

Exploring fibroblast heterogeneity in the formation of TLS in the context of primary Sjögrens syndrome

By

SABA ASAM



UNIVERSITY OF
BIRMINGHAM

A thesis submitted to the University of Birmingham for
the degree of Doctor of Philosophy

Rheumatology Research Group
Institute of Inflammation and Ageing
College of Medical and Dental Sciences
Submitted September 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The immune system is vital in the protection and maintenance of the body in a healthy state against extracellular pathogens, tissue injury and self-antigens (autoimmunity). Tertiary lymphoid structures (TLS) are non-encapsulated immune cell aggregates which can arise in response to pathogenic insult, cancer and tissue transplant in order to support antigen-specific immune responses in non-immune organs. Contrastingly, in the context of inflammation and autoimmunity, TLS can form within permissive tissues and cause tissue damage and destruction. The *Barone* group has explored the lymphoid function and phenotype of the *immunofibroblast* which contribute to the formation and organisation of these lymphocytic aggregates in the formation and organisation of TLS. Originally grouped coarsely, fibroblasts have been more recently recognized as a largely heterogeneous population defined by uniquely assigned phenotypes and functions.

In this thesis, we deconvolute the stromal cell heterogeneity using scRNAseq in human primary Sjogren's syndrome (pSS) salivary glands (SG) and inflamed murine SG and explore how these cell populations possibly underpin TLS diversity and organisation. We apply bioinformatic tools to explore mRNA gene expression patterns within cell populations, identifying a CCL21⁺ CCL19⁺ "*immunopericyte*" population which has not yet been explored in the context of TLS prior. We also explore the possible signalling mechanisms engaged within the *immunofibroblast* and "*immunopericyte*" populations using the inducible TLS mouse model. This work has created a cellular atlas of the human and mouse pSS SG and the different cell populations present within these primary tissues. This allows for future work to validate the mRNA dataset at the protein level and generate potential novel therapeutic targets at a cell specific level for patients.

Dedication

This is dedicated to my parents who have instilled in me that the impossible is possible and provided me with the means and support to pursue my dreams – Mum and Dad I am forever grateful for all that you have done. My dearest sister and brother who have provided me with constant support through this journey. Last but not least, my closest friends who have had to put up with my PhD as the additional guest on our social ventures – thank you for the constant support and encouragement throughout my journey.

To my supervisory team at UoB, my journey and growth was absolutely not possible without you. To my PI's, Dr. Francesca Barone and Dr. Ben Fisher, thank you for seeing my potential and persevering with me to bring out my best and maximum which I hope now stays with me in my future steps. To the amazing post-doctoral team Dr. Saba Nayar, Dr. Jason Turner and Dr. David Gardner – your support, guidance, kindness, and humour have been invaluable throughout my PhD. Specifically, Saba N thank you for the coffees and the wisdom you have instilled not just as a scientist but as a young woman too. To Jason, your dedication towards my curiosity for bioinformatics has helped me immensely to what I have achieved to date – your input is invaluable and will forever be treasured. And finally, to David and all the flow cytometry we have ever done together – thank you for your countless support and being a part of my work. In addition, the technical staff of the Barone team, Charlotte Smith, and Valentina Iannizzotto, with whom completing our *in vivo* experiments at the BMSU hold a plethora of memories of their own. Thank you both for your support in and out of the lab.

Finally, to all those with whom I shared a laugh or lent an antibody to me during my PhD in the RRG, thank you for your small but valuable contributions in my work.

Acknowledgements

This is to acknowledge the HARMONICS consortium who have funded me during the duration of my PhD studies.

Table of Contents

Abstract	2
Dedication	3
Table of Contents	5
List of Figures	9
List of Tables.....	11
List of Abbreviations.....	12
Chapter 1.0: Introduction.....	20
1.1: TLS and autoimmunity	21
1.2: TLS and SLO: Structure and function.....	22
1.2.1: Heterogeneity of TLS within tissue.....	24
1.2.2: Presentation of TLS in different autoimmune conditions.....	25
1.2.3: TLS in Rheumatoid Arthritis (RA).....	26
1.2.4: TLS in primary Sjögren's Syndrome (pSS).....	27
1.3: TLS in cancer.....	28
1.3.1: Characterization of TLS in non-small cell lung cancer (NSCLC).....	29
1.3.2: Characterization of TLS heterogeneity in cancer.....	30
1.4: TLS in immune responses to bacterial and viral infection	31
1.4.1: Characterization of TLS in murine models of bacterial infection.....	32
1.4.2: TLS in Crohn's Disease (CD).....	33
1.4.3: Characterization of TLS in murine models of bacterial infection.....	34
1.5: TLS in transplantation	35
1.5.1: Presentation of TLS as immunological hurdles in graft rejection.....	36
1.6: Similarities between SLO and TLS in formation and function.....	37
1.7: SLO Development: the crosstalk between Lymphoid Tissue inducer (LTi) and Lymphoid Tissue organiser (LTo) cells.....	40
1.7.1: Development of LNs in normal physiology.....	41
1.7.2: Retinoic Acid Signalling: the initiator signal of SLO anlagen maturation.....	43
1.7.3: LTβR Signalling: the maturation signal of resident SLO mesenchyme.....	44
1.7.4: RANKL Signalling: a second maturation signal of SLO mesenchyme.....	46
1.7.5: The role of T cells in the final steps of LN maturation.....	47
1.7.6: Identification of IL-7R as a key signalling pathway in mesenchyme maturation.....	48
1.8: Stromal cell heterogeneity in SLO supports the maintenance of T and B cell homeostasis.....	49
1.8.1: Fibroblastic Reticular Cells (FRC) and Follicular Dendritic Cells (FDC).....	50
1.8.2: T Cell Zone: Single cell characterisation of murine LN FRCs heterogeneity.....	51
1.8.3: The Hippo pathway: fate determining signal of FRCs.....	53
1.8.4: The B cell follicle contains stroma which supports the Germinal Centre (GC) Reaction.....	54
1.8.5: The Retinoic Acid Pathway: driver of the FDC phenotype.....	56
1.9: The GC reaction.....	57
1.9.1: The DZ of the GC.....	59

1.9.2: The LZ of the GC.....	60
1.9.3: The heterogeneity of FDC within the GC reaction.....	61
1.9.4: The contribution of CD4 ⁺ Tfh cells in the B cell follicle.....	62
1.10: TLS organisation and function mirror SLO in non-immunological organs.....	64
1.10.1: Comparing and contrasting TLS and SLO architecture and function.....	65
1.11: TLS Development: the crosstalk between Lymphoid Tissue initiator cells (LTin) and the resident stromal mesenchyme.....	67
1.12: The heterogeneity of fibroblasts: conserved yet unique in different disease contexts.....	68
1.12.1: scRNAseq profiling of CAFs in different types of cancers.....	69
1.12.2: scRNAseq profiling of fibroblasts in inflammatory conditions.....	72
1.13: Fibroblasts of mice and men: identifying conserved populations across species	76
1.14: The generation of the <i>immunofibroblast</i> : the <i>priming</i> step	79
1.14.1: The role of PDPN in <i>immunofibroblast</i> priming.....	80
1.14.2: The role of innate immune cells in <i>immunofibroblast</i> priming.....	81
1.15: The expansion of the <i>immunofibroblast</i>	83
1.16: <i>The maturation of the immunofibroblast</i>	84
1.16.1: The effects of pro-inflammatory cytokines on <i>immunofibroblasts</i>	85
1.16.2: The interference of the IFN signalling cascade in <i>immunofibroblast</i> maturation.....	86
1.16.3: The TNF superfamily in <i>immunofibroblast</i> maturation.....	88
1.17: Thesis Aims and Objectives.....	91
Chapter 2.0: Materials and Methods	93
Table 1: Human Flow Cytometry Antibody Panel	93
Table 2: Mouse Flow cytometry Antibody Panel	94
2.1: Tissue Digestion and flow cytometric analysis of human pSS SG and intracellular CCL21 staining	95
2.2: Single Cell RNA sequencing (scRNAseq): a novel tool bringing clarity to the chaos.....	96
2.3: Single Cell RNA sequencing of the human pSS SG and murine SG	99
2.3.1: Single-cell RNAseq processing for human pSS SG samples.....	103
2.3.2: Single-cell RNAseq processing for resting and inflamed D15 p.c. murine SG samples.....	110
2.4: Removal of doublets and ambient gene signatures in the scRNAseq dataset	113
2.5: Dimensionality reduction and normalisation of the scRNAseq data	115
2.6: Downstream analysis tools used for the biological inference of scRNAseq data.....	119
2.6.1: Cell Cluster annotation and differentially expressed gene (DEG) analysis.....	119
2.6.2: Slingshot Trajectory Analysis.....	119
2.6.3: Geneset enrichment analysis (GSEA).....	121
2.6.4: Pseudobulk Correlation.....	121
2.7: Human pSS SG Histology.....	122

2.8: Microdissection of human pSS SG and bulk sequencing analysis	123
Chapter 3: Transcriptomic characterisation of the human pSS SG	126
3.1: Deconvoluting the heterogeneous human pSS SG using scRNAseq.	129
3.2: The CD34 ⁺ CD146 ⁻ stromal and CD34 ⁻ CD146 ⁺ pericyte populations as sources of the lymphoid chemokine CCL19	132
3.3: The stromal and pericyte population: <i>ex vivo</i> validation.....	137
3.4: GC heterogeneity in pSS SG	142
3.5: Pathway enrichment analysis of the CD34 ⁺ CD146 ⁻ Immunofibroblasts and the CD34 ⁻ CD146 ⁺ Pericytes	150
3.6: The endothelium population	154
3.7: The epithelium population.....	157
3.8: The Glial cell population.....	161
3.9: T Cells of the human pSS SG.....	163
3.10: The B cell population of the human pSS	170
3.11: Plasma Cells of the pSS SG: the autoantibody factories	174
3.12: The Myeloid population	178
Chapter 3: Discussion	178
Chapter 4: Transcriptomic characterisation of the murine pSS SG.....	208
4.1: Exploring the cellular landscape of the murine SG	210
4.2: The single cell landscape: the stromal cell compartment.....	218
4.3: Fibroblasts of the resting and inflamed murine SG	220
4.4: Pericytes of the murine SG.....	230
4.5: Endothelium Population of the murine SG	236
4.6: Epithelium Population of the murine SG	240
Chapter 4: Discussion	244
Chapter 5: Final Thesis Discussion	255
5.1: Overall Discussion	255
5.2: Limitations within this work.....	263
5.3: Future experimental work	265
References.....	269
Appendix	286
1.0: Script for scRNAseq analysis	286
1.1: Set variables used throughout	287
1.2: Declare functions	288
1.3: Get genes detected in the cluster	289

1.4: Get genes detected in the other cells-----	289
1.5: Read in data	289
1.6: Retaining ensembl IDs and identifying hsapiens orthologs	290
1.7: Create seurat object.....	290
1.8: QC filtering	293
1.9: Subsetting data	294
1.10: Perform regression without cell cycle correction	294
1.11: Downsample to lowest cell number	294
1.12: Identify variable genes	295
1.13: Dimensionality reduction.....	295
1.14: Write out heatmaps showing the genes loading the first 12 components	296
1.15: Harmony for Seurat v2	297
1.16: Generating TSNE	298
1.17: Generating UMAP	298
1.18: Find clusters	298
1.19: Plotting key markers using Violin (Vln) plots	298
1.20: Generating Subclusters	299
1.21: Build clustree and validate clusters	299
1.22: Saving seurat objects	299
1.23: Fibroblast/mural enrichment rerun post fine clustering	300
1.24: Plotting cell populations.....	300
1.25: Cluster identity assignments.....	302
1.26: Selecting cluster colours	304
1.27: Exporting data for Cibersortx.....	304
1.28: GSEA plots	305
1.29: Removal of ambient signatures	306
1.30: Calculating Relative Cell Proportions	307
1.31: Proportion per cluster by Rho pos or neg.....	307
1.32: Pseudobulk and correlate between mouse and human pss cell clusters	309
1.33: Generate pseudobulk for the murine day15 fibroblasts	311
1.34: Finding orthologs and matching identities	312
1.35: Finding orthologs and matching identities.....	313

List of Figures

Chapter 1:

Figure 1A & B: Summary of the different fate determining signals forming the individual T cell area fibroblast (TRC) and B cell area stroma (FDC)

Figure 2A & B: Diagrammatic representation of the GC and the overarching chemokine signature and cell populations within the DZ and LZ

Figure 3: Diagrammatic summary of the sequential *priming*, *expansion* and *maturation* of the stromal cells in response to different cytokines and their cellular sources

Chapter 2:

Figure 1 : Diagrammatic summary of UMI to transcribed genes

Figure 2: Summary of outcome and limitations for retention of high-quality cells for downstream bioinformatic analysis

Figure 3: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

Figure 4: Summary of murine SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

Figure 5: Diagrammatic summary of generation of ambient cell signature in scRNAseq datasets

Figure 6: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

Figure 7: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

Chapter 3:

Figure 1: Global Cluster analysis of the human pSS SG

Figure 2: The CD34+ CD146- fibroblast and CD34- CD146+ mural populations

Figure 3: The CD34+ CD146- fibroblast and CD34- CD146+ mural populations

Figure 4: Microdissection analysis of the human pSS SG

Figure 5: The CD34+ CD146- fibroblast and CD34- CD146+ mural populations

Figure 6: The endothelium populations in the human pSS SG

Figure 7: The epithelium of the human pSS SG

Figure 8: The Glial cell population of the human pSS SG

Figure 9: The T cell population of the human pSS SG

Figure 10: Gene expression of receptor and ligand interactions between T cells and fibroblasts from the human pSS scRNAseq dataset

Figure 11: The B cell population in the human pSS SG

Figure 12: The Plasma Cell populations in the human pSS SG

Figure 13: The myeloid population in the human pSS SG

Figure 14: Diagrammatic summary of the possible functional relationship between the immunofibroblast and “immunopericytes” in TLS formation.

Chapter 4:

Figure 1: The global cellular landscape of the Day0 pSS murine model SG

Figure 2: The global cellular landscape of the Day 15 p.c. murine model SG

Figure 3: The fibroblast cell populations at D0 and D15 p.c.

Figure 4: Comparison of the fibroblast population between D0 and D15 p.c.

Figure 5: In vivo validation of PDPN+ CD34- Fibroblasts in the murine SG

Figure 6: Comparison of the fibroblast population between human pSS and D15 p.c. murine SG

Figure 7: The pericyte population at D0 and D15p.c. in the murine SG

Figure 8: Validation of CD34- CD146+ CCL21+ pericytes in the D15p.c. murine SG

Figure 9: The endothelium cell populations in D0 and D15p.c. murine pSS SG

Figure 10: The epithelium cell population in D0 and D15p.c. murine SG

List of Tables

Chapter 1:

Table 1: The developmental and anatomical differences between SLO and TLS

Table 2: Summary of the currently characterised fibroblast populations in the context of the LN

Table 3: Summary of the currently characterised fibroblast populations in the context of autoimmunity

Table 4: Summary of the currently characterised fibroblast populations in the context of cancer

Chapter 2:

Table 1: Human Flow cytometry antibody panel

Table 2: Mouse flow cytometry antibody panel

Table 3: Summary of current genomic applications used to investigate gene expression profiles

List of Abbreviations

Abbreviation	Full Form
ΔCt	Delta Cycle Threshold
ABCC9	ATP-binding cassette sub-family C member 9
ACKR3	Atypical chemokine receptor 3
AD	atopic dermatitis
Add3	Gamma-adducin
ADSC	Adipose Derived Stem Cell
Ag	Antigen
Agdre5/ Cd97	Adhesion G protein-coupled receptor E5
Ahnak	Desmoyokin
AICDA	cytosine deaminase
AID	Activation-induced cytidine deaminase
Aire	autoimmunity regulator
ALDH1A1	Retinal dehydrogenase 1
aly/aly	Alymphoplasia
ANKRD36C	Ankyrin repeat domain-containing protein 36C
ANOVA	Analysis of Variance
APCs	Antigen Presenting Cells
ApoExo	apoptotic exosome
Ascl3	Achaete-scute homolog 3
Asgr1	Asialoglycoprotein receptor 1
aSMA	alpha-smooth muscle actin
BAFF	B cell Activating Factor
BCL2A1	B-cell lymphoma 2-related protein A1
BCL-6	B Cell Lymphoma – 6
BCL-6	B-cell lymphoma 6
BEC	blood endothelial cells
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
C3	complement 3
C3^{-/-}	complement 3 knockout mice
C3aR	C3a anaphylatoxin receptor
CA4	carbonic anhydrase 4
CAF	Cancer associated fibroblasts
Car4	Carbonic anhydrase 4
CCL19	C-C Motif Chemokine Ligand 19
CCL21	C-C Motif Chemokine Ligand 21
CCR7	C-C Motif Chemokine Receptor 7
CD	Cluster Differentiation

CD	Crohn's disease
CD146 / MCAM	Cluster of differentiation 146/ Melanoma Cell Adhesion Molecule
CD1C	T-cell surface glycoprotein CD1c
CD27	T-cell activation antigen
CD31	Cluster of differentiation 31
CD34	Cluster of differentiation 34
CD36	cluster of differentiation 36
CD62L	L-selectin
Cd74	MHC class II-associated invariant chain
CD82	Tetraspanin-27
Cdkal1	CDK5 regulatory subunit-associated protein 1-like 1
CDKN1A	Cyclin-dependent kinase inhibitor 1
cDNA	complementary Deoxyribose Nucleic Acid
Cgref1	Cell growth regulator with EF hand domain protein 1
Ch25	cholesterol-25-hydroxylase
CIA	Collagen Induced Arthritis
Cldn6	Claudin 6
CLDN9	Claudin 6
CLEC-2	C-type lectin-like receptor 2
c-Myc	Cellular myelocytomatosis oncogene
Col15a1	Collagen alpha 1(XV) chain
COL3A1	Collagen alpha 1(III) chain
Cr2	complement receptors
CRCs	CXCL12 producing reticular cells
CSR	class switch recombination
CTLA-4	Cytotoxic T lymphocyte associated antigen 4
CX3CR1	CX3C chemokine receptor 1
CX3CR1	CX3C chemokine receptor 1
CXCL1	C-X-C Motif Chemokine Ligand 1
CXCL10	C-X-C Motif Chemokine Ligand 10
CXCL10	C-X-C motif chemokine ligand 10
CXCL11	C-X-C Motif Chemokine Ligand 11
CXCL13	C-X-C Motif Chemokine Ligand 13
CXCL16	C-X-C Motif Chemokine Ligand 16
CXCL9	C-X-C Motif Chemokine Ligand 9
CXCL9	C-X-C Motif Chemokine Ligand 9
DAF/CD55	Decay accelerating factor
DAMPs	danger associated molecular patterns
Dclk1	Doublecortin-like kinase 1
DCs	Dendritic Cells
DCs	dendritic cells
DKK3	Dickkopf-related protein 3

Dpt)	Dermatopontin
DZ	dark zone
E2A	transcription factor E2A
EAE	Experimental autoimmune encephalomyelitis
EAE	experimental autoimmune encephalomyelitis
EDTA	Ethylenediaminetetraacetic acid
EMT	epithelial to mesenchymal transition
Epas1	Endothelial PAS domain-containing protein 1
Epcam	Epithelial Cell Adhesion Molecule
EPPN2	Autotaxin
Esp4	Exocrine gland secreted peptide 4
eTACs	extrathymic Aire-expressing cells
Etoh	Ethanol
F2R	Thrombin receptor
FACS	Fluorescent activated Cell Sorting
FALCs	Fat associated lymphoid clusters
FALCs	fat associated lymphoid clusters
FAP	Fibroblast activation protein
FAPα	Fibroblast activating protein
Fc	Fragment crystallizable
FCGR2	Low affinity immunoglobulin gamma Fc region receptor II
FCS	Fetal Calf Serum
FDCs	follicular dendritic cells
Fgf7	Fibroblast growth factor 7
FOSB	FBJ murine osteosarcoma viral oncogene homolog B
Foxp3	Forkhead box P3
FRCs	Follicular Reticular Cells
Fxyd2	FXFD domain-containing ion transport regulator 2
Gata3	ATA Binding Protein 3
Gbp5	Guanylate-binding protein 5
GC	Germinal Centre
GI	gastrointestinal tracts
GIMATS	IgG plasma cells, inflammatory mononuclear phagocytes, activated T cells, and stromal cells
GPCRs	G protein coupled receptors
gsfisher	Geneset fisher
GZMA	Granzyme A
HEVs	High endothelial venules
HGSC	high grade serous ovarian cancer
HT	Hashimoto Thyroiditis
iBALT	inducible branchial associated lymphoid tissues
IBD	inflammatory bowel disease
ICAM-1	Intercellular Adhesion Molecule 1

ICB	immune checkpoint inhibitor
ICOS	Inducible Co-stimulator
ICOS ^{-/-}	Inducible Co-stimulator knockout
ICOS-L	Inducible Co-stimulator Ligand
Id2	Inhibitor of DNA binding 2
IFN- γ	Interferon gamma
Ig	Immunoglobulin
IgA	Immunoglobulin A
IgD	Immunoglobulin heavy constant delta
IGHG	immunoglobulin heavy chain gamma
IGHV	immunoglobulin heavy chain variable regions
IGKV	immunoglobulin <i>kappa</i> light chain variable regions
IGLV	immunoglobulin <i>lambda</i> light chain variable region
IHC	immunohistochemistry
IL-10	Interleukin 10
IL-13	Interleukin 13
IL-13R α 1	Interleukin 13 Receptor α 1
IL-17	Interleukin 17
IL-1 β	Interleukin 1 Beta
IL-2	Interleukin 2
IL-21	Interleukin 21
IL-22	Interleukin-22
IL-23	Interleukin-23
Il2ry	IL-2 receptor gamma
Il2ry ^{-/-}	IL-2 receptor gamma knockout mice
IL-4	Interleukin 4
IL4R	Interleukin-4 receptor
IL-5	Interleukin 5
IL-6	Interleukin 6
IL-7R	Interleukin 7 Receptor
IL-7R α	Interleukin 7 Receptor α
ILC	innate lymphoid cells
ILC1	Innate lymphoid cells 1
ILC2	innate lymphoid cells 2
ILC3	Innate Lymphoid Cells 3
iMSC	inflammatory mesenchymal stromal cell
iNALT	inducible nasal associated lymphoid tissues
Inmt	Indolamine <i>N-methyltransferase</i>
iNOS	inducible Nitric Oxide Synthase
Itga1	Integrin α -1
Itgae	Integrin α -E
ITGAX	Integrin α -X

Kcne1	Potassium voltage-gated channel subfamily E member 1
Ki-67	protein Ki-67
KRT7, KRT8 and KRT14	keratins 7,8 and 14
LEC	lymphatic endothelial cell
LFA-1	Lymphocyte function-associated antigen 1
LLSc	Lymphoid Like Stromal cell component
LMNA	lamin A
LN	Lymph Node
LPS	Lipopolysaccharide
Ltα	Lymphotoxin alpha
LTi	Lymphoid Tissue Inducers
LTin	lymphoid tissue initiator
LTo	Lymphoid Tissue Organisers
LTα	Lymphotoxin a
LTα1β2	Lymphotoxin α 1 β 2
LTβ	Lymphotoxin b
LTβR	Lymphotoxin β Receptor
LV	lymphatic vessels
LYVE-1	lymphatic vessel endothelial hyaluronan receptor 1
LYZ	Lysozyme C
LZ	light zone
MACS	Magnetic Cell Sorting
MADCAM-1	Mucosal vascular Addressin cell Adhesion Molecule 1
MALT	Mucosal Associated Lymphoid Tissues
MBP	Myelin basic protein
MEM	Minimum Essential Medium
Mfap4	Microfibril-associated glycoprotein 4
Mfge8	Milk fat globule-EGF factor 8
MFI	Median Fluorescence Intensity
MHC	Major histocompatibility complex
MHC II	Major histocompatibility complex II
MLR	Mixed Lymphocyte Reaction
MPZ	Myelin Protein Zero
mRNA	messenger Ribose Nucleic Acid
mTECS	mesenchymal Thymic Epithelial Cells
MUC16	Mucin-16
MUC5B	Mucin-5B
Mup6	Major urinary protein 6
MZB1	Marginal zone B and B1 cell-specific protein
NALT	nasal associated lymphoid tissues
NCAM1	Neural Cell Adhesion Molecule 1
NCR	natural cytotoxicity receptor

NFkb	nuclear factor kappa-light-chain-enhancer of activated B cells
NFKB1A	Nuclear factor NF-kappa-B
Nfkb	nuclear factor kappa-light-chain-enhancer
Ngf	Beta-nerve growth factor
NK	natural Killer Cells
NMCC	non-metastatic colorectal carcinoma
NO	Nitric Oxide
NOTCH3	notch homolog protein 3
NOTCH4	Neurogenic locus notch homolog protein 4
NSCLC	non-small cell lung cancer
OASIS	Optimising Assessment in Sjögren's Syndrome
Olfml3	Olfactomedin-like protein 3
PAMPs	pathogenic associated molecular patterns
PBMC	Peripheral Blood Mononuclear Cells
PBS	Phosphate buffer saline
PCR	Polymerase Chain Reaction
PC	Plasma Cells
PD-1	Programmed Death -1
PDGFRA	Platelet Derived Growth Factor Receptor Alpha
PDGFRB	Platelet Derived Growth Factor Receptor Beta
PD-L1	Programmed Death - Ligand 1
Pdp2	pyruvate dehydrogenase
PDPN	Podoplanin
Pi16	Peptidase inhibitor 16
Plt	paucity of lymph node T cells
Plxna2	Plexin-A2
PMA	Phorbol 12-myristate 13-acetate
PNAd	Peripheral node addressin
POSTN	Periostin
PRDM-1	positive regulatory domain containing 1
Prox-1	Prospero Homeobox 1
PRRs	pattern recognition receptors
pSS	primary Sjogren's Syndrome
RA	Rheumatoid Arthritis
ReA	Retinoic acid
Rag2^{-/-}	V(D)J recombination-activating protein 2
Raldh2	retinaldehyde dehydrogenase
Raldh2^{-/-}	retinaldehyde dehydrogenase knockout
RANK	Receptor activator of nuclear factor kappa B
Rank^{-/-}	Receptor activator of nuclear factor kappa-B knockout
RANK-L	Receptor activator of nuclear factor kappa-B – Ligand
RARB	retinoic acid receptor B

Rarg	Retinoic acid receptor gamma
RGS5	Regulator Of G Protein Signaling 5
RIP-LTα	Rat insulin promoter - lymphotoxin alpha
RNA	Ribonucleic Acid
Ro/SSA	Ribonucleoprotein
RORγ	RAR-related orphan nuclear receptor
RORγt	RAR-related orphan receptor gamma
RORα	RAR-related orphan receptor alpha
ROS	reactive oxygen species
scRNAseq	Single Cell RNA sequencing
Sdc1	Syndecan-1
SELP	P-selectin
SEMA3G	Semaphorin-3G
SERPINE 2	serpin peptidase inhibitor clade E member 2
Sfrp2	Secreted frizzled-related protein 2
SG	Salivary glands
SHISA8	Shisa family member 8
SHM	somatic hypermutation
SLC13A2	Solute carrier family 13 member 2
SLO	Secondary Lymphoid Organ
SMARCB1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1
SPARC	Secreted protein acidic and cysteine rich
T Cell	Thymus derived lymphocyte
TAMs	tumor-associated macrophages
Tbet	T-box expressed in T cells
T_{Cem}	T cell central memory
TCR	T Cell Receptor
Tfh	T Follicular Helper Cell
Tfr	T follicular regulatory
Tgfb2	Transforming growth factor beta-2
TGFBR1	TGF-beta receptor type-1
TGF-β	Transforming growth factor beta
Th1	T helper 1
Th17	T helper 17
Th2	T helper 2
Thy1/CD90	Thy-1 membrane glycoprotein
TIL	tumour infiltrating lymphocytes
TLS	Tertiary Lymphoid Structure
TME	Tumour Microenvironment
TNC	tenascin C
TNFR1/2	Tumour Necrosis Factor Receptor 1 / 2
TNF-α	Tumour Necrosis Factor - α

TRANCE	TNF-related activation-induced cytokine
Trance^{-/-}	TNF-related activation-induced cytokine knockout
TRC	T zone reticular cells
Treg	T regulatory
TSNE	t-distributed stochastic neighbor embedding
UC	ulcerative colitis
UMAP	Uniform Manifold Approximation and Projection
UV	ultraviolet
VCAM-1	Vascular cell adhesion protein 1
VEGFC	Vascular Endothelial Growth Factor C
VEGFD	Vascular Endothelial Growth Factor D
VLA-4	very late antigen 4
VSCs	versatile stromal cells
Vwf	von Willebrand factor
Wnt	Wingless/Integrated
Wnt5a	Protein Wnt-5a
WT	wild type
YAP/TAZ	Transcriptional coactivator YAP1/ Tafazzin
γδ	Gamma delta

Chapter 1.0: Introduction

Part of the text and one of the tables (Tables 2-4) in this chapter have been previously published in [1]. I was the principal author of this article, wrote large part of the text and generated all the relevant figures, with amendments as suggested by my co-authors in this publication.

