulation. A similar pattern was seen making a heatmap using the marker genes of the clusters generated from the Pelka *et al* scRNAseq dataset, with most samples clustering together by stimulation conditions, but the three outlier samples and pericytes remaining separate, highlighted with green boxes **(Fig 4.5.1.2).**

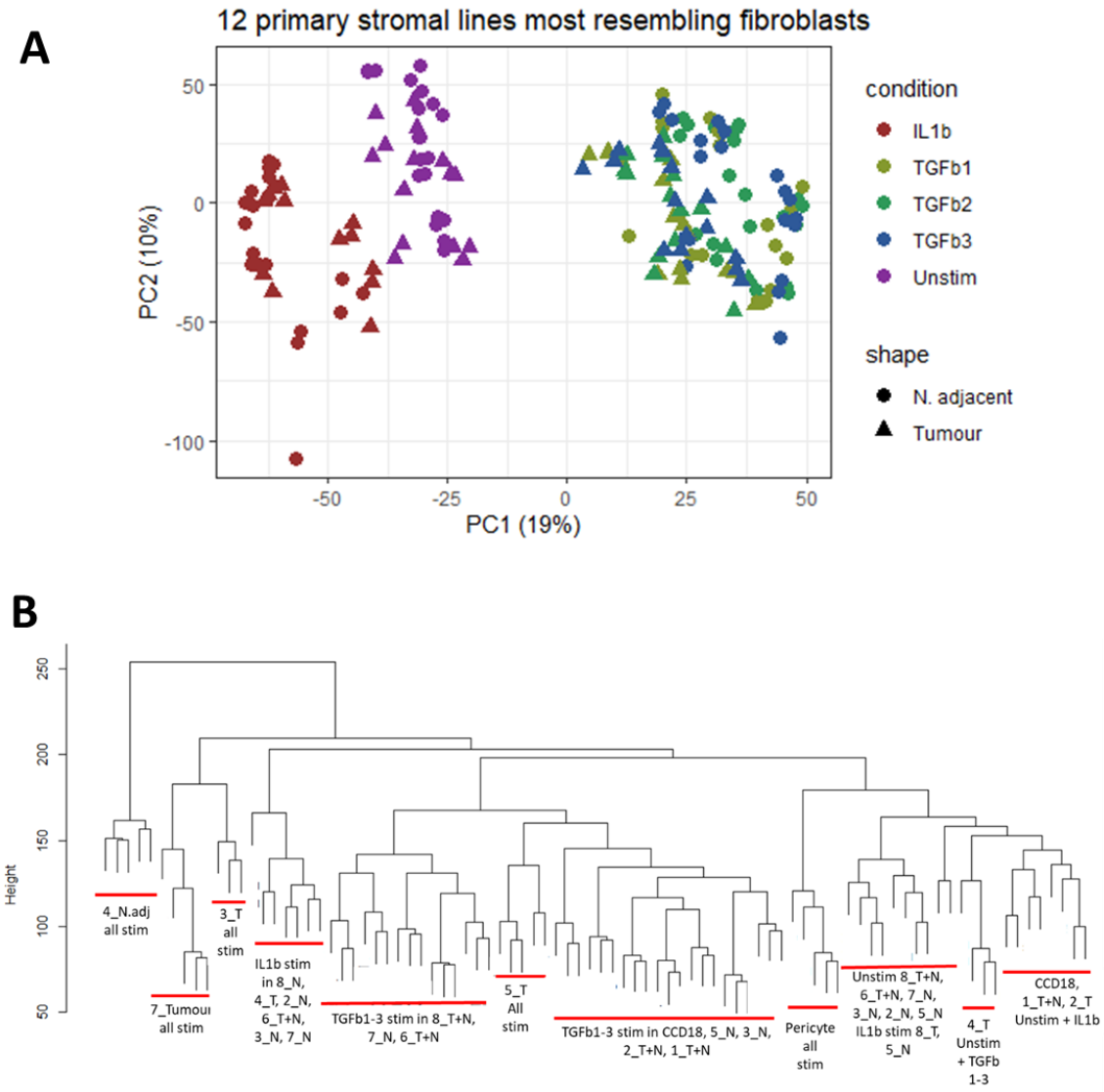


Fig 4.5.1.1 Primary stromal lines in cluster 1 cluster by stimulation conditions, other lines cluster separately.

(A) PCA of the 12 primary cell lines most resembling fibroblasts. (B) Dendogram showing 3 outlier primary cell lines, 3_Tumour, 7_Tumour and 4_N.adj, and remaining cell lines clustering largely by stimulation condition.

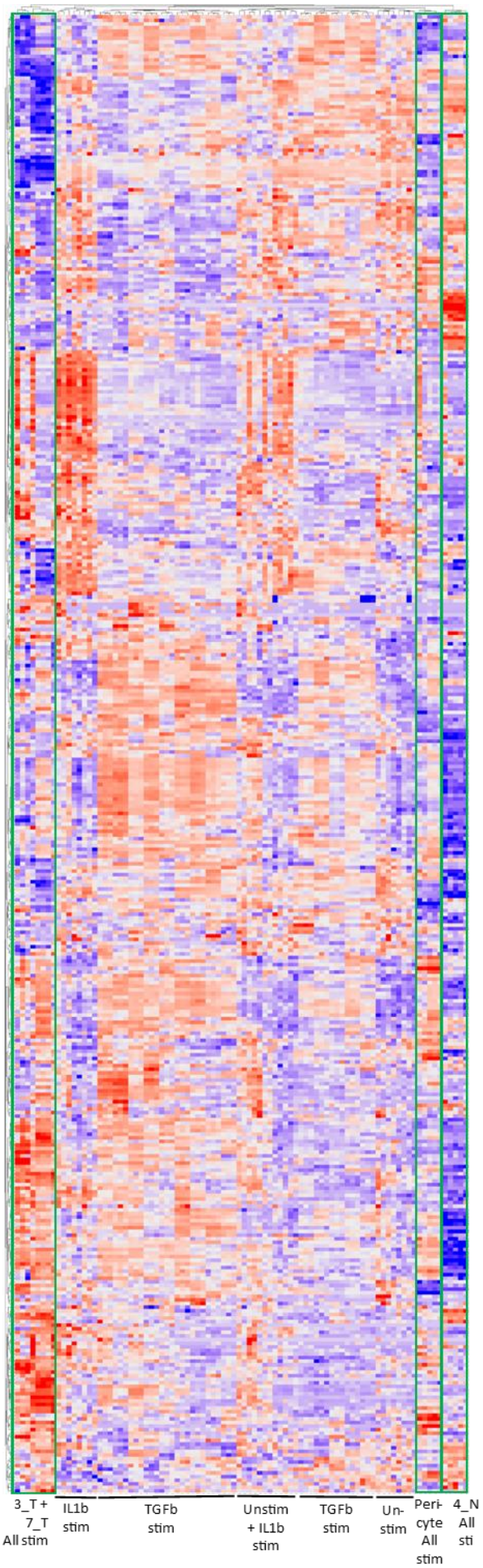


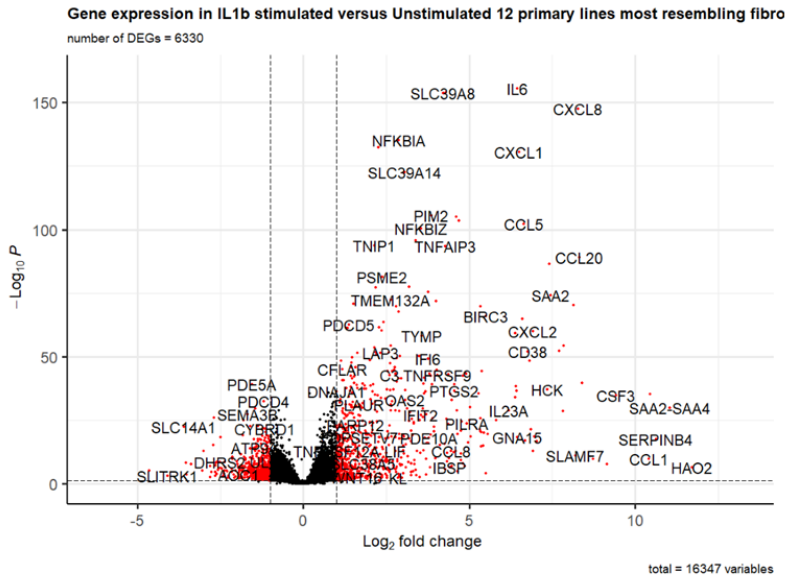
Fig 4.5.1.2 Outlier Primary lines clustering separately, remaining lines clustering by stimulation.
Heatmap of marker genes of Pelka et als stromal clusters, showing 3 primary lines (3_Tumour, 4_nadj and 7_Tumour) and pericytes cluster separately, in green boxes, whereas other lines cluster together by stimulation conditions.

4.5.1.2 IL-1β stimulation of Primary Stromal cells

With these 12 primary stromal lines that most resembled fibroblasts we looked at the effects of TGFβ1 and IL-1β stimulation. With 24-hours, 0.1ng/ml IL-1β there were 759 significantly upregulated genes with Log_2FC >1. *IL6* was the most significantly upregulated gene, with a log_2FC of 6.44. Notably, *IL-1B* itself was very strongly upregulated by IL-1β, with a log_2FC of 5.35 and p=2.19E-26, whereas *IL-1B* had not been significantly upregulated by IL-1β in the CCD18-Co fibroblast cell line. IL-1β in these 12 primary stromal lines led to significant upregulation of 357 pathways where at least 20% of the genes in that pathway were upregulated **(Fig 4.5.1.3)**. Pathways upregulated were involved in recognition of pathogens, recruitment of leucocytes and the inflammatory response.

The effects of IL-1β were also studied in Cluster 2 and Cluster 3. In cluster 2, consisting of samples 3_Tumour and 7_Tumour 583 genes were significantly upregulated with IL-1β and in Cluster 3 (4_Tumour and 4_N.adj) 235 genes were upregulated **(Fig 4.5.1.4 A-C).** Studying the overlap between GO biological processes pathways upregulated by IL-1β in the 3 separate clusters revealed pathways unique to each cluster. The 5 pathways only upregulated in Cluster 3 did not seem relevant to fibroblast cells. The pathways uniquely upregulated in Cluster 2 included those involved in T-cell, macrophage and epithelial cell differentiation, and TNF and other cytokine production **(Fig 4.5.1.4. D)**.

A



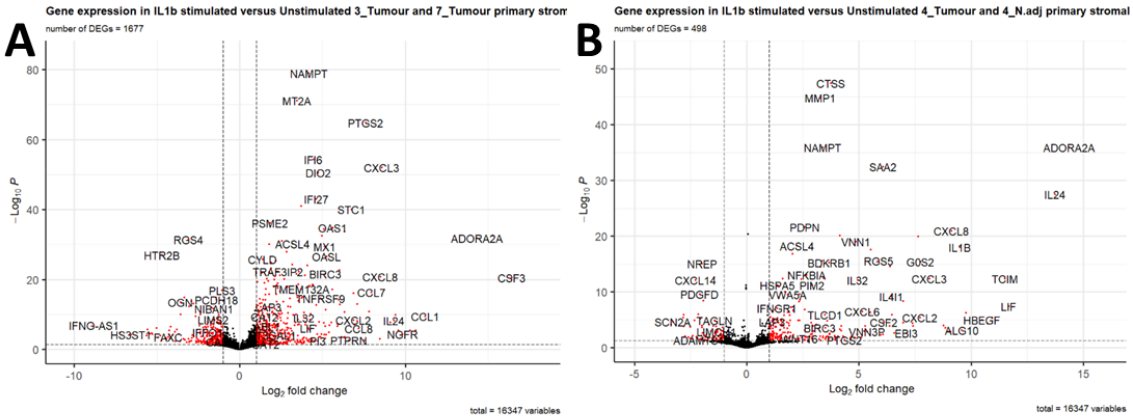
B **Genes significantly upregulated with IL1b in Primary cell lines**

Gene	Log2FC	padj
IL6	6.44	3E-156
ZC3H12A	4.30	1.4E-154
SLC39A8	4.22	2E-154
CXCL8	8.28	2.7E-148
NFKBIA	2.88	8.2E-136
SLC11A2	2.27	4.2E-133
CXCL1	6.51	2.8E-131
SLC39A14	3.06	2.3E-123
PIM2	3.87	2.8E-106
CTSS	4.61	8.5E-106
CFB	4.69	3.5E-104
CCL5	6.65	5.7E-103
NFKBIZ	3.55	3E-101
TLCD1	3.39	1.76E-96
TNIP1	2.15	1.49E-94
TNFAIP3	4.28	2.76E-94
CCL20	8.32	6.55E-90
CXCL3	7.41	2.59E-87
PSME2	2.39	5.03E-82
PTGES	3.19	3.5E-78

C

Pathway	padj	DEGs	Total genes in pathway	Ratio DEG:total
GOBP_CYTOKINE_MEDIATED_SIGNALING_PATHWAY	1.78E-42	93	341	0.2727273
GOBP_RESPONSE_TO_BACTERIUM	7.87E-29	87	432	0.2013889
GOBP_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN	3.94E-26	64	258	0.248062
GOBP_RESPONSE_TO_VIRUS	3.94E-26	71	319	0.2225705
GOBP_LEUKOCYTE_MIGRATION	5.68E-26	66	278	0.2374101
GOBP_LEUKOCYTE_CHEMOTAXIS	1.08E-24	50	164	0.304878
GOBP_NEUTROPHIL_CHEMOTAXIS	8.98E-23	33	71	0.4647887
GOBP_CELL_CHEMOTAXIS	8.98E-23	54	212	0.254717
GOBP_ADAPTIVE_IMMUNE_RESPONSE	2.93E-22	63	296	0.2128378
GOBP_GRANULOCYTE_CHEMOTAXIS	2.93E-22	36	90	0.4
GOBP_DEFENSE_RESPONSE_TO_SYMBIONT	4.84E-22	56	237	0.2362869
GOBP_POSITIVE_REGULATION_OF_DEFENSE_RESPONSE	7.49E-22	65	321	0.2024922
GOBP_GRANULOCYTE_MIGRATION	1.19E-21	38	106	0.3584906
GOBP_CELLULAR_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN	2.59E-21	46	164	0.2804878
GOBP_NEUTROPHIL_MIGRATION	6.01E-21	34	86	0.3953488
GOBP_RESPONSE_TO_CHEMOKINE	2.99E-20	29	62	0.4677419
GOBP_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS	6.34E-20	51	218	0.233945
GOBP_CELLULAR_RESPONSE_TO_BIOTIC_STIMULUS	1.38E-19	47	188	0.25
GOBP_MYELOID_LEUKOCYTE_MIGRATION	2.23E-19	44	166	0.2650602
GOBP_RESPONSE_TO_TYPE_II_INTERFERON	2.74E-19	35	102	0.3431373

Fig 4.5.1.3 IL-1β stimulation in the 12 Primary stromal cell lines resembling fibroblasts
(A) Volcano plot showing DEGs between IL-1β stimulated and unstimulated primary stromal cells. (B) The 20 most significantly upregulated genes with IL-1β, of those with log2FC >1. (C) The 20 most significantly upregulated pathways with IL-1β, of those pathways where at least 20% of the genes in the pathway are upregulated by IL-1β.



Gene	Log2FC	padj
NAMPT	4.16	9.78E-80
MT2A	3.46	7.19E-72
PTGS2	7.63	1.59E-65
IFI6	4.46	5.27E-55
CXCL3	8.59	1.04E-52
DIO2	4.74	3.09E-51
IFI27	4.62	9.41E-44
ISG15	3.71	1.13E-41
STC1	6.78	1.06E-40
PSME2	1.84	8.17E-37
OAS1	5.64	1.71E-35
ZC3H12A	5.07	4.21E-35
IL6	6.33	2.15E-34
CCL2	4.96	3.33E-33
CXCL5	15.27	3.7E-33
ADORA2A	14.33	1.65E-32
IFIT3	2.56	8.67E-32
ACSL4	3.21	6.62E-31
IRF9	1.78	9.4E-31
MX1	5.14	4.39E-30

Gene	Log2FC	padj
CTSS	3.75	3.92E-48
MMP1	3.27	1.64E-45
NAMPT	3.39	1.26E-36
ADORA2A	14.38	1.26E-36
SAA2	6.09	3.71E-33
IL24	13.75	3.17E-28
PDPN	2.59	2.27E-22
CXCL8	9.14	1.2E-21
TNFAIP3	4.16	8.46E-21
MARCHF3	7.64	1.24E-20
VNN1	4.86	1.07E-19
ACSL4	2.25	5.45E-19
IL1B	9.50	9.61E-19
CCL5	5.53	2.69E-18
AKR1B1	2.05	1.55E-17
SLC39A8	3.37	2.69E-16
RGS5	5.89	2.91E-16
G0S2	7.74	4.34E-16
BDKRB1	3.67	6.46E-16
ELF3	6.39	2.79E-15

Fig 4.5.1.4 IL-1β stimulation in Clusters 2 and 3

(A+B) Volcano plots showing DEGs between IL-1β stimulated and unstimulated primary stromal cells in Cluster 2 **(A)** and Cluster 3 **(B). (C+D)** The 20 most significantly upregulated genes with IL-1β, of those with log2FC >1, in Cluster 2 **(C)** and Cluster 3 **(D). (E)** The overlap in pathways upregulated by IL-1β in Clusters 1-3.

Twenty seven of the fifty HALLMARK pathways were significantly up or down regulated by IL-1β in at least one of the clusters of primary stromal cells. They are shown for each of the three clusters, where a normalised enrichement score is significantly altered it is shown in red, and otherwise it is shown in blue **(Fig 4.5.1.5)**. While most pathways show a similar magnitude of change across the three clusters, there are 2 notable exceptions. TGFβ signalling was up in clusters 1 and 2, but reduced in cluster 3. Myogenesis was reduced in clusters 1 and 2 and increased in cluster 3. These were both pathways we would have expected to be unchanged or reduced by IL-1β.

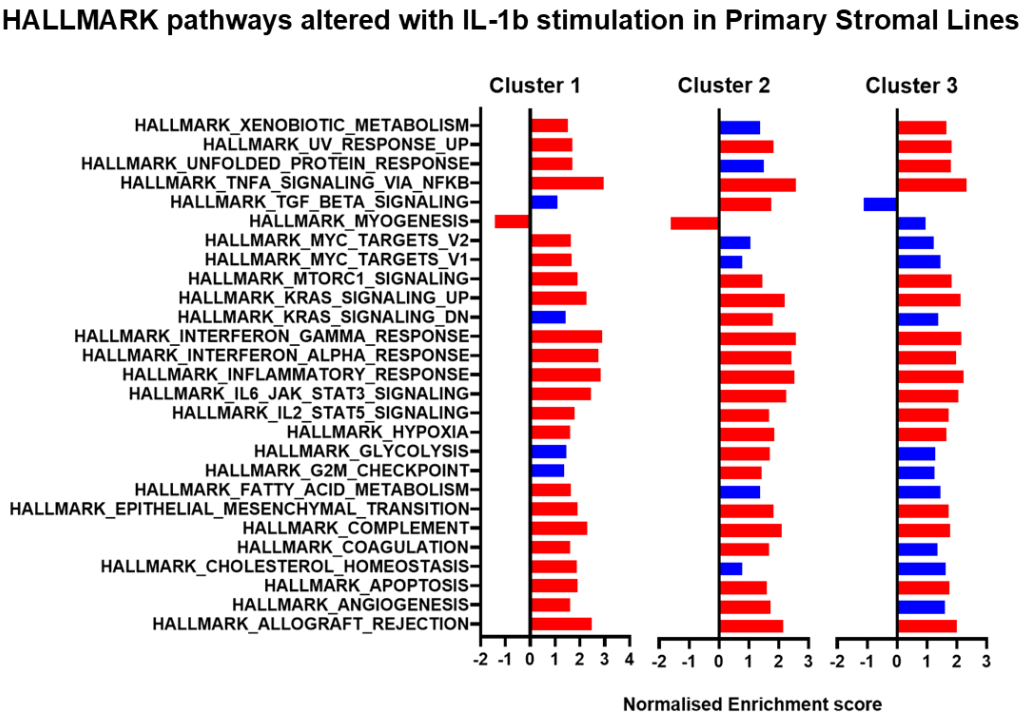


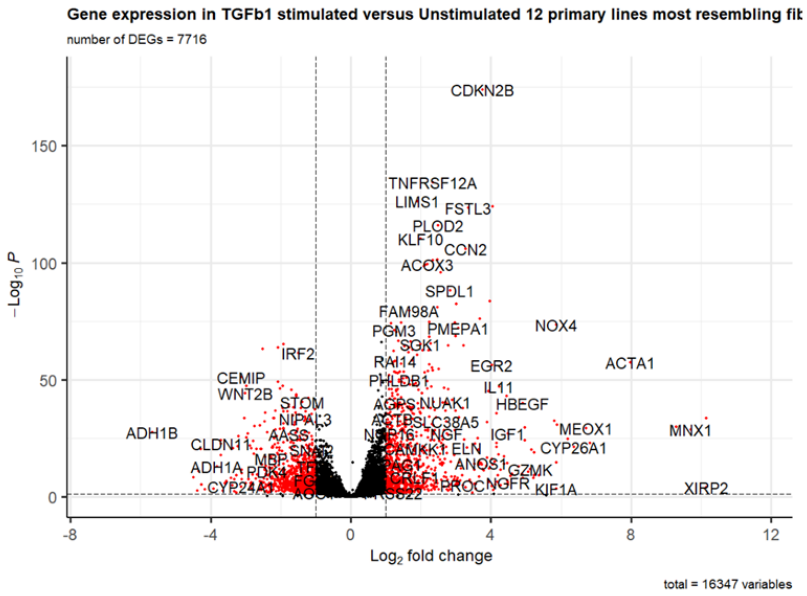
Fig 4.5.1.5 IL-1β stimulation in the three clusters of primary stromal cells
HALLMARK pathways significantly up or down-regulated with IL-1β in at least one of the three clusters of primary stromal cells. Bars showing normalised enrichment scores are red where statistically significant and blue where statistical significance is not reached.

4.5.1.3 TGFβ1 stimulation of Primary Stromal cells

With 24-hours, 5ng/ml TGFβ1 913 genes were significantly upregulated, including TGFβ1 itself with a log_2FC of 2.22, p=2.68E-63, whereas TGFβ1 had not significantly increased its own expression in the fibroblast cell line. 89 pathways were significantly upregulated with at least 20% of the genes in that pathway up **(Fig 4.5.1.6)**. Upregulated pathways with TGFβ1 are mainly related to tissue development and morphogenesis of muscles and vessels, and notably production of TGFβ pathway components, EMT and collagen organisations.

We also looked at the response to TGFβ of the Cluster 2 and Cluster 3 primary stromal lines, and **Fig 4.5.1.7** shows these: 1926 genes were upregulated in Cluster 2 and 260 in Cluster 3. This resulted in seventeen pathways being upregulated by TGFβ1 in Cluster 2 and two pathways in Cluster 3. The pathway uniquely upregulated in Cluster 3 by TGFβ1 was "Smad protein complex assembly" and the top pathways uniquely upregulated in Cluster 2 were to do with amino acid transport, IL-17 production and Th17 cell differentiation and response.