1.1: TLS and autoimmunity

Inflammation is a complex immunological phenomenon that encourages the infiltration of circulating lymphocytes to a site of infection or tissue damage. The process of inflammation involves various immune cells, lymphatic and blood vasculature as well as a mixture of pro-inflammatory cytokines and chemokines [2]. The purpose of this response is to clear the cause of infection or cellular injury and promote tissue repair. The length of time an individual sustains an inflammatory response can be categorised as acute or chronic. Acute inflammation is the initial inflammatory response that occurs to a stimulus signifying danger mainly involving components of the innate immune system, whilst chronic inflammation is a prolonged status that often comprises of both innate and adaptive immune cells. To fulfil its purpose, it is critical that the correct level of the inflammatory response is exerted. A poor inflammatory response would result in lack of clearance of pathogens and poor healing whilst too great an inflammatory response would cause excessive bystander tissue damage causing loss of function and the onset of autoimmunity. Autoimmunity is the result of an immune response which has become misdirected and lacks, the signals to direct an individual's immune cells not to target self-antigens as foreign pathogenic proteins. The onset of autoimmunity can be influenced by genetic factors or environmental factors such as smoking [3].

The molecular mediators present within the inflammatory microenvironment are able, in certain conditions, to favor the initiation of TLS development [4-6]. Importantly, chronic inflammation does not always result in TLS formation; instead, TLS have been observed to develop specifically in permissive organs, suggesting the presence of tissue-specific cues that are able to imprint the immune responses and form TLS hubs [7-11].

1.2: TLS and SLO: Structure and function

TLS are immunological structures comprising of organised immune cell populations within non immunological tissues. These structures differ from SLO as they are non-capsulated structures which arise in permissive tissues that do not have an assigned immunological function [1, 10, 11]. The formation of TLS is traditionally observed in microenvironments of an inflammatory nature which have pro-inflammatory stimuli or have an abundance of autoantigens [8]. The cellular milieu which result in the TLS formation include the resident stromal compartment including epithelial cells, fibroblasts, endothelial cells and possibly pericytes. Recently, fibroblasts have gained attention with the discovery of their immunomodulatory capabilities due to the production of lymphoid chemokines such as CXCL13, CCL19 and CCL21 and pro-inflammatory cytokines such as IL-1 [7, 12-17]. The contributions of the fibroblasts in various TLS associated diseases and inflammatory conditions have been explored by various groups and I have explored in more detail in this chapter. Additionally, resident and infiltrating immune cells further modify the underlying network of stroma to support the survival and organisation of the immune cells [18-23].

Overarchingly, the presence of TLS at specific anatomical locations allows them to be assigned unique identities which includes mucosal-associated lymphoid tissues (MALT); lymphocyte-rich clusters forming in the gastrointestinal tracts (GI); the fat-associated lymphoid clusters or FALCs, present in the adipose tissue of the mesenteries; and the nasal and inducible bronchial-associated lymphoid tissues (NALT and iBALT) [1, 7, 8, 10-12, 18, 24-31][19, 32-35]. The dependency on antigen exposure in order to shape the functional role of the TLS in disease has been demonstrated in the context of the generation of specific humoral responses against

antigen or at times autoantigen. Currently, there is no clear link of whether antigen, antigen-specific B cell maturation or autoantibody production occurs in the aforementioned sites of TLS [10]. However evidence has supported that the formation of TLS occurs in response to the display of antigens locally on both professional and non-professional APCs within an inflamed microenvironment rich in molecules such as TNF- α , IL-6, IL-17 and IL-22 amongst others [4, 7, 8, 11].

1.2.1: Heterogeneity of TLS within tissue

The molecular mediators present within the inflammatory microenvironment can favour the initiation of TLS development. Whilst TLS are not thought to be the cause of catalysing the onset of inflammation itself, but rather arise as a consequence on inflammation, the presence of TLS has been associated with poor prognosis in clinical disease. The presence of chronic inflammation does not equate to guaranteed TLS formation, instead TLS formation has been observed to occur specifically in permissive tissues such as the thyroid of patients with Hashimoto Thyroiditis (HT), the synovium of patients with Rheumatoid arthritis (RA) or the salivary glands (SG) of patients with Sjogren's Syndrome (pSS) [36-39]. The mechanistic switch that drives certain patients affected by chronic conditions to form TLS has not been elucidated and leaves an area of exploration to identify accurate biomarkers indicative of potential TLS development in individuals. Importantly, TLS do not form in all patients and, even within the same tissue, and have been characterized by the detection of different organizational features, suggesting a potential gradient, either in the quantity/quality of the antigen presented or in the intensity of the inflammatory response shaping TLS assembly. Currently, the developmental time for TLS is not known but their sequential progression of development is thought to be initiation as a small T cell dominant aggregate, which progresses to a large T and B cell foci and eventual formation to a mature functional TLS comprising an active GC. Different types of TLS have been identified in close proximity to each other within the same organ of the same patient [19], suggesting that progression of the TLS to a mature phenotype is not synchronous and homogeneous even in the same tissue. The reason of this heterogeneity could rely on the differential distribution of stimulatory signals in the tissue or in the reliance on different antigens.

1.2.2: Presentation of TLS in different autoimmune conditions

Several autoimmune diseases have now been identified to encompass TLS in their clinical presentation, which are currently being studied for use as biomarkers to predict the possible prognosis for an individual. Intriguingly, within tissue which are able to accommodate TLS development, these lymphocytic aggregates are not always observed. This suggests that the formation of TLS is governed by either the abundance of antigen present within the local environment or the pro-inflammatory milieu within the particular site of TLS establishment. In this manner, the mechanistic switch which allows TLS in particular patients but not others to form have not been formally characterised and are still being explored. In the context of autoimmunity and inflammation, TLS amplify the immune response such that tissue damage occurs due to the prolonged activation, survival, and persistence of lymphocytes within the area in which the immune response is taking place. This is due to the production of chemokines which orchestrate lymphocyte infiltration and lymphocyte survival factors [40, 41]. The expression of these factors in combination confer resistance to lymphocyte depletion therapies such as anti-CD20 Rituximab and prevent the complete dissolution of these structures [42].

1.2.3: TLS in Rheumatoid Arthritis (RA)

RA is a chronic inflammatory autoimmune condition of the synovial membrane of the joints [43]. This disease is characterised by severe tissue damage and loss of joint movement. The presence of TLS in the context of RA was initially alluded to as formation of “lymphoid nodules” and already included in the classification criteria of Rheumatoid arthritis (RA) in 1957 [44]. It was around 20 years later that the synovium layer was thought to share features of lymphoid organs such as active GC which generated Ig producing PC [45]. The thyroid and SG have large epithelium components which have also been characterised in various studies to be cellular participants responsible for the secretion of pro-inflammatory cytokines such as IL-1 β in combination with functioning as non-classical APCs.

1.2.4: TLS in primary Sjögren's Syndrome (pSS)

The autoimmune disease pSS is a multifactorial disease involving genetic and environmental factors and is characterised by dysfunctional exocrine glands which are associated with the presence of TLS within submandibular and parotid SG [46-49]. This loss of function is most commonly clinically diagnosed by dryness of the eyes and mouth, amongst other secretory tissues, and fatigue [39]. In this context, TLS present themselves as pathological entities which facilitate destruction of the saliva secreting acinar cells within the SG due to the accumulation of immune cells such as T cells, B cells and antibody secreting PC [22, 25, 48, 50-55]. The epithelial compartment has been demonstrated to play a significant role in the formation of TLS within the SG in pSS an additional permissive tissue which harbours lymphocyte aggregates in the form of TLS [9, 56, 57]. Whilst this disease may remain indolent for many years, a small proportion of pSS patients are considered at increased risk of developing B cell associated lymphomas, most notably mucosal associated lymphoid lymphomas (MALT) [48, 51]. This is due to the hyper activation of autoreactive B cells in SG which receive the necessary activation and survival signals. The lymphomatous B cell clones arise as a result of insufficient immune control which comprises aberrant antigenic activation and stimulation, coupled with the presence of increased B cell survival factor, BAFF in the SG [48, 51]. Another hallmark of pSS is the presence of circulating autoantibodies directed against the ribonucleic proteins anti-Ro/SSA and anti-La/SSB [48, 58, 59]. Based on these observations, B cells are key drivers of the pathogenicity of pSS as they can lead to development of lymphoma and become autoreactive antibody producing PC.

1.3: TLS in cancer

Cancer is a dynamic disease which encompasses many factors such as cellular mutations, tumour abundance and the TME which contribute to its complexity as a pathology requiring therapeutic intervention [60, 61].

Whilst the drivers of cancer pathology have been identified as the cancerous cells themselves, the TME within which cancer cells thrive, has recently gained attention as a therapeutic target. The TME is able to shield cancer cells physically and immunologically and provide metabolic support to malignant cells[19, 62]. The key players conferring these pro-tumorigenic functions are the cancer associated fibroblasts or CAFs and are identified by the expression of FAP, PDGFR α/β , PDPN, CD90 and α SMA [62-64]. CAFs are the most abundant stromal component in the TME, and secrete growth factors, promote angiogenesis, facilitate metastasis and regulate immune cell infiltration [62, 65]. At present, the prognostic role of TLS in cancer is debated and appears variable, based on the type of cancer within which they are found and the nature of the clinical stage of the cancer itself [19, 33, 66]. More recently, TLS formation in solid tumours has been associated in patients with a positive response to treatment with immune checkpoint inhibitors, opening a potential avenue of research aimed at favouring, rather than blocking, TLS development in the context of cancer [32].

1.3.1: Characterization of TLS in non-small cell lung cancer (NSCLC)

The presence of TLS has been associated with positive prognosis as the TLS appear to support Th1 and CD8 cytotoxic anti-tumour immune response [67]. In primary breast tumours, TLS formation has been instead associated with increased Treg recruitment, ultimately conferring poor patient prognosis [68, 69]. The increased recruitment of CD4 T cells was also uncovered by *Chaisemartin et al* who identified the classical CC19, CCL21, and CXCL13 lymphoid chemokine signature and upregulated expression of ICAM-1 and VCAM-1 within intratumoural TLS situated within lung cancers. Moreover, this study highlighted the infiltration of circulating T cells into the intratumoural TLS through *de novo* formed HEVs within the tumour mass [70]. Increased tumour infiltrating lymphocytes (TIL) within triple negative breast cancers were associated with the presence of TLS near the tumour tissue [35]. A comprehensive study within non-metastatic colorectal carcinoma (NMCC) involving 109 patients was conducted by *Posch et al* who explored the number and degree of TLS maturation using the presence of GCs as a marker of TLS maturation. This study made evident that whilst TLS were present in the majority of cancer specimens, tissues harbouring elevated BRAF mutations had increased incidence of TLS [71]. All together those findings appear to suggest that factors associated with genetic instability and potential antigenic diversification might contribute to TLS formation.

1.3.2: Characterization of TLS heterogeneity in cancer

Moreover *Kroeger et al* explored the contributions of the tumour infiltrating PC and TLS in high grade serous ovarian cancer (HGSC) [72]. Within HGSC tissues, this study identified a variety of T and B cell aggregates some of which were small and diffuse and others representing classical TLS. The location of PC within the HGSC tissue was associated to be near CD8, CD4 and CD20 TILs within the TLS. This observation inferred that the TLS may facilitate the anti-tumour response by creating the appropriate survival niche for the TILs with the enrichment of survival signals such as IL-7 and BAFF [72]. More recently cancer associated TLS formation has been investigated using scRNAseq on biopsies obtained from melanoma and sarcoma patients. Three independent research groups, *Cabrita et al*, *Helmink et al*, *Petitprez et al*, concluded that the presence of TLS in patients treated with immune checkpoint inhibitor therapy (ICB) associated with a strong B cell response and increased patient survival [19, 33, 66]. This emphasised a previously disregarded role of B cells in anti-tumour immunity. TLS functionality was determined by the output of B cells that synergise with anti-tumour T cells to kill cancer cells [66]. Intriguingly TLS heterogeneity both in terms of anatomical and functional maturation also emerges in these conditions. *Helmink et al* described the presence of TLS heterogeneity within a patient biopsy whereby mature GC⁺ TLS, which harbour proliferative B cells, co-reside with immature TLS. Meanwhile, *Petitprez et al* reported a diverse profile of a B cell repertoire suggesting that factors such as disease stage and severity might shape TLS anatomy and functional response.

1.4: TLS in immune responses to bacterial and viral infection

Individuals are constantly exposed to viral and bacterial threats which can be fatal if not cleared rapidly and effectively. The local generation of an immune response to antigen within the tissue in which the infection is present is facilitated by the presence of TLS. Consequently, in response to these pathogens TLS confer protection as demonstrated by their formation, organisation and functionality in various mouse models.

1.4.1: Characterization of TLS in murine models of bacterial infection

TLS have been identified in the upper respiratory tract in mice following influenza infection and were functional because of their ability to generate in situ differentiated anti-viral PC that secreted antibodies specific to epitopes of influenza antigen [73, 74]. On the contrary, in response to *Mycobacterium tuberculosis* infection TLS have been associated with driving the formation of granulomas – which are small aggregates of mature macrophages [75] - to prevent the further dissemination of the tuberculosis infection.

Additional studies have been used to model the dynamics of TLS in response to the two common bacteria associated with cystic fibrosis, *Pseudomonas aeruginosa* or *Staphylococcus aureus* bacteria [76]. The study carried out by *Teillud et al*, explored TLS in chronic inflammation mediated by bacterial infection by targeting the infiltrating immune cells. This study emphasised the critical contributions of the T cell compartment in the initial formation of TLS within the lungs of the infected mice. Contrastingly, in studies exploring the contributions of *Salmonella* in colitis models, *Koscsó et al* observed in a fluorescent reporter mouse and RNAseq analysis that a CXCL13⁺ CX3CR1^{high} macrophage population were enriched in TLS found within the colonic mucosa [77]. This population of APCs were responsible for the recruitment and activation of T cells at sites of *Salmonella* invasion which led to subsequent TLS formation and the generation of a mucosal IgA response.

1.4.2: TLS in Crohn's Disease (CD)

In addition to colitis, CD is another inflammatory condition of the colon which has strong associations with the imbalanced compositions of the gut microbiota [78]. The formation of TLS has also been identified in human CD patients and been observed to derive from the “creeping fat” or adipocyte population present within the mesentery. *Guedj et al* have characterised the TLS forming functions of adipocytes within CD by firstly identifying higher expression of adipocyte derived chemokines CCL19, CCL21 and CXCL13 from human mesenteric tissue, as compared to controls, which governed lymphocyte infiltration in response to bacterial insult [31]. The expression of this lymphoid stromal cell signature from resident adipocytes was revealed to be conferred by the combined effects of the pro-inflammatory cytokine TNF- α and LPS, therefore adipocytes as non-classical stromal cells are also able to participate in the formation of TLS.

1.4.3: Characterization of TLS in murine models of bacterial infection

TLS formation does not only occur in response to bacterial infection but also in the context of viral infection. Not termed TLS, induced bronchial associated lymphoid tissue (iBALT) are the organised lymphocyte structures found within the lungs following viral infection [73]. In this context, the contributions of dendritic cells (DCs) in TLS formation are not known aside from the presentation of antigen and generation of pro-inflammatory cytokines which form the inflammatory cytokine microenvironment within which TLS form. However, in the lungs of influenza infected mice *Geurts van Kessel et al* found that CD11c^{high} DCs participated in iBALT formation with the expression of CCL19, CCL21 and CXCL13 which was confirmed by the adoptive transfer of these cells [73]. Furthermore, depletion of DCs in already established iBALT structures dissolved these entities with the loss of T, B and PC – emphasising their critical contributions in providing humoral immunity in the lung.

These examples, amongst many others, emphasise the importance of TLS formation to clear pathogenic infection in synergy with the immune response generated within SLO too. Aside from the classical immune response to pathogens, TLS have also been identified in close proximity to sites of tissue transplantation.

1.5: TLS in transplantation

Whilst a critical procedure to cure certain ailments, tissue and organ transplantation present difficult immunological challenges which often result in rejection of the grafted tissue.

The generation of TLS in this context has been associated with high expression of Th17 associated cytokines, abundance of alloantigen from the grafted tissue and lack of immunoregulatory mechanisms [79-81]. At those sites TLS have been described to harbour fully mature GC able to generate PC producing anti-HLA antibodies and memory cells [79-81]. Controversial experimental reports of TLS preventing allograft tolerance are in the literature in experimental models of renal and cardiac allografts [80, 82, 83].

1.5.1: Presentation of TLS as immunological hurdles in graft rejection

Whilst kidney transplants are a lifesaving treatment for individuals, their success rate is jeopardised due to their rejection as characterised by the presence of alloreactive recipient immune cells to the site of transplantation [84]. Furthermore, the presence of PDPN⁺ vessels at sites of tissue graft were also characterised by *Kerjaschki et al*, who found an expansion of the CCL21 expressing PDPN⁺, Peripheral node addressin (PNAd⁺), Prox-1⁺ lymphatic vasculature in kidney graft biopsies. These vessels were observed to have lymphocytic infiltration comprising of Ki-67⁺ CD4 T cells and CD20 B cells. This was indicative of the presence of TLS due to the compartmentalised presence of T cells and B cells within the tissue [84].

At sites of vascular injury, the release of endothelial-derived apoptotic exosome-like vesicles (ApoExo) results in the rejection of transplanted tissue [85]. More specifically, the contributions of ApoExo in TLS formation were investigated by *Dieudé et al* who injected ApoExo in mice who had undergone allogenic aortic transplant. This study demonstrated an increased presence of TLS formation at the sites of allograft transplantation in combination with elevated levels of $\gamma\delta$ -Th17 cells and anti-nuclear antibodies. Subsequent blocking of ApoExo activity using bortezomib and the injection of ApoExo in $\gamma\delta^{-/-}$ mice resulted in the reduction of $\gamma\delta$ -Th17 cell infiltration, abundance of autoantibodies and the formation of TLS [85].

1.6: Similarities between SLO and TLS in formation and function

The molecular mechanisms underpinning the formation and maintenance of TLS within a specific tissue are not completely understood. Many efforts over the years have aimed to overlay TLS and SLO developmental factors. However, it now appears that TLS aggregation is the result of a very different process from SLO assembly. While some anatomical and developmental similarities are shared, TLS formation involves a sequence of events and signalling cascades critically different from those regulating the development of lymph node (LN), spleen, or Peyer's Patches (PP) and are summarised in table 1 [8, 23, 34, 86-88].

SLO are immunological organs which function to filter the body's lymphatic fluids and blood for pathogens and to generate an effective immune response against them. The spleen and PP provide immune surveillance to the blood and gastrointestinal tract, whilst LN are strategically situated at points of lymphatic and vasculature bifurcation around the body [86]. This anatomical positioning allows LN to screen for circulating antigens and for immune cells from surrounding tissues to drain into the lymphatic vasculature and be brought into close proximity to foreign antigens.

The efficacy with which SLO execute their immunological functions is attributed to their architectural organisation and their hallmark feature of containing specialised niches for T, B and myeloid cells [86]. The presence of these specialised lymphocyte niches allows for the efficient activation, differentiation and maintenance of lymphocytes which facilitates the generation of a specific and rapid immune response to an antigen. In order for these lymphocyte niches to exist the contribution of the underlying stromal cell compartment has been made evident.

The stromal cell heterogeneity that exists not only within SLO, but other anatomical sites, has been only recently characterised, leveraging the availability of novel platforms that enable single cell resolution. The stromal cell compartments of the murine LN have been characterised using single cell analysis by *Rodda et al*, which demonstrates the presence of phenotypically diverse stromal cell compartments with inferred unique functions [89]. In the following section I will focus on LN development and the key micro domains shaped by the underlying stromal cell compartment.

Table 1: The developmental and anatomical differences between SLO and TLS

SLO Development and anatomical features	TLS Development and anatomical features
Embryonically programmed to develop at specific anatomical locations [90, 91]	Post-natal development at immunocompromised anatomical locations in response to infection, cancer, tissue damage [11, 14, 34, 38, 92, 93]
Have an external capsule [94-96]	Are not encapsulated [10, 11]
Development and maturation take place with LT _i and LT _o interaction within the anlagen [96-99]	Development and maturation take place with inducer like cells such as infiltrating T cells and tissue resident fibroblasts. These fibroblasts undergo <i>priming, expansion</i> and <i>maturation</i> to form TLS [1, 7, 8, 10, 11]
SLOs are well networked with the vasculature and lymphatic system and transport soluble antigen and lymphocytes to sample antigen and migrate to the correct compartment [96-99]	Mature TLS generate <i>de novo</i> vasculature and lymphatics such as high endothelial venules (HEVs) which are not at the same integrity as those within SLO [9-11, 26, 100]
Have distinct organised cell compartments such as the T cell zone and the B cell zone [96-99]	Have distinct organised T and B cell zones which may not be as organised as those within the SLO [7, 8]

Table 1: Summary of the key differences and similarities between SLO and TLS development and anatomy that has been currently defined in the literature.

1.7: SLO Development: the crosstalk between Lymphoid Tissue inducer (LTi) and Lymphoid Tissue organiser (LTo) cells

The underlying crosstalk between resident stromal cells and infiltrating immune cells is critical for the exchange of correct molecular cues during SLO development. In a similar manner, during TLS formation the resident mesenchyme interacts with infiltrating immune cells which modify the stromal compartment such that TLS development, maturation and organisation can take place. These developmental patterns are aligned during SLO and TLS formation. For this reason, I will explore SLO formation in this chapter section prior to the cellular interactions in TLS development.

1.7.1: Development of LNs in normal physiology

The development of LN in mice and humans occurs during the embryological and early post-natal stages of life [90, 101, 102] and is driven by the prenatal interaction between a specific population of CD45⁺ CD3⁻ CD4⁺ CCR7⁺ CXCR5⁺ RANK⁺ RANKL⁺ IL-7Ra⁺ LTα₁β₂⁺ RORγ⁺ (CD4⁻ RORc⁺ in humans) lymphoid tissue inducer cells (LTi) and non-haematopoietic mesenchymal or endothelial derived CD45⁻ CD3⁻ CD4⁻ PDGFRα⁺ PDGFRβ⁺ LTβR⁺ lymphoid tissue organiser cells (LTo) of the LN anlagen [2, 90, 95, 103-105]. The events leading to the formation of the LN were first proposed by Florence Sabin in 1902 who demonstrated that the lymphatic system developed after the blood system [106]. Here, Sabin suggested that budding of lymph sacs from venous endothelial cells during embryogenesis and subsequent LN anlagen establishment occurred as a result of surrounding PDPN⁺ CD31⁺ lymphatic endothelial cell (LEC) populations prior to colonisation with LTi cells. In this model, PDPN⁻ CD31⁺ blood endothelial cells (BEC) within the cardinal vein in mice at embryonic day 9 (E9) – E9.5 start differentiating into LEC upon expression of the lymphatic vessel endothelial hyaluronan receptor 1 (LYVE-1). Sabin then described that the co-expression of the transcription factor Prox-1⁺ with LYVE-1 led to the budding of BEC and their subsequent differentiation to LEC at E11-E12. Newly formed LEC migrate to predetermined sites in the body in a CCL21, VEGFC and VEGFD dependant manner [107]. This leads to the formation of the lymph anlagen which begins its maturation at E12-E13 with colonisation with LTi cells [106] The CD4⁺ CD3⁻ IL-7Rα⁺ LTi cells within the mouse originate from the fetal liver at E12. This population of cells was demonstrated to migrate through the circulatory system in response to chemotactic cues at sites of LN development [99]. The profile of the human homologue of murine LTi includes CD4^{-/low} CD56⁻ IL-7Rα⁺ and the transcription factor retinoic acid receptor - related orphan receptor

C (RORC)⁺ cells. The ability of lymphocytes to move in response to a lymphoid chemotactic gradient has been demonstrated by the lack of the majority of LN in CXCR5^{-/-} and CXCL13^{-/-} mice and the identification of smaller, less populated, and disorganised LN in CCR7^{-/-} and CCL19, CCL21 deficient mice [99, 108]. These observations demonstrate the dependency of these lymphocytes on the chemokines to migrate to the LN anlagen and their contribution to the anlagen development and maturation. In subsequent studies after Sabin, investigation in Prox1^{-/-} mice demonstrated a lack of lymph sacs but the accumulation of LT_i at sites of the LN anlagen [106]. This suggested that even in the absence of differentiated LEC, LT_i were still receiving chemotactic signals guiding their migration to the site of LN development. To this end, it was then suggested that LT_i recruitment was driven by the LTβ_r mediated expression of CXCL13 from mesenchymal derived LT_o cells within the LN anlagen [109]. In this instance however, it was demonstrated that neurons innervating the LN anlagen were able to regulate the expression of CXCL13 independent of LTβ_r signalling as CXCL13 expression was still detected in *Lta*^{-/-} mice [110, 111]. CXCL13 was demonstrated to be a downstream product of the retinoic acid (ReA) signalling cascade [111]. The LTβ_r and ReA signalling cascades are two critical, amongst others, cellular mechanisms that result in stromal cells acquiring an immunomodulatory phenotype.

1.7.2: Retinoic Acid Signalling: the initiator signal of SLO anlagen maturation

As mentioned previously, the colonisation of the LN anlagen with CXCR5⁺ LTi is regulated by the production of the ReA dependant lymphoid chemokine CXCL13 [111]. The molecule ReA is a metabolite generated by the enzymatic processing of vitamin A by the enzyme retinaldehyde dehydrogenase (Raldh2) [94, 96, 111] in nerves localised close to the LN anlagen and was demonstrated by *Meibus et al*, as an essential molecule inducing the expression of CXCL13 by anlagen resident LTo [112].