A



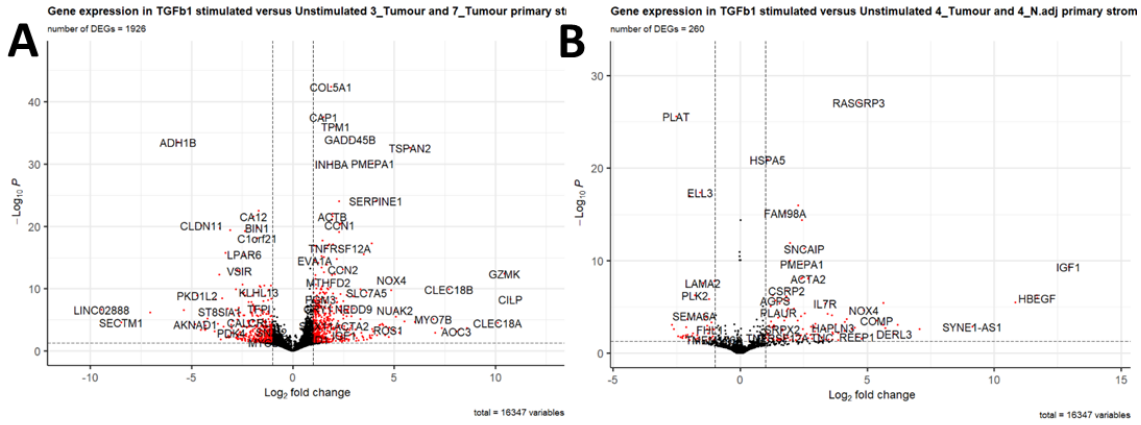
B **Genes significantly upregulated with TGFb1 in Primary cell lines**

Gene	Log2FC	padj
CDKN2B	3.759197	1.2E-174
TNFRS-F12A	2.348086	6.4E-135
LIMS1	1.891821	7.1E-127
SERPINE1	4.045164	1.4E-124
FSTL3	3.362662	3.2E-124
PLOD2	2.501596	9.6E-117
KLF10	1.99888	7.1E-111
CCN2	3.289462	1.1E-106
AHCYL1	2.473143	5E-102
PPME1	2.347426	9E-102
ACOX3	2.180048	4.2E-100
TPM1	2.138843	1.3E-99
DLC1	2.099599	4.2E-99
CTPS1	2.577415	1.76E-96
SPDL1	2.849194	5.55E-89
MYOSLID	3.971895	3.76E-84
SERPINE2	3.020114	5.48E-83
TUFT1	2.474744	1.12E-81
FAM98A	1.6682	3.51E-80

C **Pathways significantly upregulated with TGFb1 in Primary cell lines**

Pathway	padj	DEGs	Total genes in pathway	Ratio DEG:total
GOBP_MUSCLE_ORGAN_MORPHOGENESIS	9.29E-08	19	58	0.3275862
GOBP_CARDIAC_VENTRICLE_DEVELOPMENT	1.99E-07	25	104	0.2403846
GOBP_CARDIAC_CHAMBER_DEVELOPMENT	2.25E-07	29	139	0.2086331
GOBP_CARDIAC_CHAMBER_MORPHOGENESIS	3.44E-07	24	101	0.2376238
GOBP_CARDIAC_MUSCLE_TISSUE_MORPHOGENESIS	3.05E-06	15	45	0.3333333
GOBP_CARDIAC_VENTRICLE_MORPHOGENESIS	3.47E-05	15	54	0.2777778
GOBP_VENTRICULAR_CARDIAC_MUSCLE_TISSUE_DEVELOPMENT	4.78E-05	13	42	0.3095238
GOBP_ODONTOGENESIS	4.78E-05	20	96	0.2083333
GOBP_ARTERY_DEVELOPMENT	5.90E-05	19	89	0.2134831
GOBP_HOMOTYPIC_CELL_CELL_ADHESION	8.66E-05	17	75	0.2266667
GOBP_PATHWAY_RESTRICTED_SMAD_PROTEIN_PHOSPHORYLATION	1.08E-04	14	53	0.2641509
GOBP_MESENCHYME_MORPHOGENESIS	1.31E-04	14	54	0.2592593
GOBP_TRANSFORMING_GROWTH_FACTOR_BETA_PRODUCTION	1.62E-04	11	34	0.3235294
GOBP_VENTRICULAR_CARDIAC_MUSCLE_TISSUE_MORPHOGENESIS	2.15E-04	11	35	0.3142857
GOBP_REGULATION_OF_BIOMINERAL_TISSUE_DEVELOPMENT	2.15E-04	17	81	0.2098765
GOBP_VENTRICULAR_SEPTUM_DEVELOPMENT	2.80E-04	15	66	0.2272727
GOBP_COLLAGEN_FIBRIL_ORGANIZATION	3.41E-04	14	59	0.2372881
GOBP_REGULATION_OF_GLIOGENESIS	3.57E-04	16	76	0.2105263
GOBP_REGULATION_OF_GLIAL_CELL_DIFFERENTIATION	3.70E-04	13	52	0.25
GOBP_OUTFLOW_TRACT_MORPHOGENESIS	5.08E-04	15	70	0.2142857

Fig 4.5.1.6 TGFβ1 stimulation in the 12 Primary stromal cell lines most resembling fibroblasts
(A) Volcano plot showing DEGs between TGFβ1 stimulated and unstimulated in primary stromal cells. **(B)** The 20 most significantly upregulated genes with TGFβ1, of those with log2FC >1. **(C)** The 20 most significantly upregulated pathways with TGFβ1, of those pathways where at least 20% of the genes in the pathway are upregulated by TGFβ1.



C

Genes significantly upregulated with TGFb1 in Cluster 2: 3_Tumour and 7_Tumour

Gene	Log2FC	padj
COL5A1	1.89	4.62E-43
CAP1	1.53	3.12E-38
TPM1	2.12	9.69E-37
GADD45B	2.82	1.2E-34
TSPAN2	5.79	2.8E-33
PMEPA1	3.96	8.14E-31
INHBA	1.92	1.01E-30
SERPINE1	4.12	8.68E-25
SKIL	2.29	1.05E-24
SPHK1	1.97	9.46E-23
ACTB	1.97	2.75E-22
LOXL3	1.94	8.91E-22
PSPH	2.37	3.63E-21
CCN1	2.31	6.85E-21
KLF10	2.29	9.17E-20
SQLE	1.48	1.88E-18
FOXS1	3.90	5.34E-18
SCD	1.98	7.5E-18
DDAH1	1.80	7.52E-18
PEDS1	1.73	1.43E-17

D

Genes significantly upregulated with TGFb1 in Cluster 3: 4_N.adj and 4_Tumour

Gene	Log2FC	padj
RASGRP3	4.675945	8.23E-28
HSPA5	1.097586	1.3E-21
SLC46A3	2.284976	1.09E-16
FAM98A	1.78512	7.24E-16
PDPN	2.430441	4.34E-15
SGK1	1.979001	1.34E-12
SNCAIP	2.540039	5.31E-12
PLOD2	1.968778	1.02E-10
PMEPA1	2.435702	2.49E-10
IGF1	12.94207	4.86E-10
CCN2	2.704102	7.06E-09
INHBA	2.453555	7.65E-09
SERPINE2	2.402521	9E-09
ACTA2	2.680545	9.14E-09
CSRP2	1.843802	1.86E-07
COCH	1.299277	5.81E-07
SCG2	11.68566	6.19E-07
MFAP2	1.825173	7.59E-07
HBEGF	11.70024	1.22E-06
ITGA1	1.665868	1.47E-06

E

Overlap in pathways significantly upregulated with TGFβ1 in the 3 clusters of primary stromal cells

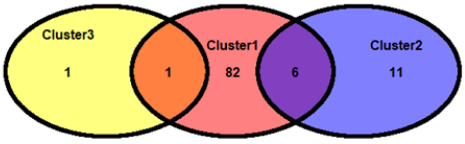


Fig 4.5.1.7 TGFβ1 stimulation in Clusters 2 and 3

(A+B) Volcano plots showing DEGs between TGFβ1 stimulated and unstimulated primary stromal cells in Cluster 2 **(A)** and Cluster 3 **(B). (C+D)** The 20 most significantly upregulated genes with TGFβ1, of those with log2FC >1, in Cluster 2 **(C)** and Cluster 3 **(D). (E)** The overlap in pathways upregulated by TGFβ1 in Clusters 1-3.

Nineteen of the fifty HALLMARK pathways were significantly up or down regulated by TGFβ1 in at least one of the clusters of primary stromal cells. They are shown for each of the three clusters, where a normalised enrichment score is significantly altered it is shown in red, and otherwise it is shown in blue **(Fig 4.5.1.8)**. Most pathways show a similar magnitude of change across the three clusters, in the two pathways where opposing directions of change are seen between clusters, in only one of the three clusters was the change statistically significant.

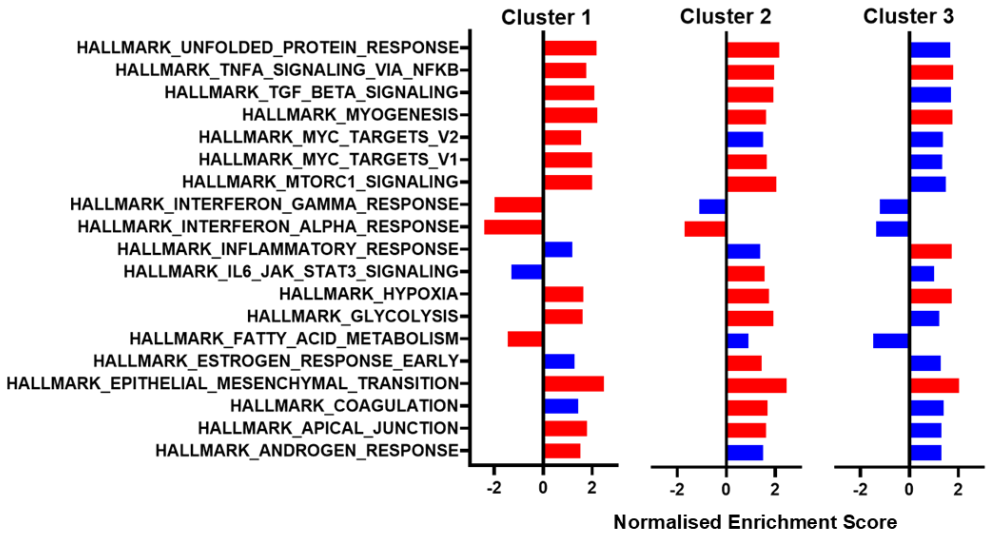


Fig 4.5.1.8 TGFβ1 stimulation in the three clusters of primary stromal cells
HALLMARK pathways significantly up or down-regulated with TGFβ1 in at least one of the three clusters of primary stromal cells. Bars showing normalised enrichment scores are red where statistically significant and blue where statistical significance is not reached.

4.5.2 Comparing Cluster 1 Primary Stromal cells by Tumour Site

We then went on to study differences between the tumour and normal adjacent cells, using the 5 pairs of cell lines where both the tumour and normal resembled fibroblasts. Firstly, the unstimulated tumour and normal adjacent cells between these 5 pairs were compared. There were 548 differentially expressed genes between tumour and normal. Looking at GO biological processes these were no pathways in which there was significant upregulation in the tumour compared to normal adjacent in at least 10% of genes in that pathway. There was significant upregulation of 3 pathways in the normal adjacent compared to tumour stromal cells; 'Retinol metabolic process', 'Retinoic acid metabolic process' and 'Ethanol metabolic process' **(Fig 4.5.2.1)**. Looking at HALLMARK pathways, none were significantly enriched between tumour and normal adjacent stromal cells.

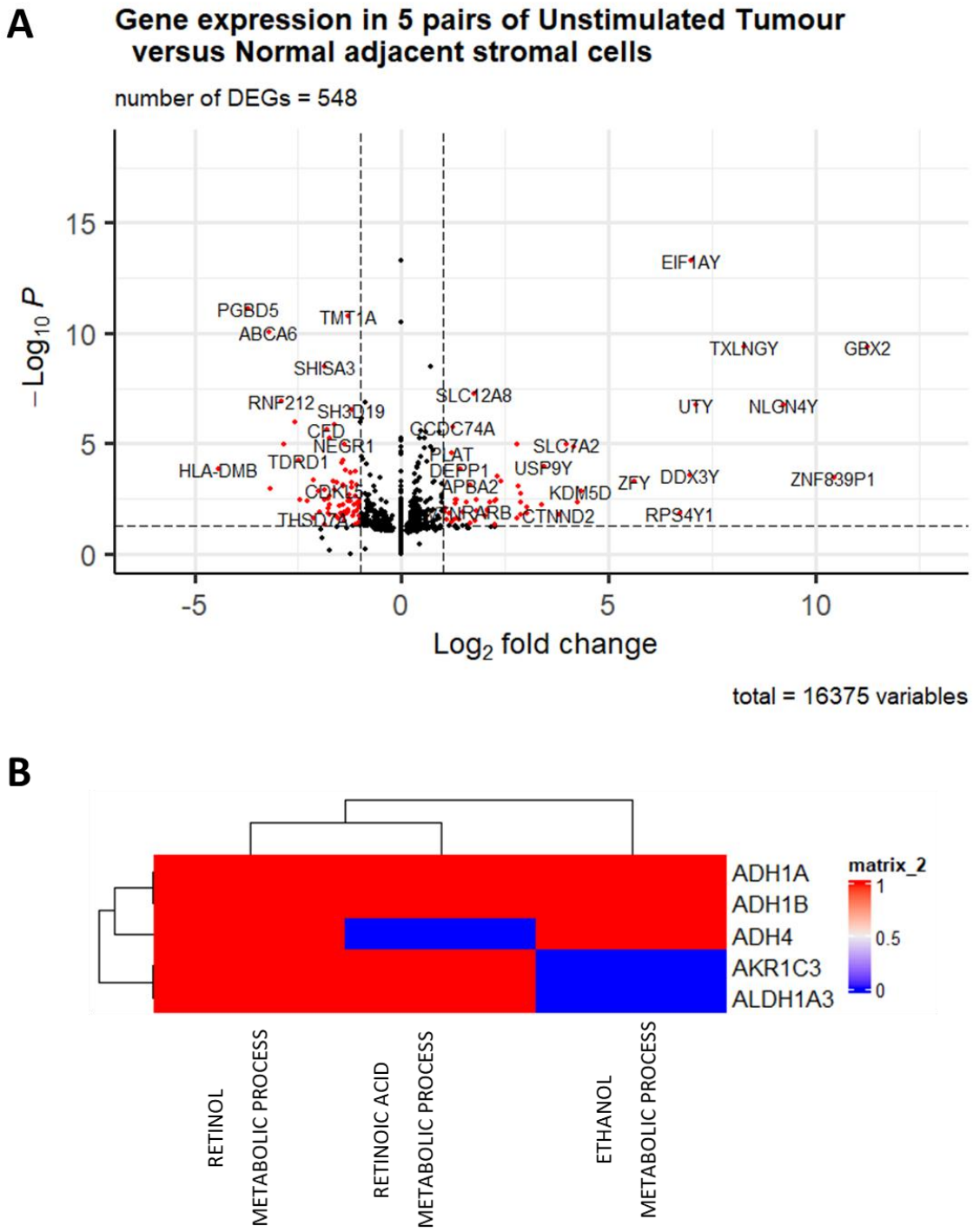


Fig 4.5.2.1 Comparison of Unstimulated primary tumour and normal adjacent stromal cells, focussing on 5 pairs.

(A) Volcano plot showing DEGs between tumour and normal adjacent stromal cells. (B) Heatmap of pathways and genes within that pathway significantly upregulated in normal adjacent compared to tumour stromal cells.

Response to each of the cytokines in tumour and normal adjacent was then assessed in the 5 pairs, using the analysis where the paired nature of the sample was imputed. This identified 16 genes responding differentially by site to 0.1ng/ml IL-1β, 16 genes for 5ng/ml TGFβ1, 11 with 5ng/ml TGFβ2 and 15 with 5ng/ml TGFβ3, all for 24 hours **(Table 4.5.2.1).** The majority of these genes were lowly expressed and the changes in tumour or normal adjacent stromal cells were not consistent across the 5 samples. With TGFβ1, 13 out of 16 genes identified were lowly expressed and did not alter consistently across the 5 lines. *ANKRD46, CDK8* and *NAA25* appear to go up consistently in N.adj, but not tumour stromal cells **(Fig 4.5.2.2 A-C)**. *ANKRD46* (Ankyrin repeat domain 46) has low cancer specificity and is not prognostic[181]. *CDK8* is an oncogene and often amplified in CRC. N-alpha-acetyltransferase 25 (NAA25) acetylated methionine residues and is expressed in the colon and many other tissues and has low cell type specificity. It is expressed in most cancer types, but is associated with poor prognosis in renal, liver and endometrial cancer[181]. Looking at the genes altered with TGFβ2, TGFβ3 or IL-1β, again, all were either lowly expressed (<10), or with inconsistent effects across the 5 lines, or differences seen were in the baseline, unstimulated cells rather than in the expression level seen after stimulation.

Table 4.5.2.1. Genes differentially regulated in tumour and normal adjacent stromal cells by cytokine stimulation, in the 5 pairs of lines most likely to be fibroblasts.

TGFβ1	**TGFβ2**	**TGFβ3**	**IL-1β**
ISM2	LPAR5	SPOCK2	DAZL
MMP28	Metazoa_SRP	TRH	KIF1A
KCNU1	CAMK1G	LPAR5	OPCML
SST	ZNF343	SPINK13	KL
IGLC3	PABIR3	SAA4	DOK7
DEPDC1P2	IGHA2	CAPN6	MNX1
IRF8	MAGEC1	CEP162	MNX-AS1
CDK8	KL	FOXP1-IT	HAS3
THSD7A	H19	IGLC3	PKDCC
NAA25	MEPCE	IGHA1	PI15
TMEM86B	SST	IGHA2	ITGB1P1
TBC1D4		C1QC	MAGEC1
CCDC85A		TIMM9	DHRS7C
DOCK10		RTN4RL1	IGLC3
ANKRD46		CDK8	ARHGDIG
ENPEP			SIGLEC17P

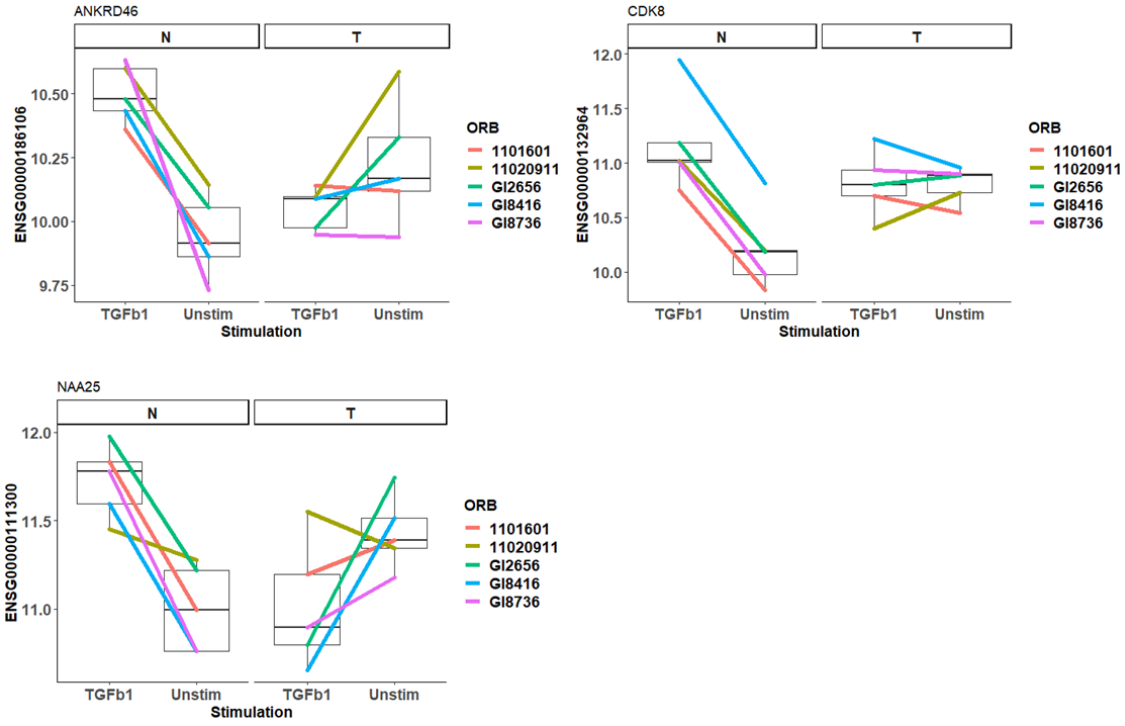


Fig 4.5.2.2 Genes differentially regulated by TGFβ1 in 5 pairs of tumour and normal adjacent stromal cells.

Boxplots showing abundance of transcripts in tumour and normal adjacent stromal cells with and without TGFβ1. **(A)** ANKRD46, **(B)** CDK8 and **(C)** NAA25.

4.5.3 Responses in Individual Cell Lines

We performed analysis of DEGs with 5ng/ml TGFβ1 and 0.1ng/ml IL-1β for 24 hours versus unstimulated in each of the sixteen primary stromal lines individually. There was huge variation in response, with between 253 and 4437 DEGs with TGFβ1 and with 329 to 3367 DEGs with IL-1β **(Table 4.5.3.1)**.

Table 4.5.3.1: Numbers of DEGs (log FC > +/- 1 and p < 0.05) with TGFβ1 and IL-1β in each individual primary stromal line

Line number	**Tumour TGFβ1 DEGs**	**Tumour IL-1β DEGs**	**N. adjacent TGFβ1 DEGs**	**N. adjacent IL-1β DEGs**
1	2577	1471	2348	1747
2	1756	3367	3229	1017
3	1338	1217	4437	2945
4	857	329	253	1170
5	1126	1474	3014	2004
6	2285	1591	1994	1739
7	3214	2916	3646	2748
8	1661	827	3403	1687

We decided to see if there was any correlation between number of DEGs with TGFβ1 and IL-1β for each of the 16 primary lines individually, there was significant positive correlation, R-squared = 0.3195, p=0.0225 **(Fig 4.5.3.1 A)**. This suggests that some lines are generally more responsive to stimulation, whatever that cytokine might be. This may be related to how well the cells grew, as 4_Tumour, 4_N,adj and 3_Tumour were some of the slowest growing lines and these had fewest DEGs.

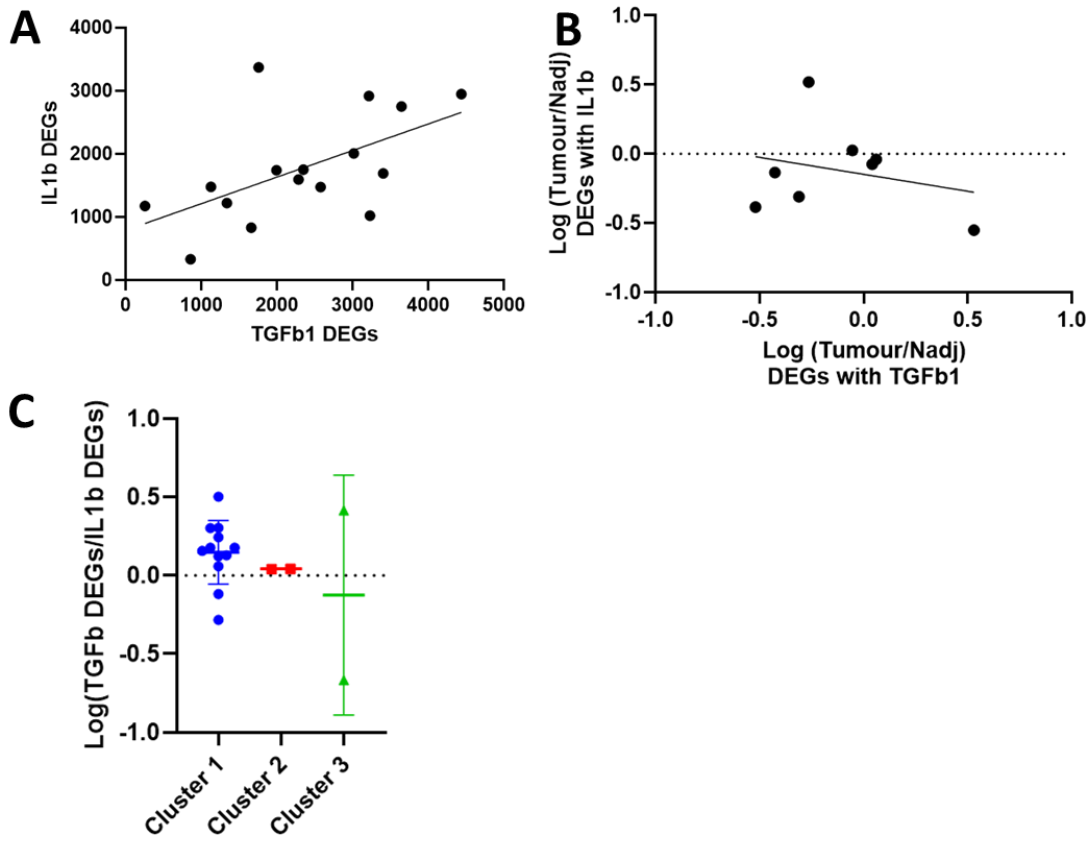


Fig 4.5.3.1. Magnitude of response to TGFβ1 and IL-1β varies by primary cell line
(A) Relationship between number of DEGs with each of TGFβ1 and IL-1β stimulation compared to unstimulated for each of the 16 primary stromal lines. (B) Correlation between the ratio of DEGs between tumour and normal with TGFβ1 versus IL-1β stimulation for the eight pairs of primary stromal lines. (C) For each of the 16 lines, log of the ratio between TGFβ1 and IL-1β DEGs, by cluster.