The ability of ReA to induce CXCL13 expression was demonstrated at embryonic day E13.5 in mesenchymal cells from *Lta*^{-/-} embryos, that, once treated with ReA were able to express CXCL13 in a retinoic acid receptor B (RARβ) receptor manner [112]. Analysis of *Raldh2*^{-/-} mice demonstrated an absence of LN anlagen at E13.5, suggesting that the metabolism of Vitamin A to ReA at early embryonic LN development is non-dispensable for the infiltration of the LN anlagen by LTi to drive LN maturation. The early expression of CXCL13 initiated by ReA initiates the infiltration of LTi to the LN anlagen such that the critical cellular crosstalk with the LN anlagen resident LTo ensues leading to LN maturation [112]. The role of ReA in the formation of specific stromal cell populations known as FDCs has also been investigated and I will discuss further in this chapter.

1.7.3: LT β R Signalling: the maturation signal of resident SLO mesenchyme

One of the key interactions between LT i and LT o is mediated by the engagement of the LT β R on the LT o with the cell surface bound LT $\alpha_1\beta_2$ on LT i [14, 23, 87, 88, 94, 96, 98, 101, 105, 113]. The engagement of LT β R and LT $\alpha_1\beta_2$ results in the downstream activation of the canonical and non-canonical nuclear factor kappa-light-chain-enhancer of activated B cells (Nf κ B) pathway [13, 104, 109, 114]. The downstream effects of each arm of this signalling pathway result in the expression of cell adhesion molecules such as ICAM-1 and VCAM-1 or the lymphoid chemokines CCL21, CCL19 and CXCL13 [13, 104, 109, 114]. The importance of LT β R signalling in LN development was highlighted in LT β R^{-/-} mice with the observation of the lack of any LN present [23, 94, 96, 98].

The downstream effects of the LT β R signalling cascade involves the translocation of the canonical Nf κ B RelA/p50 complex from the cytosol to the nucleus which induces the gene expression of ICAM-1 and VCAM-1, which, amongst others, are markers of stromal cell activation [13, 104, 109, 114]. The expression of ICAM-1 and VCAM-1 on LT o enables the tethering of infiltrating immune cells expressing lymphocyte function associated antigen (LFA-1⁺) or very late antigen 4 (VLA-4⁺). This facilitates the interaction between the two cell populations such that the exchange of key signals can take place [23, 94, 96, 98]. On the other hand, the alternative Nf κ B RelB/p100 mediates the expression of the lymphoid chemokines CXCL13, CCL19 and CCL21 which orchestrate the migration and compartmentalisation of CXCR5⁺ B and CCR7⁺ CD4⁺ T cells in the LN ([23, 94, 96, 98]. The migration of infiltrating immune cells in response to chemotactic cues of CCL19 and CCL21 of the T cell area and CXCL13 of the B cell area gives rise to the hallmark segregated T and B cell domains [23, 94, 96, 98].

We will further discuss how TLS development, and more specifically differentiation of TLS stromal cells, has a different reliance than SLO on the signals mediated by LT β R signalling in section 1.10 and 1.11. Additionally, we also explore how these differences are critical to shape TLS biology and carry a series of therapeutic implications to address TLS associated diseases.

1.7.4: RANKL Signalling: a second maturation signal of SLO mesenchyme

In addition to $LT\beta$ r signalling, a second signal which was identified to be important in driving the maturation of the LN anlagen was of the receptor activator of $Nf\kappa B$ (RANK); also known as TNFSF11 with its ligand, RANK-L [115]. It has been demonstrated that expression of RANK-L and $LT\alpha_1\beta_2$ on LTi and the subsequent engagement of these molecules, with their cognate receptors RANK and $LT\beta$ r expressed on LTo cells, mediates the maturation of the stromal organiser cell population [115]. This maturation is recognised in one way, by the increased expression of RANK-L on the LTo which by the establishment of a positive feedback loop, amplifies the expression of RANK and $LT\beta$ R [115]. This positive feedback system increases the expression of lymphoid chemokines i.e., CXCL13, which amplifies the recruitment of $CXCR5^+$ cells [115]. Aside from governing the migration of LTi and lymphocytes, lymphoid chemokines also promote the expression of cell surface $LT\alpha_1\beta_2$ on LTi [108]. Further amplifying the continued expression of the lymphoid chemokines and resulting in full stromal cell maturation.

1.7.5: The role of T cells in the final steps of LN maturation

In the final stages of mouse LN development, migration of mature lymphocytes into the maturing LN anlagen generates the formation of adult LN. The infiltration of immune cells to the LN occurs through movement within lymphatic vessels (LV) or HEVs which are found integrated within a SLO [94]. HEVs are formed from cuboidal endothelium and permit the entry of L-selectin^{high} lymphocytes into LNs from the blood upon interaction with their ligand PNAd [103]. On the contrary, the trafficking of lymphocytes into the spleen occurs along a network of perivascular tracks as HEV are absent in this SLO [116]. In a study by *Chauveau et al*, it was demonstrated that circulating T cells migrate on perivascular tracks that are coated with a subset of ERT7⁺ CCL21⁺ stromal cells i.e. T cell zone FRC [116]. This cellular interaction between the pericytes and CCL21⁺ stromal cells formed a T cell highway enabling their organised entry and movement within the spleen [116].

1.7.6: Identification of IL-7R as a key signalling pathway in mesenchyme maturation

It is understood that T and B cells assume the role of LTi in post-natal LN development, however of the two lymphocyte subsets, T cells have been shown to influence LN organisation more than B cells [96, 101]. This was made evident in mice carrying a natural mutation known as *paucity of lymph node T cells (plt)*, characterised by defective expression of CCL19 and CCL21, and the presentation of small and disorganised LN, due to the failed recruitment of the T cells within the anlagen and later in post-natal life [96, 101]. Furthermore, the IL-7R and IL-7 interaction was demonstrated to be critical during a unique stage in post-natal LN development [117]. To this end, *Il2ry*^{-/-}, *Rag2*^{-/-} mice presented aggregates of LTi at E15.5 however, this LN primordia in *Il2ry*^{-/-}, *Rag2*^{-/-} was lost by E18.5 [117]. This suggests that within this 7-day window there was a signal required carried by infiltrating lymphocytes which allowed for the maintenance and persistence of the development of the postnatal LN. The transfer of T or natural killer (NK) cells into *Il2ry*^{-/-} or *Rag2*^{-/-} incurred LN rescue whilst the transfer of B cells did not have this effect. Immediately, this suggests that a commonality exists between T and NK and this was the expression of the IL-7 receptor [117]. Stromal cells have been demonstrated to be a source of IL-7, a key T cell survival factor and the engagement of this signalling axis results in the production of growth factors which support survival, growth, and maintenance of the LN stromal cells. Thus, T and NK lymphocytes may replace the role of embryonic LTi cells at post-natal development and continue the maturation of the LN mesenchyme. The formation of these lymphocyte specific regions within SLO creates specialised compartments promoting the activation, differentiation and survival of T and B lymphocytes.

1.8: Stromal cell heterogeneity in SLO supports the maintenance of T and B cell homeostasis

Primarily, SLO function as immunological filters, providing an adequate microenvironment facilitating the interaction between naïve immune cells in search for their cognate receptor antigen, subsequently resulting in their complete activation and differentiation [23, 86, 88, 94, 96]. Whilst soluble antigens can enter SLO through the lymphatic vasculature, others are transported in conjunction with APCs such as DC [23, 86, 88, 94, 96]. By bringing immune cells and their potential cognate Ag into close proximity, SLO are local outlets housing adaptive immune responses generating activated and differentiated effector immune cells and memory cells. Aside from this, SLO provide an additional immunological screen in the context of tolerance [118]. Where self-reactive immune cells may have escaped the initial purge during the establishment of central tolerance in the thymus [86, 119-121]. The ability of SLO to execute these functions is attributed to the range of fibroblast subsets, which reside in the hallmark T cell area and B cell follicle – which contains the GC reaction [122, 123]. The exclusivity of these fibroblasts creates specialised micro domains within the SLO, which influences the migration, retention, activation and survival of T and B lymphocytes. Using single cell RNA sequencing methods, the heterogeneity within the murine LN mesenchyme has been phenotypically characterised with some attribution to their possible function in this SLO.

1.8.1: Fibroblastic Reticular Cells (FRC) and Follicular Dendritic Cells (FDC)

Lymphocyte segregation in diverse areas is accompanied with the presence of specific stromal cells which are able to support the activation, maturation, and survival of these cells as they infiltrate the LN. Whilst heterogeneity within the stromal cell population has always been acknowledged, recent advances in transcriptomic methodologies has uncovered distinct subsets of fibroblasts with unique phenotypical and functional signatures. Current single cell RNA sequencing analysis undertaken by *Roddha et al*, found several fibroblast populations within a mouse LN. To this end I will discuss the current stromal cell landscape within the murine LN and how T and B lymphocytes are supported within their unique compartments.

1.8.2: T Cell Zone: Single cell characterisation of murine LN FRCs heterogeneity

Characterised as CD45⁻ gp38⁺ CD31⁻ CCL19⁺ CCL21⁺ IL-7⁺ ER-TR7⁺ fibroblastic reticular cells (FRCs), also known as T zone reticular cells (TRCs), exclusively reside in the T cell cortex of the LN [30, 89, 124-130]. Within the umbrella of FRC, multiple subsets were characterised by single cell RNA sequencing which characterised and described 9 non-endothelial stromal cell subsets within an immunised mouse LN [89]. Of the subsets that were identified, the T cell zone FRCs were distinguished as being the classical CCL21⁺ CCL19^{high} IL-7⁺ stromal cells or alternatively CCL21⁺ CCL19^{low} IL-7⁺ stromal cells [89]. The CCL21⁺ CCL19^{high} IL-7⁺ subset fulfils the classical function of recruiting and maintaining CD4⁺ IL-7 α ⁺ CCR7⁺ naïve T cells and DCs in a L-selectin (CD62L), VLA-4, LFA-1 guided manner from HEVs [124]. These FRCs also form the substratum for CCR7⁺ naïve T cells as they navigate the paracortex domain in a CCL19/CCL21 gradient in search for their cognate receptor antigen in the context of DCs [30, 89, 124-130].

The CLEC-2⁺ DCs also utilise this FRC network to migrate by binding to the PDPN⁺ FRCs [126] thus increasing the chances of cognate antigen and receptor recognition by the naïve T cells. The other subset identified by single cell RNA analysis was characterised as CCL21⁺ CCL19^{low} IL-7⁺ FRCs which have been associated with CXCL9 and CXCL10 expression [89]. These cytokines are involved in the expression of MHC genes, and due to the nature of this specific FRC cluster it has been suggested that this population of FRCs perhaps function to induce tolerance in CD4 and CD8 T cells in the periphery.

The ability of FRCs to participate in the maintenance of immunological tolerance has been alluded to previously with the identification of FRC being able to present MHC II complexed self-peptides of DC origin [118, 126]. The presentation of self-antigen by FRC results in T cell

anergy or apoptosis leading to the establishment of peripheral tolerance [118, 126]. Aside from presenting DC derived self-antigens, *Gardner et al*, have identified a stromal subset that mainly resides in the T cell area assigning them as extrathymic Aire-expressing cells (eTACs) [118]. These eTACs express the autoimmunity regulator (*aire*) gene which is classically found in the thymus and is involved in inducing central tolerance [118].

Additional distinct FRC subsets were identified within the follicle T zone interface and were characterised as cholesterol-25-hydroxylase (Ch25)⁺ CCL19^{low} with high differential gene expression of the lymphoid chemokines CCL21, CXCL13 and lymphocyte survival factors IL-7 and BAFF [89]. To this end, this FRC compartment displays the expression of appropriate molecules to support the migration, retention and survival of both T and B cells.

1.8.3: The Hippo pathway: fate determining signal of FRCs

The fate determining signal driving the acquisition of the SLO mesenchyme to a FRC phenotype was recently demonstrated by *Choi et al* to be as a result of the Hippo signalling pathway [131]. *Ludewig et al* showed that by modulating the expression of the YAP/TAZ transcription factors which are the downstream effectors of the Hippo pathway, the ability of CCL21 and CCL19 expression by FRC in YAP/TAZ^{FRCΔ} mouse LN was impaired. This was followed by a loss of lymphocyte compartmentalisation in the LN and demonstrated that the Hippo pathway is responsible for the expression of T cell chemokines. In addition, the induction of Hippo pathway signalling in a C3H10T1/2 cell line resulted in the loss of an adipogenic signature and expressed enhanced FRC markers, thus emphasising the role of the Hippo pathway in driving mesenchymal cells to an FRC phenotype. *Choi et al* also demonstrated that the Hippo signalling pathway preceded the LTβr pathway as the shuttling of the YAP/TAZ proteins from the nucleus to the cytoplasm was observed following engagement of the LTβR pathway. This resulted in the suppression of the Hippo signalling pathway in the FRC but the maintenance and continued maturation of the FRC phenotype.

1.8.4: The B cell follicle contains stroma which supports the Germinal Centre (GC) Reaction

Neighbouring the T cell area of the LN is the B cell follicle. Within the B cell area within a SLO it is possible to identify both primary and secondary B cell follicles [92, 101, 133]. The primary B cell follicles are mainly inhabited by CXCL12 producing reticular cells (CRCs) [89] whilst the secondary follicles, which contains the GC reaction are populated with CD45⁻ gp38⁺ CD21⁺ CD35⁺ BP-3⁺ CXCL13⁺ FDCs [133, 134]. FDC are entitled as such, due to their DC like appearance and the ability to present intact antigen to naïve B lymphocytes through their fragment crystallizable (FC) and complement receptors (Cr2) in complexes known as iccosomes [123]. FDC produce the CXCR5⁺ B cell oxysterol chemokine CXCL13 and the B cell survival factor, BAFF [22, 40] which drives the formation of the B cell specific niche, in conjunction with T follicular helper (Tfh) cells [123]. The polarisation of the B cell area is mediated by CXCL13, and its importance in this was made evident from observations made from CXCL13^{-/-} mice displaying not only reduced B cell migration to the B cell follicle, but also the lack of FDC themselves [102, 108]. This was attributed to two key points. Firstly, CXCL13 can favour the expression of LT α β ₂ on the surface of LT_i and B cells, an important signalling molecule in stromal maturation. The importance of LT β r signalling in FDC maturation was further highlighted with LT β R blockade using LT β R-Ig in adult mice, which resulted in the loss of functional GC FDC [40]. This demonstrated that the engagement of LT β r signalling is critical for FDC development. Secondly, CXCL13 is also responsible for the migration of CXCR5⁺ B cells to the B cell follicle. The absence of CXCL13 prevents the infiltration of B cells which physically inhibits the crosstalk between B cells and B cell supporting stroma to take place [Ansel, 2002 #194, 102]. Additionally, this highlights the dependency of the fibroblasts on infiltrating immune cells to

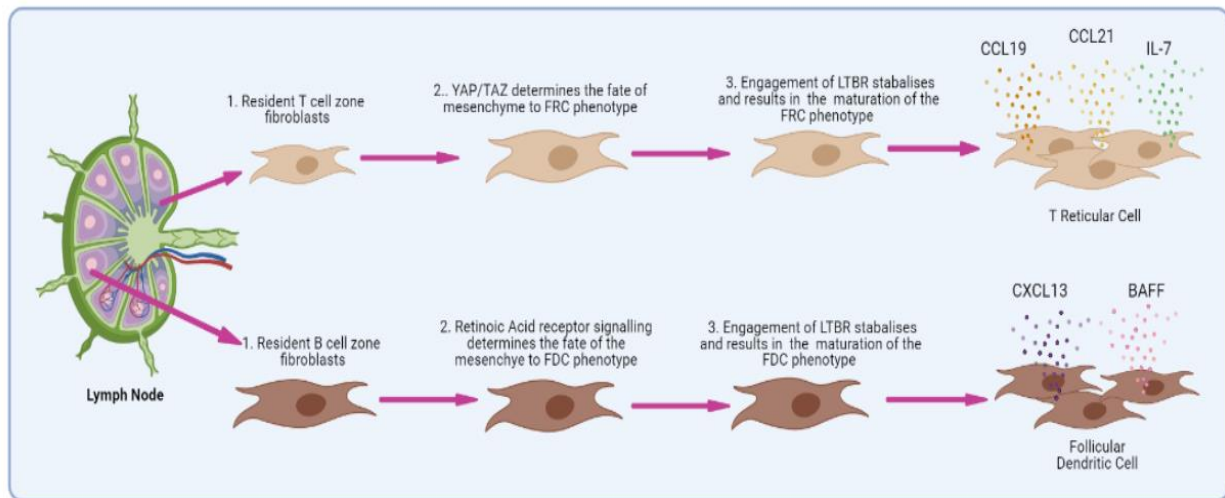
maintain their matured status and effector function. Thus, emphasising the importance of investigating the bi-directional crosstalk between lymphocytes and stromal cells.

1.8.5: The Retinoic Acid Pathway: driver of the FDC phenotype

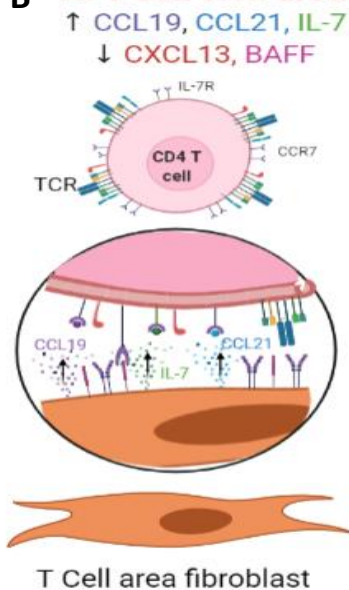
Whilst the LT β R signalling cascade has been demonstrated to result in CXCL13 expression the fate determining signal for differentiation of mesenchymal precursor cells to FDC in the murine LN has defined by RA signalling [112]. *Mebius et al* demonstrated that embryonic mesenchymal cells treated with exogenous RA and subsequently with LT β r agonists retained their CXCL13 expressing FDC phenotype and did not express the FRC chemokines CCL19 and CCL21 [112]. Expanding on the previous finding of the importance of the RA signalling in LN development, *Koning et al* further demonstrated that the source of RA signalling in post-natal development of FDCs were neighbouring endothelial cells [112]. This suggests that engagement of the RA signalling cascade drives the stromal cell precursor to a FDC identity and the engagement of LT β r signalling is necessary for the further maturation of the FDC not to determine its cell fate. Conversely, engagement of LT β R signalling prior to RA signalling resulted in the upregulation of CCL19, CCL21 and IL-7 of embryonic mesenchymal cells – a chemokine signature aligned with T cell supportive FRC stroma [112]. In this study, the Hippo pathway was not explored in the context of understanding its contribution in FDC formation (this is summarised in Fig 1). However, these study highlights how different fate determining signals contribute to the acquisition of the specific lymphoid phenotype of the specific fibroblast populations in SLO and may warrant their contribution in TLS associated fibroblast activation and differentiation.

Figure 1: Summary of the different fate determining signals forming the individual T cell area fibroblast (TRC) and the B cell area stroma (FDC).

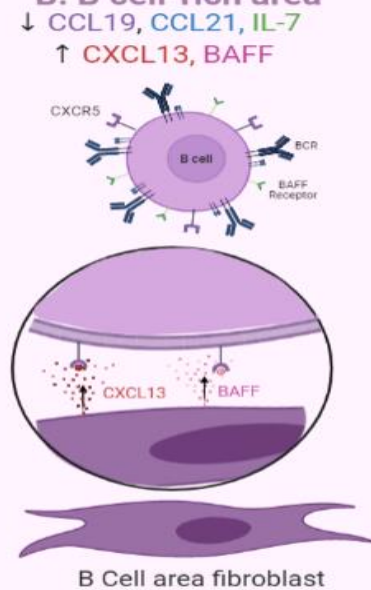
A



B A. T cell-rich area



B. B cell-rich area



Created in BioRender.com bio

Figure 1: A: Summary of the two different fate determining pathways which have been explored by *Choi et al* and *Koning et al*. The top panel summarises the engagement of the YAP/TAZ pathway prior to LTβR engagement induces the FRC phenotype with the expression of CCL19, CCL21 and IL-7. The bottom panel demonstrates the FDC phenotype is determined by the engagement of the RA signalling cascade prior to LTβR engagement resulting in the expression of the B cell chemokine CXCL13 and their survival factor BAFF. **B:** Diagrammatic summary of the individual phenotypical properties of the T cell area and B cell area stromal cells (Images created using *Biorender*).

1.9: The GC reaction

To generate an effective humoral response to an antigen, antibody secreting, and memory B cells must be generated and expanded rapidly and effectively. Microstructures termed GCs are distinct and specialised regions within the secondary B cell follicle that facilitate this process, supporting B cell somatic hyper mutation, clonal expansion, and affinity-based maturation to generate high affinity antibodies to a particular antigen (Fig 2) [123, 135]. Several key molecules such as, B-cell lymphoma 6 (BCL-6), transcription factor E2A (E2A), Activation-induced cytidine deaminase (AID), Cellular myelocytomatosis oncogene (c-Myc) have been demonstrated to be essential in the development and maintenance of a functional GC reaction [122]. The dysregulated activity of these proteins has been reported to alter the functional dynamics within the GC resulting in the formation of B cell malignancies such as Burkitt's lymphoma [122]. Similar to what has been described in the broader context of the LN, GC formation and function is supported by a complex and organized network of stromal cells which contribute to the functional organisation and compartmentalisation of this microstructure into its dark zone (DZ) and light zone (LZ).

1.9.1: The DZ of the GC

The development of a GC occurs within the B cell follicle following a T cell-dependant immune response and begins with the formation of the DZ near the T zone followed by the formation of the LZ within a SLO. The DZ and LZ within the GC are histologically distinct and serve an individual purpose during the mature GC reaction [122]. During a GC reaction, activated naïve B cells enter the DZ, and are characterised as CXCR4^{high}, the receptor for the migratory signal CXCL12, and CXCR5^{low} centroblasts [122]. High in BCL-6 expression, the DZ is an active site of centroblast proliferation and genetic modification; occurring in the form of somatic hypermutation (SHM) and class switch recombination (CSR) which generates B cell receptor diversity [18, 38, 122, 123, 135, 136]. This genetic diversity is mediated by the enzyme AID [18, 38, 122, 123, 135, 136]. The function of AID to generate genetic diversity is tightly regulated, as the introduction of point mutations and double strand DNA breaks resolved as SMH or CSR have the potential to be precursors of oncogenic genomic instability. The elevated expression of BCL-6 provides centroblasts with the “tolerance” towards genomic editing by suppressing the activation of cellular distress signals i.e. activation of p53. Upon sufficient proliferation and generation of BCR genetic diversity, centroblasts migrate from the DZ to the LZ in a CXCL13 dependant manner and are identified as CXCR4^{low} CXCR5^{high} centrocytes [122].

1.9.2: The LZ of the GC

The LZ contains phenotypically different B cells identified as centrocytes as they migrate out of the DZ in response to a CXCL13 gradient, that, within the LZ undergo CSR and high affinity maturation [122, 135]. FDC reside specifically in the LZ of a GC in the secondary B cell follicle and present intact antigen to LZ B cells to facilitate the selection of high antibody affinity producing B cells [10, 90, 122, 135]. The process of high affinity maturation results in the retention and subsequent clonal expansion of B cells that are able to bind antigen presented by FDC at the correct affinity, subsequently allowing them to acquire co-stimulatory and survival signals compared to B cells that do not express a high affinity antigen receptor [122, 135]. Centrocytes that are selected can achieve terminal differentiation to positive regulatory domain containing 1 (PRDM-1)⁺ PC or memory B cells [122, 135]. The movement of B cells through the GC reaction is not linear, as B cells from the LZ are able to recirculate to the DZ to undergo further rounds of SMH in order to generate a BCR with the optimal affinity for a particular antigen.

1.9.3: The heterogeneity of FDC within the GC reaction

Recent transcriptomic analysis has characterised the stromal landscape supporting the GC reaction which identified two FDC subsets which were determined by the increased expression of Cr2 or CXCL12 [136]. In this study by *Pikor et al*, transcriptional and cell specific gene deletion techniques were used to characterise the stromal landscape supporting germinal centre responses in LN of naïve or immunised Cxcl13-Cre/dTomato⁺/EYFP mice. In this study, they identified 7 Milk fat globule-EGF factor 8 (Mfge8)⁺ CXCL13⁺ murine fibroblast subsets in the LN. *Pikor et al* demonstrated that these subsets pre-exist in non-inflammatory conditions and increased in number in response to inflammation. Furthermore, they characterised LZ and DZ FDC from immunised mice and identified a differential transcriptional status between the subsets. This included the elevated expression of the chemokines CXCL1, CXCL16 and IL-6. The topology of LZ and DZ FDC was governed by CXCL12 expression as it was observed that CD21⁺ CD35⁺ LZ FDC in CXCL12^{-/-} mice were scattered throughout the GC as opposed to being concentrated in the LZ as per in wild type mice [136]. This was attributed to the fact that CXCR4⁺ B cells carrying surface LTα₁β₂ were unable to migrate in response to CXCL12 and therefore not engage in cellular crosstalk with the B cell area reticular cells. Thus, resident B cell stromal cells were not remodelled to contribute to the normal functioning and organisation of the GC.

1.9.4: The contribution of CD4⁺ Tfh cells in the B cell follicle

Aside from the FDC which produce CXCL13, the B cell survival factor BAFF and present intact antigen, the GC reaction is also shaped by the presence of a subset of CD4⁺ T cells identified as CXCR5⁺ ICOS^{high} PD-1^{high} BCL-6⁺ CD40⁺ IL-21⁺ Tfh in the LZ [122, 123]. Tfh development within SLO is mediated by the completion of the 3-step model involving the interaction of antigen loaded DC at the extra follicular space, followed by TCR engagement with peptide on GC B cells at the T-B cell border and their subsequent migration into GCs in SLO as they mature [137]. This CD4 T cell subset actively participates in the selection of high affinity antibody producing B cells by producing IL-21 [137] and CXCL13, stimulating further migration and retention of centrocytes [138]. This interaction of B cells with FDC and antigen specific Tfh results in clones which terminally differentiate to PC or memory B cells [137, 138]. In addition, their regulatory counterpart, T follicular regulatory (Tfr) cells do not produce the helper cytokines IL-21 and IL-4 [139] and function to monitor GC output by preventing the movement of autoreactive or cancerous B cells generated within the GC. They also limit Tfh numbers within GC to prevent the prolonged survival and maturation of defective B cells [139].

In essence, within the B cell compartment of a SLO the presence of specialised stromal cells and CD4 T cells work together to enable the maturation and survival of infiltrating B cells. This interaction between B cell stroma and infiltrating B cells and resident CD4 T cells emphasises the importance of the different molecular interactions taking place which participate in the activation, differentiation, and survival of B cells.