We then looked at the tumour compared to normal adjacent response to TGFβ1 and IL-1β for each of the eight pairs, and plotted the relative response to the two cytokines against each other. We predicted there would be negative correlation, that each tumour would either be hyper-responsive to TGFβ1 or IL-1β compared to normal adjacent, but not both. There was negative correlation, although it was not statistically significant **(Fig 4.5.3.1 B),** and the logFC for tumour compared to normal adjacent response was small for all sample with both cytokines. Next we hypothesised that primary stromal lines from some clusters may be more responsive to a particular cytokine. Therefore, for each of the 16 primary stromal lines, we looked at the ratio of DEGs with TGFβ1 and IL-1β by cluster. There was no association with cluster, but given the small numbers in clusters 2 and 3 this is unsurprising **(Fig 4.5.3.1 C).**

4.6 Mapping to Pelka scRNAseq

4.6.1 Manual correlation of DEGs with cytokine stimulation and Pelka scRNAseq cluster marker genes

I decided to see if there was any correlation between my DEGs with TGFβ1 and IL1β stimulation compared to unstimulated and the marker genes for the Pelka *et al.* scRNAseq clusters, as this may indicate which clusters were responding to these cytokines *in vivo*. I looked at the 200 marker genes of each Pelka cluster, which are the genes with the smallest p-values, and selected from these those with a logR >1. I then looked at overlap between these gene lists and the DEGs (defined as Log2FC > +/- 1 and p<0.05) with TGFβ1 **(Table 4.6.1.1)** and IL-1β **(Table 4.6.1.2)** compared to unstimulated in my primary stromal lines. There was greater overlap in my TGFβ1 DEGs and the myofibroblast and GREM1+ CAF clusters, and with IL-1β there was a gradually increasing percentage overlap from the myofibroblast cluster through to the MMP3+ CAF cluster. This indication that cytokine stimulation can lead my cells to resemble the scRNAseq clusters made me decide to do some more formal alignments using existing algorithms.

Table 4.6.1.1 Correlation between genes upregulated with TGFβ1 in our primary stromal lines and marker genes of the Pelka *et al.* clusters.

Cluster	Overlapping genes	Total genes in cluster	Percentage overlap
cS26 (myCAF)	24	87	27.6
cS27 (CXCL14+ myCAF)	19	122	15.6
cS28 (GREM1+ overlap)	32	111	28.8
cS29 (MMP3+ iCAF)	22	91	24.2
cS30 (CCL8+)	11	121	9.1
cS31 (stem-like CAF)	9	128	7.0

Table 4.6.1.2 Correlation between genes upregulated with IL-1β in our primary stromal lines and marker genes of the Pelka *et al.* clusters.

Cluster	Overlapping genes	Total genes in cluster	Percentage overlap
cS26 (myCAF)	3	87	3.4
cS27 (CXCL14+ myCAF)	7	122	5.7
cS28 (GREM1+ overlap)	15	111	13.5
cS29 (MMP3+ iCAF)	34	91	37.4
cS30 (CCL8+)	11	121	9.1
cS31 (stem-like CAF)	9	128	7.0

4.6.2 Multi-subject Single-cell Deconvolution (MuSiC)

4.6.2.1 Mapping of all samples using MuSiC

We mapped our bulk RNAseq data to the Pelka *et al* scRNAseq dataset using the Multi-subject Single-cell Deconvolution (MuSiC) algorithm, to see which clusters our primary stromal cells resembled with and without cytokine stimulation. A proportion overlap between each of our samples and each of the gene signatures from the 11 fibroblast and 6 pericyte clusters was calculated, and for each sample the proportions added up to 1. The results were averaged for each of our technical replicates, then the overlap for the similar clusters were combined; the five cancer pericyte cluster, the two stem-like normal fibroblast clusters, and the two BMP-producing normal fibroblast clusters. The overall distribution of the samples is shown in **Fig 4.6.2.1 A.** The CCL8+ normal fibroblast cluster is absent from the graph, as it had a concordance of 0 in every sample in all stimulation conditions.

Interestingly our unstimulated hPC-PL pericyte line had much better concordance with some of the fibroblast clusters than the pericytes, it shared 13.2% overlap with the pericyte clusters combined, but showed 47.7% overlap with the GREM1+ CAF cluster. The unstimulated CCD18-Co fibroblast line on the other hand showed lower 5.6% overlap with the pericyte clusters, but its highest concordance of 30.6% was with the CXCL14+ CAF cluster. Every single unstimulated sample showed its highest overlap with a fibroblast cluster. We noticed that the sample with greatest concordance with the pericyte clusters was 7_Tumour, which was one of our Cluster 2 samples previously thought to be more pericyte-like in nature. The other sample in Cluster 2 was 3_Tumour, which showed one of the highest overlaps with stem-cell like normal fibroblasts and interestingly no overlap at all with either the myofibroblast or CXCL14+ CAF clusters. 7_Tumour also showed very low overlap with these clusters. The unstimulated samples showing greatest overlap with the BMP-producing normal fibroblasts were 4_Tumour and 4_N.adj, which were the samples in the separate cluster 3, therefore these samples are likely to be telocytes. None of our unstimulated samples showed more than 50% alignment with any scRNAseq cluster.

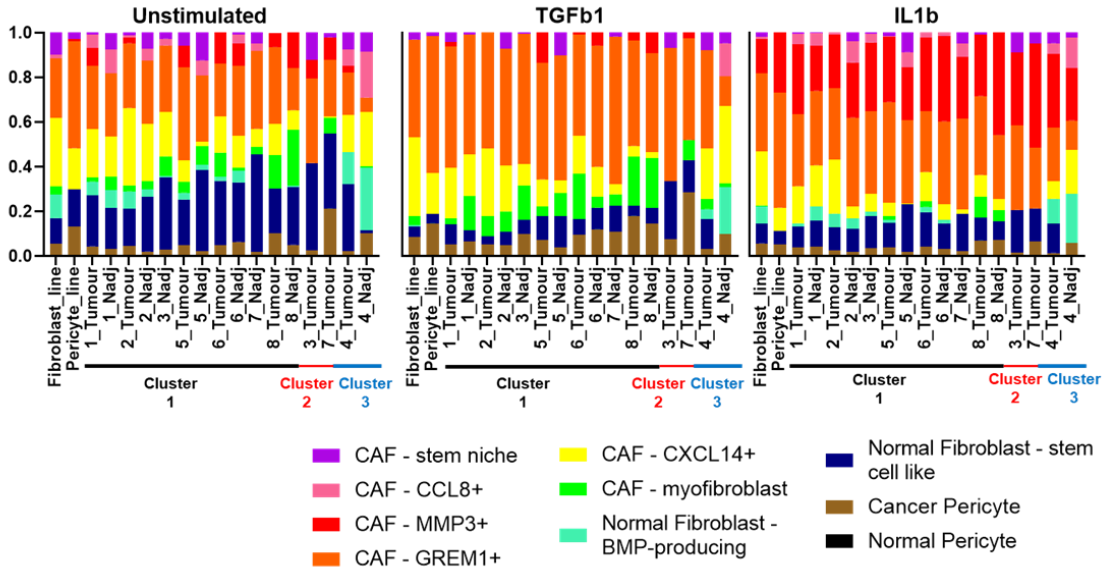


Fig 4.6.2.1 Alignment of each sample with the scRNAseq data from Pelka *et al* using the MuSiC algorithm

Samples are fitted to the pericyte and fibroblast clusters from the Pelka *et al.* scRNAseq dataset so that the total concordance makes 1. Cell lines are shown first, and then the 16 primary stromal lines are ordered by cluster, shown by stimulation conditions.

4.6.2.2 Mapping of the 12 primary stromal lines in Cluster 1 to the scRNAseq using MuSiC

As TGFβ is expected to promote a myCAF phenotype we hypothesised we would see greater concordance between the TGFβ stimulated samples and the myofibroblast and CXCL14+ CAF scRNAseq clusters than the unstimulated or IL-1β stimulated samples. Similarly, we expected to see greater overlap between our IL-1β stimulated samples and the iCAF-like MMP3+ CAF cluster than either unstimulated or TGFβ stimulated samples. We therefore studied cytokine response to the overlap with the Pelka *et al* clusters for the 12 primary stromal cell lines in our Cluster 1 which were most likely to represent fibroblasts, **Fig 4.6.2.2 A** shows the average of the 12 samples. A paired one-way Anova was performed, comparing concordance with each cluster for TGFβ and IL-1β stimulation compared to unstimulated, and where there is a statistically significant difference in concordance with a cluster a barplot is shown in **Fig 4.6.2.2 B** and full results are shown in **Table 4.6.2.2**.

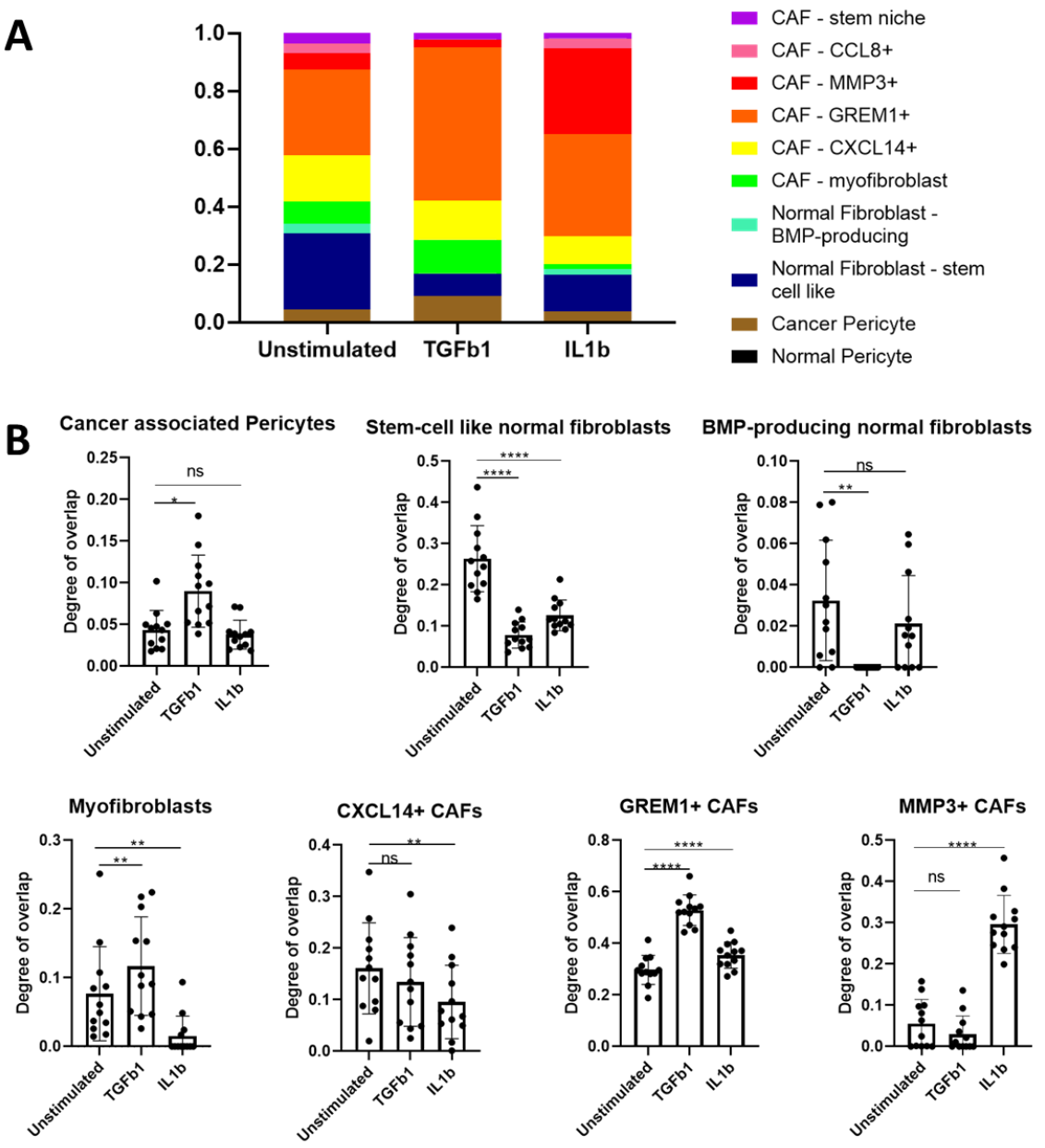


Fig 4.6.2.2 Cytokine stimulation alters alignment with the scRNAseq clusters from Pelka *et al* using the MuSiC algorithm

The 12 primary stromal lines in Cluster 1 are fitted to the pericyte and fibroblast clusters from the Pelka *et al.* scRNAseq dataset so that the total concordance makes 1. (A) The averages for each stimulation condition. (B) Alignment with each scRNAseq cluster with TGFβ1 and IL1β stimulation compared to unstimulated. Statistics with one-way ANOVA (* p<0.05; ** p<0.01, **** p<0.0001, ns = no significance).

Table 4.6.2.2: Comparing the overlap of unstimulated, TGFβ1 and IL-1β stimulated primary stromal lines with Pelka *et al*'s single cell clusters using the MuSiC algorithm. Only the 12 lines most likely to be fibroblasts included in analysis, and mapped to pericyte and fibroblast clusters from the single cell dataset.

Clusters	Unstimulated	TGFβ1 stimulation		IL-1β stimulation	
	Mean	Mean	P-value v unstim	Mean	P-value v. unstim
cS31 CAF stem niche	0.03661	0.02096	0.1223	0.01811	*0.0343
cS30 CAF CCL8+	0.03303	0.000	*0.0108	0.03485	0.9508
cS29 CAF MMP3+	0.05542	0.02960	0.1114	0.2961	****<0.0001
cS28 CAF GREM1+	0.2975	0.5292	****<0.0001	0.09564	****<0.0001
cS27 CAF CXCL14+	0.1608	0.1346	0.1236	0.09564	**0.0037
cS26 CAF myofibroblast	0.07690	0.1169	**0.0081	0.01531	**0.0031
cS 25 N.fibro	0.000	0	NA	0	NA
cS23-24 N.fibro BMP producing	0.03247	0.000	**0.0050	0.02123	0.1019
cS21-22 N.fibro stem-like	0.2637	0.07864	****<0.0001	0.1262	****<0.0001
cS16-20 Cancer pericyte	0.04352	0.09006	***0.0006	0.03797	0.4141
cS15 normal pericyte	1.833e-005	0	0.2887	0	0.2887

As hypothesised, we did see much greater overlap between the MMP3+ CAF cluster and the IL-1β stimulated samples, confirming that this cluster represents an activation state of fibroblasts responding to IL-1β; average concordance with MMP3+ CAFs was 0.055 for unstimulated and 0.296 for IL-1β stimulated samples ($p<0.0001$). There was also increased overlap between GREM1+ CAFs and IL-1β stimulated primary stromal cells than unstimulated, concordance 0.355 rather than 0.298 ($p<0.0001$), and reduced overlap between the IL1β-stimulated primary stromal cells compared to unstimulated and CXCL14+ CAFs (0.096 versus 0.161, $p=0.0037$), Myofibroblasts (0.015 versus 0.077, $p=0.0031$), the stem-cell like normal fibroblasts (0.126 versus 0.264, $p<0.0001$) and the stem-cell niche CAFs (0.018 versus0.0366, $p=0.0343$) **(Fig**

4.6.2.2). These results suggest IL-1β is driving cells away from the normal fibroblast and myCAF phenotypes towards iCAF identities.

On comparing the TGFβ1 stimulated samples to the unstimulated samples, there was greater concordance with the myofibroblasts (0.117 versus 0.077, p=0.0081), but surprisingly no statistically significant difference in overlap with the CXCL14+ CAFs. This suggests that either TGFβ1 is not they key driver of the myCAF phenotype, or the Pelka CXCL14+ CAFs are not myCAFs. As with IL-1β , TGFβ1 stimulated cells did show greater overlap with the GREM1+ CAF cluster than unstimulated cells (0.529 versus 0.298, p<0.0001). TGFβ1 stimulated cells also showed greater overlap than unstimulated to the cancer-associated pericyte clusters (0.090 versus 0.044, p=0.0006), suggesting that these cell types could be interchangeable depending on the microenvironment they are in**.** As with IL-1β , TGFβ stimulated cells also showed lower overlap with the stem cell-like normal fibroblast clusters than unstimulated (0.079 versus 0.264, p<0.0001), and also gave lower overlap with the BMP-producing normal fibroblasts (0 versus 0.0325, p=0.005) suggesting stimulation with TGFβ1 is driving fibroblasts away from their normal state to a pathological phenotype. TGFβ1 stimulation also reduced overlap with the CCL8+ producing CAF (0 versus 0.033, p=0.0108) **(Fig 4.6.2.2 B)**.

4.6.2.3 Identity of primary stromal cells and cytokine response

Given we know there's enormous heterogeneity between the primary stromal cells, we considered if the identity of the unstimulated sample may determine its response to cytokine stimulation. Therefore, for each of the 16 primary unstimulated stromal cell lines, the overlap with each scRNAseq CAF cluster of interest was compared to the number of DEGs when that line was stimulated with IL-1β of TGFβ1. I wondered if, as published work had suggested CAFs may be hyper-responsive to stimulation with these cytokines, there may be a potentiating effect where fibroblasts that are already myCAF-like respond more to TGFβ and those that are already iCAF-like respond more to IL-1β, therefore driving cells to more disparate phenotypes. As there is strong correlation between the numbers of genes differentially expressed by each cytokine across the lines previously shown in **Fig 4.5.3.1 A**, I used the ratio of DEGs with TGFβ1 to DEGs with IL-1β for this.

There was a trend towards higher TGFβ1 DEG:IL-1β DEGs with higher unstimulated myofibroblast cluster identity with an R-squared >0.1 **(Fig 4.6.2.3 A)**, although this was not statistically significant. This suggests that cells previously stimulated by TGFβ1 to give a myofibroblast phenotype may become more responsive to TGFβ1 in the future. I did not see any correlation between the relative number of DEGs after stimulation with each cytokine and identity of the unstimulated primary cells as any of the other clusters **(Fig 4.6.2.3 B-D),** all other R-squared values were <0.1, suggesting either that fibroblasts resembling these clusters do not become relatively more responsive to either cytokine, or that my primary stromal cells simply don't resemble these clusters. There was also no correlation between absolute numbers of DEGs altered with each cytokine and its correlation with each scRNAseq cluster.

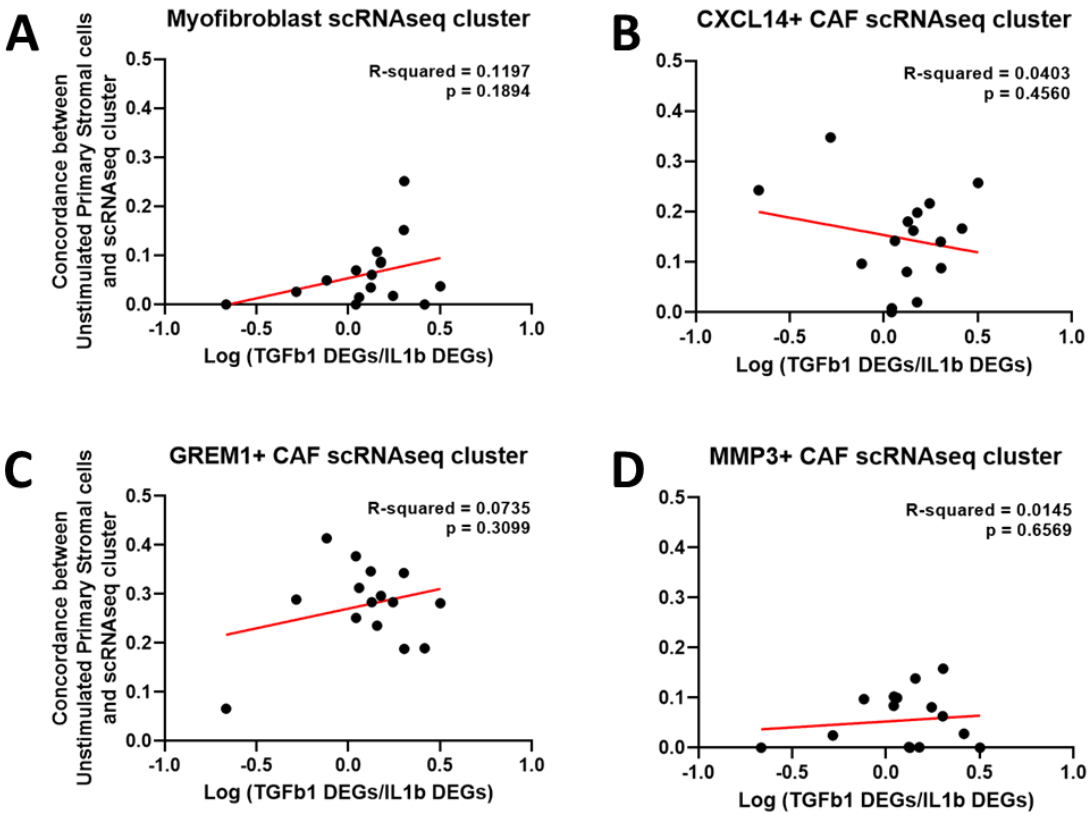


Fig 4.6.2.3 Unstimulated primary stromal cell phenotype and relative responsiveness to TGFβ1 and IL1β

(A-D) Studying all 16 primary stromal lines, the concordance of the unstimulated line with clusters of interest from the Pelka *et al* scRNAseq dataset against the ratio of number of DEGs with each of TGFβ and IL-1β stimulation compared to unstimulated. (A) Myofibroblasts (B) CXCL14+ CAFs (C) GREM1+ CAFs and (D) MMP3+ CAFs. Data analysed by simple linear regression.

4.6.3 Mapping with the SingleR algorithm

As MuSiC wasn't the perfect tool to map my data to the Pelka *et al* dataset, due to it being designed for bulk RNAseq data on tissue samples with multiple cell types, whereas my cell lines should be homogenous, we decided an additional mapping tool should be studied. We therefore decided to assume my cells lines were homogenous and were therefore each behaving as a single cell and use SingleR to define them as a Pelka *et al* cluster.

Fig 4.6.3.1 A shows how each line mapped to Pelka *et als* clusters by stimulation condition with SingleR. Of the unstimulated cell lines, the CCD18-Co fibroblasts mapped most closely to cancer pericytes proliferating, whereas the unstimulated pericytes mapped most closely to CXCL14+ CAFs. Notably, both of our 'Cluster 2' unstimulated primary lines 3_Tumour and 7_Tumour mapped to cancer pericytes proliferating, whereas the 'Cluster 3' unstimulated primary lines mapped to CXCL14+ CAFs and BMP-producing normal fibroblasts. The 'Cluster 1' unstimulated primary lines mapped to a mixture of cancer pericytes proliferating and CXCL14+, GREM1+ and MMP3+ CAFs. Of note, the two that mapped to GREM1+ CAFs came from the same patient 8_Tumour and 8_N.adj, and the two that mapped to MMP3+ CAF also mapped to the same patient 1_Tumour and 1_N.adj. In total for 5 out of the 8 samples the tumour and normal adjacent unstimulated sample mapped to the same cluster, suggesting we were picking up inter-patient differences rather than something more specific to the cells. However, the delta.next value, indicating the degree of confidence with which the samples matched, was low for all samples, suggesting that they were not able to match accurately to any cluster.

Analyses of the cytokine stimulated cells found every IL-1β stimulated sample, whether from a cell line or primary stromal cells, mapped most closely to the MMP3+ CAFs. This concurs with the findings of the MuSiC algorithm and, as this is the iCAF cluster, supports IL-1β driving this phenotype. The CCD-18-Co fibroblasts and hPC-PL pericytes stimulated with TGFβ1 all mapped to pericytes proliferating with SingleR. The primary lines stimulated with TGFβ1 mapped to a

mixture of cancer pericytes proliferating and CXCL14+, GREM1+ and MMP3+ CAFs, with the most frequently matched group being the proliferating pericytes. **Fig 4.6.3.1 B** summarises the mapping of the primary stromal cells with the differing cytokine stimulations, there was very little change in assignment of the primary stromal lines between TGFβ1 and unstimulated, 10 of the 16 primary lines remained allocated to the same groups.