Figure 2: Diagrammatic representation of the GC and the overarching chemokine signature and cell populations within the DZ and LZ

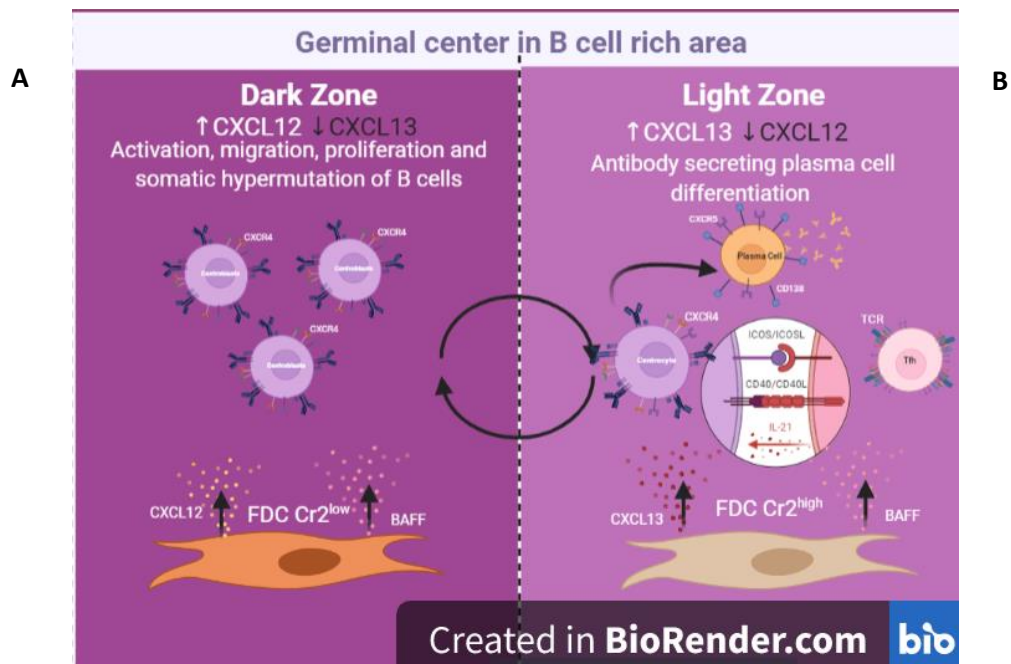


Figure 2: Diagrammatic summary of the GC with the DZ and LZ and the individual functions these regions execute. **A:** The DZ harbours CR2 low fibroblasts and is the site where centrocytes undergo various phenotypical changes prior to their migration to the LZ. **B:** The LZ contains CD2 high FDCs as demonstrated by *Pikor et al* where centroblasts from the DZ acquire a centroblast phenotype and with Tfh receive the necessary signals to become PCs (Images created using *Biorender*).

1.10: TLS organisation and function mirror SLO in non-immunological organs

The concept of TLS was introduced by Louis Picker and Eugene Butcher in 1992, as the third lymphoid compartment of the immune system [140]. TLS are anatomical entities comprised of immune cell aggregates present within non-immunological organs that can be induced in an inflammatory microenvironment or in response to exogenous stimuli. These immune cell aggregates are assigned individual identities depending on their location. Mucosa associated lymphoid tissues (MALT) comprises the lymphocyte rich clusters present in the gastrointestinal tracts (GI), the fat associated lymphoid clusters or FALCs form in the adipose tissue of the mesenteries, the inducible nasal and branchial associated lymphoid tissues (iNALT and iBALT) encompass the aggregates detected in the respiratory tract from the nose to the lung parenchyma [16, 141-143]. In the context of inflammation and autoimmunity, TLS are largely considered pathogenic as their persistence in the inflamed tissue associates with progressive disease, local generation of autoantibodies and tissue damage. In the context of cancer, TLS role is more debated and recently high-profile publications suggest that TLS formation in proximity to malignancy represent a positive prognostic factor in response to therapy.

1.10.1: Comparing and contrasting TLS and SLO architecture and function

TLS lack some of the architectural features that define primary and SLO which include the absence of a capsule and reduced structural organisation [10, 92] yet provide a suitable microenvironment for activated T and B cells to reside within tissue lacking immunological competency. The anatomy of TLS in its cellular composition is highly variable and plastic, likely reflecting different maturational stages. Items consistently described include dense B cell aggregation and segregation between the T and B cell zone, expression of lymphoid chemokines and specialisation of TLS associated vasculature [10, 11]. The presence of histologically defined GC with hallmarks of B cell proliferation and differentiation of follicular dendritic cells is typical of the largest and more mature aggregates. The spatial development of TLS has been observed in proximity to vasculature or epithelial structures, often associated with pericytes, smooth muscle cells and myofibroblasts [11] [1, 3, 92], thus suggesting that the resident vasculature facilitates the extravasation of lymphocytes from the circulation into the inflamed or infected organ and that pericytes may play a role in TLS establishment [144].

The molecular mechanisms which initiate the formation of TLS and maintains their persistence within a tissue are not completely understood. However, similarly to SLO the role of the underlying fibroblast compartment is emerging as a driver in TLS establishment. It has been demonstrated using *in vitro* and *in vivo* techniques that tissue resident fibroblasts, under inflammatory conditions and in response to pathogenic insults and pro-inflammatory cytokine *priming*, are capable of acquiring an *immunofibroblast* phenotype [8]. This consists in the upregulation of adhesion molecules, and the ability to produce chemo attractants classically detected in SLO such as CCL19, CCL21 and CXCL13 [8]. This phenomenon is followed by

circulating T and B cell recruitment, cross talk mediated by lymphotoxin, stabilization of the fibroblast phenotype in the aggregates and full establishment of the TLS in the tissue.

In the following section I will discuss the studies undertaken to understand the nature of TLS development and the molecular cues currently known to initiate and drive this process.

1.11: TLS Development: the crosstalk between Lymphoid Tissue initiator cells (LTin) and the resident stromal mesenchyme

The formation of TLS shares some similarity to SLO development, however the sequence of events and signalling cascades that are involved differ from that regulating SLO development. In SLO, the mesenchyme compartment does not have an assigned functional identity other than to become the supportive stroma of an immunological organ [90, 95, 96, 145-147]. Contrastingly, TLS formation occurs within organs harbouring postnatal differentiated stroma ascribed to provide anatomical support and form the extracellular matrix of the organ [3, 8, 11, 92, 145, 148-150]. However, it has become clear that postnatal fibroblasts inhabiting mature organs can retain the plasticity to alter their functional phenotype. In this context Krausgruber et al demonstrated that fibroblasts possess the epigenetic machinery poised to transcribe genes downstream of proinflammatory cytokine stimulation [151].

Whilst SLO development is programmed embryonically [90, 95, 96, 145-147] TLS develop in postnatal life in response to molecular cues, mediated by pathogen associated molecular patterns (PAMPs), pro-inflammatory cytokines or cellular contact with lymphocytes largely absent during embryonic lymphoid neogenesis [8, 10-12]. The secretory cues, such as proinflammatory cytokines belong to the tumour necrosis factor (TNF) and interferon (IFN) family and may be sourced from neighbouring stromal cells responding to pathogenic insult as well as infiltrating lymphocytes at the site of infection or inflammation. In this manner, the ability of the resident fibroblasts to respond to these inflammatory cues generates a heterogenous landscape of TLS associated stromal cells which we have explored in our current work. Independent of our work, this fibroblast heterogeneity has been addressed in the context of cancer and inflammation using the novel tool of scRNAseq.

1.12: The heterogeneity of fibroblasts: conserved yet unique in different disease contexts

The availability of advanced high throughput techniques has highlighted novel and critical roles for fibroblasts in diseases, not limited to the spatial arrangement of neighbouring structural cells and immune cells. This has led to the exploration of this cell population and their emerging immunomodulatory properties in different physiological contexts such as cancer and inflammation which I will explore further.

1.12.1: scRNAseq profiling of CAFs in different types of cancers

The multiple functionality and heterogeneity of fibroblasts in cancer or cancer associated fibroblasts (CAFs) has been deconvoluted using scRNAseq [152]. *Davidson et al* described 3 distinct CAF populations and the cellular machinery they possessed to support melanoma growth in an *in vivo* model. The different CAF populations were referred to individually as immune, desmoplastic, or contractile based on their differential gene expression profile [152]. Characterising the CAF landscape, *Davidson et al* were able to elucidate the potential functional abilities of the specific immune stroma population in ensuing immune response within the cancer. The exploration of a curated receptor-ligand scRNA database identified the receptor ligand axis of complement component C3 – which strongly demarcated the “immune stroma population” and its receptor C3aR, expressed on macrophages. They were then able to pharmacologically interfere with this signalling axis which, not only disrupted immune cell infiltration, but slowed tumour growth *in vivo*. This observation emphasised how the immune-stromal cell crosstalk can dramatically influence tumour progression.

Additionally, the analysis of multiple myeloma biopsies and matched healthy controls using scRNAseq identified the presence of 2 inflammatory mesenchymal stromal cell (iMSC) populations, which were absent in the healthy tissues, and were enriched for genes associated with tumour survival and immune modulation [153]. Furthermore, these iMSC were demonstrated to arise in response to *in vitro* stimulation with IL-1 β and TNF- α which were inferred to derive from the CD8 and neutrophil population within the scRNAseq dataset [153]. This study elegantly demonstrated the critical differences in the microenvironment of healthy tissue vs the TME which was driven by the presence of specific fibroblast populations.

In a similar manner, in bladder urothelial carcinoma, the presence of two distinct fibroblast populations identified as PDGFR α ⁺ CAFs and Regulator Of G Protein Signaling 5 (RGS5⁺) myo-CAFs were reported [154]. PDGFR α ⁺ CAFs were associated with the expression of a plethora of cytokines and chemokines including IL-6, CXCL12, and CXCL14, a phenotype similar to immunomodulatory fibroblasts identified in inflammatory conditions [8, 43, 154-157]. Using cellular interaction inference tools within the scRNAseq data, *Chen et al* identified the critical interaction of this cell population with the CXCR4⁺ tumor-associated macrophages (TAMs).

1.12.2: scRNAseq profiling of fibroblasts in inflammatory conditions

The engagement of the C3 and C3aR complement cascade has also been described in the context of the contribution to the priming of immunomodulatory synovial fibroblasts [158]. The dependency of the inflammatory synovial fibroblasts on this signalling cascade was made evident with the observation of reduced tissue priming in the bone marrow chimera models in C3^{-/-} mice [158]. Currently, fibroblast heterogeneity has also been described to be critical in the pathogenesis of various diseases such as RA, UC, and ageing [43 Kinchen, Smilie et, Martin et al]. Two distinct fibroblast subsets have already been identified in the synovium joint of RA as either joint destructive Fap α^+ or pro-inflammatory FAP α^+ Thy1⁻ [43]. *Brenner et al* demonstrated that the developmental signals driving this pattern of fibroblast clustering involved the Neurogenic locus notch homolog protein 3 (NOTCH3) gradient which was governed by neighbouring endothelial cells [155-157]. Leveraging the comparison between human biopsies of healthy and diseased individuals these studies identified the emergence of a population of proinflammatory fibroblasts in human diseased samples, which were not identified within healthy individual tissue. *Kinchen et al* reported the presence of a population named S4, annotated as being involved in cytokine signalling pathways and positive regulation of T cell activation [156]. In support of this observation, *Smilie et al* also identified a population of “inflammation-associated fibroblasts,” which were unique to UC and shared gene expression signatures with CAFs [157]. *Martin et al*, focusing specifically on Crohn's disease, identified a conserved cellular module, termed GIMATS comprising of IgG PC, inflammatory monocytes, and activated T cells and stromal cells, characteristically defined in a subset of patients with poor responses to anti-TNF therapy [159]. Additional studies, within the skin, identified the alterations of fibroblast function in an age-dependent manner [160]. In this

study, *Solé-Boldo et al* revealed the loss of the ability of fibroblast priming with age, using scRNAseq studies skin biopsies from young and old male donors. A delayed transition from the G1/2 phase of the cell cycle in the older fibroblasts in comparison with the younger counterpart was observed. Further interrogation of this phenomenon lead to the identification of an age-dependent loss of cell-to-cell interaction and the development of a gene signature in older fibroblasts suggestive of an increased susceptibility of this population to the accumulation of reactive oxygen species (ROS) [160]. The authors suggest the possibility that this signature could contribute to the poor immunological responses generated in the older population as a result of the inability of the “old fibroblasts” to modulate the immune response in vivo. The pathogenic immunofibroblast in pSS thought to drive TLS biology have been phenotyped as PDPN⁺ CD34⁺ cells but have now been characterised using scRNAseq analysis. A summary of the differing fibroblast populations identified within immune organs, and sites of inflammation and cancer can be seen in Tables 1-3.

Table 2: Summary of the currently characterised fibroblast populations in the context of LN

Characterisation of fibroblast subsets in LN	Function
CCL19^{high} / CCL19^{low} T Zone Reticular Cells (TRC)	- CCL19^{high} TRCs express PDPN to support DC motility and LN expansion [1] [2] - CCL19^{low} TRCs at T zone interface express cholesterol-25-hydroxylase guide activated B cells, T cells and DCs to the follicle T zone interface
Cxcl9⁺ T Zone Reticular Cells (TRC)	- Cxcl9⁺ TRCs an activated subset of CCL19^{high} TRCs tolerize T cells in a resting LN and position CXCR3⁺ cells during an immune response - Cxcl9⁺ TRC DEGs were dominated by IFN-γ genes Gbp4 and Gbp5 [1]
Ennp2⁺ Marginal Zone Reticular Cells (MRC)	Ennp2⁺ MRCs influence lymphocyte antigen interaction with RANKL on follicle edge of SCS [1][3]
Itga7⁺ Perivascular Cells	- Non-adventitial PvCs support multiple functions of blood vessels - PvC subset which was characterised as Cnn1⁺ Itga7⁺ Pdpn⁺ enriched for thrombospondin (Thbs1), Il-34 and endothelin-A receptor, suggesting a role in endothelial support [1]
CD34⁺ Stromal Cells	- Recent work characterised PDPN⁺ BST1⁺ ACTA2⁺ CD31⁺ CD34⁺ SCs surrounding large vessels in the LN medullary cords as adventitial cells[4] - Enriched for CD248 expression which is reported in the resting LN capsule and required for LN expansion after immunization [1]
Nur77⁺ Stromal Cells	- Consist of activated cells from other subsets - Distinguished by expression of early response genes downstream of TLR or cytokine stimulation and included: Nur77, Fosb, Fos, JunB, Egr1, Ikba, Zfp36 [1]
Follicular Dendritic Cells (FDC)	- Two FDC subsets: FDC1 and FDC2 - LZ FDC1 = increased Cr2 and CXCL13 - DZ FDC2 = increased CXCL12 and IL-6, CXCL16 in response to inflammation [1][3][5]

Table 1: Summary of the currently characterised fibroblasts and their unique markers in within lymphoid tissues,

Table 3: Summary of the currently characterised fibroblast populations in the context of autoimmunity

Characterisation of fibroblast subsets in autoimmunity	Function
Fapα⁺ Thy1⁺ Synovial Fibroblasts in Rheumatoid Arthritis	- Phenotyped as the “immune effector” population resident within the synovial lining layer - Adoptive transfer of this population resulted in persistence in inflammatory arthritis [6]
Fapα⁺ Thy1⁻ Synovial Fibroblasts in Rheumatoid Arthritis	- Phenotyped as the “destructive fibroblasts” resident within the synovial lining layer - Adoptive transfer of this population resulted in damage of bone and cartilage with little modulation of the inflammatory signature [6]
PDPN⁺ CD34⁺ Salivary gland fibroblasts in primary Sjögren’s Syndrome	The compartmentalisation of function with ICAM^{high} VCAM^{high} expression and enrichment for the expression of the lymphocyte survival factors IL-7 and BAFF [7]
PDPN⁺ CD34⁻ Salivary gland fibroblasts in primary Sjögren’s Syndrome	Displayed enrichment for the expression of the lymphoid chemokines CCL19, CCL21 and CXCL13 [7]
SOX6⁺ colonic fibroblasts in inflammatory bowel disease	High expression of TGFβ ligands BMP2, BMP5, non-canonical Wnt ligands (Wnt5a, Wnt5b) and the secreted Wnt antagonist FRZB therefore involved in the mediation of repair and regeneration of damaged tissue [8-10]
PDPN⁺ CCL19⁺ IL33⁺ colonic fibroblasts in inflammatory bowel disease	- This stromal population was expanded in IBD disease as it was not identified in healthy colonic tissue and aligned to stromal cells underlying TLS. - The inflammation associated fibroblasts in UC had a gene signature for inflammatory gene expression of IL-1R1, TNFSF11, IL-13RA2. [8-10]

Table 2: Summary of the currently characterised fibroblasts and their unique markers in, autoimmune disease

Table 4: Summary of the currently characterised fibroblast populations in the context of cancer

Characterisation of fibroblast subsets in cancer	Function
PDPN⁺ CD34⁺ CAFs in melanoma	- Engages in immune cell crosstalk via upregulation of Cxcl12, Csf1, Ccl8, IL6ra, Il6st and components of the complement cascade - Identified stromal immune interactions between C3/CXCL12/CSF1 expressing stromal cells and macrophages which expressed C3AR1, CXCR4, CSFR1 [11]
PDPN⁺ CD34^{low} CAFs in melanoma	- An intermediate population of stromal cells that expressed PDPN and PDGFRα but low expression of acta2 and CD34 - Uniquely expressed the genes Postn & TNC and was associated with a fibrotic matrix [11]
PDPN⁻ CD34⁻ CD146⁺ myoCAFs in melanoma	Acta2^{high} contractile stromal had pericyte associated markers Ng2, CD146, Rgs5 - Also shared the expression of Col1a1, Cola1a2 with the neighbouring PDPN⁺ fibroblasts [11]
PDGFRα⁺ IL-6⁺ CAFs in bladder carcinoma	- PDGFRα⁺ express various cytokines and chemokines including CXCL12, IL6, CXCL14, CXCL1, CXCL2 and were termed inflammatory CAFs iCAFs had enrichment of cytokine-cytokine receptor interaction pathway - Also had increased proliferation ability as compared to non inflammatory CAFs [12]
PDGFRα⁻ RGS5⁺ myoCAFs in bladder carcinoma	RGS5⁺ stromal cells have been characterised to be similar myoCAFs in melanoma [12]
Sox9⁺ developmental CAFs in breast cancer	- dCAFs distinguished by the expression of genes related to stem cells Scrg1, Sox9, Sox10 - Found during early developmental stages of cancer involved in the development and morphogenesis of TME [13]

Table 3: Summary of the currently characterised fibroblasts and their unique markers in, cancer

1.13: Fibroblasts of mice and men: identifying conserved populations across species

This concept of the conservation of fibroblast subsets in different disease concepts has been actively investigated in inflammatory conditions. The use of this transcriptomic technology was applied in a cross-disease tissue comparison in order to try and understand whether these pathogenic fibroblasts were present in different diseases of inflammatory nature at various anatomical locations. Our group has contributed to an extensive effort aimed to characterised shared and differential features of fibroblasts across tissues isolated from Rheumatoid synovium, intestine of patients affected by inflammatory bowel disease, lung of patients with chronic interstitial lung disease and SG of patients with pSS [161]. Isolated fibroblasts were analysed with scRNAseq to generate a multi-tissue atlas which identified key stromal cell populations unique to each diseased tissue [161]. In this study, the authors integrated the scRNAseq fibroblast data to address two key ideas which included identifying genes signatures which were either shared or unique to the fibroblasts that were being explored [161]. Intriguingly, they identified two fibroblast populations identified as CXCL10⁺ CCL19⁺ immune interacting and SPARC⁺ COL3A1⁺ vascular interacting fibroblasts were expanded in all inflamed tissues across the 4 diseases at the mentioned anatomical sites [161]. These findings were then validated by exploring an additional independent dataset of an inflammatory skin condition, atopic dermatitis (AD) which demonstrated an abundance of the CXCL10⁺ CCL19⁺ immune interacting population but not of the SPARC⁺ COL3A1⁺ vascular interacting population. This observation was associated with the reduced vascularisation of the skin samples utilised within the dataset [161]. The study then went further to explore whether the CXCL10⁺ CCL19⁺ immune interacting and SPARC⁺ COL3A1⁺ vascular interacting fibroblasts were conserved

between human and mice tissues, in which case it was demonstrated that these two populations were identified in the bleomycin treated lung but only one or the other was present in mouse models of arthritis and colitis [161].

The exploration of cross species fibroblast characterisation was conducted within a different study conducted by *Buechler et al*, which took a cross species pan tissue approach looking at the presence of conserved and unique populations of stromal cells in different organs in steady and perturbed states in both mouse and human tissue [162]. Collective analysis of stromal cells from different organs within the mouse identified 2 universal fibroblast cell types that were differentiated by the expression of *Pi16* or *Col15a1* [162]. From these cell types, 5 specialised fibroblast subtypes in steady state tissues and 3 activated fibroblast subtypes were found in perturbed or diseased tissues. This study was also able to investigate whether there was any developmental link between these steady state and inflammatory induced fibroblasts in the various mouse organs. It was demonstrated that the inflammatory activated subsets could be traced back to arise from the steady state fibroblasts using transcriptomic trajectory analysis. This suggested that the universal steady state stromal population functions as a core group of cells from which specialised or activated cells could be derived depending on the context of microenvironmental cues. Further investigation of human sample data from pancreatic cancer patients identified genes specific to the universal fibroblasts clusters and activated fibroblast clusters found within the mouse [162]. This data demonstrated that not only can we begin to characterise the heterogeneity of stromal cells in various organs and disease contexts but also revealed that fibroblast subtypes are conserved between species. The availability of high-resolution transcriptomic technology has allowed for the characterisation of fibroblast

populations, infer developmental trajectories and various cellular interactions with different cell compartments.

From a cellular perspective, the identity of the underlying fibroblast compartment which becomes remodelled in response to inflammation to drive TLS development has not been defined yet at a single cell level. Our group and others have been able to demonstrate that fibroblasts in the context of TLS are also extremely heterogeneous and possess key immunological functions to support the immune infiltrate in murine and human SG [8]. We further demonstrated that the formation of the TLS network requires a multistep process that involves a phase of priming and a phase of stabilization in order to achieve the immune phenotype and is summarised in Figure 3.

1.14: The generation of the *immunofibroblast*: the *priming* step

One of the crucial phenotypical changes occurring during the *priming* step is the increased expression of the family of adhesion molecules such as intercellular adhesion molecule -1 (ICAM-1) and vascular cell adhesion protein (VCAM-1) and of the mucin-like glycoproteins, podoplanin (PDPN or gp38) [7]. The upregulation of the adhesion molecules facilitates the physical interaction of infiltrating T and B cells which express the complementary integrins such as lymphocyte function associated antigen (LFA-1) or very late antigen -4 (VLA-4) [70, 163]. The expression of ICAM-1 and VCAM-1 was characterised on human pSS SG patient samples coupled with additional *activation* markers such as PDPN and FAP [8]. Furthermore, *in vitro* stimulation of human pSS fibroblasts with the pro-inflammatory cytokines IL-13, TNF and LT α 1 β 2 resulted in the upregulation of ICAM-1 and VCAM-1 demonstrating the *priming* and *activation* of these fibroblasts [8]. This study also demonstrated the compartmentalisation of human pSS fibroblast function in a CD34 expression dependant manner. Here, it was observed that PDPN⁺ CD34⁺ and PDPN⁺ CD34⁻ fibroblasts had different functional properties. In this instance, the PDPN⁺ CD34⁺ fibroblasts expressed the lymphocyte survival factors BAFF and IL-7 whilst the PDPN⁺ CD34⁻ expressed the lymphoid chemokines CCL19, CLL21 and CXCL13 [8].

1.14.1: The role of PDPN in *immunofibroblast* priming

The induction of PDPN within a cytokine milieu has been demonstrated by *Ekwall et al* and *Honam et al* in the presence of potent proinflammatory cytokines such as TNF- α , IFN- γ , IL-1 and IL-6 [164, 165]. The expression of PDPN has also been characterised in other tissues (gut, lung and heart) and in different disease contexts such as autoimmunity and cancer [8, 130, 166]. Within the SLO, PDPN⁺ FRCs have been characterised to interact with its cognate receptor Clec-2 which is expressed on DC and facilitates their movement in the T cell area [126]. The stabilization of the interaction with these difference cell types allows for the exchange of molecular cues responsible for the broader activation and maturation of the stromal compartment and, on the other side, of the recruited lymphocytes. The *priming* step also includes the transient expression of lymphoid chemokines, in particular T cell chemokine CCL19 which is responsible for the initial wave of lymphocyte infiltration. Interestingly, smaller aggregates of lymphocytes detected in humans present mainly T cell infiltration, supporting the concept that T lymphocytes are the first cells recruited in the TLS anlagen. In this way, these infiltrating T cells, amongst other immune cells, are assigned as being the lymphoid tissue initiator (LTin) cells that carry signals on their surface, or produce in a secretory manner, resulting in the maturation of the underlying stroma such that it acquires a lymphoid stroma like phenotype.

1.14.2: The role of innate immune cells in *immunofibroblast priming*

Differently from SLO, mature fibroblast *priming* can occur independently from the LT β R signalling cascade and might occur prior to and independently from the interaction with infiltrating adaptive immune cells [8, 10, 11]. Resident immune cells such as resident innate lymphoid cells (ILC) and DC have all been described as potential providers of signals able to induce immunofibroblast *priming* [119, 167].

Resident DC populations are able to produce similar pro-inflammatory cytokines such as TNF- α , IL-23, as well as chemokines [168]. Furthermore, the presence of DC in the inflamed microenvironment can facilitate the induction of TLS [169]. In a classical immune response, DCs fulfil their function as classical APCs by transporting and presenting antigen (Ag) epitopes to the adaptive immune cells within the optimised environment of an SLO [169]. The presence of Ag has been associated with the formation of TLS in CCR7^{-/-} mice, as the inability of Ag loaded DC to migrate to local SLO resulted in the spontaneous formation of TLS [169]. These TLS were resolved with the reconstitution of CCR7⁺ DC emphasising that the retention of Ag within tissues can result in the formation of TLS [169]. The presence of various pro-inflammatory cytokines from resident cells presents at sites of TLS development coupled with the availability of Ag can initiate the cascade events leading to the formation of the *immunofibroblasts*. However, the current sequence of events that occur in TLS formation consist of the *priming* of the resident stroma and its *maturation* which allows for the acquisition of the functional properties similar to lymphoid stromal cells found in SLO.

Innate lymphoid cells include NK cells, ILC1s, ILC2s, ILC3s and LTi cells demonstrated to participate in the *priming* of resident stromal cells [93]. The soluble cytokines produced by ILCs

can interact with their corresponding receptors which are expressed on the stromal cells. The ILC population of cells have been mapped in the context of T lymphocytes as they share similar transcription factors and cytokine profiles [170]. ILC1s mirror CD4 Th1 cells with the expression of the transcription factor Tbet, while ILC2s are aligned to CD4 Th2 cells which produce IL-5 and IL-13 in a Gata3 and ROR α dependant manner [170]. The presence of resident ILCs in SG were demonstrated to function as a cellular source of IL-13 which regulated the phenotype of PDPN⁺ *immunofibroblast* progenitor populations [8]. Characterisation of the source of IL-13 in a mouse model of TLS identified ILCs to be the source of this pro-inflammatory cytokine. Furthermore, the *in vitro* challenge of human pSS fibroblasts with this cytokine resulted in the upregulation of adhesion markers ICAM-1 and VCAM-1 which are indicative of cell *priming* [8].

ILC3 cells can be further divided into two populations dependant on the expression of the natural cytotoxicity receptor (NCR) and are dependent on the transcription factor ROR γ t and express the pro-inflammatory cytokines IL-17 and IL-22 [98, 111, 112]. A key member of the IL-23 family of cytokines, IL-17 has been explored in its contribution to the development of TLS at various anatomical locations [4, 10, 171]. The reliance on Th17 for TLS formation has been also described in the meninges of experimental autoimmune encephalomyelitis (EAE) [6, 172]. The production of IL-17 derived from Th17 cells has been shown to prime lung fibroblasts prior to iBALT formation promoting an LT- α independent expression of CXCL13 [4, 141]. Whilst IL-17 has been demonstrated in *activation* of the stromal cell compartment in certain sites, it does not appear to play a significant role in SG. On the contrary, IL-22, a cytokine also bellowing to the IL-17 family, has been shown to drive the proliferation and subsequent expansion of the fibroblast compartment [7].

1.15: The expansion of the immunofibroblast

After the resident mesenchyme has undergone *priming* and *activation* a subsequent phase of expansion and proliferation occurs. One of the key cytokines demonstrated to drive this expansion is the member of the IL-10 cytokine family, IL-22 [7]. In both physiological and pathological contexts, IL-22 induces the proliferation and survival of epithelial cells, suggesting it has a stroma supportive cytokine. Furthermore, the expression of IL-22 has been associated with the production of autoantibodies in the context of pSS [173]. To this end, *Barone et al* explored the effects of IL-22 using a pSS mouse model and observed that T cell derived IL-22 was instrumental in the preservation of TLS organisation and expression of the lymphoid chemokines. This was highlighted by the failure of CXCL13 and CXCL12 expression in the SG of IL-22^{-/-} mice which consequently resulted in the impaired formation of TLS [7]. However, further exploration of the effects of IL-22 on human pSS fibroblasts showed that whilst this cytokine was unable to induce immunofibroblast *priming*, it was able to give these cells a proliferative ability. This observation was recapitulated in the murine SG where IL-22^{-/-} and IL-22R^{-/-} mice were unable to expand their PDPN⁺ stromal population. The expansion of this *immunofibroblast* population reflects the expansion phase of the LTo mesenchyme during SLO development [104]. Nonetheless, aside from the *priming* and *proliferative* effects of individual cytokines, lymphocytes and in particular T cells are also required for the phase of expansion and for full maturation of the mesenchyme [8].

1.16: The *maturation of the immunofibroblast*

Following the priming step which is LT β R independent, the maturation of the stroma becomes dependent on this signalling cascade to acquire its lymphoid-like phenotype. This maturation is characterised by the expression of lymphoid chemokines CXCL13, CCL19, CCL21 and lymphocyte survival factors IL-7 and BAFF [7, 20, 118]. This expression of the lymphoid chemokines mediates the constant influx of T and B cells which, coupled with the expression of their corresponding survival factors, and continues the stromal maturation process. Within TLS development, the immunofibroblasts become dependent on the infiltrating lymphocytes to continue their maturation and remodelling, much like LTo in SLO development. In this context, lymphocytes assume the role normally supported by the LTi in embryonic SLO and are often LTin cells in the context of TLS formation [5, 36, 174]. To this end, LTin can exert their effects to mature the stromal cell population by production of various pro-inflammatory cytokines which I will describe below.

1.16.1: The effects of pro-inflammatory cytokines on *immunofibroblasts*

The different signals induced by the LTin are sourced from various cellular sources and families of molecules. In the proinflammatory cytokine milieu within the inflamed tissue many other cytokines derived from infiltrating adaptive or innate immune cells, or resident stromal cells, can contribute to the priming of the pre-TLS supportive stroma. In addition, the different cytokines appear to induce different biological effects at different locations.

We have demonstrated that the Th2 cytokines IL-4 and IL-13, through the engagement of IL-4R α receptor, can induce the upregulation of ICAM-1 and VCAM-1 in synergy with other cytokines [8]. However, this upregulation was not seen in the expression of PDPN suggesting that different pro-inflammatory cytokines differentially regulate the expression of markers which are indicative of stromal *priming*. Our group also explored the effects of IL-22, another Th17 derived cytokine, and identified that this cytokine regulated the proliferation of gp38+ immunofibroblasts in TLS forming within mucosal tissue and CXCL13 expression [7].

An additional member of the IL-23 family of cytokines, IL-17 has been explored in its contribution to the development of TLS at various anatomical locations [5, 6, 175]. The production of IL-17 derived from Th17 cells has been shown to prime lung fibroblasts prior to iBALT formation by promoting the LT- α independent expression of CXCL13 [4]. The formation of TLS by Th17 cells was also identified in the meninges of experimental autoimmune encephalomyelitis (EAE) which acquired a lymphoid like phenotype [6].

1.16.2: The interference of the IFN signalling cascade in *immunofibroblast* maturation

Another family of cytokines known to participate in early immune responses to pathogens, tissue damage and partake in the propagation of inflammation is the IFN superfamily. The engagement of the IFN family occurs in response to the recognition of various viral and bacterial sourced pathogenic associated molecular patterns (PAMPs) and danger associated molecular patterns (DAMPs) such as endogenously released molecules from distressed or necrotic cells, which are recognised through pattern recognition receptors (PRRs) [176]. The production of IFNs results in the release of proinflammatory cytokines by infiltrating immune cells and the presentation of antigen by upregulation of major histocompatibility complexes (MHC) on both lymphocytes and stromal cells [177].

Three classes of IFNs exist which are structurally different yet function to perpetuate the inflammatory response. The IFN- γ signature can be sourced from various resident and infiltrating innate and adaptive lymphocytes such as macrophages, neutrophils, activated CD4 and CD8 T cells and has been shown to contribute to the priming of stromal cells through upregulation of ICAM-1 and VCAM-1 but not PDPN [178, 179]. Interestingly, the type 1 IFN family has recently been demonstrated to be a novel inducer of CXCL13 expression supporting ectopic GC establishment [180]. In this study, *Denton et al* show that the *in vitro* challenge of lung fibroblasts with IFN- α and IFN- β was sufficient to induce expression of CXCL13 mRNA suggesting that downstream signalling via this pathway could result in the acquisition of immunomodulatory fibroblast phenotype. This finding was supported further by looking for the enrichment of CXCL13 mRNA in *Ifnar1*^{-/-} mice. In this case, microarray data demonstrated the significant reduction in CXCL13 expression in pulmonary fibroblasts from *Ifnar1*^{-/-} mice.

Overall, this emphasised the role of IFNs in not only being able to *prime* fibroblasts but facilitate its maturation by inducing the expression of the B cell chemoattractant CXCL13. Like many cytokines, the IFNs do not engage alone but synergise with other proinflammatory molecules. The most classical synergy of IFNs has been associated with another class of cytokines known as the TNF superfamily which can also be produced by resident immune and stromal cells at sites of TLS formation.

1.16.3: The TNF superfamily in *immunofibroblast* maturation

The TNF superfamily has been demonstrated to participate greatly in the initiation of the inflammatory response. The classic members of this superfamily are engaged during the development of SLO such as: LT α , LT β , RANKL, and BAFF [181] and their roles are also reprised in the formation of TLS. The infiltrating lymphocytes, which migrate to sites of TLS development in response to expression of CXCL13, CCL19 and CCL21, can carry this membrane bound signals and engage with the fibroblast compartment expressing the complimentary receptor or ligands.

TNF- α is a proinflammatory cytokine that has a limited role during lymphoid neogenesis but is a key driver during TLS development. TNF- α is a critical molecule functioning as the bodies “fire alarm” in response to pathogenic insult or tissue damage [182], and exerts its signalling function by binding to one of two receptors TNFR1 - involved in perpetuating inflammation, or TNFR2 - which is involved in repair and cellular homeostasis. The contribution of TNF- α in the formation of TLS has been documented with the induction of stromal cell *priming*, at times in synergy with IFN- γ , with the upregulation of various adhesion molecules and to some extent the transient expression of CXCL13 [171]. This observation was demonstrated by the ability of lung fibroblasts of virally challenged mice to not only upregulate the expression of adhesion molecules but express CXCL13 [180]. This was concomitant with the observation of GCs amongst these fibroblasts which had become activated in response to myeloid cell derived IFN- γ expression [180]. Moreover, gingival fibroblasts have been demonstrated to upregulate ICAM-1 and VCAM-1 expression upon stimulation with TNF- α or IFN- γ , reinforcing the *priming* and *activating* effects of these pro-inflammatory cytokines on resting mesenchyme cells [183].

Collectively, there are many signals present within the inflamed microenvironment within which TLS formation occurs that can initiate the *priming* of the fibroblast compartment independent of any cellular crosstalk. This crosstalk becomes more critical during the maturation stage which requires prolonged and effective communication between the lymphocyte and fibroblasts whilst TLS assembly takes place.

Figure 3: Diagrammatic summary of the sequential *priming*, *expansion* and *maturation* of the stromal cells in response to different cytokines and their cellular sources

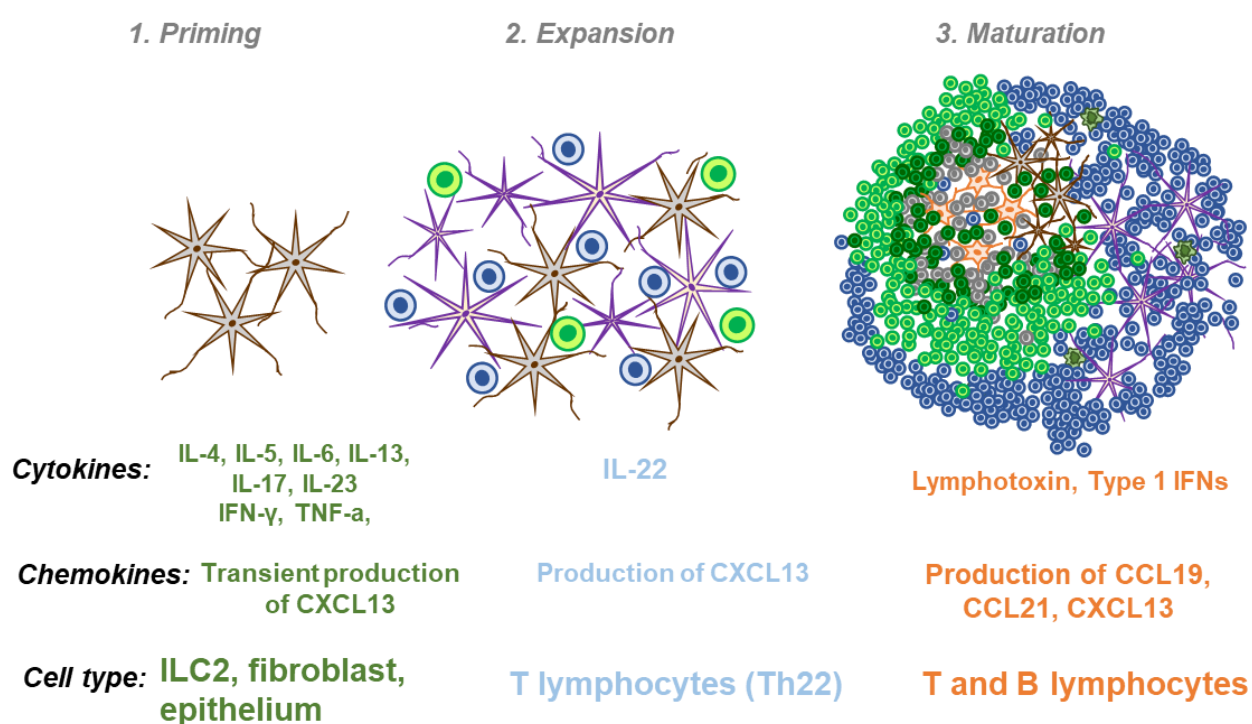


Figure 3: Diagrammatic summary of the 3-step model of immunofibroblast formation. This figure summarises the different cytokines, chemokines and cell populations which are involved in each of the following steps of Priming, expansion, and maturation.

1.17: Thesis Aims and Objectives

In the introduction we have discussed the function and organisation of TLS in both health and disease, the different cellular compartments of TLS and the signals involved in their formation and development in comparison to SLO. We have also addressed the emerging importance and nature of the stromal heterogeneity which exists and underpins the complexity of TLS development and function as explored in different diseases contexts of inflammation and cancer. With the development of advanced transcriptomic technologies, the ability to resolve the cellular heterogeneity has become more feasible. This not only facilitates our understanding of the cellular signals exchanged amidst the crosstalk of different cell compartments but also helps resolve the complexity of TLS heterogeneity in disease.

In this thesis I aimed to resolve the cellular organization of TLS and to explore signals and pathways underpinning the role of these structures in a model of autoimmune disease using scRNAseq in combination with *in vivo* and *in vitro* techniques to characterise the cellular landscape of the human pSS SG with mature TLS.

More specifically in our human pSS SG dataset I have explored 3 key aims which include:

- 1) Exploring and characterise the fibroblast heterogeneity present within the inflamed human pSS SG (Chapter 3: Transcriptomic characterisation of the human pSS SG) and murine pSS SG model using scRNAseq (Chapter 4: Transcriptomic characterisation of the murine pSS SG).
- 2) Exploring the possible signalling pathways that lead to the formation of potentially pathogenic *immunofibroblast* and *immunopericyte* populations which may contribute to the formation of TLS using bioinformatic tools within the scRNAseq dataset.

- 3) Explore and validate our mRNA data at the protein level using human pSS SG tissue and murine tissue from either WT or knockout mice to support our findings of the CD34⁺ CD146⁻ immunofibroblasts and the CD34⁻ CD146⁺ *“immunopericyte”* from the scRNAseq dataset.

In my thesis I have discussed and described the computational advantages and limitations of scRNAseq, and the different tools used to begin exploring the data we have generated (Chapter 2). I have used scRNAseq to explore the cellular heterogeneity within the human pSS SG in order to begin understanding the contributions of different cell populations in the possible development of TLS (Chapter 3). To further our exploration, I have also applied scRNAseq to murine SG derived from an inducible TLS model which has been characterised and explored in a functional manner [8].

Chapter 2.0: Materials and Methods

Table 1: Human Flow Cytometry Antibody Panel

Marker	Clone	Isotype	Conjugate	Dilution (μl)	Supplier
<i>CD36</i>	5-271	Mouse IgG2a, κ	APC/Fire 750	1/100	Biolegend
<i>TNC</i>	4C8MS	IgG1	Unconjugated	1/100	Biotechne
<i>CD146</i>	P1H12	Mouse IgG2a, κ	BV605	1/100	Biolegend
<i>CD34</i>	581	Mouse IgG1, κ	BV510	1/50	Biolegend
<i>CCL21</i>	N/A	Polyclonal Goat IgG	Unconjugated	1/100	Biotechne
<i>Anti-Goat 2^o Antibody</i>	N/A	Donkey / IgG	Alexa Fleur 647	1/100	Life Technologies

Table 2: Mouse Flow cytometry Antibody Panel

Marker	Clone	Isotype	Conjugate	Dilution (μl)	Supplier
<i>CD34</i>	Ram34	Rat / IgG2a, kappa	Alexa Fleur 700	1/100	Life Technologies
<i>CD146</i>	ME-9F1	Rat IgG2a	Pe-Cy7	1/100	Biolegend
<i>CCL21</i>	N/A	Polyclonal Goat IgG	unconjugated	1/50	Biotechne
<i>Anti- Goat 2^oAntibody</i>	N/A	Polyclonal Donkey / IgG	PE	1/100	Life Technologies
<i>CD31</i>	390	Rat IgG2a, κ	BV421	1/200	Biolegend
<i>Epcam</i>	G8.8	Rat IgG2a, κ	BV711	1/400	Biolegend
<i>CD45</i>	30-F11	Rat / IgG2b, kappa	PerCP-Cyanine5.5	1/300	Life Technologies
<i>PDPN</i>	8.1.1	Syrian Hamster IgG	PerCP-Cyanine5.5	1/100	Biolegend
<i>Viability</i>	N/A	Cellular Stain	Apc-Cy7	1/1000	eBioscience

2.1: Tissue Digestion and flow cytometric analysis of human pSS SG and intracellular CCL21 staining

Human SG tissue used in these validation studies were acquired from within the Optimising Assessment in Sjögren's Syndrome (OASIS) cohort at the University of Birmingham, under ethics number 10-018. New patients over the age of 18 are recruited into this cohort after demonstrating symptoms for pSS or with a diagnosis of this condition. The patient samples used here had a physician formally diagnosis them as having pSS and fulfilled the 2016 ACR/EULAR classification criteria. All recruited patients had provided written informed consent to participate in the study, and the study itself was approved by the Wales Research Ethics Committee 7 (WREC 7) formerly Dyfed Powys REC; 13/WA/0392 [184]. The SG tissue was mechanically fragmented using dissection scissors whilst in 2 ml RPMI constituted with 2% FCS, 0.8mg/ml Dispase, 0.2mg/ml Collagenase P and 0.1mg/ml DNase I. The SG pieces were then incubated in the enzyme mix at 37°C in a water bath with constant agitation for 20 min, Afterwards, the SG samples were further mechanically digested by pipetting to disrupt tissue and harvest more cells. Following this, the sample was allowed to settle for 1 min prior to removal of the enzyme mix through a 70um mesh strainer. The filtrate was collected in ice cold MACs buffer (0.5% (w/v) BSA, 2 mM EDTA in PBS) and kept on ice. The remaining human SG sample was incubated for a further 10 min in fresh enzyme mix and recovered in the same manner after 20 min incubation with the enzyme mix. Once the enzymatic digestion was completed the cells were centrifuged at 1800 rpm for 5 min and washed twice in MACS buffer. Following this, the cell pellet was stained in 100 ul of the according antibody cocktail using the antibodies and their concentrations as demonstrated in Table 1 and process for flow cytometry analysis as previously described.

2.2: Single Cell RNA sequencing (scRNAseq) pipeline

By harnessing an insight into the fundamental concept of biology – *transcription and translation* - gene expression profiling of particular cells presents many opportunities for targeted, patient stratified therapies. The study of transcriptomics has been essential in driving the understanding of the functional and phenotypical nature of cells and tissue in the context of health and disease. In the field of transcriptomics, 3 technologies have been extensively used to study gene expression: quantitative reverse transcription polymerase chain reaction (qRT-PCR), microarrays and ribonucleic acid (RNA) sequencing (Table 1) [185].

The limitations of qRT-PCR and microarrays consist of the inability to generate a gene expression profile at a single cell level in a cost and labour effective manner. In addition to these techniques, bulk sequencing also identifies gene signatures at a coarse level but scRNAseq has overpowered these tools with its high cellular resolution with gene expression being characterised at the single cell level [185, 186]. This technique allows one to infer many parameters including, but not limited to, cell abundance, developmental trajectories, engagement of specific signalling cascades and cellular interactions which may otherwise not be observed using preceding technologies [185].

In this work we have used scRNAseq, in combination with other experimental techniques, to deconvolute the heterogeneity of the human pSS SG. In addition to this, we have also applied the same transcriptomic analysis on the inflamed and resting murine SG model to investigate the possible cellular differences which may arise within the two different physiologically different SGs. In chapter 2.0, I will discuss in detail the pipeline of analysis used to generate the scRNAseq datasets for both the human and mouse tissues. Here I will discuss the various

steps undertaken to assure that the data used for biological analysis has been curated accordingly and the differing packages used for the downstream analysis in my work

Table 3: Summary of different genomic technologies and their limitations as compared to scRNAseq

<i>Technology</i>	<i>Methodology</i>	<i>Limitations for progression to single cell platform</i>
<i>qRT-PCR</i>	● Measure of accumulation of amplified genetic material using fluorescent hybridisation tools [185] ● Sensitive and robust method of gene expression quantification	● Low throughput ● High cost per cell ● High labour ● Selection of target genes is based on previous knowledge therefore potential for biased analysis
<i>Microarrays</i>	● Use pre-designed RNA probes to capture transcripts of interest [185]	● Coverage of non-coding regions and splice variants has a high cost ● Sensitivity is limited ● Would require a large amount of starting material
<i>RNA-seq</i>	● Use of next generation sequencing (NGS) to identify and quantify RNA in a sample at a given time [185]	● Enables unbiased profiling of whole transcriptome ● Cost dependant on sample size and sequencing depth

Table 3: Summary of the required outcome and the limitations that exist with the initial QC step of the data which one needs to consider to retain biologically relevant cells in the dataset whilst excluding cells which may not contribute any biological signal

2.3: Single Cell RNA sequencing of the human pSS SG and murine SG

The generation of single cell data from the murine and human SG is a multistep process involving the generation of a single cell suspension, which I have described in Materials and Methods section 2.1. Following this, the cells within the suspension are captured in single droplets called Gel Bead-In Emulsions (GEMs), followed by the cDNA amplification, library construction and finally sequencing[186]. For the human single cell dataset, sequencing was conducted at the University of Oxford using the 10x 3' Chromium Sequencing platform at a 50,000 read depth. Briefly, this droplet based assay captures individual cells in microfluidic droplets which also contain the reagents required for *cDNA amplification and reverse transcription* [186]. Collectively, these reagents induce the disruption of the cellular membrane, capture of the intracellular mRNA and reverse transcription to cDNA which is then amplified with repeated cycles of transcription and generation of cDNA. In our dataset, the captured mRNA is also labelled with *unique molecular identifiers (UMIs)*. Since the cDNA is amplified prior to the sequencing process, UMIs distinguish copies of amplified mRNA or cDNA from the same molecule from additional mRNA molecules that have been transcribed and are derived from the same gene (Figure 1). Once the libraries have been sequenced the raw data is generated in the form of base call files (Bcl). These files are converted to fastq files which are then used in the appropriate downstream analysis software such as R studio for subsequent bioinformatic analysis (Figure 2).

The data is then aligned to the subsequent referencing genome for either human or murine datasets. This allows for the genome and genes to be generated from the sequenced data. The dataset is then batch corrected to remove and technical variation from sample processing

and to normalise the genes across the samples before the data is mined for biological analysis.

The datasets were further curated to minimise the ambient gene signature using *Scrubble* and *SoupX* algorithms from the dataset which masks the mRNA expression of statistically significant differentially expressed genes of interest.

Finally the datasets are then explored using various bioinformatic tools such as Geneset enrichment analysis (GSEA), trajectory analysis, pathway enrichment and cell:cell interaction analysis. I will describe these packages in more detail further in this chapter.

Figure 1 : Diagrammatic summary of UMI to transcribed genes

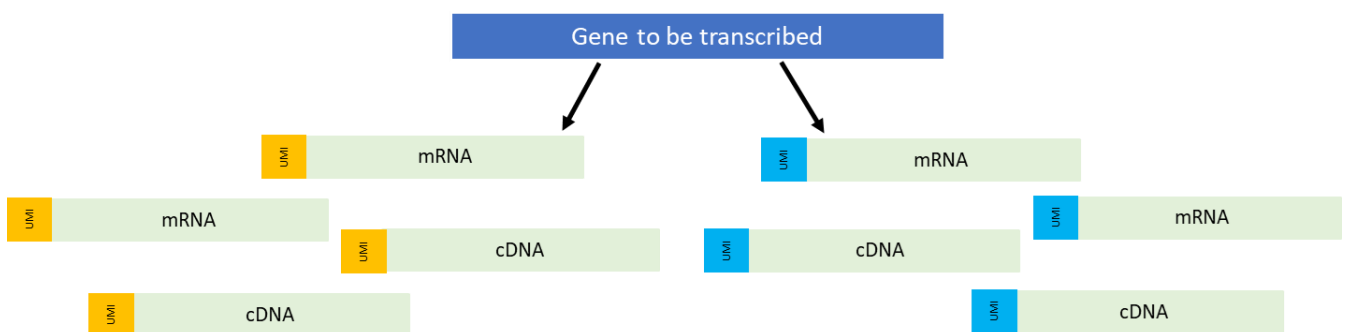


Figure 1: Diagrammatic summary of the concept of UMI generation in the scRNAseq datasets.

Transcription of a particular gene results in the tagging of a unique molecular identifier which distinguishes between different mRNA strands deriving from the same gene. This prevents the assumption of mRNA abundance for a particular gene as the individual transcription is accounted for

Figure 2: Summary of outcome and limitations for retention of high-quality cells for downstream bioinformatic analysis

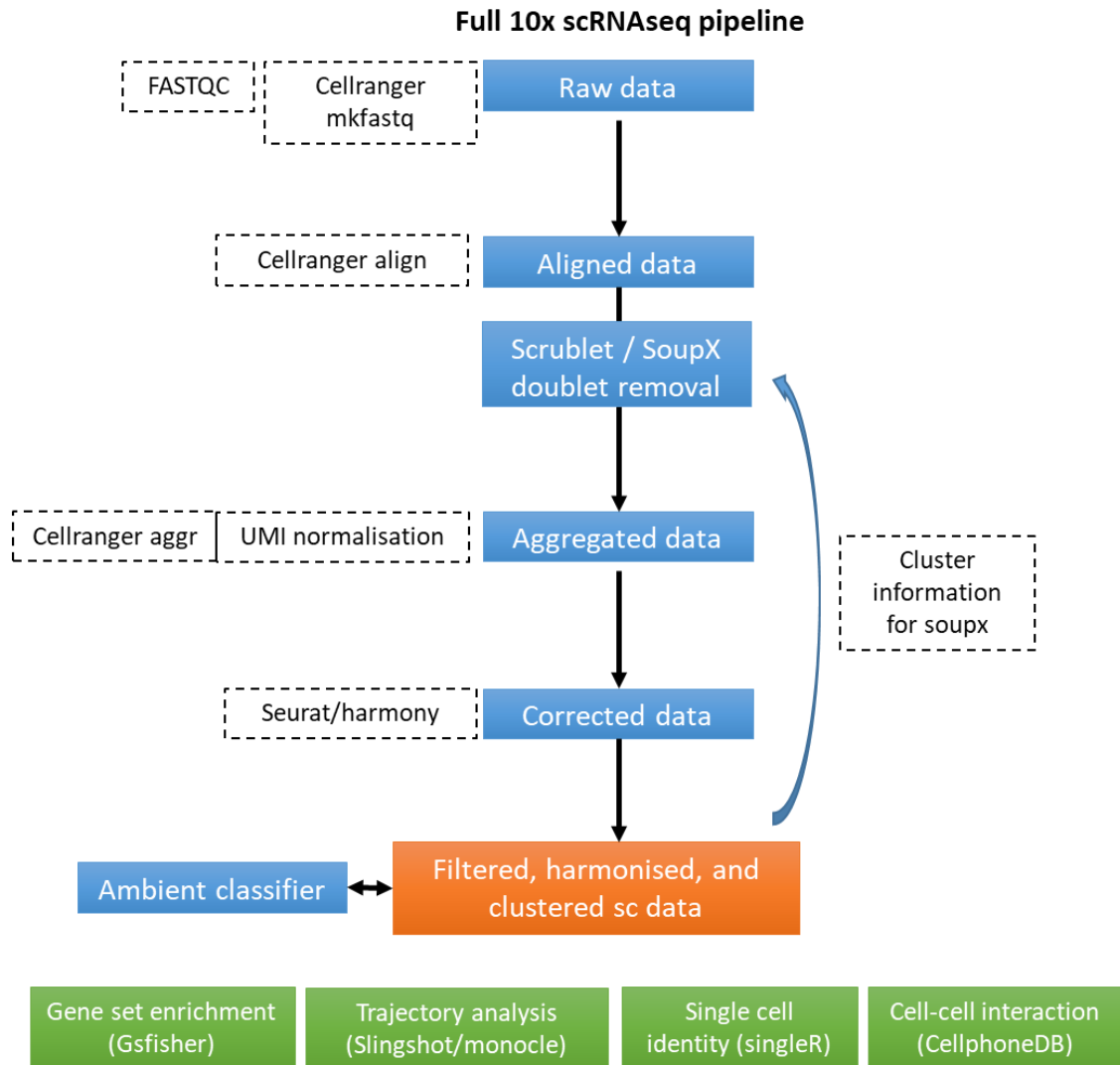


Figure 4: Flow chart of the analysis pipeline used for the human pSS salivary gland and the murine D0 and D15 p.c. salivary gland scRNAseq sequencing datasets. The blue boxes represent the methodologies applied within the pipeline, dotted boxed denote the packages used to carry out a step of analysis. The orange box indicates the “prepared” data ready for downstream analysis as summarised in the green coloured boxes.