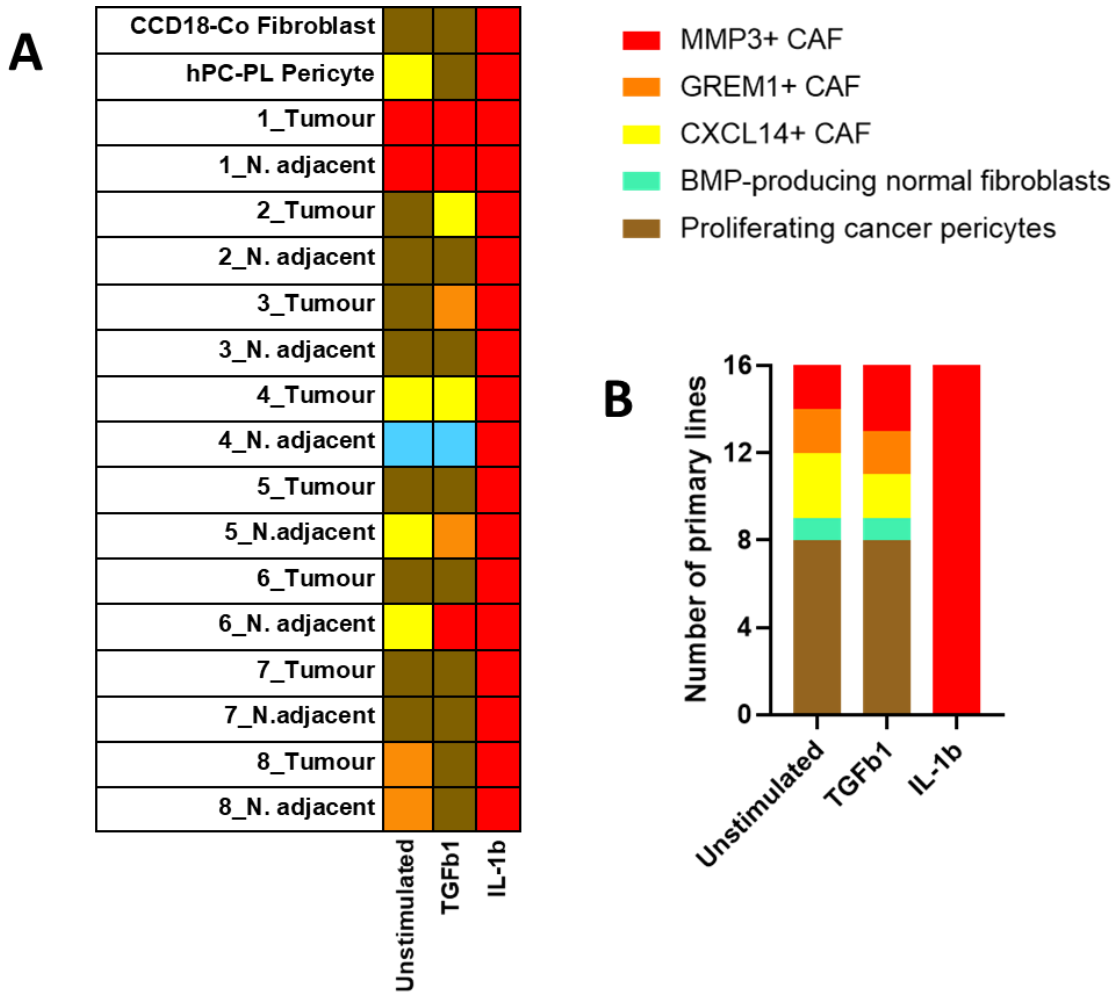


Fig 4.6.3.1 Alignment of each primary stromal sample with the scRNAseq data from Pelka *et al* using the Single R algorithm

Samples are assumed to be homogenous and behaving as a single cell and fitted to the pericyte and fibroblast clusters from the Pelka *et al.* scRNAseq dataset using the SingleR algorithm. Each sample in each stimulation condition is fitted to one cluster. (A) The assigned cluster for each cell line either unstimulated, or TGFβ1 or IL-1β stimulated. (B) The frequency of these clusters across the 16 primary stromal cell lines.

4.7 Conclusions

We found that primary stromal cells primarily clustered by the cytokine stimulation they had received; unstimulated versus IL-1β versus TGFβ. Within these groups the tumour and normal adjacent stromal cells from the same patient often, but not always, clustered more closely together than primary lines from different individuals. We saw no evidence of unstimulated cells clustering by the site they came from, tumour versus normal adjacent, or tumour versus normal adjacent stromal cells responding differently to cytokine stimulation. We also saw no clustering by isoform within the TGFβ stimulated samples.

The fact that tumour fibroblasts did not appear to be hyper-responsive to cytokines compared to normal adjacent differs to the findings seen by Koncina *et al* and Vizoso *et al*. However, in the work by Koncina *et al* there was an unusually poor response of the normal adjacent cells to IL-1β, they only identified 46 DEGs in this group compared to 829 for CAFs. In my primary lines I found between 329 and 3367 DEGs with IL-1β. In both my study and Koncina's we used an adjusted p-value of 0.05 to define an alteration as significant. They also used a FC >1, and a cut-off base mean expression of 50 to avoid inclusion of lowly expressed genes with large fold changes, but small absolute differences. We however, used FC shrink to avoid this problem giving a more graded reduction in weighting to lowly expressed genes, and looked for an absolute $\log_2$FC >1, but these differing significance criteria should not lead to huge differences in numbers of DEGs seen. Vizoso *et al* used 4 pairs of matched fibroblasts, and the hyper-responsiveness with TGFβ1 was not consistent across the 4 pairs.

We did see that the primary stromal cells stimulated with IL-1β more closely mapped to the iCAF-like MMP3+ cluster from the Pelka *et al* dataset than other clusters, using both the MuSiC and SingleR algorithms, and that there was considerable overlap between the DEGs with IL-1β and the marker genes for this cluster. This supports the hypothesis that IL-1β is driving the iCAF phenotype.

Of the three scRNAseq clusters expressing a myCAF-like contractile programme, TGFβ stimulation led to increased concordance with two: myofibroblasts and GREM1+ CAFs. This suggests TGFβ is driving these pathological phenotypes. The primary stromal cells stimulated with TGFβ did not map more closely to the CXCL14+ cluster than the unstimulated ones.

Studying cytokine response in the 12 primary stromal lines in Cluster 1, TGFβ1 stimulation significantly increased *TGFB1* expression and IL-1β stimulation increased *IL-1B* expression. However, we did not find further evidence that fibroblasts become more responsive to a cytokine they have previously been stimulated with.

4.8 Limitations and Future Work

Expanding stromal cells from resection tissue led to hugely heterogeneous samples being studied, making any comparison between sites challenging. Within each primary cell line there was competition between clones to become the dominant cell type, this meant within each primary line we were selecting for cells that grew faster. This could have led to many forms of bias in the primary cell lines. The normal adjacent stromal cells grew slightly faster than the tumour stromal cells, this would mean that any unchanged, normal fibroblasts on the edges of my tumour may become over-represented in our tumour samples, thus masking any differences between tumour and normal stromal cells. There is also considerable overlap in the iCAF phenotype and senescence-associated secretory phenotype of fibroblast, thought to limit ECM deposition in wound healing; MMP3[184], SERPINE1[107] and IL-6[108] have been used as a markers' for both. Senescent cells of this phenotype would generally stop replicating, therefore, this cell type may be under-represented in our primary cell lines. The phenotype of cells in culture may also alter significantly over time and particularly as they undergo multiple rounds of passage, therefore *ex vivo* cells should be analysed as rapidly as possible. However, some studies have cultured iCAFs *ex vivo* and maintained their inflammatory phenotype[100].

To overcome the problems of having alternative cell types growing out, the selective pressures required to generate a primary line, and changes that may occur in the cells on passage, it may be possible to use fluorescence-activated cell sorting (FACS) of the single cell suspension on the day of CRC resection to identify fibroblasts for immediate cytokine stimulation. We were unable to do this as our lab's filter for sorting is only 100um, and as many fibroblasts are larger than this they have been damaged with a lot of cell death when sorting has been attempted. Even if we were able to filter larger cells, there is not a single marker both sensitive and specific for all fibroblasts to use for sorting. Additionally, we only receive small resection samples, meaning that the initial numbers of cells are very low, and therefore we need to expand our primary lines before undertaking cytokine stimulation.

Some of my primary stromal cells, 1-4 tumour and normal, were frozen before undergoing cytokine stimulation. While this was done with the aim of batching cytokine stimulation and given the infrequency with which we had access to primary samples, this may have altered their phenotype further and it was notable that defrosted cells grew more slowly than those that had no undergone freeze-thaw and also were some of the poorest responders to cytokine stimulation with fewest DEGs. They were also stimulated after a high number of passages, between P5 and P8, due to re-expansion being required after defrosting. Therefore, with samples 5-8 I optimised my protocol so that cytokine stimulation was performed as soon as cells had expanded for there to be sufficient numbers, prior to freezing. With these cells stimulation was performed on passages P3 to P5.

Trying to ascertain the cell type identity of the primary stromal cells proved challenging, as I was using immortalised fibroblast and pericyte cell lines as a comparator, and many primary lines appeared to resemble fibroblasts in some measures, but not others. Comparison to these cell lines is not ideal given that after multiple passages they will no longer be 'normal' human cells. It is also possible that fibroblasts are inherently plastic and our primary cells may change their phenotype to that more similar to a pericyte or smooth muscle cell depending on their environment.

It was possible that two of my primary cell lines from the tumour may have been epithelial cells that had undergone EMT, given their high expression of KRT16. It would therefore be helpful when establishing my primary stromal lines, to keep the epithelial cells removed during the EDTA washes to test for tumour mutations. I would therefore be able to assess for the same mutations in my stromal cells, and if absent this would confirm there was not contamination from the neoplastic cells themselves.

There is no algorithm designed specifically for mapping bulk RNAseq data from a primary cell line grown *in vitro* to an existing scRNAseq dataset. The MuSiC algorithm is designed for mapping resection specimens with multiple cells types undergoing immediate bulk RNAseq to a scRNAseq dataset, not map a homogenous line to the scRNAseq dataset. SingleR is designed to map single cell transcriptomes to bulk transcriptomes of known cell types, therefore we were using it backwards mapping the bulk transcriptomes of individual cell types back to the scRNAseq data. Therefore, the use of these algorithms in the situation we have employed them in has not been validated.

TGFβ stimulation did not lead to increased concordance with all the scRNAseq clusters expressing a myCAF-like contractile programme. This suggests that there is an alternative driver of the contractile programme in the CXCL14+ CAF cluster. Alternatively, previous work has suggested that growing cells in 2D culture can promote a myCAF phenotype[98], this may therefore mask any additional effect of TGFβ. It is also possible that these cells had been exposed to large amounts of TGFβ *in vivo* and therefore could not generate a normal, additional response for this reason.

The Single R algorithm found that the unstimulated samples from the same patient tended to map to the same clusters. This is not necessarily an indication that the mapping wasn't functioning correctly, looking at the Pelka *et al* dataset, not all patients had cells from every cluster, certain clusters were more abundant from some patients even in the normal samples.

When planning the experiment, it was decided to use one timepoint for both the TGFβ and IL-1β stimulated samples; this was so that these could both be compared to the same unstimulated samples that had also be starved of FCS for the same amount of time prior to and for the duration the cytokine stimulation. It also meant that all cells of one line were plated out and collected together, reducing variability between the samples with different stimulation

conditions. However, the 24-hour timepoint chosen may not have been ideal for assessing upregulation of myofibroblast extracellular matrix genes, as these are known to take several days to upregulate. For example *ACTA2* is most upregulated by PCR at 4-5 days treatment with TGFβ1[185].

5 Comparison of the TGFβ isoforms on PBMCs and fibroblasts

5.1 Introduction to TGFβ Isoforms

Three TGFβ isoforms exist in humans, encoded by separate genes, but with high sequence homology[138]. All are latent when secreted from the cell and require activation in the ECM to exert their effects. Individual isoform knock-out studies in mice reveal differing roles in development; with lack of TGFβ1 leading to widespread inflammation and neonatal death, loss of TGFβ2 leading to widespread developmental abnormalities across multiple organs, and deficiency of TGFβ3 resulting in cleft lip and palate[121]. The three isoforms are also differentially expressed in the normal colonic mucosa; with very low TGFβ1 expression, TGFβ2 staining in the columnar epithelium, but not in the crypt cells, and TGFβ3 expression throughout[121].

The vast majority of published work on TGFβ *in vitro* uses only TGFβ1. For TGFβ1, RNA-sequencing (RNAseq) has characterised TGFβ-response signatures (TBRS) for CD4+ T-cells, fibroblasts, endothelial cells and macrophages[112] and studies of specific TGFβ1 effects have been performed in numerous other cells types. CD4+ T-cells are the only cell type for which the isoforms have been extensively compared, with the finding of greater IL-22 and IL-23R expression in TGFβ3 versus TGFβ1-stimulated Th17 cells[143].

Studies of the contribution of each isoform to poor prognosis in CRC have led to differing conclusions, with some suggesting TGFβ1 and 3 are most strongly associated with worse outcome[112], and others that TGFβ2 is the main driver of metastatic disease[52]. However, these studies are hampered by a lack of protein-based tools with specificity for the advertised isoform[121, 146].

The aim of this chapter is to compare the effects of the three isoforms of TGFβ on fibroblasts and other cell types critical in shaping the TME, particularly CD8+ T-cells and NK-cells. We hypothesise that the isoforms will have differing effects and that the poor prognostic effects of TGFβ in CRC may be exerted by specific isoforms. Knowledge of this area could potentially guide more specific targeting for cancer treatment.

5.2 Isoform expression and CMS groupings

We first looked in the S:CORT dataset, to see if particular TGFβ isoforms are associated with specific CMS subgroups or patients outcomes. It should however be noted that these are z-scores, giving an individual sample's expression in terms of standard deviations from the mean of the whole groups score. Therefore, comparisons between expression of the isoforms cannot be made directly, and bigger variation in standard deviation may be seen with a lowly expressed gene. This showed that TGFβ1 expression was fairly consistent across CMS subgroups, with CMS4 having levels no higher than the other groups. TGFβ2 and TGFβ3 were however more highly expressed in CMS4 than the other CMS groups, and the difference between groups was greatest for TGFβ3 **(Fig 5.2.1.1 A)**. When correlating F-TBRS with micro-array expression of each of the isoforms, there was no correlation with TGFβ1, whereas TGFβ2 and TGFβ3 expression levels did correlate with F-TBRS, despite TGFβ1 having been used to generate this signature **(Fig 5.2.1.1 B)**. The expression of each isoform was compared for patients who had and had not had a recurrence of their cancer, but there was no difference in expression levels depending on disease recurrence for any of the isoforms. Similarly, there was no difference in expression of any of the isoforms of TGFβ between patients who were alive and deceased.

There were differences in which cell types were producing each of the TGFβ isoforms, from the Pelka *et al* scRNAseq study[103]. *TGFB1* was more highly expressed, and produced by myeloid cells, fibroblasts, pericytes and endothelial cells, and at a lower level from NK cells/ILCs. *TGFB2* and *TGFB3* were only expressed in cells from the stromal compartment **(Fig 5.2.1.2 A+B)**. *TGFB2* came mainly from pericytes, whereas *TGFB3* was produced by stem-like CAFs, and to a lesser extent by GREM1+ CAFs and pericytes **(Fig 5.2.1.2 C).**

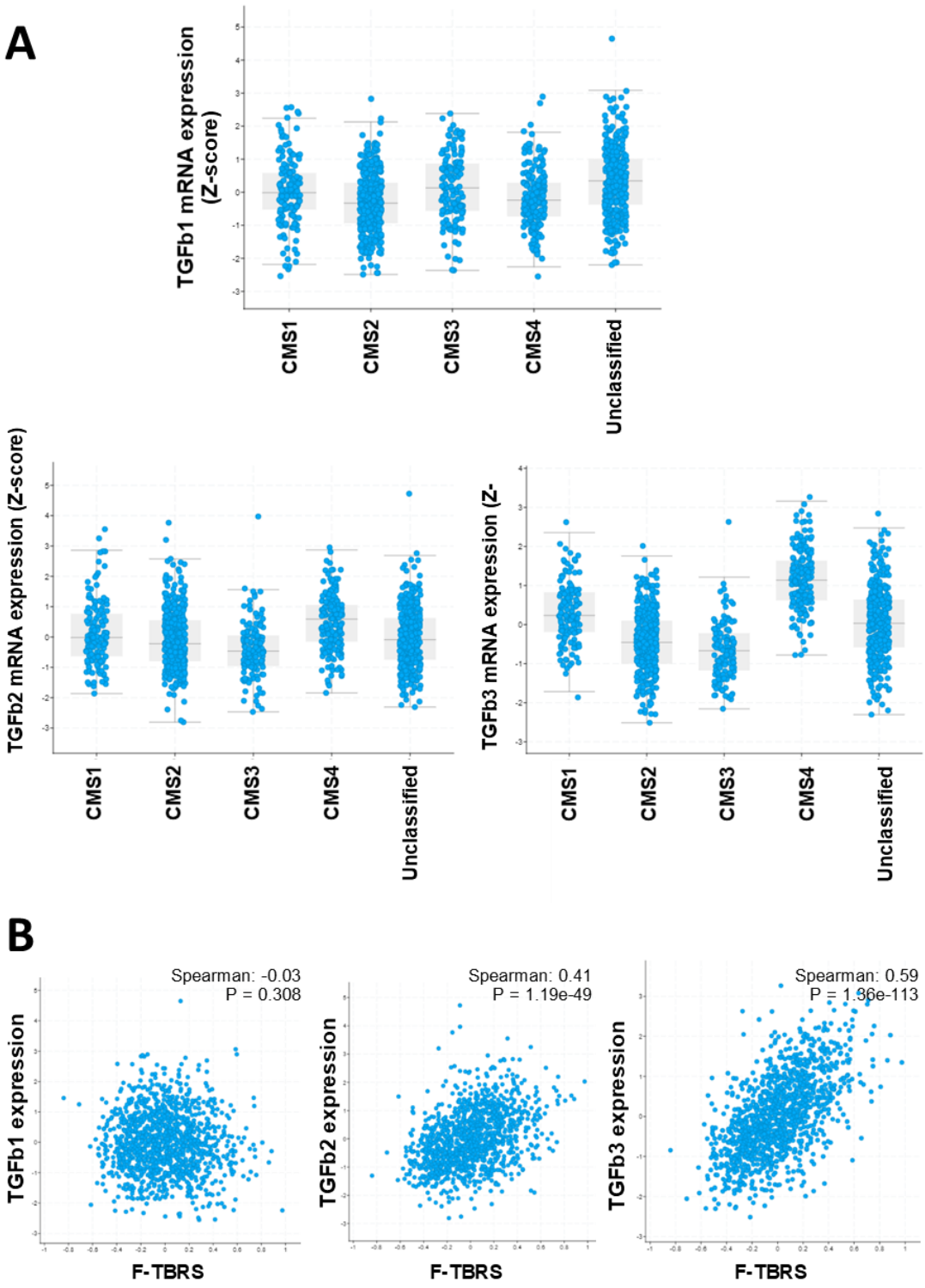


Fig 5.2.1.1 TGFβ3 is more strongly associated with CMS4 and F-TBRS than the other TGFβ isoforms

(A) Z-score from microarray for each of the TGFβ isoforms by conventional CMS grouping from micro-array. (B) Z-score from microarray for each of the TGFβ isoforms plotted against F-TBRS from micro-array. Data analysed by Spearman's rank correlation.

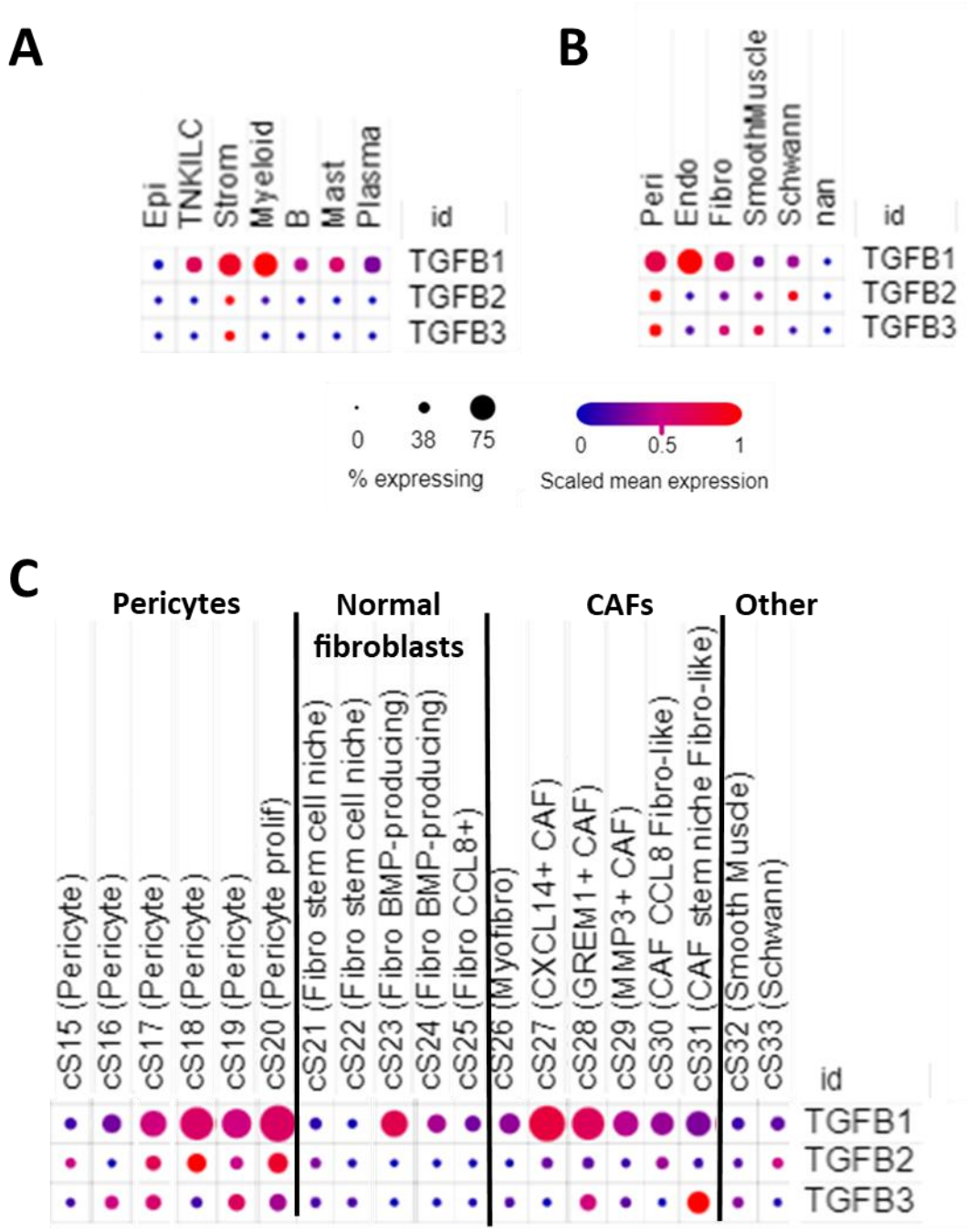


Fig 5.2.1.2 TGFβ2 and 3 are only produced by the stromal cells

(A) Expression of each of the TGFβ isoforms in Pelka *et als* broadest clusters of all cells, showing TGFβ 2 and 3 are only produced from the stromal compartment. (B+C) Expression of each of the TGFβ isoforms in the stromal cells, using broad (B) and precise (C) clustering. TGFβ1 is widely expressed, but TGFβ2 is produced by pericytes and TGFβ3 by stem-like and GREM1+ CAFs and pericytes.

5.3 The effects of the 3 TGFβ isoforms on PBMCs

CD8+ T-cells and NK cells are both components of the immune system with potential cytotoxic activity against CRC, but this ability is supressed by TGFβ1[53-57, 186, 187]. As the actions of TGFβ2 and 3 are unknown on these cell types, the effects of the 3 isoforms were compared.

5.3.1 Proliferation of PBMCs

Published studies have conflicting data on whether TGFβ1 reduces proliferation of NK cells, with one study finding mice lacking TGFβ receptors specifically on NK cells have higher NK cell numbers[188], whereas another showed lack of TGFβ receptors on NK cells did not alter numbers of this cell type[189]. PBMCs isolated from healthy blood donors were labelled with CTV to study proliferation and stimulated with αCD3 and IL-2, with or without TGFβ. Gating enabled identification of CD4+ T-cells, CD8+ T-cells and NK cells **(Fig 5.3.1.1 A)** and representative histograms of CTV dilution after 3 days are shown **(Fig 5.3.1.1.B)**. No significant difference in proliferation was observed with any of the TGFβ isoforms on days 3, 5 or 7 **(Fig 5.3.1.1 C)**.