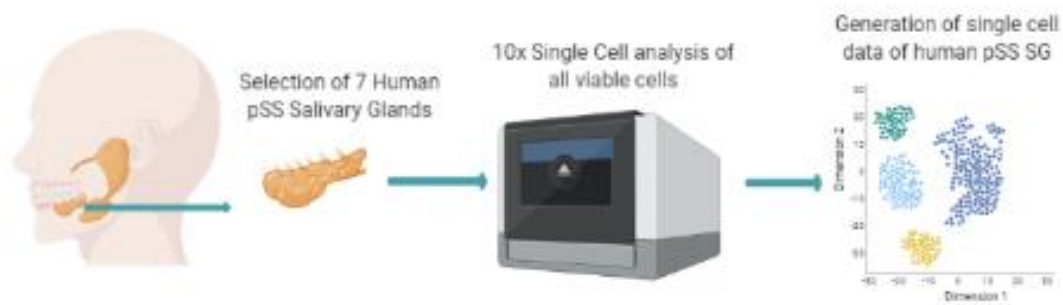
2.3.1: Single-cell RNAseq processing for human pSS SG samples

The single cell experiment was conducted on 7 pSS human SG at University of Oxford (**Fig 3A**). Prior to running through the 10X genomics facility, samples were digested, and dead cells were removed using STEMCELL Annexin IV Dead Cell Removal Kit. Cells were then encapsulated in a single droplet-based system using the 10X Genomics Chromium System. Once the sequencing libraries had been generated cells were subjected to Illumina sequencing at 50,000 reads per cell. Following this, the alignment and aggregation was performed using the CellRanger Pipeline (v.3.0.2) and human reference sequences (GRCh38 v1.2.0, supplied with CellRanger) median genes per cell are sample specific as according to the cellranger output files (**Fig 3B-H**). In order to account for index hopping between samples, duplicated barcodes were removed from the aggregated reads. Data analysis after proceeding the aforementioned quality control steps was performed using Seurat R package (v.2.3.4, R v.3.6.1). Cells with 20% or more expression of mitochondrial reads or fewer than 200 genes were excluded from the analysis (**Fig 3 I -K**). Cells were downsampled to the lowest common denominator for the disease group. The complete number of cells retained were $n= 17542$ from all biological replicates. From the exclusion of the decided mitochondrial and minimum gene threshold from the number of cells we had, we retained the first 30 principal components for t-distributed stochastic neighbor embedding (TSNE) projection and clustering analysis (resolution set to 0.5 for the full analysis of clusters and subsequent cell clusters). Only genes found in 25% of cells (either within or outside the cluster of interest) and that showed a minimum log fold difference of 0.25 were tested for differential expression. Data was corrected for batch effect as human pSS samples were run on 10x Chromium platform in 2 separate runs using Harmony (V 1.0). Markers used to assign cluster identity were selected

based on the significance ($P < 0.01$). The TSNE map and Uniform Manifold Approximation and Projection (UMAP) was generated using the “TSNE” and “UMAP” function in Seurat using the Louvain algorithm for shared nearest neighbour clustering. Gene set enrichment was completed using the GitHub package gsfisher in Seurat and data was statistically tested using Fishers Exact test. Differentially expressed genes were statistically tested using non-linear regression.

Figure 3: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

A



B

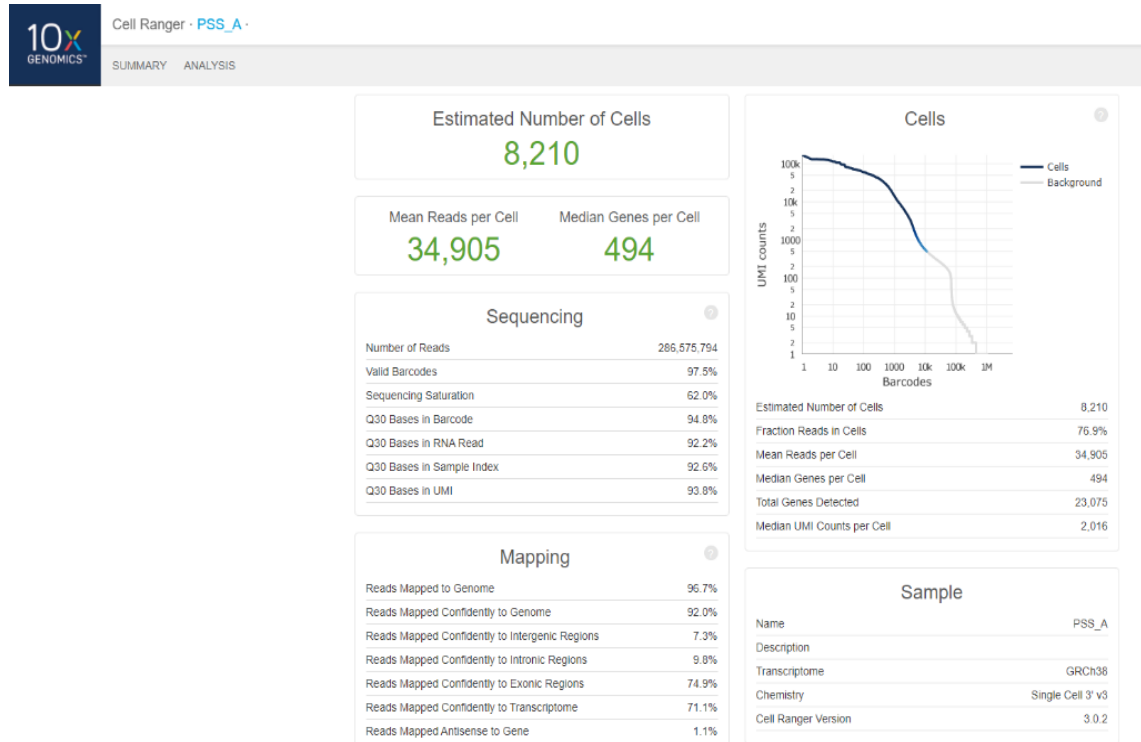


Figure 3: A) Flow of the work process of the 7 selected human pSS SG for 10x scRNAseq processing. **B)** CellRanger output summarising the QC of the sequenced data of all the human SG samples. This includes key metrics including estimated number of cells, median genes per cell, number of reads in sequencing, and cell profile after processing

Figure 3: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

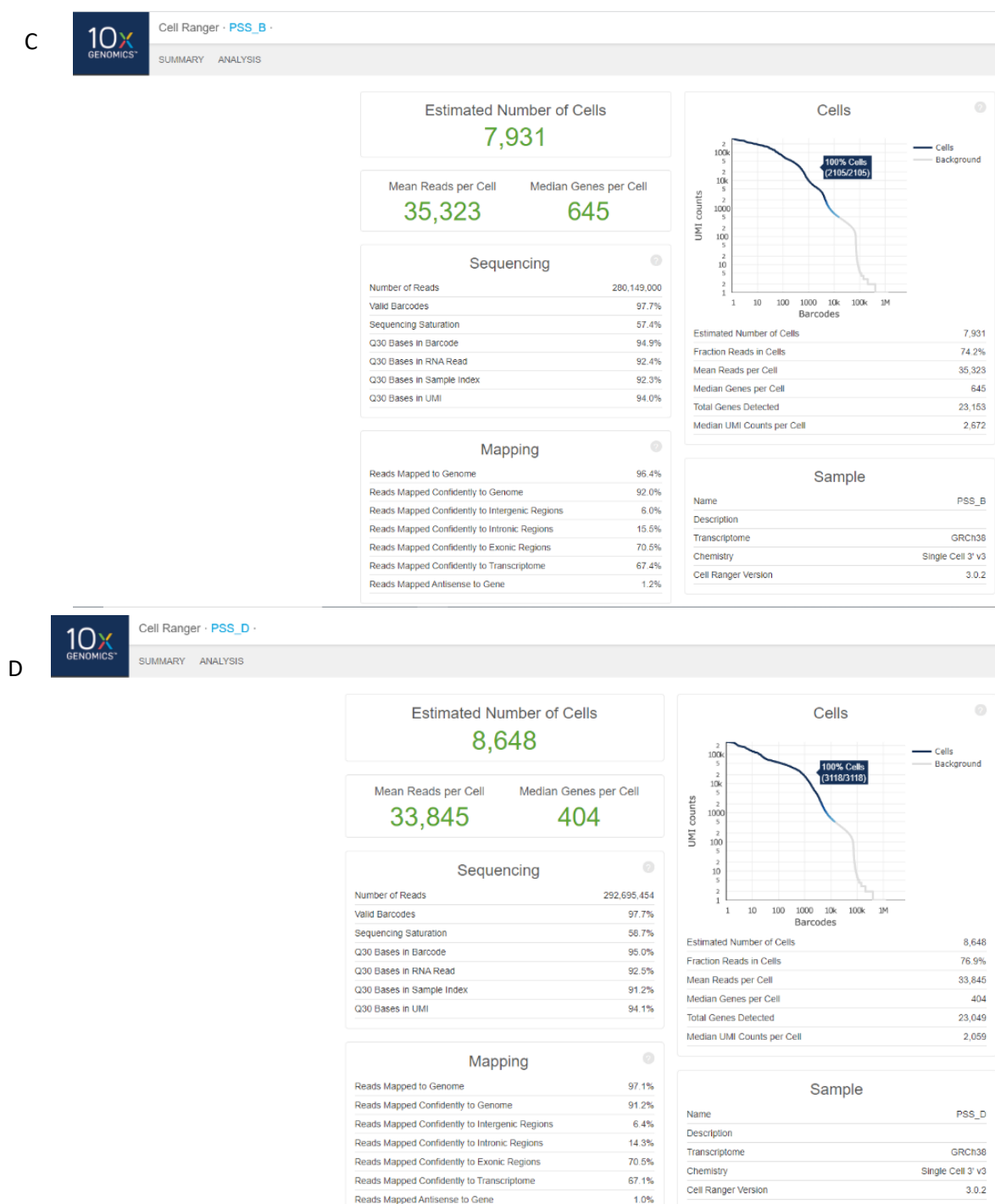
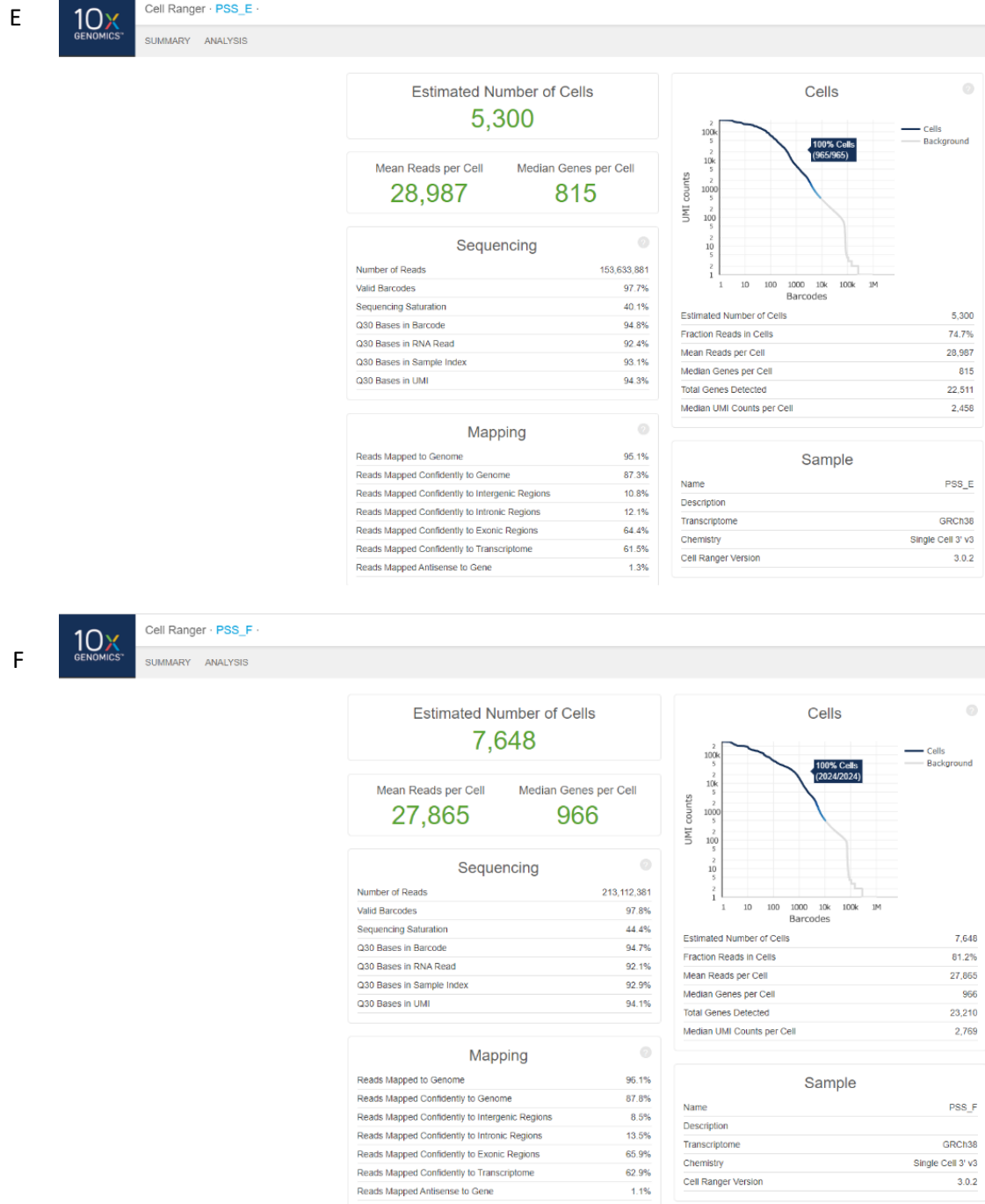


Figure 3: C,D) CellRanger output summarising the QC of the sequenced data of all the human SG samples.

This includes key metrics including estimated number of cells, median genes per cell, number of reads in sequencing, and cell profile after processing

Figure 3: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis



Estimated Number of Cells

7,648

Mean Reads per Cell

27,865

Median Genes per Cell

966

Sequencing

Number of Reads	213,112,381
Valid Barcodes	97.8%
Sequencing Saturation	44.4%
Q30 Bases in Barcode	94.7%
Q30 Bases in RNA Read	92.1%
Q30 Bases in Sample Index	92.9%
Q30 Bases in UMI	94.1%

Mapping

Reads Mapped to Genome	96.1%
Reads Mapped Confidently to Genome	87.8%
Reads Mapped Confidently to Intergenic Regions	8.5%
Reads Mapped Confidently to Intronic Regions	13.5%
Reads Mapped Confidently to Exonic Regions	65.9%
Reads Mapped Confidently to Transcriptome	62.9%
Reads Mapped Antisense to Gene	1.1%

Cells

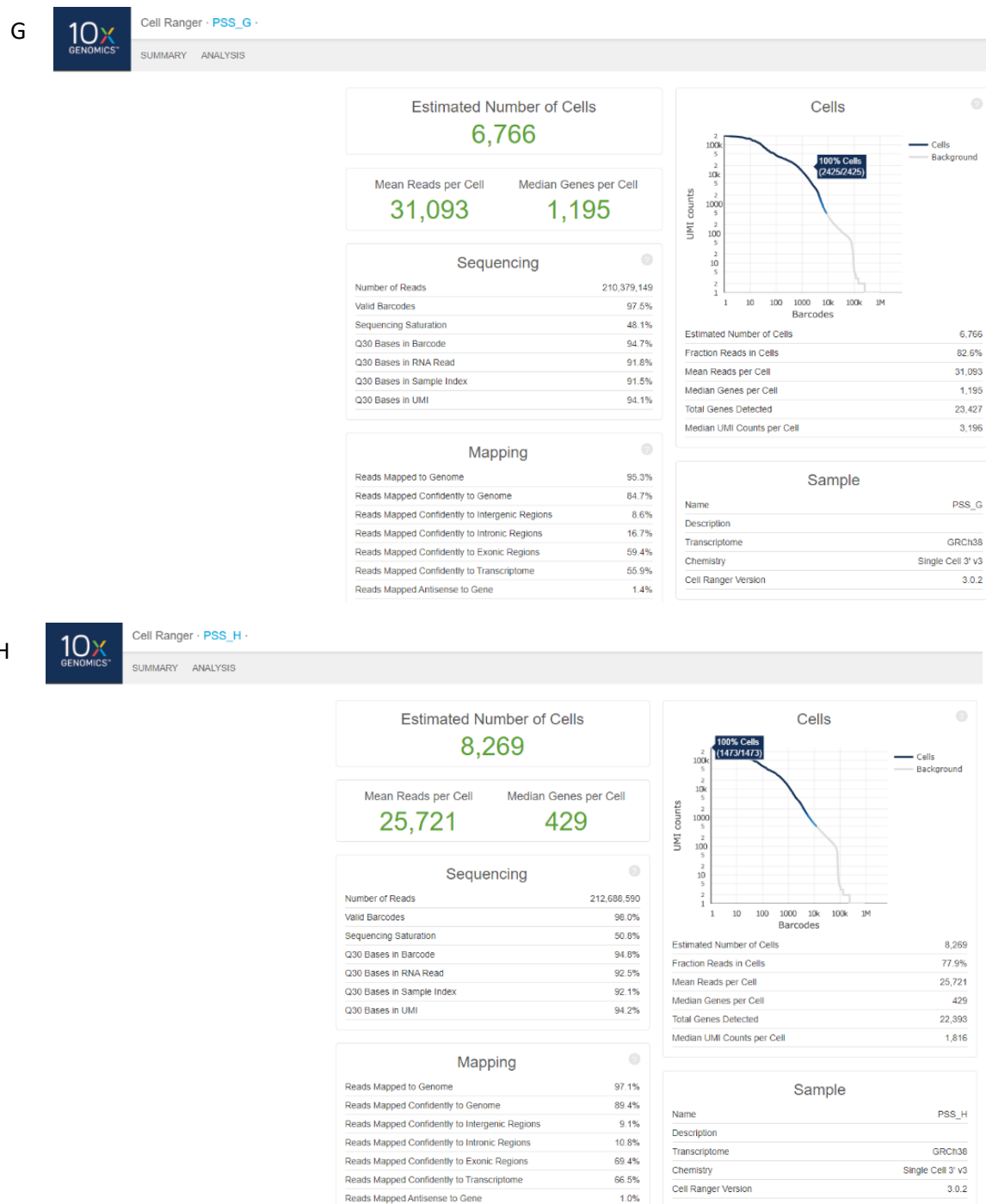
Estimated Number of Cells	7,648
Fraction Reads in Cells	81.2%
Mean Reads per Cell	27,865
Median Genes per Cell	966
Total Genes Detected	23,210
Median UMI Counts per Cell	2,769

Sample

Name	PSS_F
Description	
Transcriptome	GRCh38
Chemistry	Single Cell 3' v3
Cell Ranger Version	3.0.2

Figure 3: E,F) CellRanger output summarising the QC of the sequenced data of all the human SG samples. This includes key metrics including estimated number of cells, median genes per cell, number of reads in sequencing, and cell profile after processing

Figure 3: Summary of SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis



Estimated Number of Cells

8,269

Mean Reads per Cell

25,721

Median Genes per Cell

429

Sequencing

Number of Reads	212,686,590
Valid Barcodes	96.0%
Sequencing Saturation	50.8%
Q30 Bases in Barcode	94.8%
Q30 Bases in RNA Read	92.5%
Q30 Bases in Sample Index	92.1%
Q30 Bases in UMI	94.2%

Mapping

Reads Mapped to Genome	97.1%
Reads Mapped Confidently to Genome	89.4%
Reads Mapped Confidently to Intergenic Regions	9.1%
Reads Mapped Confidently to Intronic Regions	10.8%
Reads Mapped Confidently to Exonic Regions	69.4%
Reads Mapped Confidently to Transcriptome	66.5%
Reads Mapped Antisense to Gene	1.0%

Cells

Estimated Number of Cells	8,269
Fraction Reads in Cells	77.9%
Mean Reads per Cell	25,721
Median Genes per Cell	429
Total Genes Detected	22,393
Median UMI Counts per Cell	1,816

Sample

Name	PSS_H
Description	
Transcriptome	GRCh38
Chemistry	Single Cell 3' v3
Cell Ranger Version	3.0.2

Figure 3: G,H) CellRanger output summarising the QC of the sequenced data of all the human SG samples. This includes key metrics including estimated number of cells, median genes per cell, number of reads in sequencing, and cell profile after processing

Figure 3: Dot plots of the number (n) of Genes, UMI and percentage of mitochondrial genes across the 7 human pSS SG samples

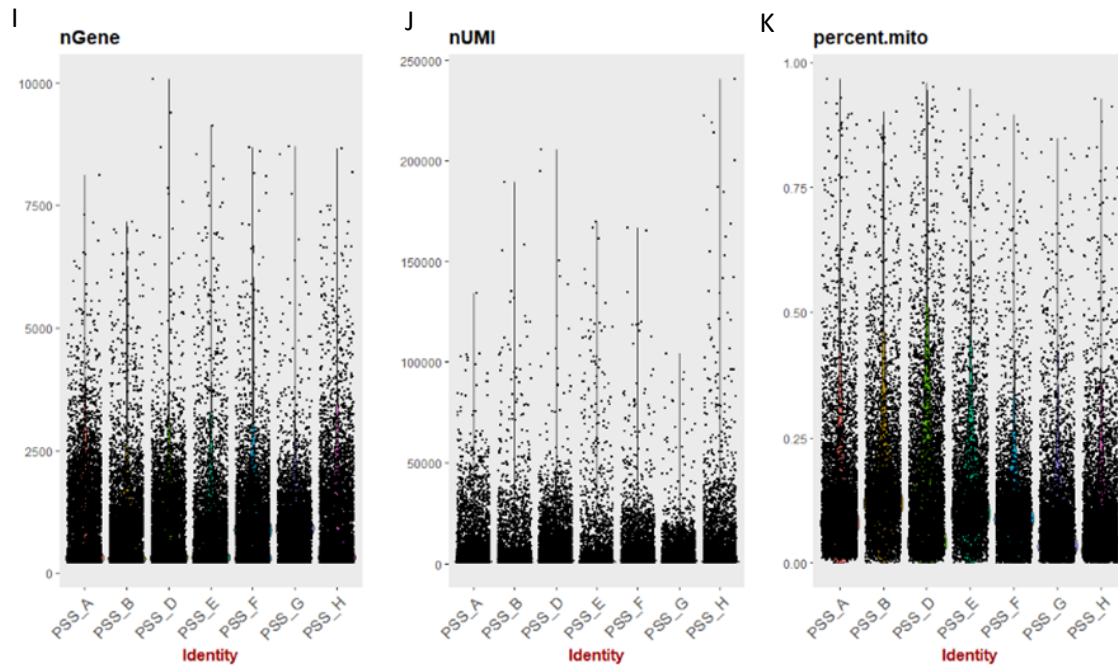


Figure 3 I-K: Flow of the work process for the 7 human pSS SG samples for scRNAseq analysis. Key metrics including estimated number of cells, median genes per cell, number of reads in sequencing, and cell profile after processing. Plots demonstrating the number of genes (nGene), number of unique molecular identifiers (nUMI), and percentage of mitochondrial genes (percent.mito) per sample

2.3.2: Single-cell RNAseq processing for resting and inflamed D15 p.c. murine SG samples

To generate the mouse dataset, we used 6 female C57BL/6J were purchased and maintained at Biomedical Service Unit (BMSU) under pathogenic free conditions prior to SG cannulation.

Cannulation: 3 mice were administered ketamine/domitor anaesthesia prior to procedure and 3 mice were retained as Day 0 control mice (**Fig 4A**). The submandibular glands of female mice were intraductally cannulated with 10^8 - 10^9 p.f.u of replication-defective adenovirus (Adv5). After cannulation, mice were recovered from anaesthesia and at specific time point post cannulation were culled by terminal anaesthesia using 200ul-300ul of Euthasol^R Vet 400 mg/ml Solution for Injection. **Digestion of mouse SG to generate single cell suspension:**

Harvested SG were processed to generate smaller tissue pieces and digested for 40 min at 37°C in 2ml RMPI 1640 (Sigma) containing: collagenase D (3.7 mg/ml: Roche), DNase I (30 ug/ml: Sigma), and 2% (v/v) fetal calf serum (FCS). The suspension was pipetted after 15 minutes to remove large cell aggregates. The cell suspension was digested for a further 20 minutes. The cell suspension was then passed through a 70 um mesh filter and washed twice with phosphate buff saline (PBS) (Sigma). **Non-viable cell removal:** The cell suspension was then processed using STEMCELLTM Annexin IV Dead Cell Removal Kit. The viable cell suspension was then taken for single cell processing. **Single Cell Processing:** Single cell analysis was conducted at the Genomics Lab Facility, University of Birmingham using the NextSeq 500 (Illumina). Raw data was processed and aligned in a manner identical to the human single cell dataset. The complete number of cells retained were 1534 from all biological replicates at Day 0 and 2248 from all biological replicates at Day 15.