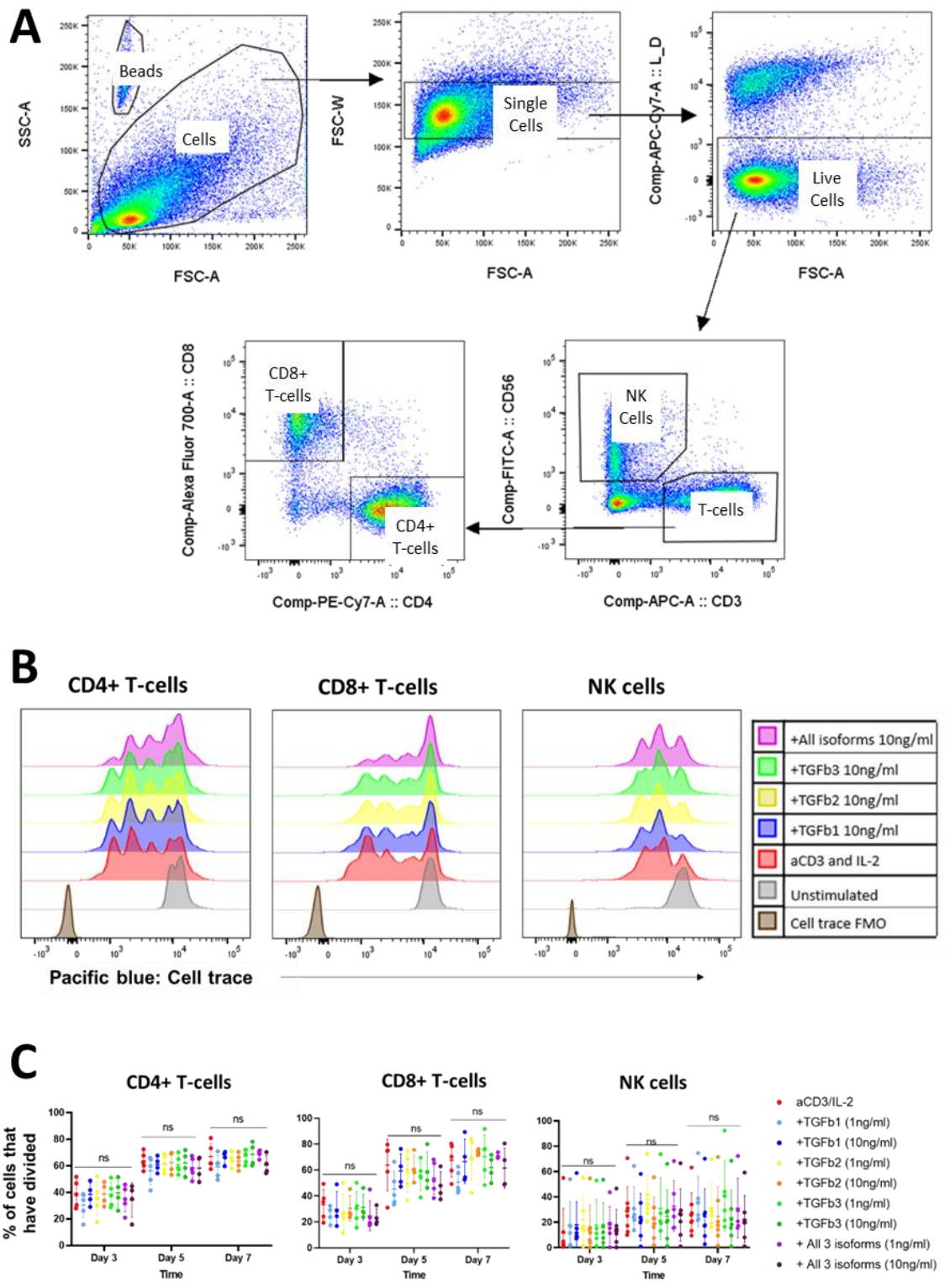


Fig. 5.3.1.1 Similar effects of TGFβ1-3 on proliferation of T-cells and NK cells.

PBMCs were unstimulated, or treated with αCD3 and IL-2, with or without TGFβ at a range of doses, for 3, 5 or 7 days, n=5. **(A)** Gating strategy to identify NK-cells and CD4+ and CD8+ T-cells. **(B)** Representative histograms of cell trace on CD4+ T-cells, CD8+ T-cells and NK cells after 3 days. **(C)** Percentage of CD4+ T-cells, CD8+ T-cells and NK-cells that have divided over the time-course.

5.3.2 Activation of PBMCs

The activation state of PBMCs from healthy donors stimulated with TGFβ was assessed by quantifying NKG2D, CD16, CD25 and CD69. NKG2D is an activating receptor on NK cells known to be decreased by TGFβ1[55]; CD16 is an Fc receptor involved in antibody-dependant cellular cytotoxicity (ADCC) although it has not previously been found to change in response to TGFβ1[55]; CD25 is part of the IL-2 receptor and becomes upregulated when T-cells are activated; and CD69 is one of the earliest markers to be upregulated when T-cells and NK-cells become activated**.**

On NK cells, there were greater changes in expression of these four surface markers with TGFβ at day 3 than later timepoints, relative to those samples with αCD3 and IL-2 alone **(Fig 5.3.2.1. A).** All 3 isoforms of TGFβ at 10ng/ml significantly reduced expression of NKG2D at day 3 ($p=0.0243$ for TGFβ1, $p=0.0093$ for TGFβ2, $p=0.0044$ for TGFβ3 and $p=0.0029$ for the combination), but no statistically significant differences were seen between the isoforms at either dose level **(Fig 5.3.2.1 B+C).** All isoforms of TGFβ reduced CD16 expression at day 3 (p values for the higher doses ranging from 0.0014 to 0.0066 depending on the isoform), but no differences were observed between isoforms **(Fig 5.3.2.1 B+D).** All TGFβ isoforms prevented upregulation of CD69 expression at day 3 ($p<0.0001$ for 10ng/ml of all isoforms), with no significant differences between isoforms **(Fig 5.3.2.1 B+E).**

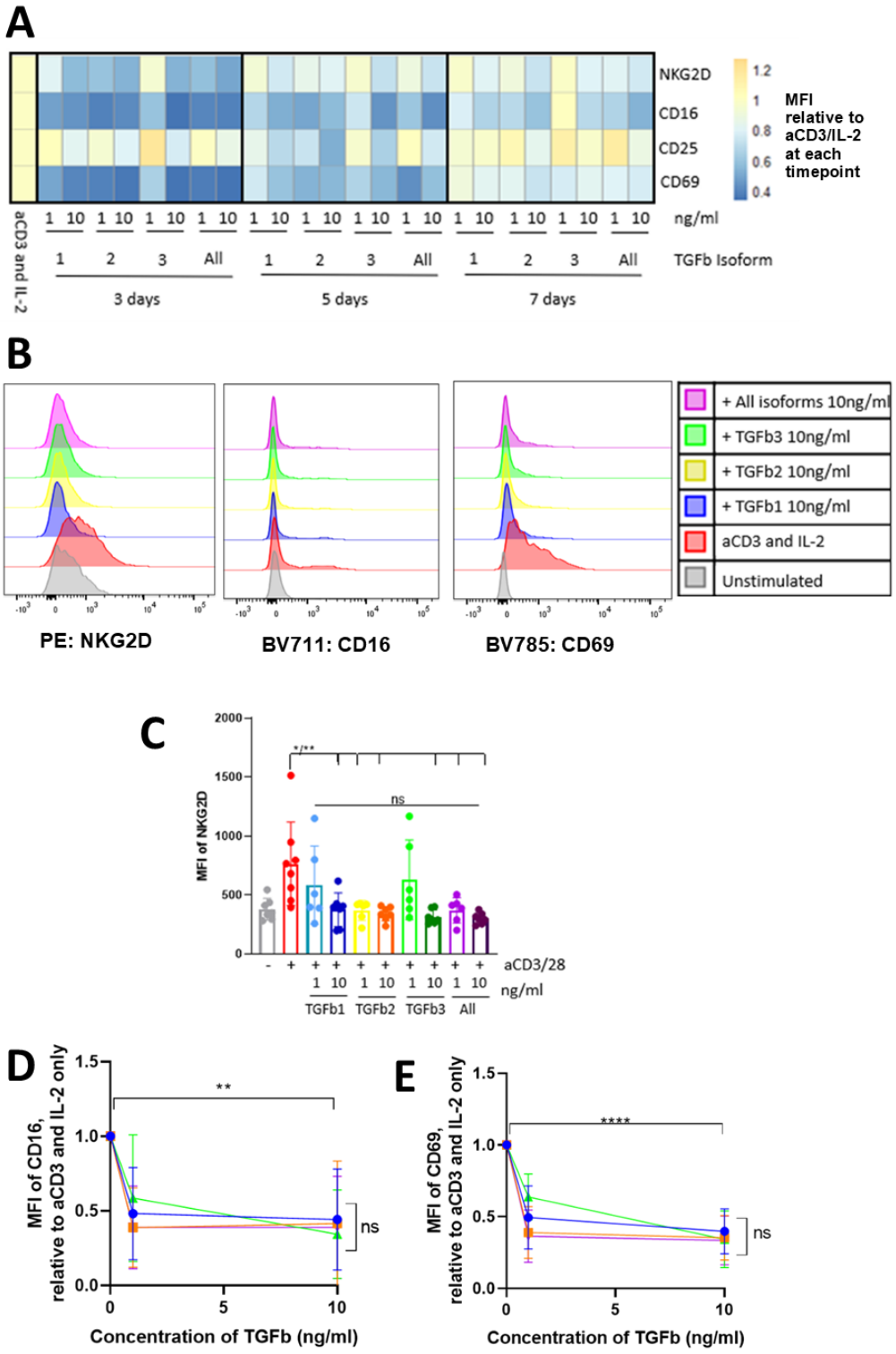


Fig. 5.3.2.1 Similar effects of TGFβ1-3 on receptor expression on NK cells.

PBMCs were unstimulated, or treated with αCD3 and IL-2, with or without TGFβ at a range of doses, for 3 days, n=6. Receptor expression on NK-cells was measured by flow cytometry. **(A)** Heatmap showing the effects of TGFβ with αCD3 and IL-2, fold-change relative to αCD3 and IL-2 alone at each timepoint, on four surface markers (rows), over a time-course. **(B)** Representative histograms of NKG2D, CD16 and CD69 on NK cells. **(C)** Median Fluorescence Intensity (MFI) of NKG2D on NK-cells. **(D-E)** Median CD16 **(D)** and CD69 **(E)** expression on NK-cells treated with TGFβ and αCD3 and IL-2, relative to those treated with αCD3 and IL-2 alone. (ns= non-significant, *=p<0.05, **=p<0.01, ***=p<0.001, ****=p<0.0001).

On CD8+ T-cells the only surface receptor studied that changed significantly with TGFβ was CD16, and the greatest effect was seen after 3 days (p values ranging from 0.0097 to <0.0001). Again, no significant differences were seen between isoforms at either dose level **(Fig 5.3.2.2 A-C).**

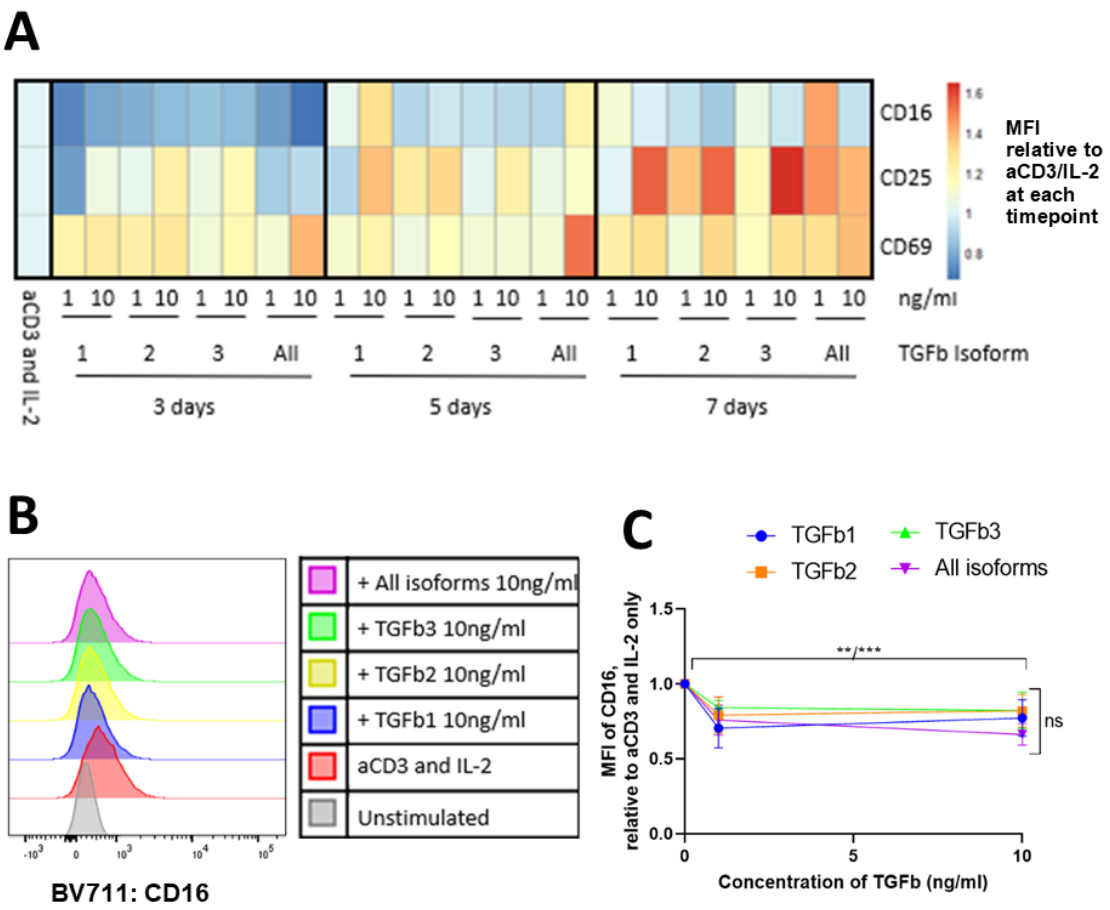


Fig. 5.3.2.2. Similar effects of TGFβ1-3 on receptor expression on CD8+ T-cells.
PBMCs were unstimulated, or treated with αCD3 and IL-2, with or without TGFβ at a range of doses, for 3 days, n=6. Receptor expression was measured by flow cytometry. (A) Heatmap showing the effects of TGFβ with αCD3 and IL-2, relative to αCD3 and IL-2 alone, on three surface markers (rows), over a time-course. (B) Representative histogram of CD16 on CD8+ T-cells. (C) Median CD16 expression on CD8+ T-cells treated with TGFβ and αCD3 and IL-2, relative to those treated with αCD3 and IL-2 alone. (ns= non-significant, *=p<0.05, **=p<0.01, ***=p<0.001).

5.3.3 Checkpoint expression on T-cells

TGFβ inhibition, targeting all three isoforms, is known to increase expression of the immune checkpoints PD-1 and CTLA-4, leading to combinations targeting both TGFβ and checkpoints being trialled. The direct effects of the individual isoforms on PD-1, CTLA-4 and TIGIT was therefore analysed in CD4+ and CD8+ T-cells. PD-1 was significantly increased on CD4+ T-cells by TGFβ, ($p<0.0001$ for each isoform at 5ng/ml compared to no TGFβ stimulation), but no statistically significant differences were seen between isoforms **(Fig 5.3.3.1 A+B)**. No other differences were seen in checkpoint expression with TGFβ1-3 on CD4+ or CD8+ T-cells.

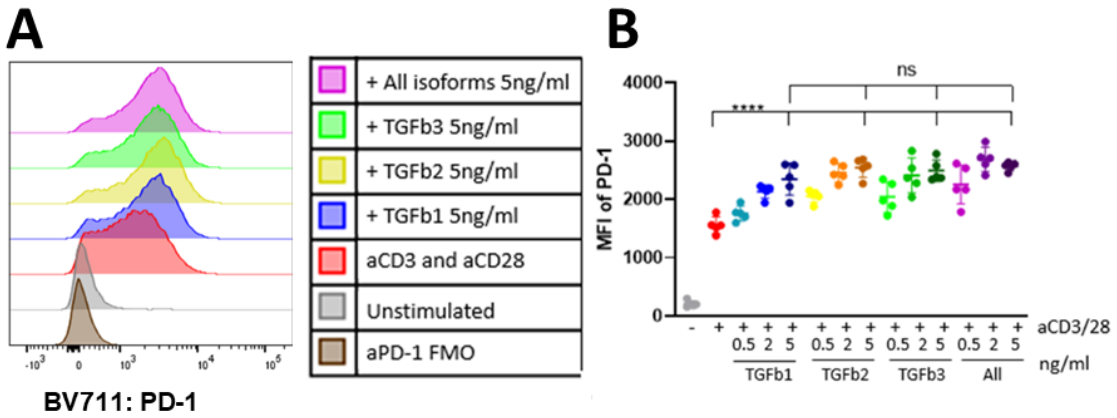


Fig. 5.3.3.1 Similar effects of TGFβ1-3 on checkpoint expression in T-cells.
PBMCs were unstimulated, or treated with αCD3 and αCD28, with or without TGFβ at a range of doses, for 3 days, n=5. **(A)** Representative histogram of PD-1 on CD4+ T-cells. **(B)** MFI of PD-1 on CD4+ T-cells measured by flow cytometry (ns=non-significant, ****=$p<0.0001$).

5.4 Comparison of the TGFβ isoforms on CCD18-Co fibroblasts.

A TBRS has been derived from RNA-sequencing of TGFβ1-stimulated ccd18-Co colonic fibroblasts, but little attention has been given to the effects of TGFβ2/3. One study found that TGFβ1 and 2 similarly upregulate αSMA and COL1A1 from ocular fibroblasts[190], but this has not been confirmed using fibroblasts from other anatomical locations. I compared the effect of the three isoforms on ccd-18Co fibroblasts.

Responsiveness of the ccd-18Co fibroblasts to the three isoforms was confirmed by western blotting for phosphorylated Smad2, with maximal phosphorylation after 30-60 minutes for all isoforms (**Fig 5.4.1.1 A+B**).

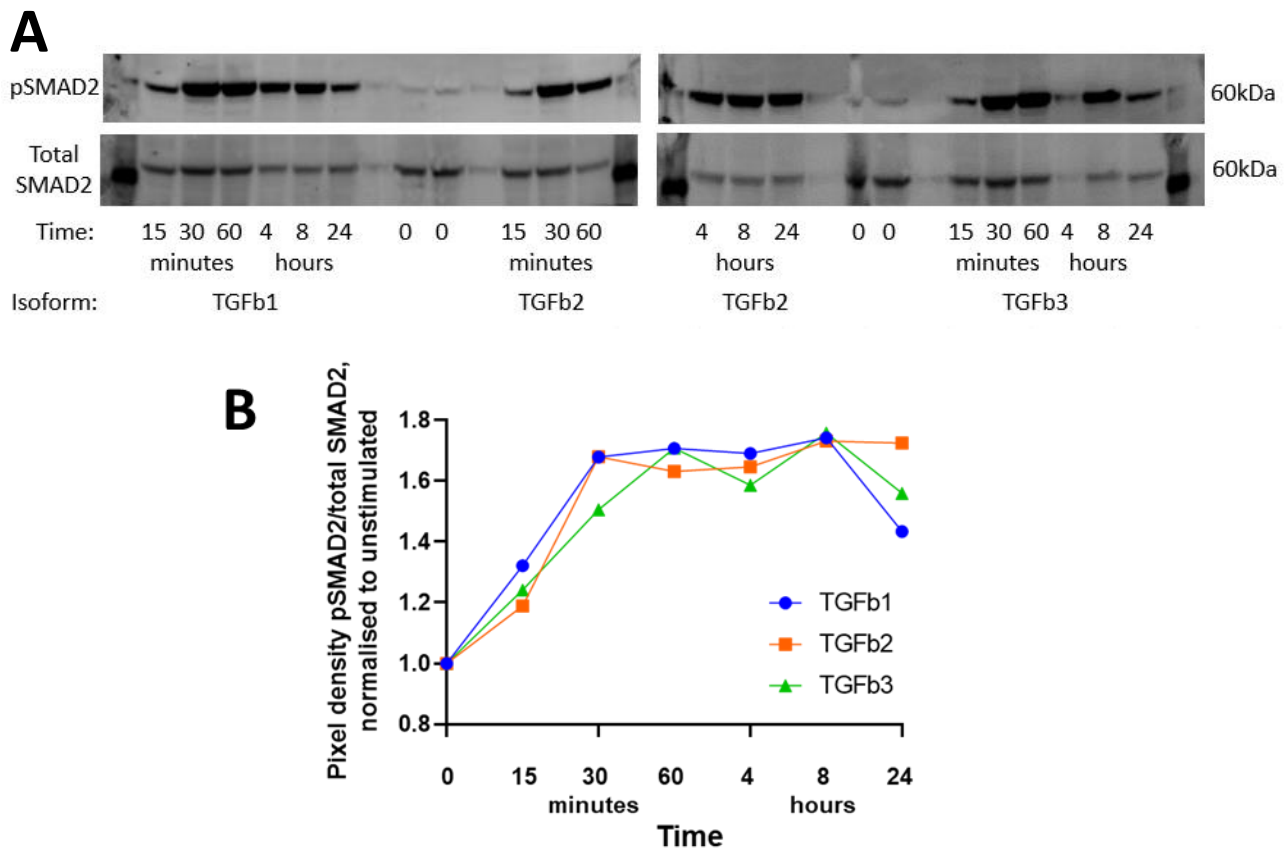


Fig. 5.4.1.1. Similar effects of TGFβ 1-3 on phosphorylation of SMAD2 and proliferation of CCD-18Co colonic fibroblasts.
(A) Western blot for pSMAD2 and total SMAD2 either without stimulation or after stimulation with TGFβ 1, 2 or 3. **(B)** Pixel density analysis of pSMAD2 relative to total SMAD2, normalised to time 0, from fig A, n=1.

Next, expression of a selection of genes known to be increased by TGFβ1 in the TBRS[112] (*IL6, IL11, TIMP3, VEGFA*), ECM genes (*COL1A1* and *ACTA2*), a chemokine known to be reduced by TGFβ1 (*CXCL12*) and the 3 isoforms themselves were probed by RT-qPCR, when the cells were either unstimulated, stimulated with one of the TGFβ isoforms or a combination of the three, or treated with Vactosertib, a TGFβ receptor inhibitor **(Fig 5.4.1.2 A)**. Vactosertib was used to prevent the effects of any TGFβ in the culture medium or produced by the fibroblasts themselves during the experiment. No significant changes were seen with Vactosertib compared to unstimulated, suggesting the effects of any TGFβ in the culture medium or produced by the cells themselves were minimal. Many known effects of TGFβ1 were replicated, but there were no significant differences between TGFβ isoforms in their effects on expression of the genes analysed, with the exception of *IL11*. All isoforms significantly increased *IL11* compared to unstimulated (p ranging from <0.0001 to 0.0308), but at the 8-hour timepoint TGFβ3 gave significantly greater increases than TGFβ 1 or 2 (p=0.0041 and 0.0049 respectively) **(Fig 5.4.1.2 B)**. However, when *IL11* expression was confirmed at the protein level by ELISA there were no statistically significant differences between the isoforms **(Fig 5.4.1.2 C)**. With the higher concentrations of TGFβ (2ng/ml and 5ng/ml), IL-11 was significantly higher for all isoforms compared with unstimulated cells (p ranging from 0.0141 to 0.0328).

All isoforms of TGFβ trended towards increased mRNA expression of TGFβ1 and TGFβ2 **(Fig 5.4.1.2 A)**, suggesting an autocrine feedback loop. TGFβ1 has been similarly reported to increase TGFβ1 mRNA 18-fold in CAFs isolated from CRC patients[191]. Conversely, I found that stimulation with any isoform of TGFβ reduced mRNA expression of TGFβ3, suggestive of a negative feedback mechanism. Looking at the mRNA expression of all three isoforms in my unstimulated samples, there is substantial TGFβ1 expression, whereas TGFβ2 and TGFβ3 are low. This may provide a continuous baseline level of TGFβ1 production by fibroblasts.

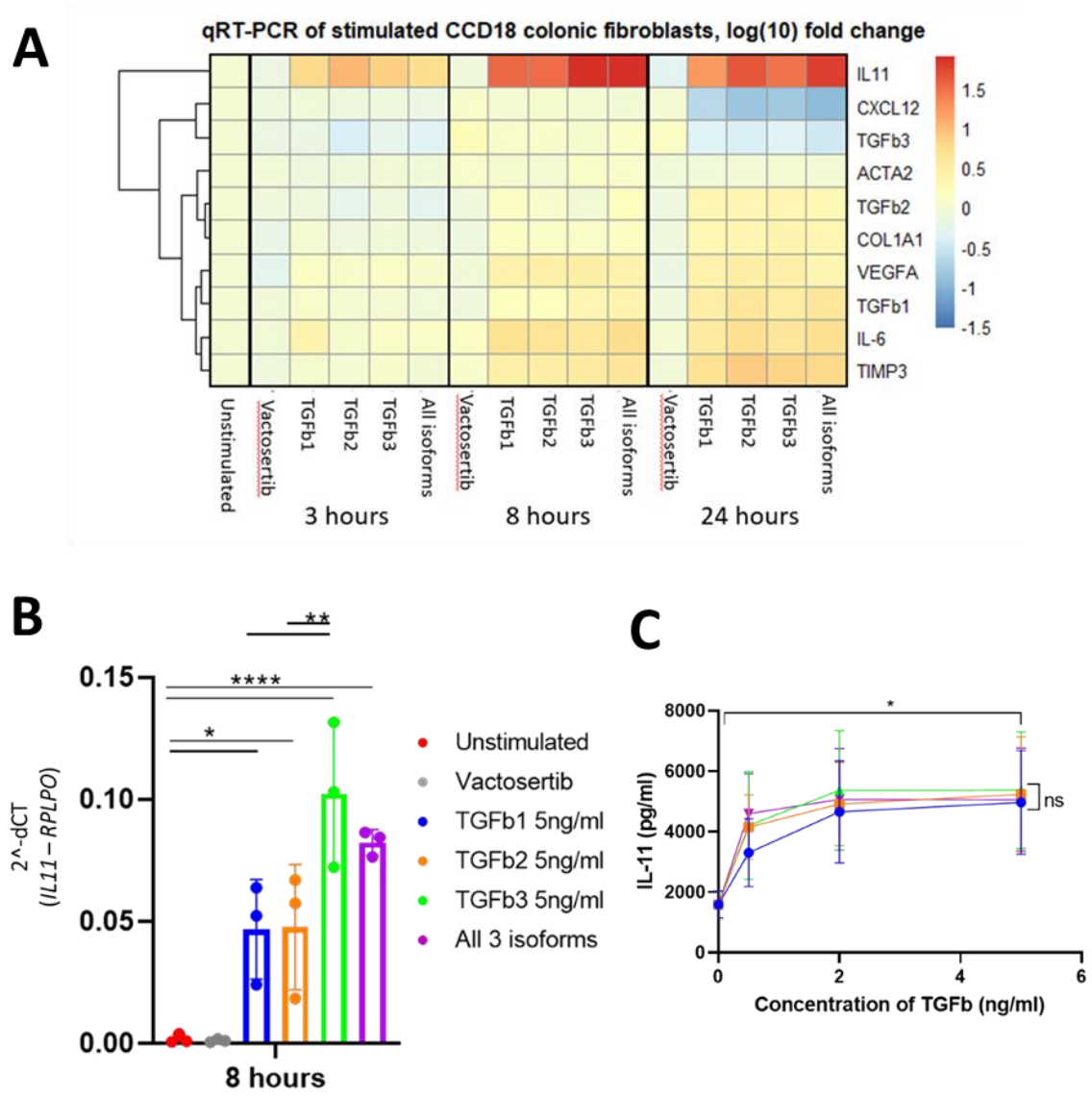


Fig. 5.4.1.2. Similar effects of TGFβ 1-3 on gene expression in CCD-18Co colonic fibroblasts.
(A) Heat map showing the effect of 5ng/ml of the TGFβ isoforms at 3 time-points (columns) on the mRNA expression by RT-qPCR of 8 selected genes (rows). Results are shown relative to unstimulated samples, n=3. (B) Expression of *IL11* by RT-qPCR. Each of the three data-points represents a separate experiment, which is the average of two biological replicates, with technical replicates for PCR in triplicate. (C) ELISA for IL-11 following 48 hours stimulation with TGFβ 1, 2 or 3. N=3, each data-point is a separate experiment, which is the average of three biological replicates.

To test whether TGFβ affected cell proliferation we used the MTT test. All 3 TGFβ isoforms similarly supressed total metabolic activity on an MTT test **(Fig 5.4.1.3 A)**. At 96 hours the difference between the unstimulated samples and the higher concentrations (2ng/ml or 5ng/ml) of all isoforms was highly significant ($p < 0.0001$). There were no significant differences between the isoforms. It is possible that the reduced formazan generation measured on MTT test was

due to decreased cell survival or metabolic activity, rather than proliferation, so RT-qPCR was performed for *MKI67* on ccd-18Co fibroblasts treated for 24 hours with 5ng/ml TGFβ 1, 2 or 3, which showed a small, but significant, reduction with all isoforms (p=0.0282-0.0421 depending on the isoform) **(Fig 5.4.1.3 B).** This is consistent with a previous study where fibroblasts lacking a TGFβR2 were found to have increased Ki67 compared to control fibroblasts, suggesting TGFβ limits proliferation[192]**.**

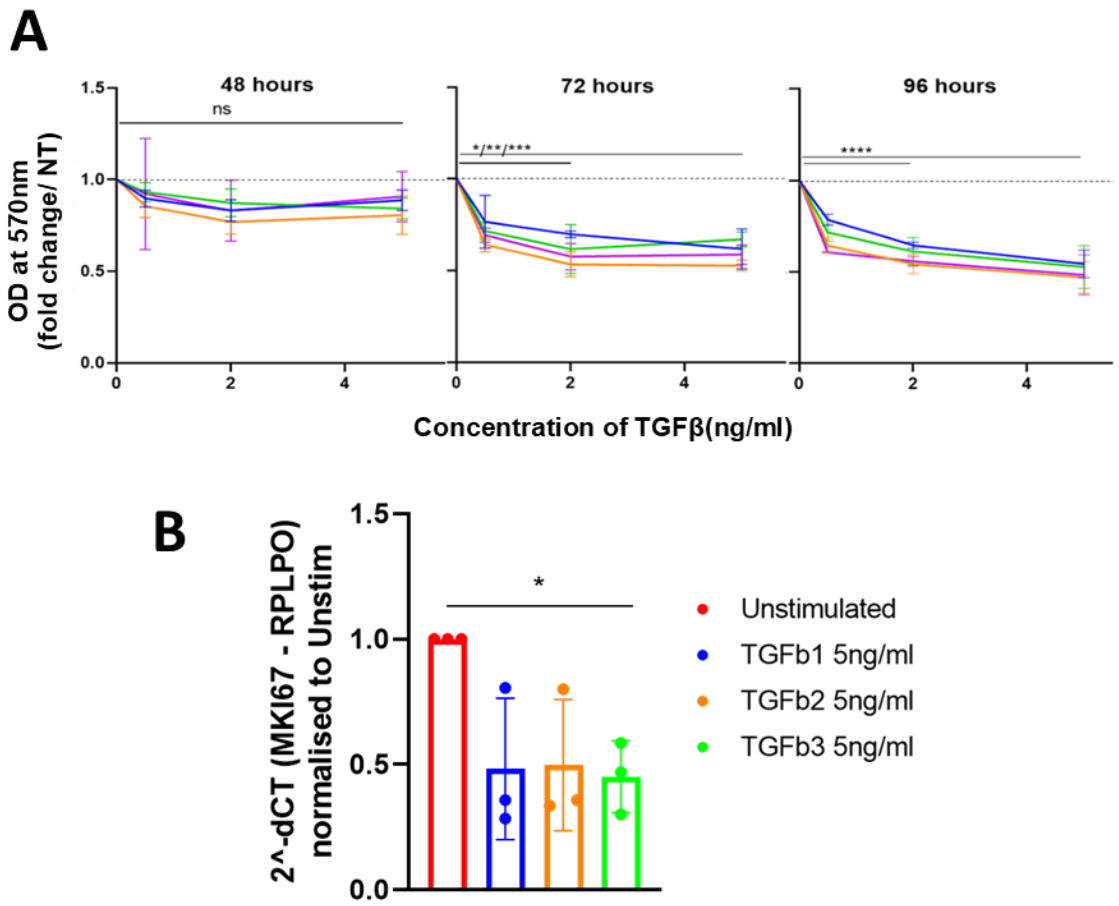


Fig. 5.4.1.3. Similar effects of TGFβ1-3 on proliferation of CCD-18Co colonic fibroblasts.
(D) MTT test for number of viable cells. Cells were stimulated with TGFβ 1, 2 or 3, over a 96-hour time-course. Results are normalised for unstimulated cells at each timepoint, n=3. (E) Expression of *MKI67* by RT-qPCR. Data from three independent experiments, and analysed by one-way Anova (ns= non-significant, *=p<0.05, **=p<0.01, ***=p<0.001, ****=p<0.0001).

5.5 Comparison of the Isoforms in Primary CRC cell lines

To be able to conclude a lack of difference between the three isoforms of TGFβ in fibroblasts an unbiased approach was required, to identify possible effects of TGFβ2 or 3 that are not occurring with TGFβ1 and therefore would not be considered as targets to measure specifically. Therefore, the effects of the three isoforms of TGFβ were compared by RNA sequencing in the ccd-18Co fibroblast line, a pericyte line and primary stromal cells from both CRC tumour and matched normal adjacent.

Differential gene expression was studied in TGFβ1 versus TGFβ2 stimulated, TGFβ1 versus TGFβ3 stimulated, and TGFβ2 versus TGFβ3 stimulated cells. These are shown as volcano plots for the CCD18-Co fibroblasts **(Fig 5.5.1.1)**. **Table 5.5.1.1** shows the number of DEGs for each of the 4 cell types for each of the 3 TGFβ comparisons, defined as having a log FC of > +/- 1 and a p value of <0.05. These genes are listed in **Table 5.5.1.2** under the cell type in which they were differentially altered by TGFβ.

Table 5.5.1.1 Numbers of DEGs between stimulation with the different TGFβ isoforms across 4 cell types

	TGFβ1 v 2	**TGFβ1 v 3**	**TGFβ2 v 3**
CCD18-Co Fibroblast line	2	3	12
hPL-PC Pericyte line	4	9	7
Primary tumour stromal cells	0	0	0
Primary normal adjacent stromal cells	3	6	0

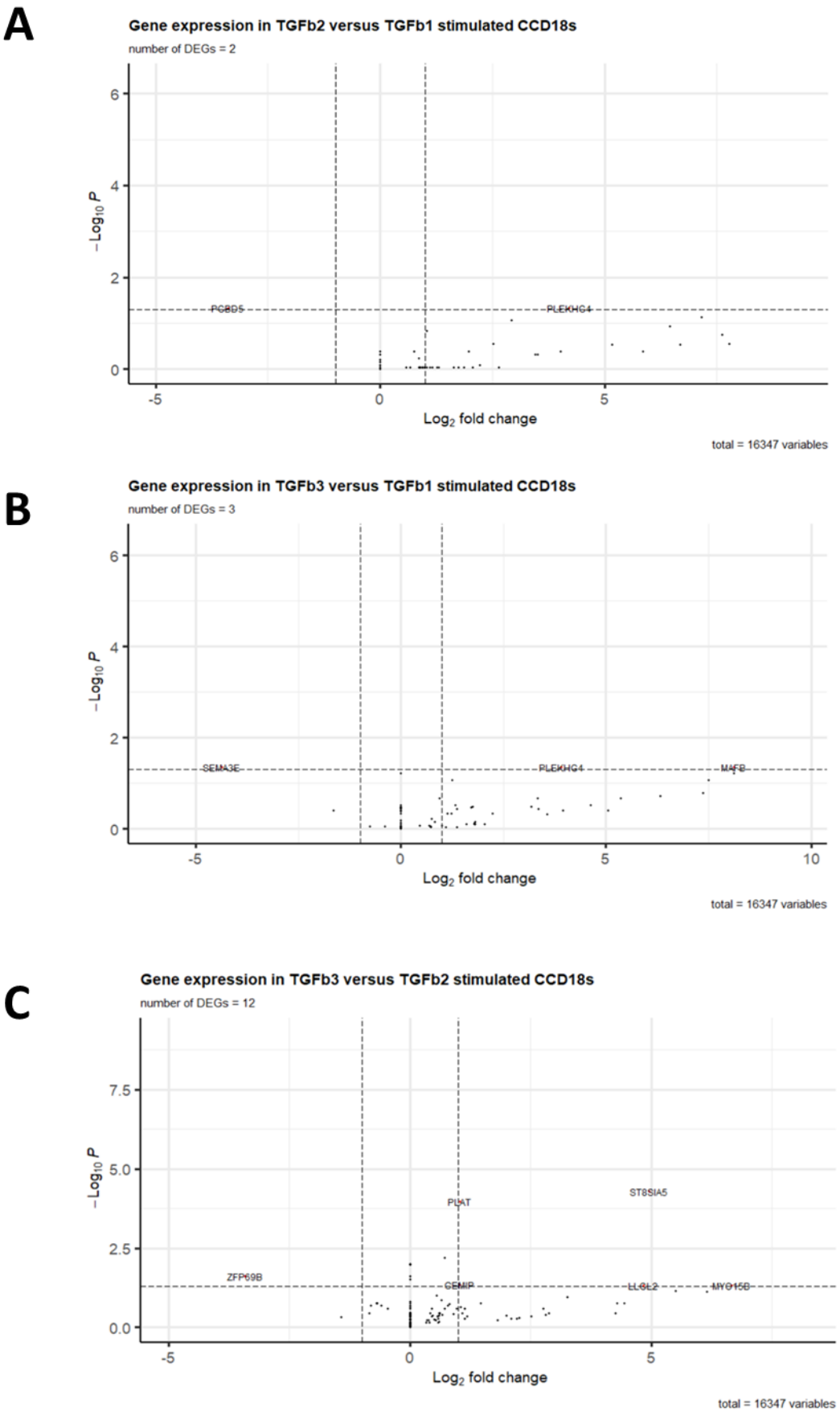


Fig 5.5.1.1 Volcano plots showing DEGs between different isoforms of TGFβ in CCD18-Co fibroblasts

(A) TGFβ1 v TGFβ2 **(B)** TGFβ1 v TGFβ3, **(C)** TGFβ2 v TGFβ3.

Table 5.5.1.2 Genes differentially altered by the TGFβ isoforms

CCD18-Co Fibroblasts	hPL-PC Pericytes	Primary stromal cells
PGBD5	INO80D	USP43
PLEKHG4	RAP2A	B3GALT5
SEMA3E	GTF3C2-AS2	RFPL3S
MAFB	MATR3	GDF7
LLGL2	DBP	SPTBN2
CELSR1	TSPAN11	COL11A2
ST8SIA5	PARD3B	ARHGAP8
CEMIP	SPATA6	ISL1-DT
HOMER2	ZNF12	Metazoa_SRP
PLAT	ZNF704	
TMEM71	KCNAB3	
ALDH1A1	COL4A6	
ZNF562	FBN1-DT	
ZFP69B	TXLNGY	
PTPN20	USP24	
MYO15B	PRR34	
	LINC02982	
	WEE2-AS1	
	LINC01614	

The numbers of identified DEGs in each scenario are very small given 16,347 genes were measured in RNA sequencing, and on the volcano plots these marked genes are of borderline significance. Many of the genes identified are not known to be expressed in fibroblasts and boxplots of expression of these genes show low abundance of <10, and therefore these are likely to be chance differences rather than real findings. The only genes identified at significantly differing levels after stimulation with the TGFβ isoforms where in at least one condition average expression was >10 were PLAT, ALDH1A1 and CEMIP, which were higher in TGFβ3 than TGFβ2 stimulated CCD18 fibroblasts, and MATR3, which was higher in TGFβ2 than TGFβ1 stimulated pericytes **(Fig 5.5.1.2).** PLAT is tissue-type plasminogen activator and is generally described a being expressed by endothelial cells, ALDH1A1 is aldehyde dehydrogenase 1 family member A1 and has been associated with poor prognosis and liver metastases in CRC[193], and CEMIP is cell migration inducing hyaluronidase 1, which is active in fibroblasts and has also been found to promote CRC liver metastases[194]. However, to find such small numbers of genes differentially

altered by the TGFβ isoforms, when hundreds of DEGs overlap between them, suggests that in reality the different isoforms would have no real significant differences in effect.

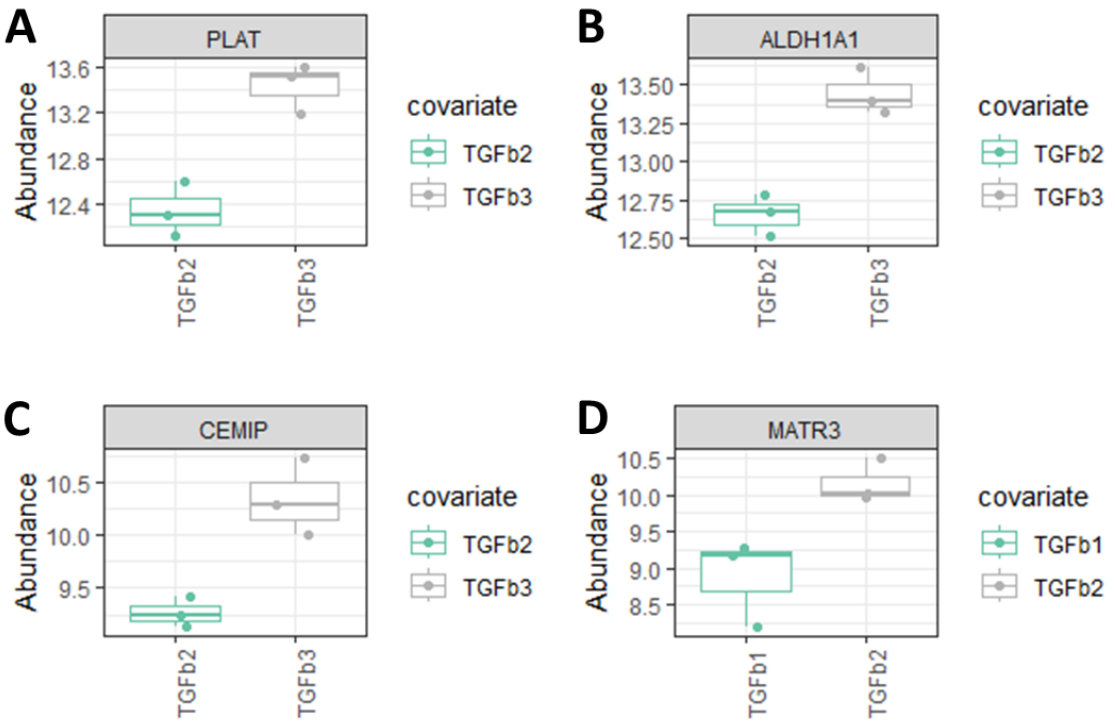


Fig 5.5.1.2 Gene expression from RNAseq of genes differentially expression after stimulation with different isoforms of TGFβ
(A) PLAT, (B) ALDH1A1 and (C) CEMIP in TGFβ2 and TGFβ3 stimulated CCD18-Co fibroblasts. (D) MATR3 in TGFβ1 and TGFβ2 stimulated Pericytes.

5.6 Conclusions from TGFβ Isoforms Comparison

No significant differences were observed between the effects of TGFβ1, 2 and 3 in all measures studied, on fibroblasts, T-cells and NK-cells. Many known effects of TGFβ1 have been successfully replicated, but the magnitude of these was similar for all isoforms. While there has been some interest in specific inhibition of TGFβ isoforms in CRC, these results do not support this approach. To only inhibit specific isoforms might still leave the tumour exposed to the detrimental effects of the remaining un-inhibited isoform, possibly as a result of redundancy.

5.7 Limitations and Future Work

The analysis of the effects of the TGFβ isoforms on PBMCs only looked at a few specific outcomes, future work in this area relevant to CRC should include looking at markers of CD8 cytotoxic potential, such as production of Granzyme B or IFNg. An unbiased approach such as RNA-sequencing or proteomics would provide more conclusive evidence of whether the isoforms do have differing roles in these cells.

While the study of the TGFβ isoforms in fibroblasts was broader, utilising RNAseq across cell lines, tumour and normal adjacent cells, differences between the effects of the isoforms may remain. Firstly, as well as altering the transcriptional program within a cell, TGFβ also signals through non-classical pathways, impacting on the activation status of many existing proteins present in the cell cytoplasm, these changes would not be detected through RNAseq. Secondly, growing fibroblasts in a monolayer is known to convert them to a more myofibroblast like phenotype[98], which may mask TGFβ effects. Thirdly, stimulation with one isoform of TGFβ can lead to effects on the expression of the other isoforms, and therefore these experiments were not necessarily analysing the effect of one isoform in isolation. However, over the short time periods used in most of these experiments, it is unlikely that significant production of the other isoforms would occur, and any TGFβ produced would be in a latent form and require activation. Fourthly, we studied the effects of TGFβ alone, in reality these cells would be responding to many cytokines and TGFβ may modify the response to these, therefore the isoforms could potentially have differing effects in altering the response to other cytokines.

6 Discussion

6.1 Summary

Colorectal cancer incidence is increasing in younger patients and CRC remains a major cause of cancer-related mortality[2]. High stromal content has been associated with worse outcomes in CRC, however, there is increasing awareness that CRC, like most cancer types, is a heterogenous disease and that successful new treatments will need to be targeted to the underlying cellular and molecular phenotype of the tumour. Many studies have found different clusters of fibroblasts on scRNAseq, though few have identified these *in situ* or correlated them with other tumour characteristics. Therefore, improved methods to stratify and treat CRC depending on the TME landscape are urgently needed.

In this thesis we aimed to develop a deeper understanding of fibroblasts within colorectal cancer. In particular to understand how different fibroblast subsets develop in response to cytokine stimulation, where they locate within the tumour and how these associate with other tumour characteristics such as immune infiltrate and treatment response.

From our experimental results, further analysis of existing datasets, and review of the research literature, we hypothesise that there are 3 main groups of fibroblasts in CRC, all of which most likely derive from the normal telocytes. We also hypothesise that there are a separate group of activated smooth-muscle cells, often confusingly described as myofibroblasts. Our work suggests that the relative abundance of these different stromal cell subsets will vary greatly between individuals with CRC, depending on the nature of the neoplastic cells themselves and the other cell types present in the tumour micro-environment.

6.2 CD56+ CAFs

The first group of CAFs, which we will term CD56+ CAFs, are a group most closely related to normal telocytes and are not myCAFs or iCAFs. They remain in close proximity to the neoplastic cells forming a border around the tumour nests. They continue to express some markers that are seen in normal telocytes, such as CXCL14, and additionally express markers only seen in disease such as CD56, which we predict are induced by the cancer cells themselves. Evidence for this is our finding of CD56 forming a ring around the tumour nests, that *NCAM1* is most often expressed in Pelka *et al* CXCL14+ cluster, and that they also found CXCL14+ fibroblasts in close association with epithelial cells. Our PCR data suggests that CD56 expression is not triggered by either IL-1β or TGFβ1, if anything these cytokines reduce its expression, and our RNAseq data shows that stimulation with these cytokines *in vitro* does not make fibroblasts map any more closely to the CXCL14+ cluster.

While the driver of CD56 upregulation on fibroblasts in cancer has not been determined, possibilities include an effect of direct contact with the cancer cells[195] or due to cytokine stimulation, for example, TNFα has been shown to upregulate NCAM on mesenchymal stem cells *in vitro*[196]. A recent preprint has identified the Hippo pathway in breast cancer cells as inducing *NCAM1* upregulation in αSMA+ CAFs (Moroishi *et al*, 2024). In this study the CD56+ fibroblasts also often expressed CXCL14+, concurring with the Pelka data showing NCAM expression in the CXCL14+ cluster.

The role CD56+ CAFs play in CRC is also to be determined. Moroishi *et al.* found that increased NCAM1+ αSMA+ CAFs secrete more TGFβ1 and 2 than other CAFs leading to T-cell exhaustion. Another study found that CD56 expression leads to increased cell motility in mesenchymal stem cells compared to knock-out[196]. The ability of mesenchymal cells to migrate could either lead the neoplastic cells to migrate with them directly, or may prepare the soil for future invasion by neoplastic cells. When CD56 is expressed on Schwann cells it is known to induce cancer cell invasion and metastasis. Direct cell-cell contact[195] between neoplastic cells and CD56 on the

Schwann cells leads to the cancer cells forming protrusions and moving towards the Schwann cells, and also detachment from other neoplastic cells due to intercalation of the cancer cells by Schwann cell protrusions. Whether cancer cells directly interact with and move towards fibroblasts expressing CD56 *in vivo* remains to be determined.

Alternatively, the ring of CD56+ fibroblasts around the tumour may be providing factors promoting survival and differentiation of the cancer cells, but actually limiting their ability to migrate and invade. In this scenario it may only be cancer cells that develop the potential to survive without these supporting cells that have the capacity to bud off the tumour nest and enter surrounding tissue. In this situation CD56+ fibroblasts may be partially protective, those neoplastic cells that are able to survive in their absence may migrate more rapidly if they were depleted.

We predict the CD56+ CAF subset to be present most frequently in CMS2 disease (Fig 6.2.1). We found the border of CD56+ CAFs around the tumour nests more frequently in CMS 2 and 3 tumours than 1 and 4. Additionally, in the Pelka *et al* dataset, a significantly higher proportion of fibroblasts and pericytes were CXCL14+ CAFs in MSS CMS2 tumours.