Figure 4: Summary of murine SG processing and QC metrics of the raw data prior to downstream bioinformatic analysis

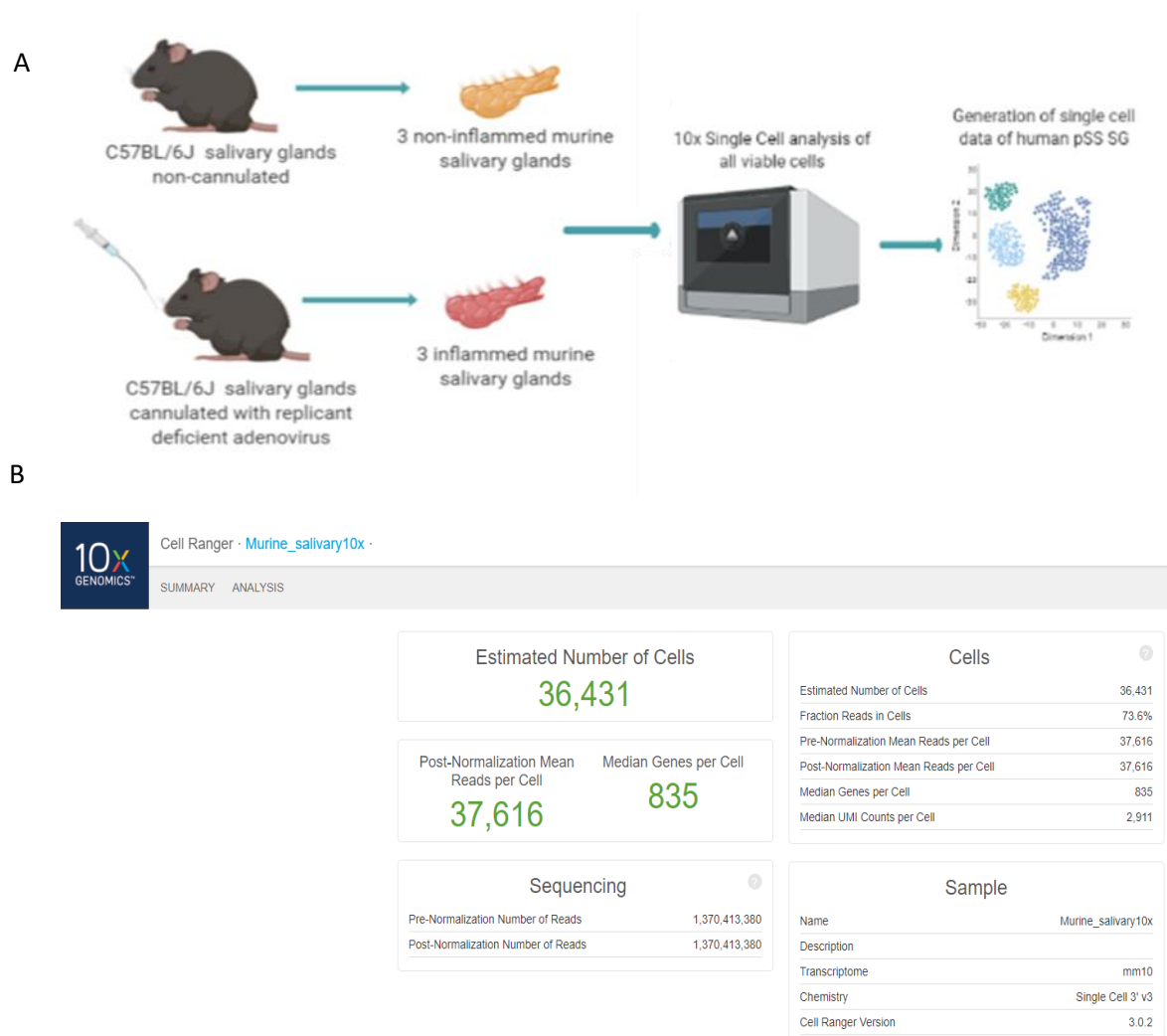


Figure 4: A) Flow of the work process of three non-inflamed and 3 inflamed murine SGs for scRNAseq analysis. **B)** CellRanger output summarising the QC of the sequenced data of all the murine SG samples. This includes key metrics including estimated number of cells, median genes per cell, number of reads in sequencing, and cell profile after processing.

Figure 4: Plots of murine SG QC metrics of the raw data prior to downstream bioinformatic analysis

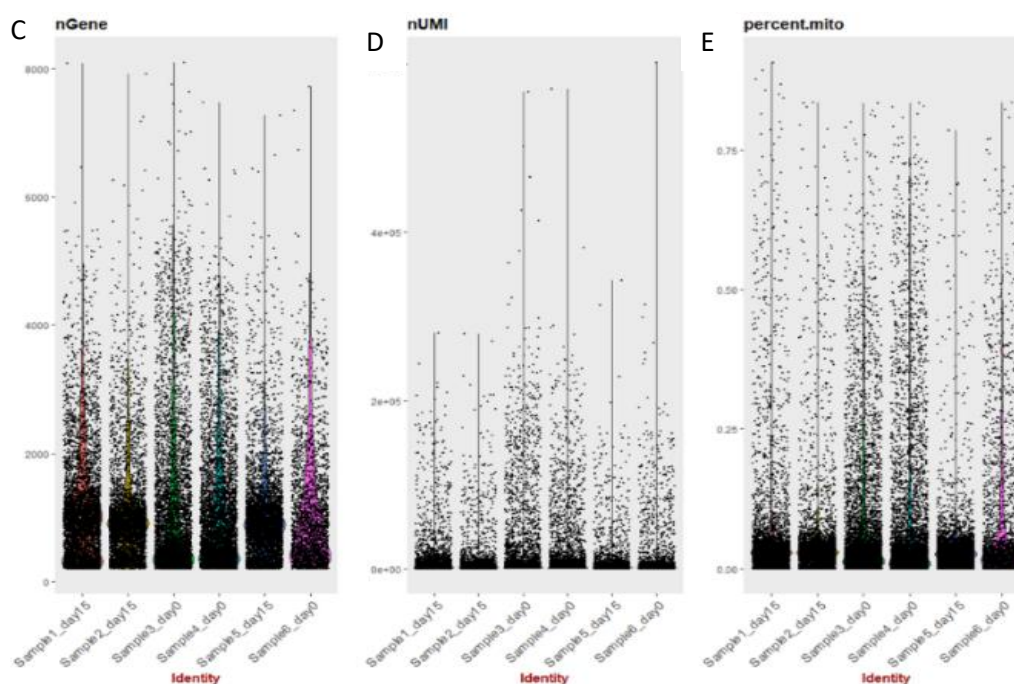


Figure 4: C, D, E) Plots demonstrating the number of genes (nGene), number of unique molecular identifiers (nUMI), and percentage of mitochondrial genes (percent.mito) per sample. F, G) Principal component analysis of Day 0 and Day 15 samples.

2.4: Removal of doublets and ambient gene signatures in the scRNAseq dataset

Following this initial QC of the aligned data, technical artefacts within the data may be generated as a consequence of two key factors. Firstly, droplet based scRNAseq aims to encapsulate a single cell within a microfluidic droplet [185, 186, 190, 192]. However, often this method generates *multiplet* artefacts whereby two or more cells become encapsulated within the same droplet. In contrast, droplets may contain no cells which results in the sequencing of cell free RNA from the input solution which becomes incorporated into the droplet instead [193]. In order to address these artefacts we have incorporated the use of *single cell remover of doublets (Scrublet)* and *SoupX* algorithms in our analysis pipeline [193, 194]. The identification of such clusters which present themselves as “noise” rather than biological data does not occur *a priori* to the visualisation of the data. In essence, the screening of the gene expression and manual identification of typical cell types is undertaken such that cell clusters appearing difficult to *classify* are then further investigated before being assigned as technical artefacts. I will briefly discuss the algorithms applied in our dataset to address these issues.

Figure 5: Diagrammatic summary of generation of ambient cell signature in scRNAseq datasets

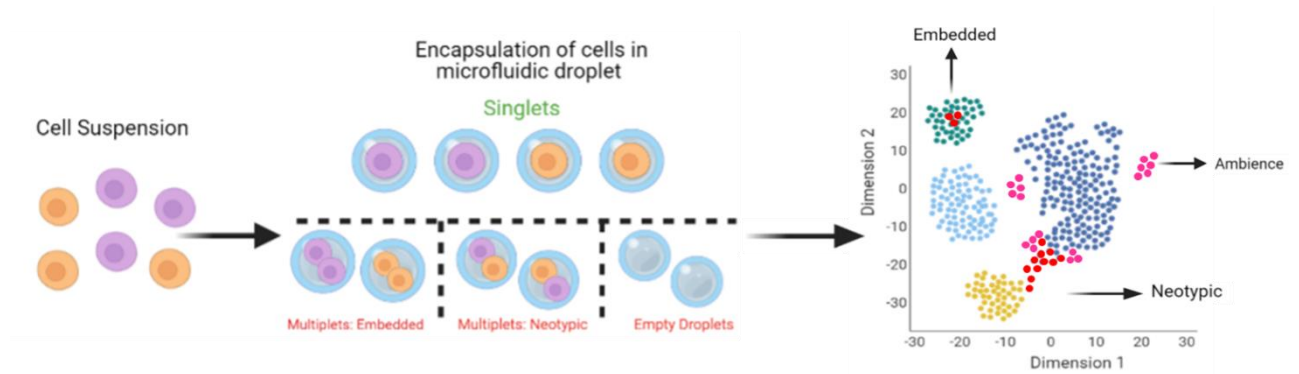


Figure 5 : A) Single cell suspension of the sample of interest prior to cell loading for 10x Chromium processing. B) Top panel: capture of single cells in individual microdroplets. B) Bottom panel encapsulation of multiple identical cells, multiple different cells or no cells in a single microfluidic droplet. C) Positioning of either embedded, neotypic or ambient signatures during data visualisations.

2.5: Dimensionality reduction and normalisation of the scRNAseq data

The datasets generated from scRNAseq are *high dimensional* and require *dimensional reduction* [186]. The concept of *dimensionality*, in the context of scRNAseq, refers to the many different features the dataset contains, in this instance the features refer to the different genes captured. This reduction in dimension is undertaken in order to facilitate downstream analysis whilst simultaneously visualising the data and reducing its noise [190]. In order to be able to visualise the data, the first step in *dimensionality reduction* is *feature selection* [190]. *Feature selection* comprises of the retention of the highly variable genes that “drive” the variability which arises within the dataset. This *high dimensionality* associated with this dataset can interfere with the statistical robustness of the observations seen, also known as the “the *curse of dimensionality*” a term first coined by Richard Bellman [191]. This “*curse*” essentially mathematically models the concept that increased number of features make it difficult to cluster observations effectively [191]. This occurs as a result of the increased dimensions diluting the differences between observations and thus making them appear as if they are all equidistant from one another in mathematical space. As a consequence, the clusters which would appear equally similar, or different, resulting in the loss of the capture of the heterogeneity present within the dataset. For this reason *dimensionality reduction* of the dataset is necessary in order to unveil the prominent underlying trends present within the data which allow for further biological exploration [186].

Following the initial QC, the datasets for the murine and human SGs were first visualised in its dimensionally reduced form as a PCA (Fig 6A and 6B). The PCA visualises the top principal components (PCs) that capture the variability within the dataset, with the most relevant PCs

in our dataset being between 1 and 30. This summarisation of the data following the relevant QC analysis demonstrated that the variation occurring within the data set was not greatly driven by one sample and that the correction of any batch variation was successful following the tissue processing and sequencing of the different murine SG samples we had used. Following the generation of the PCAs, the current methodology of visualising scRNAseq data comprises of the use of uniform manifold approximation and projection (UMAP). The visualisation of the data in this *dimensionally reduced space* captures the differentially expressed genes driving the clustering of the cells. The use of UMAPs has become common integrated practise in scRNAseq analysis due to greater practical applications such as UMAPs can be used to process datasets without the necessary requirement of the initial PCA. Therefore, we have used UMAPs to visualise the scRNAseq datasets used within this work.

[186][189]Once the data set has been normalised as previously described, variability may still be present as a result of batch effect. Batch effect arises when the cell suspension samples are processed as distinct groups due their processing within differing sequencing lanes within a flow cell or the capture of tissue at different time points. These differences in the sample acquisition and sample processing can induce variation in the detection of the genetic signature. In this instance, the human scRNAseq has utilised 7 SG for the pSS and the murine scRNAseq has used 3 resting and 3 inflamed SG which required bioinformatic batch correction in the datasets. The *batch correction* and *integration* of this data was conducted using the package *Harmony* version 1,0 [195].

Figure 6: PCA of the human pSS SG dataset after normalisation using Harmony

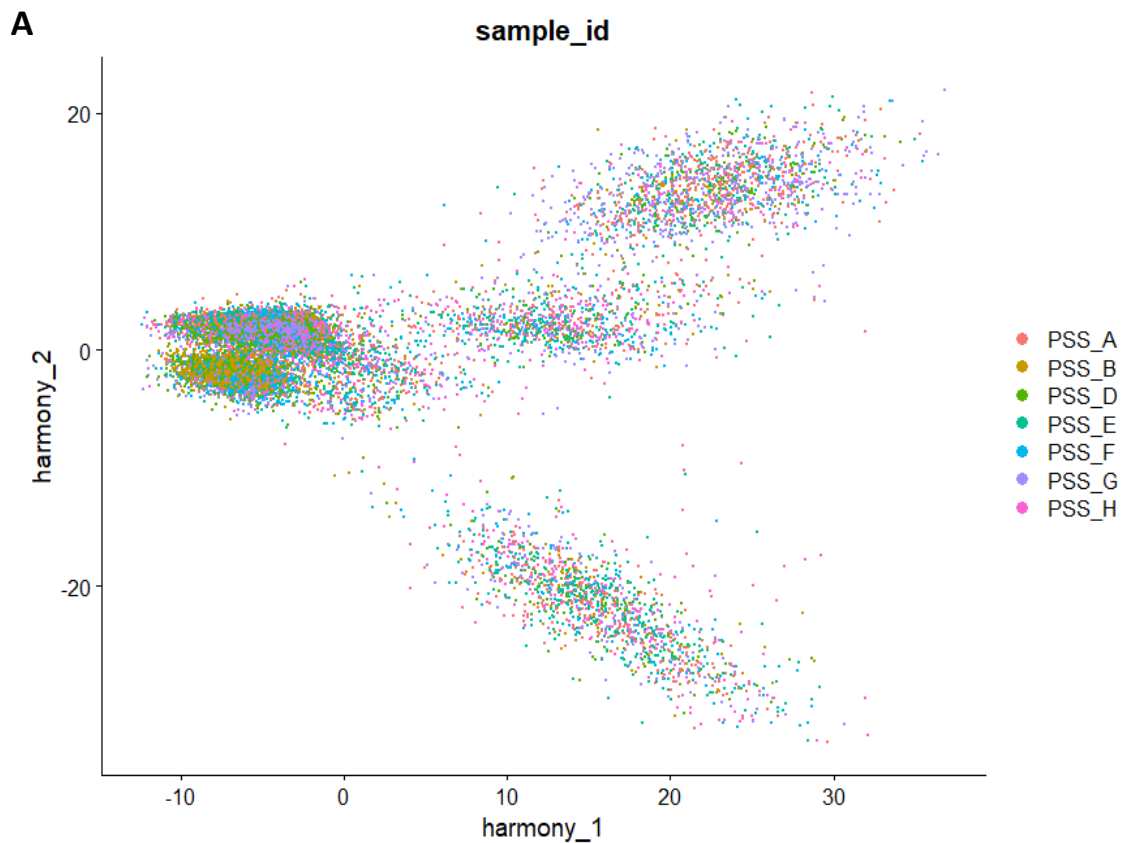


Figure 6: F) PCA plot demonstrating the overlap of human pSS samples following batch correction using the tool Harmony, to ensure that variability in the data is overwhelmingly being driven by all the samples used within this study and not one particular biological contributor.

Figure 6: PCA of the murine pSS SG dataset after normalisation using Harmony

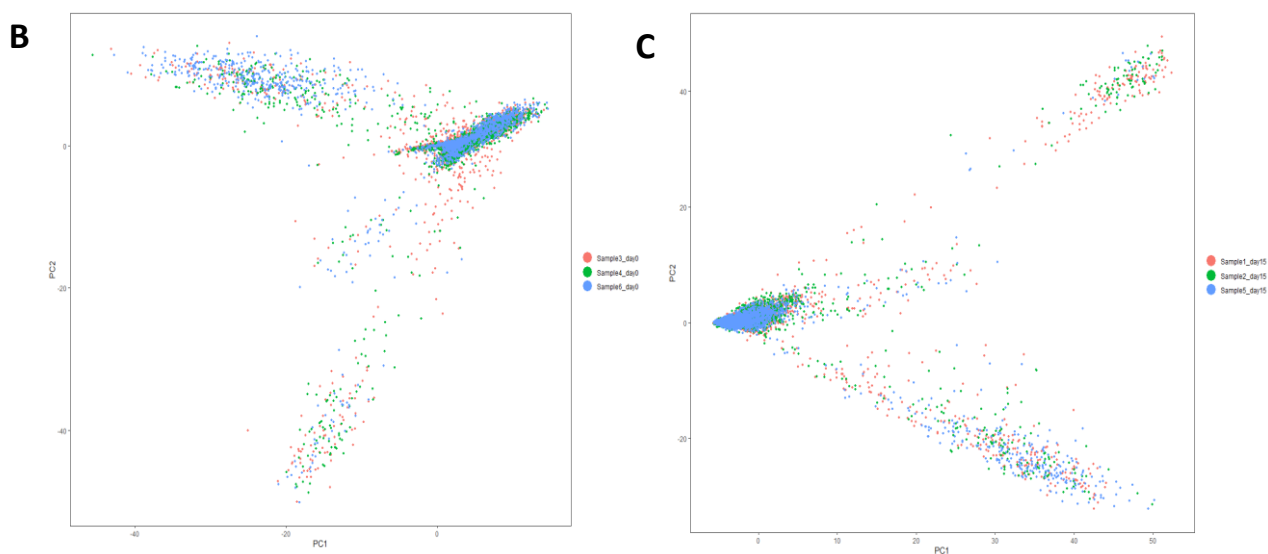


Figure 6: B,C) PCA plot demonstrating the overlap of murine pSS samples at D0 and D15p.c. following batch correction using the tool Harmony, to ensure that variability in the data is overwhelmingly being driven by all the samples used within this study and not one particular biological contributor

[2.6: Downstream analysis tools used for the biological inference of scRNAseq data](#)

Both datasets were analysed using established scRNAseq tools in Seurat_2.3.4 using R version 3.6.1 to assign cell cluster identity, identify differential gene expression, infer trajectory development and engagement of signalling pathways using pathway enrichment analysis. Below I will describe the different computational techniques applied within Seurat to successfully analyse the scRNAseq datasets using the tools.

[2.6.1: Cell Cluster annotation and differentially expressed gene \(DEG\) analysis](#)

Cluster analysis and annotation were conducted using feature plots and violin plots within the *Seurat* analysis package. These plots demonstrated the expression of given mRNA within the cells captured within the dataset. Absence of plots denotes to the absence of the mRNA within the dataset being explored. DEGs are determined as those with a P value of <0.05 [186]. DEG lists for each cell population were generated by reclustering cell groups from the global UMAP to generate individual population UMAPs and gene lists (see appendix 1.21). This allowed for statistically significant genes to be identified within gene lists.

[2.6.2: Slingshot Trajectory Analysis](#)

Trajectory analysis was completed using the *Slingshot* version 1.2.0 package. The purpose of applying the *Slingshot* algorithm is to use selected clusters of cells and to infer global structure of the data and identify developmental lineages using *pseudotime* within the global structure of the cell clusters we have selected for investigation. The visualisation of these *pseudotime* lineages allows for the inference of potential developmental relationships between cell

clusters based on expression patterns of genes on a mathematically calculated trajectory [196].

[2.6.3: Geneset enrichment analysis \(GSEA\)](#)

The GSEA algorithm version `gsfisher_0.2`, infers for possible signalling cascades which maybe active in specific cell clusters within the scRNAseq datasets. This tool functions to identify pathways which maybe enriched in a gene list in a statistically significant manner by highlighting specific genes (or features) which are present within cell populations. These genes are then matched to a pathway analysis database which corresponds to the possible pathways that the genes may belong to. Following this analysis, it is necessary to manually explore the signalling cascades associate with cell populations i.e. confirming the gene and protein expression of relevant receptors within the cell populations such that it is possible for the identified signalling pathways to be activated within cells [185, 186].

[2.6.4: Pseudobulk Correlation](#)

This analysis package allows for the differential gene expression analysis between conditions of different cell types. In the context of our data, we wanted to explore whether there were any similarities or differences in the mRNA expression captured between the murine SG subsets within the inflamed and resting SG. In this instance, the cells of interest are subsetted and differential expression analysis is applied to them. Following this the raw counts of the cells to be used for differential expression are extracted and aggregated together. This allows for the similarities and differences in mRNA gene expression profiles to be identified resulting in the comparison of cell mRNA profiles depending on activation state or the anatomical location of cell populations in question [185, 186, 196].

2.7: Human pSS SG Histology

Indirect detection by fluorescence was based on the Opal™ Multiplex immunohistochemistry (IHC) method (Akoya) and performed on the Autostainer Leica Bond RX. Samples were sequentially stained with anti-human PDPN (Abd Serotech), CD146 (Sigma), CD36 (Cell Signalling), CD82 (Cell Signalling) CD34 (Leica Microsystems)(Leica) and CD90 (Cell Signalling). Staining was performed using the Opal™ 7-Color Automatic IHC Kit (Akoya) according to the manufacturer's recommendations. Signal amplification and covalent binding of fluorophore was achieved by using a tyramide signalling amplification reagent (included in the Opal kit) that is conjugated with a different fluorophore for each cycle. Multispectral images were acquired at ×20 magnification using the Vectra Polaris Automated Quantitative Pathology Imaging System (Akoya). MoTIF settings were used for multispectral image acquisition. Multispectral image processing of multiplex IHC stains was performed using Phenochart (version 1.0.11/Akoya) and inForm Image Analysis Software (version 2.3, Akoya).

2.8: Microdissection of human pSS SG and bulk sequencing analysis

The selected tissue samples for microdissection processing were acquired from patients recruited within the OASIS cohort at the University of Birmingham, under ethics number 10-018. Preparation of the samples for microdissection: all surfaces of the cryostat and lab bench were thoroughly cleaned with 70% Etoh and sterilised with ultraviolet (UV) light for at least 30 min. Containers used to keep store the collected samples and the filter paper used to separate slides (Membrane slides, Carl Zeiss™ 1.0 PEN NF 41590-9081-000) must also be treated with UV for 30 min and kept sterile before handling the tissue. To cut tissue for microdissection, the UV sterilised membrane slides were maintained at a cool temperature within the cryostat prior to sample cutting. The tissue was mounted on a cryostat block with as little mounting media (OCT16-004004) as possible to minimise interference with laser cutter during microdissection. Sections were cut at 8 µm thickness and transferred to the membrane slides. Once the section had been mounted onto the membrane slides, the membrane slide was stored in its carrier box and in dry ice until final storage at -80oC prior to staining with Cresyl violet acetate (SIGMA, C5042). Preparation of Cresyl violet acetate is required one day prior to staining the mounted tissues. Preparation of the CresylViolet (Sigma, C5042), (1% solution) included dissolving the powder in 80% ethanol which had been made using sterile RNase-free water in a MSC Class II ventilator system. This solution was mixed overnight on a shaker at 4oC to ensure that the CresylViolet (Sigma, C5042) was completely dissolved and the solution was then passed through a 70µm strainer before use. The slides were first dipped in RNase-free water and suspended for 2 min in 70% Etoh, also prepared in RNase-free water, to facilitate the removal of OCT on the tissue. The slides were then

suspended in the CresylViolet 1% solution for 30 secs to 1 min and excess solution was tapped off. The slides were then processed for 1 min in 100% Etoh and this step was repeated three times. The PALM slides were then left to dry until the membrane was observed to be completely dry. The PALM slides were then stored at -800C in a parafilm sealed 50 ml falcon until the samples were ready for microdissection. Prior to microdissection, the PALM slides were brought to room temperature for 10 min. Microdissection was completed using the PALM Robo Software V.4.6 software using the brightfield “AxioCam CC1” setting. The objective lens used to micro dissect SG sections was 10X whilst tonsil was set at 5X. The area of interest was then identified using the ocular eye piece. Once the region had been selected the laser was aligned to the section that was to be cut. The collection tubes required for the capture of the microdissected tissue contained 20 ul of RLT buffer with B-mercaptanol, ensuring the lid was kept dry, and tube was kept on dry ice prior to use. The collection tubes were then mounted onto the Collector Set 200 CMII and set to the correct position to collect the microdissected sample as they were cut by applying the calibration procedure in order to collect tissue in the centre of the lid of the collection tube which were stored at -80oC prior to RNA extraction. RNA extraction from the microdissected samples was completed as per the protocols provided using the RNeasy FFPE kit (cat. no. 74404). RNA extracted from these samples were then subjected to Bulk sequencing at The University of Birmingham Genomics facility and processed using the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Cat. Nos. 634888, 634889, 634890, 634891, 634892, 634893, and 634894). Fastq files from multiple flowcells were concatenated by sample and sorted to ensure correct read paring. Quality was assessed using FastQC (v0.11.5). Samples were then aligned using the STAR aligner (v2.6.0) to the human assembly GRCh38. A count matrix was generated using feature Counts (Subread v1.5.3). All

further analysis was completed in DESeq2 (v1.24.0, Or v3.6.1) and genes expression was statistically tested using the 3-way comparison ratio test.

Chapter 3: Transcriptomic characterisation of the human pSS SG

The presence of SG lymphocytic aggregates is also clinically used to histologically diagnose pSS [197]. The presence of TLS in the SG is a typical feature of pSS, classically associated with inflammation and loss of organ function, resulting in dryness of the mouth, eyes and other exocrine gland dysfunction [22, 39, 198]. Moreover, TLS are believed to support the production of antibody secreting PC which produce the hallmark pSS anti-nuclear autoantibodies Ro/SSA and La/SSB [199], whose presence associates with disease severity [59].

Intriguingly, the presence of TLS has been also associated with disease progression, with patients affected by highly organised TLS which harbour GCs and is associated with an elevated risk of developing B cell lymphomas [39, 48, 200]. Mechanistically, the presence of GC like structures, within the TLS, is believed to provide B cells a pro-survival microenvironment which promotes B cell aberrant mutation and proliferation [200] [39, 48, 200]. The underlying fibroblast compartment of TLS have been suggested to govern the formation of these structures as their immunomodulatory properties have been uncovered in different anatomical and disease contexts [8, 151, 154]. In this manner, the Barone group has previously demonstrated, using ex vivo and in vivo approaches, the critical role of fibroblasts in the context of TLS formation and maintenance [7, 8]. To explore the biology of TLS formation in an ex vivo manner, Nayar et al utilised human pSS SG biopsies to retrieve human pSS fibroblasts. This aided in the characterisation in the phenotype and function of this stromal compartment in response to various cytokine stimulations as reported by Nayar et al. [8]. Furthermore, access to primary human pSS samples has allowed for the expansion of

understanding TLS within the human SG using histological approaches which demonstrate the possible organisation of T and B cells within the human SG. Additionally, in order to complement and explore TLS formation and development utilisation of the inducible TLS mouse model to address specific questions to explore the heterogeneity of the fibroblasts underpinning TLS formation and function [8, 93].

In our previous work we have identified and characterised the compartmentalisation of fibroblast function and heterogeneity in the human pSS SG prior to the use of transcriptomic analysis [8]. In this manner, we explored the expression patterns of key fibroblast activating markers such as PDPN, FAP, ICAM-1, VCAM-1 and CD34 of pSS patient SG samples and identified a homogenous pattern of expression for these markers [8]. Moreover, CD34 expression was enriched amongst the patient samples and from which 2 populations of PDPN+ fibroblasts were identified as being PDPN+ CD34+ or PDPN+ CD34- [8]. Intriguingly, further exploration of these PDPN+ CD34+ or PDPN+ CD34- populations demonstrated differential expression of the lymphoid chemokines within PDPN+ CD34+ cells and the lymphocyte survival factors IL-7 and BAFF within the PDPN+ CD34- population [8]. Collectively, this in vivo work gave us the first insight into the compartmentalisation of function of human pSS SG fibroblasts. Until now, the heterogeneity of the cellular landscape of the human pSS SG has not been characterised using single cell RNA sequencing analysis. Furthermore, whilst anatomical differences between TLOs and SLO GC have been described, the transcriptomic profiling of these GC structures from the different immunological structures has not been characterised. Exploration of this avenue could uncover the engagement of specific pathways which may be exclusive to TLS formation that could be therapeutically targeted to either induce or dissolve these structures.