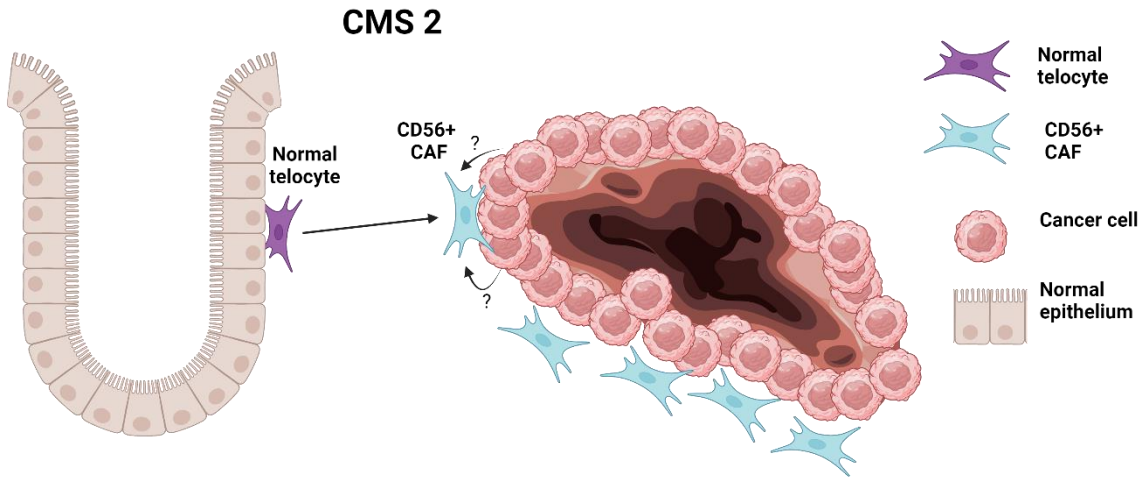


Fig 6.2.1 CMS2 CRC with CD56+ CAFs surrounding the tumour nests
Normal telocyte fibroblasts differentiate into CD56+ CXCL14+ CAFs as a result of direct contact with, or due to factors released from, the neoplastic cells. Therefore, these cells form a rim around the tumour nests. This occurs most frequently in CMS2 disease.

Recently CMS 2 and 3 tumours have been further subdivided into PDS 1 and 3 disease. The PDS1 tumours are faster growing and the PDS3 tumours less active, but the PDS3 tumours actually have worse prognosis. While we have not studied PDS group, we predict that these CAFs would be most commonly seen in the slow-growing CMS2-3, PDS3 group of tumours. We hypothesise that, as they are similar to normal telocytes, they help support survival of these slow-growing neoplastic cells which are most similar to normal colonic epithelial cells, and are probably protecting them from immune destruction and possibly bolstering a slow growing tumours' innate resistance to cancer treatment. We expect that faster-growing, more aggressive, PDS1 type neoplastic cells would no longer require the support of these telocyte-like cells for survival and invade more quickly than this layer of CAFs can form around them. If this were to prove to be the case, targeting CD56+ fibroblasts for cancer treatment in a PDS3 CRC may directly lead to death of the neoplastic cells and permit immune cell entry to the tumour, in cancers that are often resistant to conventional chemotherapy treatment.

6.3 iCAFs

The second group, inflammatory CAFs, differentiate in response to IL-1β. They are therefore found in close proximity to IL-1β producing cells, particularly neutrophils, and that as immune cells enter the tumour through the vasculature, are often found around the vessels. These cells then produce more cytokines and chemokines, recruiting more immune cells and potentiating the inflammatory nature of the tumour. We hypothesise that these CAFs are seen more often in tumours with a large immune infiltrate, such as CMS1 and CMS3 tumours **(Fig 6.3.1),** which are considered to be one group of CRCs by Joanito *et al*[22].

Evidence for this are our findings on immunofluorescence staining of 1) iCAFs being found at higher levels in CMS3 tumours, 2) IL-6 positive stromal cells being seen more frequently in neutrophil high tumours, 3) these cells often being located close to neutrophils and to a lesser extent T-cells, and 4) iCAFs being seen in close proximity to vessels. Our data confirmed that IL-1β stimulated primary stromal cells map more closely to Pelka *et als.* MMP3+ CAFs. Pelka *et al.* also saw MMP3+ CAFs in close proximity to tumour vessels and at the luminal border, where a higher concentration of vessels are found[180]. There may be more neutrophils at this location due to increased bacterial infiltration from the gut lumen.

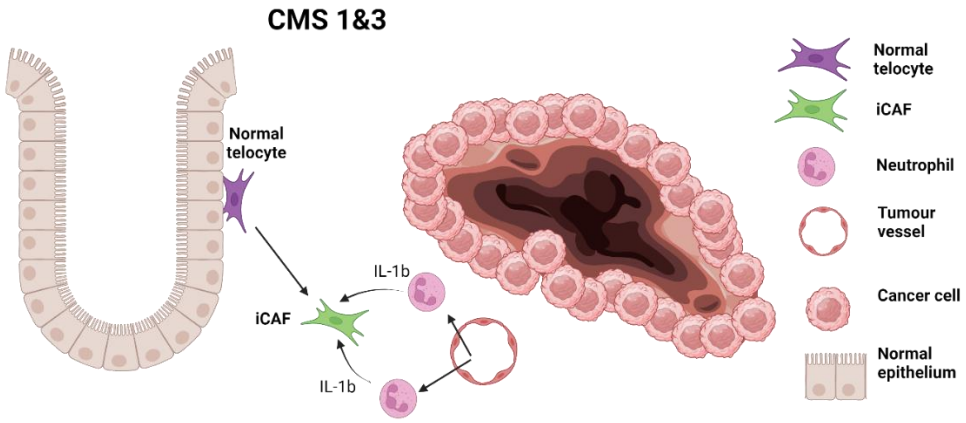


Fig 6.3.1 CMS1&3 CRC with iCAFs associated with neutrophils and vessels
Normal telocyte fibroblasts differentiate into iCAFs on exposure to IL-1β , produced by neutrophils entering the tumour from the vessels. This occurs most frequently in CMS1 and 3 CRC.

6.4 GREM1+ CAFs

We hypothesise that these cells are also derived from telocytes and are CAFs responding to TGFβ, which is present throughout the TME, but without the signals leading to CD56+ CAF or iCAF differentiation **(Fig 6.4.1)**. Evidence for this is that TGFβ stimulation of our primary cells results in them mapping more closely to the Pelka *et al* GREM1+ CAF subset.

We expect that the CD56+, GREM1+ and inflammatory CAF subsets are probably temporary responses to the local cytokine and surrounding cells. Therefore, these phenotypes are probably interchangeable as conditions within the TME change. Evidence for this is that our tumour stromal cells did not differ from normal adjacent stromal cells, suggesting any differences they may have had *in vitro* are lost as soon as they are transferred to a new environment.

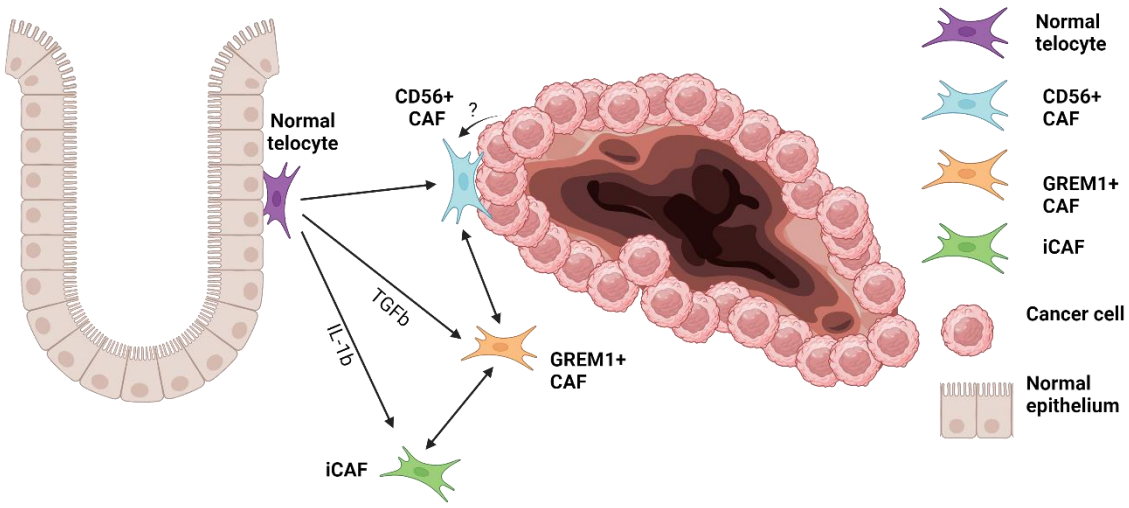


Fig 6.4.1 GREM1+ CAFs induced by TGFβ
Normal telocyte fibroblasts differentiate into GREM1+ CAFs on exposure to TGFβ, found at high levels within the TME. This phenotype can change to that of a CD56+ CAF or iCAF, if the cells are exposed to different tumour niches or cytokines within the microenvironment.

6.5 myCAF-like stromal cells

The final group, myCAF-like stromal cells, are found deep within the stromal tracts of the tumour and where the tumour invades the underlying muscle, the muscularis mucosa and muscularis propria. They express contractile gene expression programmes and smooth muscle markers such as *MYH11, TAGLN* and *CNN1*, but do not express fibroblast markers such as *PDGFRA* and *PDPN*. We hypothesise that, rather than being fibroblasts, these are smooth muscle cells responding to the high levels of TGFβ seen within the TME. The TGFβ within a tumour comes largely from the endothelial cells, fibroblasts, pericytes and macrophages, which we predict enter the smooth muscle layer alongside the invasive neoplastic cells. **(Fig 6.5.1)**.

Evidence for this is that we found myCAF-like stromal cells more frequently in CMS4 disease than other subtypes, and they were found further away from neoplastic cells, immune cells and vessels on immunohistofluorescence imaging, deep in connective tissue. Primary stromal cells stimulated with TGFβ did map more closely to the Pelka *et al* myofibroblast cluster. CMS4 is associated with poorer prognosis, although it is unclear whether this is causal; the myCAF-like stromal cells in CMS4 could be contributing to greater invasion and metastases, but, alternatively, the finding of more myCAF-like stromal cells in the tumour may be an effect of a tumour of any underlying CMS group having already invaded deeper into the underlying muscle layer and therefore more smooth muscle cells being sampled. Of note, Joanito *et al* described all tumours as an underlying CMS2 or CMS1/3, and described either of these as having the potential to become stromal high.

The current nomenclature is very confusing, as while many studies use the term 'myofibroblast' different criteria are used for defining this group. In some studies fibroblast markers are seen in the myofibroblast group, in others the cells described as myCAFs are probably smooth muscle or pericytes. Defined criteria will be needed in the field to be able to draw clear, direct comparisons between studies.

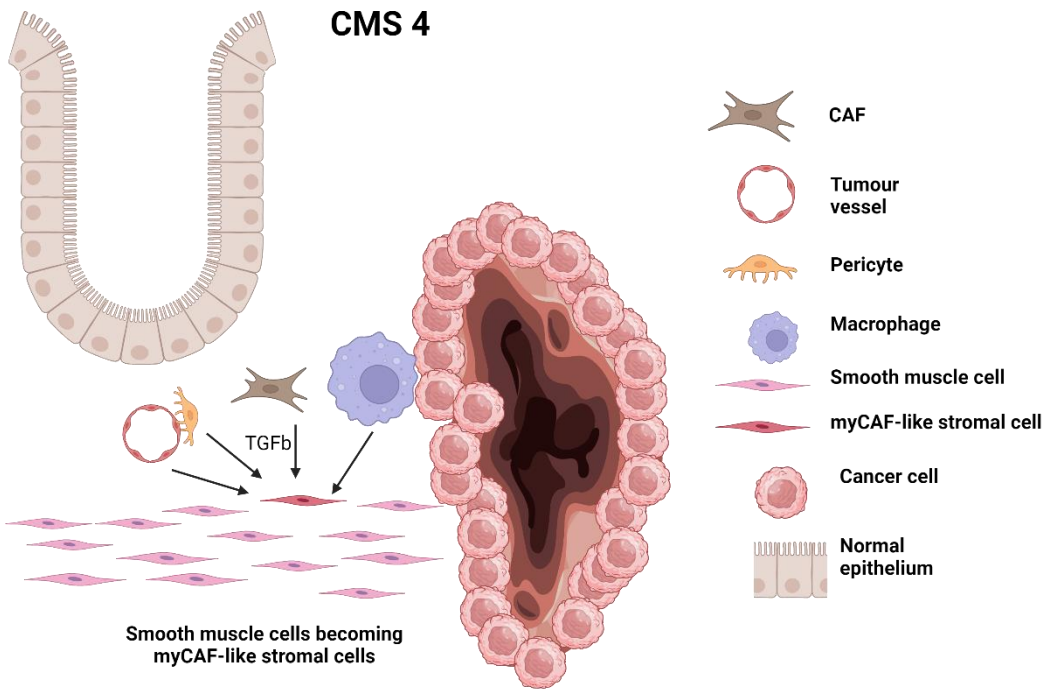


Fig 6.5.1 CMS4 CRC with myCAF-like stromal cells derived from smooth muscle
Smooth muscle cells become myCAF-like stromal cells as a response to high concentrations of TGFβ; these are most often found in CMS 4 disease where the tumour has invaded more deeply, into the muscularis mucosa and muscularis propria. The main producers of TGFβ in the TME are endothelial cells, pericytes, fibroblasts and macrophages.

6.6 Further Work

We would like to explore this hypothesis further by performing hyperplex imaging on CRC sections to confirm cell type identify of all four of these stromal subsets. We plan to have at least two markers of fibroblasts, PDPN and PDGFRα, alongside smooth muscle markers such as MYH11, and pericyte markers such as RGS5. We will also have at least one marker for each cancer associated stromal subset. We would also like to confirm co-localisation between the specific fibroblast subsets and other cell types, such as neutrophils, across the same slides, and therefore stain for immune markers, such as CD45, CD66b and CD3, on the same slide.

We would like to expand to a larger dataset, to test our hypothesis that specific fibroblast subsets are associated with CMS groups. This could be done through exploration of other existing scRNAseq datasets in CRC. It would also be interesting to study whether any of these stromal cell types are associated with PDS groupings or outcomes for patients.

We would also like to investigate how CD56 expression is induced in fibroblasts around the epithelium, the functional role of this fibroblast subsets and how it impacts upon disease course. The first of these questions could be addressed in vitro, measuring CD56 expression on fibroblasts after exposure to conditioned supernatants from CRC cell lines or on co-culture with CRC cell lines. Invasion assays could be performed with CRC cell lines grown with fibroblasts with and without CD56 upregulation.

6.7 Implications for Treatment

The finding that the fibroblast subsets are seen at different frequencies in the CMS groups could guide treatment targeting these cells. The CMS1/3 tumours enriched for iCAFs may respond to treatment with anti-IL-1 such as Anakinra. Logically, we would assume that the more invasive CMS4 tumours may be more likely to respond to anti-TGFβ treatments, although where tumours are labelled CMS4 as a results of sampling bias they may not be any more likely to respond than other subsets. It is currently unclear what impact the CD56+ fibroblasts, seen most frequently in CMS2 disease, are having on prognosis, and therefore whether they are a target for cancer treatment.

6.8 Conclusions

Many trials of agents aiming to disrupt the detrimental role of fibroblasts in CRC are ongoing. It will be important to characterise the stromal landscape at baseline in biopsies from patients enrolled in these trials, to understand which tumour types are responding to treatment. It will also be important to understand how the tumour microenvironment alters in response to treatment, to understand resistance mechanisms and to identify early biomarkers of which patients may get durable benefit.

7 References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Sinicrope FA. Increasing Incidence of Early-Onset Colorectal Cancer. N Engl J Med 2022;386:1547-1558.
3. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin 2018;68:250-281.
4. Jess T, Rungoe C, Peyrin-Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clin Gastroenterol Hepatol 2012;10:639-45.
5. Gehart H, Clevers H. Tales from the crypt: new insights into intestinal stem cells. Nat Rev Gastroenterol Hepatol 2019;16:19-34.
6. Weiser MR. AJCC 8th Edition: Colorectal Cancer. Ann Surg Oncol 2018;25:1454-1455.
7. Akkoca AN, Yanık S, Ozdemir ZT, et al. TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma. Int J Clin Exp Med 2014;7:2828-35.
8. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759-67.
9. Yamane L, Scapulatempo-Neto C, Reis RM, et al. Serrated pathway in colorectal carcinogenesis. World J Gastroenterol 2014;20:2634-40.
10. Pai RK, Bettington M, Srivastava A, et al. An update on the morphology and molecular pathology of serrated colorectal polyps and associated carcinomas. Mod Pathol 2019;32:1390-1415.
11. Williams MJ, Werner B, Barnes CP, et al. Identification of neutral tumor evolution across cancer types. Nat Genet 2016;48:238-244.
12. Sottoriva A, Kang H, Ma Z, et al. A Big Bang model of human colorectal tumor growth. Nat Genet 2015;47:209-16.
13. Clevers H, Nusse R. Wnt/β-catenin signaling and disease. Cell 2012;149:1192-205.
14. Fang JY, Richardson BC. The MAPK signalling pathways and colorectal cancer. Lancet Oncol 2005;6:322-7.
15. Saeed H, Leibowitz BJ, Zhang L, et al. Targeting Myc-driven stress addiction in colorectal cancer. Drug Resist Updat 2023;69:100963.
16. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther 2013;7:57-68.
17. Arakawa H. Netrin-1 and its receptors in tumorigenesis. Nat Rev Cancer 2004;4:978-87.
18. Stoffel EM, Koeppe E, Everett J, et al. Germline Genetic Features of Young Individuals With Colorectal Cancer. Gastroenterology 2018;154:897-905.e1.
19. Win AK, Jenkins MA, Dowty JG, et al. Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. Cancer Epidemiol Biomarkers Prev 2017;26:404-412.
20. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350-6.
21. La Vecchia S, Sebastián C. Metabolic pathways regulating colorectal cancer initiation and progression. Semin Cell Dev Biol 2020;98:63-70.
22. Joanito I, Wirapati P, Zhao N, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. Nat Genet 2022.
23. Marisa L, Blum Y, Taieb J, et al. Intratumor CMS Heterogeneity Impacts Patient Prognosis in Localized Colon Cancer. Clin Cancer Res 2021;27:4768-4780.
24. Khaliq AM, Erdogan C, Kurt Z, et al. Refining colorectal cancer classification and clinical stratification through a single-cell atlas. Genome Biol 2022;23:113.

25. Brouwer NP, Webbink L, Haddad TS, et al. Transcriptomics and proteomics reveal distinct biology for lymph node metastases and tumour deposits in colorectal cancer. J Pathol 2023.
26. Dunne PD, McArt DG, Bradley CA, et al. Challenging the Cancer Molecular Stratification Dogma: Intratumoral Heterogeneity Undermines Consensus Molecular Subtypes and Potential Diagnostic Value in Colorectal Cancer. Clin Cancer Res 2016;22:4095-104.
27. Malla SB, Byrne RM, Lafarge MW, et al. Pathway level subtyping identifies a slow-cycling biological phenotype associated with poor clinical outcomes in colorectal cancer. Nat Genet 2024;56:458-472.
28. Eide PW, Bruun J, Lothe RA, et al. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Sci Rep 2017;7:16618.
29. Isella C, Brundu F, Bellomo SE, et al. Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. Nat Commun 2017;8:15107.
30. Lao VV, Grady WM. Epigenetics and colorectal cancer. Nat Rev Gastroenterol Hepatol 2011;8:686-700.
31. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330-7.
32. Fleming M, Ravula S, Tatishchev SF, et al. Colorectal carcinoma: Pathologic aspects. Journal of Gastrointestinal Oncology; Vol 3, No 3 (September 01, 2012): Journal of Gastrointestinal Oncology (Current Topics in Gastrointestinal and Liver Pathology) 2012.
33. Budinská E, Hrivňáková M, Ivkovic TC, et al. Molecular portraits of colorectal cancer morphological regions. Elife 2023;12.
34. Sirinukunwattana K, Domingo E, Richman SD, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut 2021;70:544-554.
35. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet 2018;391:2128-2139.
36. Maoz A, Dennis M, Greenson JK. The Crohn's-Like Lymphoid Reaction to Colorectal Cancer-Tertiary Lymphoid Structures With Immunologic and Potentially Therapeutic Relevance in Colorectal Cancer. Front Immunol 2019;10:1884.
37. Labianca R, Nordlinger B, Beretta GD, et al. Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2013;24 Suppl 6:vi64-72.
38. Sargent D, Sobrero A, Grothey A, et al. Evidence for cure by adjuvant therapy in colon cancer: observations based on individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol 2009;27:872-7.
39. André T, Meyerhardt J, Iveson T, et al. Effect of duration of adjuvant chemotherapy for patients with stage III colon cancer (IDEA collaboration): final results from a prospective, pooled analysis of six randomised, phase 3 trials. Lancet Oncol 2020;21:1620-1629.
40. Cervantes A, Adam R, Roselló S, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2023;34:10-32.
41. Magrin L, Fanale D, Brando C, et al. POLE, POLD1, and NTHL1: the last but not the least hereditary cancer-predisposing genes. Oncogene 2021;40:5893-5901.
42. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509-20.
43. Martini G, Dienstmann R, Ros J, et al. Molecular subtypes and the evolution of treatment management in metastatic colorectal cancer. Ther Adv Med Oncol 2020;12:1758835920936089.
44. Kopetz S, Grothey A, Yaeger R, et al. Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer. N Engl J Med 2019;381:1632-1643.

45. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
46. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell 2010;140:883-99.
47. Colotta F, Allavena P, Sica A, et al. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis 2009;30:1073-81.
48. Gasteiger G, D'Osualdo A, Schubert DA, et al. Cellular Innate Immunity: An Old Game with New Players. J Innate Immun 2017;9:111-125.
49. Pang Y, Gara SK, Achyut BR, et al. TGF-beta signaling in myeloid cells is required for tumor metastasis. Cancer Discov 2013;3:936-51.
50. Nandan D, Reiner NE. TGF-beta attenuates the class II transactivator and reveals an accessory pathway of IFN-gamma action. J Immunol 1997;158:1095-101.
51. Papaspyridonos M, Matei I, Huang Y, et al. Id1 suppresses anti-tumour immune responses and promotes tumour progression by impairing myeloid cell maturation. Nat Commun 2015;6:6840.
52. Jackstadt R, van Hooff SR, Leach JD, et al. Epithelial NOTCH Signaling Rewires the Tumor Microenvironment of Colorectal Cancer to Drive Poor-Prognosis Subtypes and Metastasis. Cancer Cell 2019;36:319-336.e7.
53. Donatelli SS, Zhou JM, Gilvary DL, et al. TGF-β-inducible microRNA-183 silences tumor-associated natural killer cells. Proc Natl Acad Sci U S A 2014;111:4203-8.
54. Yu J, Wei M, Becknell B, et al. Pro- and antiinflammatory cytokine signaling: reciprocal antagonism regulates interferon-gamma production by human natural killer cells. Immunity 2006;24:575-90.
55. Castriconi R, Cantoni C, Della Chiesa M, et al. Transforming growth factor beta 1 inhibits expression of NKp30 and NKG2D receptors: consequences for the NK-mediated killing of dendritic cells. Proc Natl Acad Sci U S A 2003;100:4120-5.
56. Lee JC, Lee KM, Kim DW, et al. Elevated TGF-beta1 secretion and down-modulation of NKG2D underlies impaired NK cytotoxicity in cancer patients. J Immunol 2004;172:7335-40.
57. Thomas DA, Massagué J. TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. Cancer Cell 2005;8:369-80.
58. Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. Science 1996;271:1734-6.
59. Ribas A. Releasing the Brakes on Cancer Immunotherapy. N Engl J Med 2015;373:1490-2.
60. Littman DR. Releasing the Brakes on Cancer Immunotherapy. Cell 2015;162:1186-90.
61. Borgeaud M, Sandoval J, Obeid M, et al. Novel targets for immune-checkpoint inhibition in cancer. Cancer Treat Rev 2023;120:102614.
62. DuPage M, Bluestone JA. Harnessing the plasticity of CD4(+) T cells to treat immune-mediated disease. Nat Rev Immunol 2016;16:149-63.
63. Batlle E, Massague J. Transforming Growth Factor-beta Signaling in Immunity and Cancer. Immunity 2019;50:924-940.
64. Zhou L, Lopes JE, Chong MM, et al. TGF-beta-induced Foxp3 inhibits T(H)17 cell differentiation by antagonizing RORgammat function. Nature 2008;453:236-40.
65. Calon A, Lonardo E, Berenguer-Llergo A, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet 2015;47:320-9.
66. Isella C, Terrasi A, Bellomo SE, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet 2015;47:312-9.
67. Fitzgerald AA, Weiner LM. The role of fibroblast activation protein in health and malignancy. Cancer Metastasis Rev 2020;39:783-803.
68. Khaliq AM, Kurt Z, Grunvald MW, et al. Redefining colorectal cancer classification and clinical stratification through a single-cell atlas. bioRxiv 2021:2021.02.02.429256.
69. Kraman M, Bambrough PJ, Arnold JN, et al. Suppression of antitumor immunity by stromal cells expressing fibroblast activation protein-alpha. Science 2010;330:827-30.

70. Santos AM, Jung J, Aziz N, et al. Targeting fibroblast activation protein inhibits tumor stromagenesis and growth in mice. J Clin Invest 2009;119:3613-25.
71. Ostermann E, Garin-Chesa P, Heider KH, et al. Effective immunoconjugate therapy in cancer models targeting a serine protease of tumor fibroblasts. Clin Cancer Res 2008;14:4584-92.
72. Fang J, Xiao L, Joo KI, et al. A potent immunotoxin targeting fibroblast activation protein for treatment of breast cancer in mice. Int J Cancer 2016;138:1013-23.
73. Loeffler M, Krüger JA, Niethammer AG, et al. Targeting tumor-associated fibroblasts improves cancer chemotherapy by increasing intratumoral drug uptake. J Clin Invest 2006;116:1955-62.
74. Duperret EK, Trautz A, Ammons D, et al. Alteration of the Tumor Stroma Using a Consensus DNA Vaccine Targeting Fibroblast Activation Protein (FAP) Synergizes with Antitumor Vaccine Therapy in Mice. Clin Cancer Res 2018;24:1190-1201.
75. Lo A, Wang LS, Scholler J, et al. Tumor-Promoting Desmoplasia Is Disrupted by Depleting FAP-Expressing Stromal Cells. Cancer Res 2015;75:2800-2810.
76. Wang LC, Lo A, Scholler J, et al. Targeting fibroblast activation protein in tumor stroma with chimeric antigen receptor T cells can inhibit tumor growth and augment host immunity without severe toxicity. Cancer Immunol Res 2014;2:154-66.
77. Kakarla S, Chow KK, Mata M, et al. Antitumor effects of chimeric receptor engineered human T cells directed to tumor stroma. Mol Ther 2013;21:1611-20.
78. Scott AM, Wiseman G, Welt S, et al. A Phase I dose-escalation study of sibrotuzumab in patients with advanced or metastatic fibroblast activation protein-positive cancer. Clin Cancer Res 2003;9:1639-47.
79. Hofheinz RD, al-Batran SE, Hartmann F, et al. Stromal antigen targeting by a humanised monoclonal antibody: an early phase II trial of sibrotuzumab in patients with metastatic colorectal cancer. Onkologie 2003;26:44-8.
80. Narra K, Mullins SR, Lee HO, et al. Phase II trial of single agent Val-boroPro (Talabostat) inhibiting Fibroblast Activation Protein in patients with metastatic colorectal cancer. Cancer Biol Ther 2007;6:1691-9.
81. Roberts EW, Deonarine A, Jones JO, et al. Depletion of stromal cells expressing fibroblast activation protein-α from skeletal muscle and bone marrow results in cachexia and anemia. J Exp Med 2013;210:1137-51.
82. Tran E, Chinnasamy D, Yu Z, et al. Immune targeting of fibroblast activation protein triggers recognition of multipotent bone marrow stromal cells and cachexia. J Exp Med 2013;210:1125-35.
83. Özdemir BC, Pentcheva-Hoang T, Carstens JL, et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. Cancer Cell 2014;25:719-34.
84. Chen Y, Kim J, Yang S, et al. Type I collagen deletion in αSMA(+) myofibroblasts augments immune suppression and accelerates progression of pancreatic cancer. Cancer Cell 2021;39:548-565.e6.
85. Friedrich M, Pohin M, Jackson MA, et al. IL-1-driven stromal-neutrophil interactions define a subset of patients with inflammatory bowel disease that does not respond to therapies. Nat Med 2021;27:1970-1981.
86. Kinchen J, Chen HH, Parikh K, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. Cell 2018;175:372-386.e17.
87. Kalluri R. The biology and function of fibroblasts in cancer. Nat Rev Cancer 2016;16:582-98.
88. Korsunsky I, Wei K, Pohin M, et al. Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. bioRxiv 2021:2021.01.11.426253.
89. Friedrich M, Pohin M, Jackson MA, et al. IL-1-driven stromal-neutrophil interaction in deep ulcers defines a pathotype of therapy non-responsive inflammatory bowel disease. bioRxiv 2021:2021.02.05.429804.

90. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 2019;178:714-730.e22.
91. Buechler MB, Pradhan RN, Krishnamurty AT, et al. Cross-tissue organization of the fibroblast lineage. Nature 2021;593:575-579.
92. Croft AP, Campos J, Jansen K, et al. Distinct fibroblast subsets drive inflammation and damage in arthritis. Nature 2019;570:246-251.
93. Dominguez CX, Müller S, Keerthivasan S, et al. Single-Cell RNA Sequencing Reveals Stromal Evolution into LRRC15(+) Myofibroblasts as a Determinant of Patient Response to Cancer Immunotherapy. Cancer Discov 2020;10:232-253.
94. Grauel AL, Nguyen B, Ruddy D, et al. TGFβ-blockade uncovers stromal plasticity in tumors by revealing the existence of a subset of interferon-licensed fibroblasts. Nat Commun 2020;11:6315.
95. Kieffer Y, Hocine HR, Gentric G, et al. Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. Cancer Discov 2020;10:1330-1351.
96. Hanley C, Waise S, Parker R, et al. Spatially discrete signalling niches regulate fibroblast heterogeneity in human lung cancer. bioRxiv 2020:2020.06.08.134270.
97. Elyada E, Bolisetty M, Laise P, et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov 2019;9:1102-1123.
98. Öhlund D, Handly-Santana A, Biffi G, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med 2017;214:579-596.
99. Costa A, Kieffer Y, Scholer-Dahirel A, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer Cell 2018;33:463-479.e10.
100. Su S, Chen J, Yao H, et al. CD10(+)GPR77(+) Cancer-Associated Fibroblasts Promote Cancer Formation and Chemoresistance by Sustaining Cancer Stemness. Cell 2018;172:841-856.e16.
101. Li H, Courtois ET, Sengupta D, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 2017;49:708-718.
102. Lee HO, Hong Y, Etlioglu HE, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet 2020;52:594-603.
103. Pelka K, Hofree M, Chen JH, et al. Spatially organized multicellular immune hubs in human colorectal cancer. Cell 2021;184:4734-4752.e20.
104. Wang F, Long J, Li L, et al. Single-cell and spatial transcriptome analysis reveals the cellular heterogeneity of liver metastatic colorectal cancer. Sci Adv 2023;9:eadf5464.
105. Sahai E, Astsaturov I, Cukierman E, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer 2020;20:174-186.
106. Chen Z, Zhou L, Liu L, et al. Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. Nat Commun 2020;11:5077.
107. Mellone M, Hanley CJ, Thirdborough S, et al. Induction of fibroblast senescence generates a non-fibrogenic myofibroblast phenotype that differentially impacts on cancer prognosis. Aging (Albany NY) 2016;9:114-132.
108. Demaria M, O'Leary MN, Chang J, et al. Cellular Senescence Promotes Adverse Effects of Chemotherapy and Cancer Relapse. Cancer Discov 2017;7:165-176.
109. Grout JA, Sirven P, Leader AM, et al. Spatial positioning and matrix programs of cancer-associated fibroblasts promote T cell exclusion in human lung tumors. bioRxiv 2022:2022.01.20.476763.
110. Kobayashi H, Gieniec KA, Wright JA, et al. The Balance of Stromal BMP Signaling Mediated by GREM1 and ISLR Drives Colorectal Carcinogenesis. Gastroenterology 2021;160:1224-1239.e30.

111. Orimo A, Gupta PB, Sgroi DC, et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. Cell 2005;121:335-48.
112. Calon A, Espinet E, Palomo-Ponce S, et al. Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. Cancer Cell 2012;22:571-84.
113. Biffi G, Oni TE, Spielman B, et al. IL1-Induced JAK/STAT Signaling Is Antagonized by TGFβ to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma. Cancer Discov 2019;9:282-301.
114. Nicolas AM, Pesic M, Engel E, et al. Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. Cancer Cell 2022;40:168-184.e13.
115. Tjomsland V, Spångeus A, Välilä J, et al. Interleukin 1α sustains the expression of inflammatory factors in human pancreatic cancer microenvironment by targeting cancer-associated fibroblasts. Neoplasia 2011;13:664-75.
116. Lopez-Castejon G, Brough D. Understanding the mechanism of IL-1β secretion. Cytokine Growth Factor Rev 2011;22:189-95.
117. Wang Y, Wang K, Han GC, et al. Neutrophil infiltration favors colitis-associated tumorigenesis by activating the interleukin-1 (IL-1)/IL-6 axis. Mucosal Immunol 2014;7:1106-15.
118. Janney A, Powrie F, Mann EH. Host-microbiota maladaptation in colorectal cancer. Nature 2020;585:509-517.
119. Eigenbrod T, Park JH, Harder J, et al. Cutting edge: critical role for mesothelial cells in necrosis-induced inflammation through the recognition of IL-1 alpha released from dying cells. J Immunol 2008;181:8194-8.
120. Weber A, Wasiliew P, Kracht M. Interleukin-1 (IL-1) pathway. Sci Signal 2010;3:cm1.
121. Bellone G, Carbone A, Tibaudi D, et al. Differential expression of transforming growth factors-beta1, -beta2 and -beta3 in human colon carcinoma. Eur J Cancer 2001;37:224-33.
122. Pasche B. Role of transforming growth factor beta in cancer. J Cell Physiol 2001;186:153-68.
123. Medici D, Hay ED, Olsen BR. Snail and Slug promote epithelial-mesenchymal transition through beta-catenin-T-cell factor-4-dependent expression of transforming growth factor-beta3. Mol Biol Cell 2008;19:4875-87.
124. Liu YP, Zhu HF, Liu DL, et al. DcR3 induces epithelial-mesenchymal transition through activation of the TGF-beta3/SMAD signaling pathway in CRC. Oncotarget 2016;7:77306-77318.
125. Achyut BR, Bader DA, Robles AI, et al. Inflammation-mediated genetic and epigenetic alterations drive cancer development in the neighboring epithelium upon stromal abrogation of TGF-β signaling. PLoS Genet 2013;9:e1003251.
126. Morris JC, Tan AR, Olencki TE, et al. Phase I study of GC1008 (fresolimumab): a human anti-transforming growth factor-beta (TGFbeta) monoclonal antibody in patients with advanced malignant melanoma or renal cell carcinoma. PLoS One 2014;9:e90353.
127. Keedy VL, Bauer TM, Clarke JM, et al. Association of TGF-β responsive signature with anti-tumor effect of vactosertib, a potent, oral TGF-β receptor type I (TGFBRI) inhibitor in patients with advanced solid tumors. 2018;36:3031-3031.
128. Melisi D, Garcia-Carbonero R, Macarulla T, et al. Galunisertib plus gemcitabine vs. gemcitabine for first-line treatment of patients with unresectable pancreatic cancer. Br J Cancer 2018;119:1208-1214.
129. Holmgaard RB, Schaer DA, Li Y, et al. Targeting the TGFbeta pathway with galunisertib, a TGFbetaRI small molecule inhibitor, promotes anti-tumor immunity leading to durable, complete responses, as monotherapy and in combination with checkpoint blockade. J Immunother Cancer 2018;6:47.
130. Mariathasan S, Turley SJ, Nickles D, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 2018;554:544-548.

131. Vanpouille-Box C, Diamond JM, Pilones KA, et al. TGFβ Is a Master Regulator of Radiation Therapy-Induced Antitumor Immunity. Cancer Res 2015;75:2232-42.
132. Metelli A, Wu BX, Fugle CW, et al. Surface Expression of TGFβ Docking Receptor GARP Promotes Oncogenesis and Immune Tolerance in Breast Cancer. 2016;76:7106-7117.
133. Eberlein C, Kendrew J, McDaid K, et al. A human monoclonal antibody 264RAD targeting αvβ6 integrin reduces tumour growth and metastasis, and modulates key biomarkers in vivo. Oncogene 2013;32:4406-16.
134. Morris JC, Tan AR, Olencki TE, et al. Phase I study of GC1008 (fresolimumab): a human anti-transforming growth factor-beta (TGFβ) monoclonal antibody in patients with advanced malignant melanoma or renal cell carcinoma. PLoS One 2014;9:e90353.
135. Strauss J, Heery CR, Schlom J, et al. Phase I Trial of M7824 (MSB0011359C), a Bifunctional Fusion Protein Targeting PD-L1 and TGFβ, in Advanced Solid Tumors. Clin Cancer Res 2018;24:1287-1295.
136. Formenti SC, Lee P, Adams S, et al. Focal Irradiation and Systemic TGFbeta Blockade in Metastatic Breast Cancer. Clin Cancer Res 2018;24:2493-2504.
137. Anderton MJ, Mellor HR, Bell A, et al. Induction of heart valve lesions by small-molecule ALK5 inhibitors. Toxicol Pathol 2011;39:916-24.
138. Elliott RL, Blobe GC. Role of transforming growth factor Beta in human cancer. J Clin Oncol 2005;23:2078-93.
139. Shull MM, Ormsby I, Kier AB, et al. Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease. Nature 1992;359:693-9.
140. Sanford LP, Ormsby I, Gittenberger-de Groot AC, et al. TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. Development 1997;124:2659-70.
141. Proetzel G, Pawlowski SA, Wiles MV, et al. Transforming growth factor-beta 3 is required for secondary palate fusion. Nat Genet 1995;11:409-14.
142. Chia ZJ, Cao YN, Little PJ, et al. Transforming growth factor-β receptors: versatile mechanisms of ligand activation. Acta Pharmacol Sin 2024;45:1337-1348.
143. Lee Y, Awasthi A, Yosef N, et al. Induction and molecular signature of pathogenic TH17 cells. Nat Immunol 2012;13:991-9.
144. Cheifetz S, Weatherbee JA, Tsang ML, et al. The transforming growth factor-beta system, a complex pattern of cross-reactive ligands and receptors. Cell 1987;48:409-15.
145. Massagué J. The transforming growth factor-beta family. Annu Rev Cell Biol 1990;6:597-641.
146. Riemenschneider MJ, Hirblinger M, Vollmann-Zwerenz A, et al. TGF-ß isoforms in cancer: Immunohistochemical expression and Smad-pathway-activity-analysis in thirteen major tumor types with a critical appraisal of antibody specificity and immunohistochemistry assay validity. Oncotarget 2015;6:26770-81.
147. Zhang S. The role of transforming growth factor beta in T helper 17 differentiation. Immunology 2018;155:24-35.
148. Bhattacharya A, Al-Sammarraie N, Gebere MG, et al. Myocardial TGFβ2 Is Required for Atrioventricular Cushion Remodeling and Myocardial Development. J Cardiovasc Dev Dis 2021;8.
149. Herbertz S, Sawyer JS, Stauber AJ, et al. Clinical development of galunisertib (LY2157299 monohydrate), a small molecule inhibitor of transforming growth factor-beta signaling pathway. Drug Des Devel Ther 2015;9:4479-99.
150. Bogdahn U, Hau P, Stockhammer G, et al. Targeted therapy for high-grade glioma with the TGF-β2 inhibitor trabedersen: results of a randomized and controlled phase IIb study. Neuro Oncol 2011;13:132-42.
151. Gurluler E, Tumay LV, Guner OS, et al. Oncostatin-M as a novel biomarker in colon cancer patients and its association with clinicopathologic variables. Eur Rev Med Pharmacol Sci 2014;18:2042-7.

152. Li B, Wang F, Ma C, et al. Predictive value of IL-18 and IL-10 in the prognosis of patients with colorectal cancer. Oncol Lett 2019;18:713-719.
153. Cui G, Yuan A, Pang Z, et al. Contribution of IL-33 to the Pathogenesis of Colorectal Cancer. Front Oncol 2018;8:561.
154. Joosten SPJ, Spaargaren M, Clevers H, et al. Hepatocyte growth factor/MET and CD44 in colorectal cancer: partners in tumorigenesis and therapy resistance. Biochim Biophys Acta Rev Cancer 2020;1874:188437.
155. Hanley CJ, Mellone M, Ford K, et al. Targeting the Myofibroblastic Cancer-Associated Fibroblast Phenotype Through Inhibition of NOX4. J Natl Cancer Inst 2018;110:109-20.
156. Ford K, Hanley CJ, Mellone M, et al. NOX4 Inhibition Potentiates Immunotherapy by Overcoming Cancer-Associated Fibroblast-Mediated CD8 T-cell Exclusion from Tumors. Cancer Res 2020;80:1846-1860.
157. Bechtel W, McGoohan S, Zeisberg EM, et al. Methylation determines fibroblast activation and fibrogenesis in the kidney. Nat Med 2010;16:544-50.
158. Hu M, Yao J, Cai L, et al. Distinct epigenetic changes in the stromal cells of breast cancers. Nat Genet 2005;37:899-905.
159. Jiang L, Gonda TA, Gamble MV, et al. Global hypomethylation of genomic DNA in cancer-associated myofibroblasts. Cancer Res 2008;68:9900-8.
160. Ling E, Ringel A, Sigal-Batikoff I, et al. Human Colorectal Cancer Stage-dependent Global DNA Hypomethylation of Cancer-associated Fibroblasts. Anticancer Res 2016;36:4503-7.
161. Mrazek AA, Carmical JR, Wood TG, et al. Colorectal Cancer-Associated Fibroblasts are Genotypically Distinct. Curr Cancer Ther Rev 2014;10:97-218.
162. Vizoso M, Puig M, Carmona FJ, et al. Aberrant DNA methylation in non-small cell lung cancer-associated fibroblasts. Carcinogenesis 2015;36:1453-63.
163. Koncina E, Nurmik M, Pozdeev VI, et al. IL1R1(+) cancer-associated fibroblasts drive tumor development and immunosuppression in colorectal cancer. Nat Commun 2023;14:4251.
164. Zhou Y, Bian S, Zhou X, et al. Single-Cell Multiomics Sequencing Reveals Prevalent Genomic Alterations in Tumor Stromal Cells of Human Colorectal Cancer. Cancer Cell 2020;38:818-828.e5.
165. Becker WR, Nevins SA, Chen DC, et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. Nat Genet 2022.
166. Nee K, Ma D, Nguyen QH, et al. Pre-neoplastic stromal cells drive BRCA1-mediated breast tumorigenesis. bioRxiv 2021:2021.10.20.465221.
167. Che LH, Liu JW, Huo JP, et al. A single-cell atlas of liver metastases of colorectal cancer reveals reprogramming of the tumor microenvironment in response to preoperative chemotherapy. Cell Discov 2021;7:80.
168. Giguelay A, Turtoi E, Khelaf L, et al. The landscape of cancer-associated fibroblasts in colorectal cancer liver metastases. Theranostics 2022;12:7624-7639.
169. Puram SV, Tirosh I, Parikh AS, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell 2017;171:1611-1624.e24.
170. Gong Z, Li Q, Shi J, et al. Lung fibroblasts facilitate pre-metastatic niche formation by remodeling the local immune microenvironment. Immunity 2022;55:1483-1500.e9.
171. Ye J, Guo W, Wang C, et al. Peritumoral Immune-suppressive Mechanisms Impede Intratumoral Lymphocyte Infiltration into Colorectal Cancer Liver versus Lung Metastases. Cancer Res Commun 2023;3:2082-2095.
172. Luca BA, Steen CB, Matusiak M, et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell 2021.
173. Lambrechts D, Wauters E, Boeckx B, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nat Med 2018;24:1277-1289.
174. Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 2016;352:189-96.

175. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, et al. The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. Cell 2020;181:236-249.
176. Qian J, Olbrecht S, Boeckx B, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res 2020;30:745-762.
177. Hanley CJ, Waise S, Ellis MJ, et al. Single-cell analysis reveals prognostic fibroblast subpopulations linked to molecular and immunological subtypes of lung cancer. Nat Commun 2023;14:387.
178. Fan J, Bellon M, Ju M, et al. Clinical significance of FBXW7 loss of function in human cancers. Mol Cancer 2022;21:87.
179. Germann M, Zangger N, Sauvain MO, et al. Neutrophils suppress tumor-infiltrating T cells in colon cancer via matrix metalloproteinase-mediated activation of TGFβ. EMBO Mol Med 2020;12:e10681.
180. Kather JN, Zöllner FG, Schad LR, et al. Identification of a characteristic vascular belt zone in human colorectal cancer. PLoS One 2017;12:e0171378.
181. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347:1260419.
182. van Esveld SL, Rodenburg RJ, Al-Murshedi F, et al. Mitochondrial RNA processing defect caused by a SUPV3L1 mutation in two siblings with a novel neurodegenerative syndrome. J Inherit Metab Dis 2022;45:292-307.
183. Komura M, Wang C, Ito S, et al. Simultaneous Expression of CD70 and POSTN in Cancer-Associated Fibroblasts Predicts Worse Survival of Colorectal Cancer Patients. Int J Mol Sci 2024;25.
184. Hattangady NG, Carter K, Maroni-Rana B, et al. Mapping the core senescence phenotype of primary human colon fibroblasts. Aging (Albany NY) 2024;16.
185. Pasvanis Z, Boynes A, Kong RCK, et al. Transcriptomic analysis of TGFβ-mediated fibrosis in primary human Tenon's fibroblasts. bioRxiv 2024:2024.03.09.583791.
186. Gorelik L, Flavell RA. Immune-mediated eradication of tumors through the blockade of transforming growth factor-β signaling in T cells. Nature Medicine 2001;7:1118-1122.
187. Gao Y, Souza-Fonseca-Guimaraes F, Bald T, et al. Tumor immunoevasion by the conversion of effector NK cells into type 1 innate lymphoid cells. Nat Immunol 2017;18:1004-1015.
188. Laouar Y, Sutterwala FS, Gorelik L, et al. Transforming growth factor-beta controls T helper type 1 cell development through regulation of natural killer cell interferon-gamma. Nat Immunol 2005;6:600-7.
189. Viel S, Marçais A, Guimaraes FS, et al. TGF-β inhibits the activation and functions of NK cells by repressing the mTOR pathway. Sci Signal 2016;9:ra19.
190. Kottler UB, Jünemann AG, Aigner T, et al. Comparative effects of TGF-beta 1 and TGF-beta 2 on extracellular matrix production, proliferation, migration, and collagen contraction of human Tenon's capsule fibroblasts in pseudoexfoliation and primary open-angle glaucoma. Exp Eye Res 2005;80:121-34.
191. Hawinkels LJ, Paauwe M, Verspaget HW, et al. Interaction with colon cancer cells hyperactivates TGF-β signaling in cancer-associated fibroblasts. Oncogene 2014;33:97-107.
192. Bhowmick NA, Chytil A, Plieth D, et al. TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. Science 2004;303:848-51.
193. van der Waals LM, Borel Rinkes IHM, Kranenburg O. ALDH1A1 expression is associated with poor differentiation, 'right-sidedness' and poor survival in human colorectal cancer. PLoS One 2018;13:e0205536.
194. Xu G, Zhao L, Hua Q, et al. CEMIP, acting as a scaffold protein for bridging GRAF1 and MIB1, promotes colorectal cancer metastasis via activating CDC42/MAPK pathway. Cell Death Dis 2023;14:167.
195. Deborde S, Omelchenko T, Lyubchik A, et al. Schwann cells induce cancer cell dispersion and invasion. J Clin Invest 2016;126:1538-54.

196. Shi Y, Xia YY, Wang L, et al. Neural cell adhesion molecule modulates mesenchymal stromal cell migration via activation of MAPK/ERK signaling. Exp Cell Res 2012;318:2257-67.

8 Appendices

8.1 Appendix 1

Setting for Fibroblasts subsets panel using HALO HighPlex FL v.4.2.3

Nuclear detector	
Segmentation type	Traditional
Contrast threshold	0.5
Min nuclear intensity	0
Max image brightness	1
Nuclear seg aggressiveness	0.8
Holes	TRUE
Size	10.0 – 571.7
Min nuclear roundness	0
No. of dyes	1 (DAPI)
Membrane/Cyto	
Max cytoplasm radius	1.5
Membrane segmentation aggressiveness	0.9
Cell size	10, 600
DAPI	
Nuclear positive threshold	20
% completeness	5
CD34	
Cytoplasm positive threshold	5
% completeness	40
IL-6	
Cytoplasm or nucleus positive threshold	2.5
% completeness	0
CD45	
Cytoplasm positive threshold	2.5
% completeness	20
PDPN	
Cytoplasm positive threshold	2
% completeness	0
pSMAD2	
Nuclear positive threshold	2
% completeness	40
panCK	
Cytoplasm positive threshold	1
% completeness	0

8.2 Appendix 2

Settings used in HALO HighPlex FL v.4.2.3 to detect the nuclei, perform cell segmentation and define positivity for each marker

Nuclear detector	
Segmentation type	Traditional
Contrast threshold	0.5
Min nuclear intensity	0.03
Max image brightness	1
Nuclear seg aggressiveness	1
Holes	TRUE
Size	11.3 – 571.7
Min nuclear roundness	0
No. of dyes	1 (DAPI)
Membrane/Cyto	
Max cytoplasm radius	2.3
Membrane segmentation aggressiveness	0.9
Cell size	10, 600
DAPI	
Nuclear positive threshold	20
CD45	
Cytoplasm positive threshold	10
CD66b	
Cytoplasm or nucleus positive threshold	6
CD3	
Cytoplasm positive threshold	1.8
HLA-DR	
Cytoplasm positive threshold	2
CD56	
Cytoplasm positive threshold	2.5
CD68	
Cytoplasm positive threshold	1