In the work underpinning this thesis, we have generated a single cell dataset from 7 human pSS SG biopsies containing both the haematopoietic and non-haematopoietic compartments. The cell populations were not enriched for prior to single cell processing which allowed us to preserve all the different cell types within the tissue. These two cell compartments were segregated using the immune cell marker CD45, and identities of these cell populations were more specifically assigned using canonical and established cellular markers which I will discuss further on in this chapter. We have then explored the different cell populations which are present within the inflamed SG and characterised the heterogeneity we have observed. In this chapter, I will describe and discuss the heterogeneity of the different cell compartments within the human pSS SG and the different cellular sources of lymphoid chemokine expression which includes fibroblasts and mural cells in the human pSS SG. We further our exploration of heterogeneity within TLS by observing variation in the phenotypic characteristics of GCs in the human pSS SG.

3.1: Deconvoluting the heterogeneous human pSS SG using scRNAseq.

The emergence of tools such as scRNAseq, has enabled scientists to synchronously explore and visualise the different cell populations and the possible cellular interactions they may engage in at a granular level within a tissue. These interactions can be inferred by bioinformatically available tools such as CellPhone DB amongst other bioinformatic tools [201].

However, this technology has limitations which need to be considered. In this instance, whilst scRNAseq allows the cellular heterogeneity to be explored at the gene level it does not confirm that capture of the expression of genes of interest is directly correlated with protein expression. Furthermore, the gene expression does not demonstrate the quantity of protein expression within the cell of interest. For this reason, validation at the protein level using tools such as flow cytometry and histology are critical. Moreover, scRNAseq does not capture all transcripts which are associated with classical cell populations which may be attributed to the limitations of the technology regarding the cell capture rate and the sensitivity of the chemistry of the single cell processing kit used. Currently, this approach has allowed various studies to capture and characterise the fibroblast diversity of various anatomical sites such as LNs [89] and disease states including cancers [202], yet this approach has not been conducted to unravel the fibroblast heterogeneity present within TLS in the context of the inflammatory disease pSS. In order to address this question and gain a higher resolution of the cells present within the inflamed glandular microenvironment harbouring TLS, we used scRNAseq on 7 human pSS SG biopsies (6 female and 1 male) which comprised of 4 Ro/SSA- patients (Gx47, Gx48, Gx69, Gx57) and 3 Ro/SSA+ patients (Gx50, Gx81, Gx93) (Fig 1A). The inclusion of both

Ro/SSA+ and Ro/SSA- patient subtypes in our dataset allowed us to explore the differences between proportions of cell clusters between these categories of patients, that significantly segregate clinically in terms of disease severity and prognosis [199]. The global cellular landscape of the human pSS SG identified 11 cell clusters consisting of 6 stromal cell clusters and 5 hematopoietic clusters (Fig 1B). The individual coarse identity of the cell clusters was assigned using feature plots, which visualised the expression of the classical genes associated with the biological nature of the cell type (Fig 1C).

Figure 1: Global cluster analysis of the human pSS SG

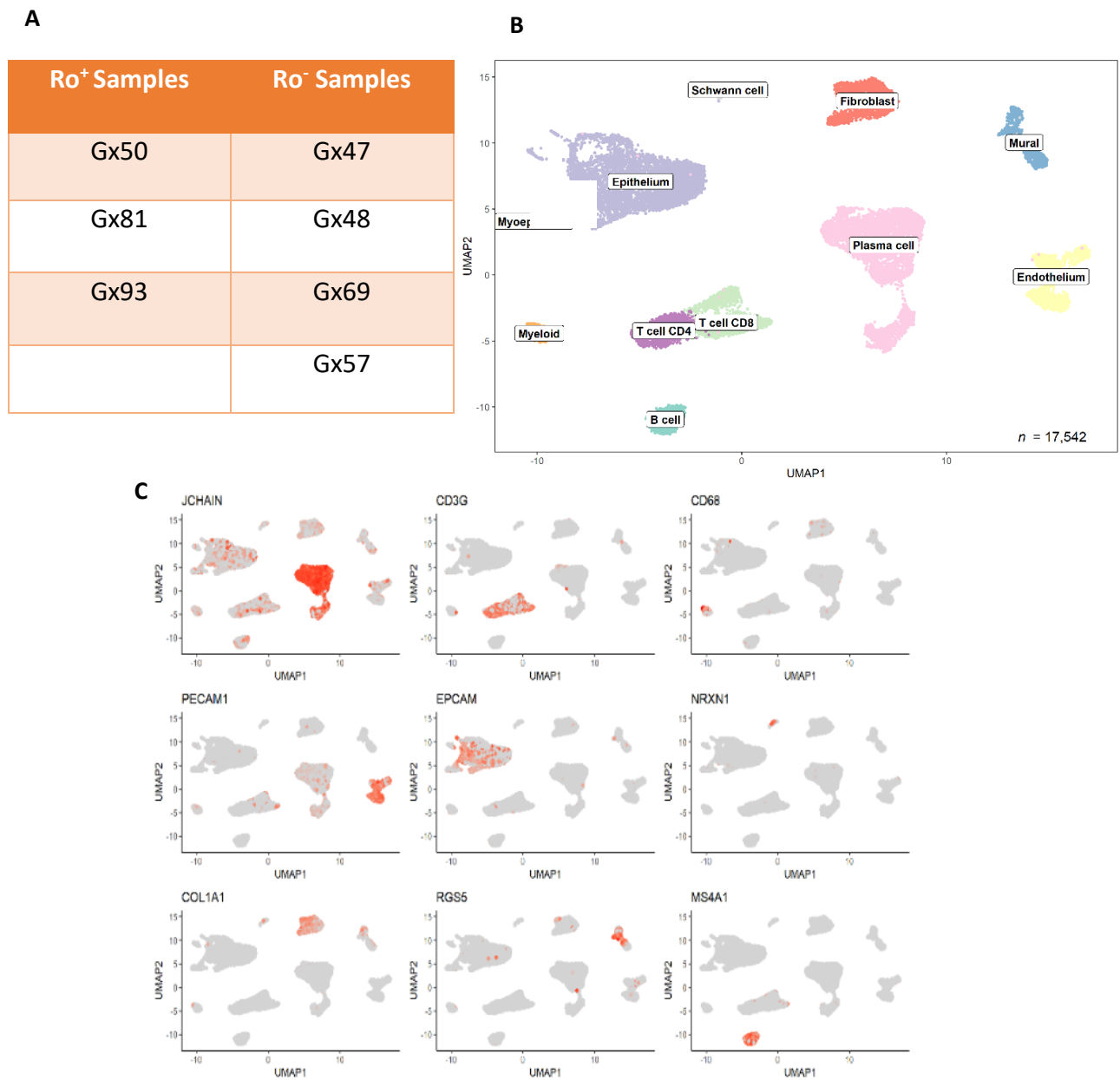


Figure 1: Characterisation of global cell populations from the human pSS SG using scRNA sequencing.

A: Summary of Ro/SSA⁺ and Ro/SSA⁻ patients in our pSS single cell dataset. **B:** UMAP of the over arching cell populations present within the human pSS SG from 17,542 cells derived from *n*= 7 human pSS SG biopsies. **C:** Feature plots of classical genes associated with the characterised cell populations.

3.2: The CD34⁺ CD146⁻ stromal and CD34⁻ CD146⁺ pericyte populations as sources of the lymphoid chemokine CCL19

Recently, the resident stromal cell compartment within inflamed or cancerous tissues, has gained interest as novel targets of immunotherapy and anti-inflammatory therapies. In the context of TLS development, the contributions of the resident epithelial, endothelial, pericytes and fibroblast populations have not been extensively characterised.

Our group has clearly demonstrated a pivotal role of stromal cells in TLS assembly and persistence at sites of permissive tissue [8] [7], showing that, in response to inflammatory cytokines, the underlying stromal compartment within the human pSS SG acquires an *activated* phenotype. TLS *immunofibroblasts* have been thereby characterised, on the basis of the expression of the activation markers PDPN, ICAM-1, VCAM-1 and FAP [8]. Two subsets of *fibroblasts* were identified within the human pSS SG characterised as PDPN⁺ CD34⁺ and PDPN⁺ CD34⁻ [8]. The function of CD34, also known as sialomucin, is that of a T cell adhesion molecule [203-205]. The expression of pattern CD34 in these human pSS fibroblasts was enriched in the SG samples for which reason CD34 was used as sorting marker to explore CD34⁺ and CD34⁻ cells within the PDPN⁺ populations.

These 2 subsets were then observed to be differentially enriched for the expression of either lymphocyte survival factors BAFF and IL-7 or the lymphoid chemokines CCL19, CCL21 and CXCL13 [8]. This finding demonstrated the presence of functional compartmentalisation in the human pSS SG population as qPCR analysis of the aforementioned survival factors and chemokines demonstrated compartmentalised expression within either the PDPN⁺ CD34⁺ or the PDPN⁻ CD34⁻ fibroblasts. Additional studies conducted within our group, have demonstrated that in response to the IL-10 cytokine family member IL-22, fibroblasts also

undergo an expansion phase *in vivo* [7]. Whilst we have been able to explore the different phenotypic and functional dynamics of fibroblasts with both *ex vivo* and *in vivo* approaches, we have lacked the transcriptional resolution which scRNAseq provides.

Using scRNAseq we have expanded these findings and defined the heterogeneity of the human pSS SG fibroblast compartment, associated with TLS formation and persistence. Cluster analysis confirmed our previous findings, validating the presence of both PDPN⁺ CD34⁺ and PDPN⁻ CD34⁻ MCAM (CD146)⁺ stromal cells. **(Fig 2A)**. Within the CD34⁺ cluster we further identified 5 different fibroblast subsets, respectively defined by the expression of the *atypical chemokine receptor 3 (ACKR3)*, *decay accelerating factor (DAF/CD55)*, *ATP-binding cassette sub-family C member 9 (ABCC9)* and *autotaxin (EPPN2)*, *serpin peptidase inhibitor clade E member 2 (SERPINE 2)*, *Thrombin receptor (F2R)*, *FBJ murine osteosarcoma viral oncogene homolog B (FOSB)*, *Cyclin-dependent kinase inhibitor 1 (CDKN1A)*, and *CCL19, TNFSF13 (BAFF)* and *tetraspanin-27 (CD82)* **(Fig 2A & 2B)**.

Using slingshot trajectory analysis and specifying the ABCC9 fibroblast population as the source cluster, three divergent trajectories were found. For all three trajectories, the path first passes from the ABCC9 cluster through the FOSB CDKN1A cluster before trifurcating with each new branch of the trajectory progressing to the remaining three clusters. Overall, this suggests that an initial population of fibroblasts may give rise to three distinct functional clusters of fibroblasts in the salivary gland, from the ABCC9 and FOSB CDKN1A populations into either the ACKR3 CD55 fibroblast, SERPINE 2 F2R fibroblast, or CCL19 TNFSF13B fibroblast population respectively **(Fig 2C)**.

Within the CD34⁻ CD146⁺ compartment we identified 4 clusters of mural cell clusters which comprised of 2 vascular smooth muscle populations, critically differentiated by the expression of the lymphoid chemokines *CCL21* and *CCL19*, the glycoprotein *tenascin C (TNC)* and VCAM-1 (**Fig 2A & 2B**).

Figure 2: The CD34⁺ CD146⁻ fibroblast and CD34⁻ CD146⁺ mural populations

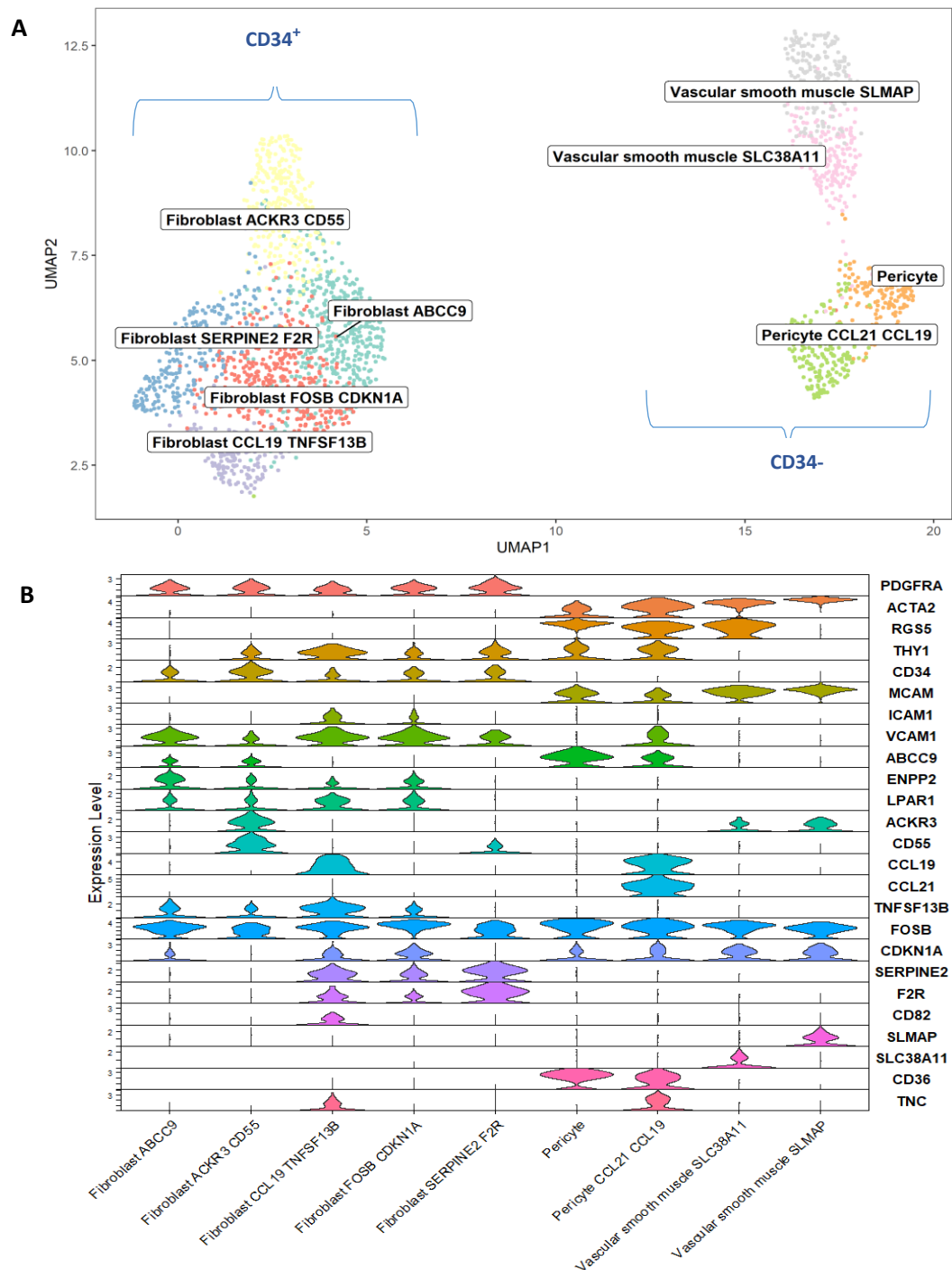


Figure 2 A,B: Exploration of the fibroblast and mural cell populations in the human pSS scRNAseq dataset **A:** UMAP of 5 Fibroblasts clusters which form the CD34⁺ population, and 2 Pericyte clusters and 2 Vascular smooth muscle cell clusters which form the CD34⁻ population. **B:** Violin plot of unique identifying genes for each subcluster.

Figure 2: The CD34⁺ CD146⁻ fibroblast and CD34⁻ CD146⁺ mural populations

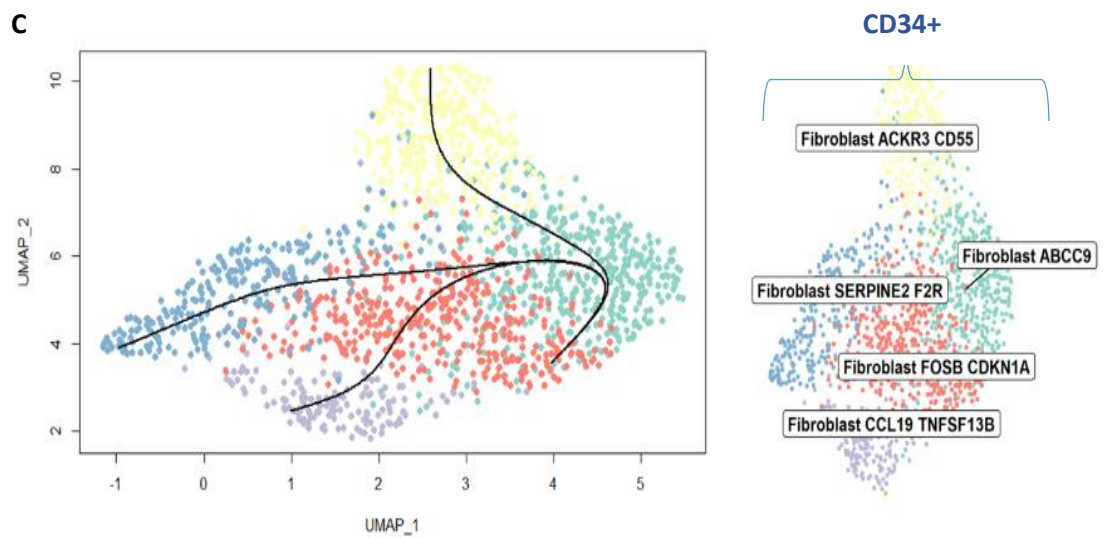


Figure 2C: Exploration of the fibroblast and mural cell populations in the human pSS scRNAseq dataset C:

Violin plots of number of features and counts of mRNA captured within each subcluster of interest. **D:**

Trajectory analysis of development within the CD34⁺ fibroblast population using RNA slingshot analysis.

3.3: The stromal and pericyte population: ex vivo validation

To validate the presence of the identified clusters in SG, we performed flow cytometry analysis on human pSS SG using markers identified in our single cell analysis (**Fig 3**). We focused on exploring the lymphoid like pericyte population using the TNC and CD36 expression within the CD146 population of the human pSS SG.

Firstly, we validated the presence of a CD34⁺ CD146⁻ and of a CD34⁻ CD146⁺ population at the protein level within the human pSS SG (**Fig 3**). Within the CD34⁻ CD146⁺ pericytes of our flow cytometric data, we identified the T cell lymphoid chemokine expressing pericyte population, using the markers CD36 and TNC which we identified the mRNA to be uniquely expressed within the “*immunopericyte*” population within our human scRNAseq analysis (**Fig 3A**). Whilst we have identified the presence of this specific pericyte population of interest, this flow cytometric analysis was conducted on one human pSS SG sample. For this reason, repeating this experiment to confirm whether these specific populations are present within every human pSS SG. In addition to validating the presence of these populations by flow cytometry, we also looked at the spatial arrangement of these populations in the human pSS SG using histology (**Fig 3B & C**). Within this analysis we focused on locating the presence of the PDPN⁺ CD34⁺ CD82⁺ *immunofibroblast* and the PDPN⁻ CD34⁻ CD146⁺ CD36⁺ pericytes. Using 6 colour microscopy, we identify the *immunofibroblast* population to be in a compact cluster within the SG (yellow) whilst the pericyte population (light blue) appears to be located around the fibroblast population and in close proximity to vascular ducts (**Fig 3B**).

These validation experiments confirmed the overarching presence of the CD34⁺ CD146⁻ fibroblast populations but moreover the presence of CD34⁻ CD146⁺ CCL21⁺ pericytes in our

flow cytometric analysis of the human pSS salivary glands. We furthered these findings by exploring the spatial arrangement of our cell populations of interest within the human pSS SG to begin understanding the topology of these populations within the inflamed tissue.

Figure 3: The CD34⁺ CD146⁻ fibroblast and CD34⁻ CD146⁺ mural populations

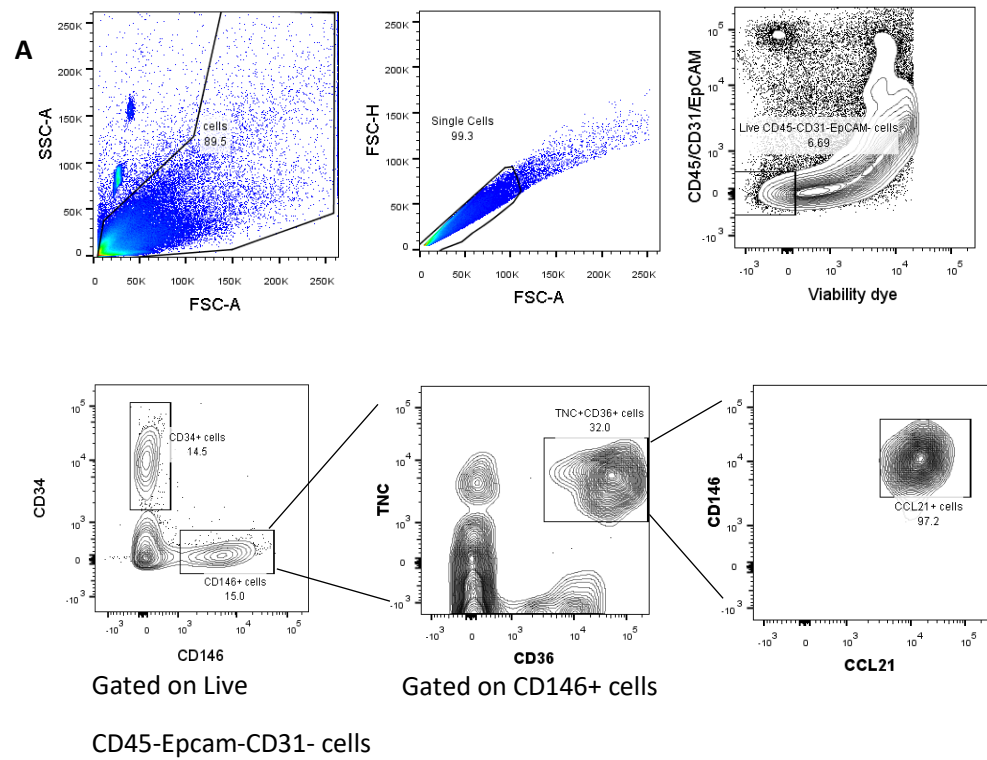


Figure 3A: Validation of the fibroblast and pericyte populations using flow cytometry and histological analysis. A: Representative flow cytometry dot plots of human pSS SG used to identify the CD34⁻ CD146⁺ CD36⁺ TNC⁺ CCL21⁺ pericyte.

Figure 3B: The CD34⁺ CD146⁻ fibroblast and CD34⁻ CD146⁺ mural populations

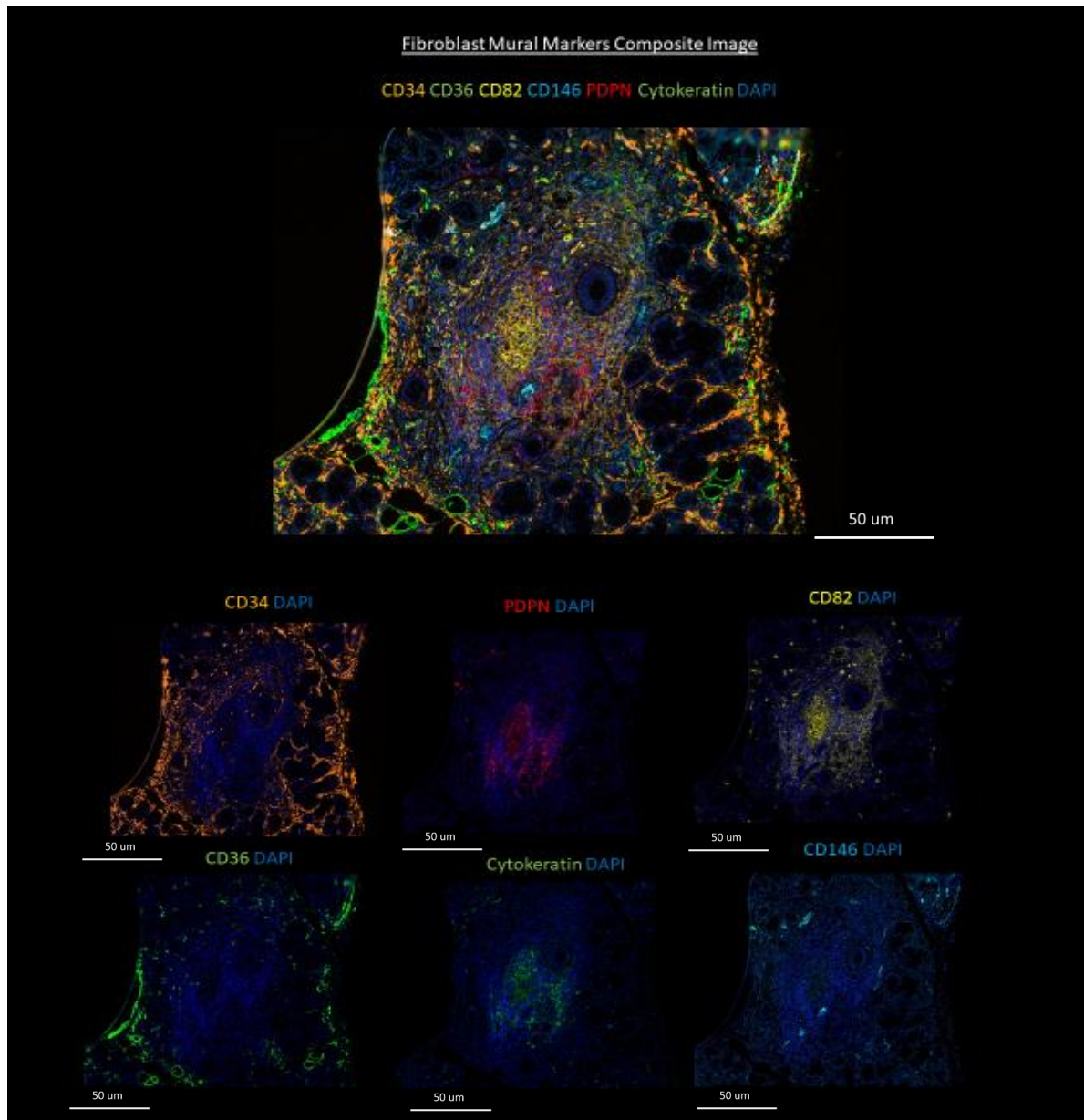


Figure 3B: Histology of the human pSS SG using the markers CD34, CD36, CD82, CD146, PDPN, Cytokeratin and the nuclei stain DAPI. This tissue histology demonstrates the expression of the genes used to demarcate the unique fibroblast populations and the *immunopericyte* population within the human pSS scRNAseq dataset. Image representative of 1 tissue sample.

Figure 3C: The CD34⁺ CD146⁻ fibroblast and CD34⁻ CD146⁺ mural populations

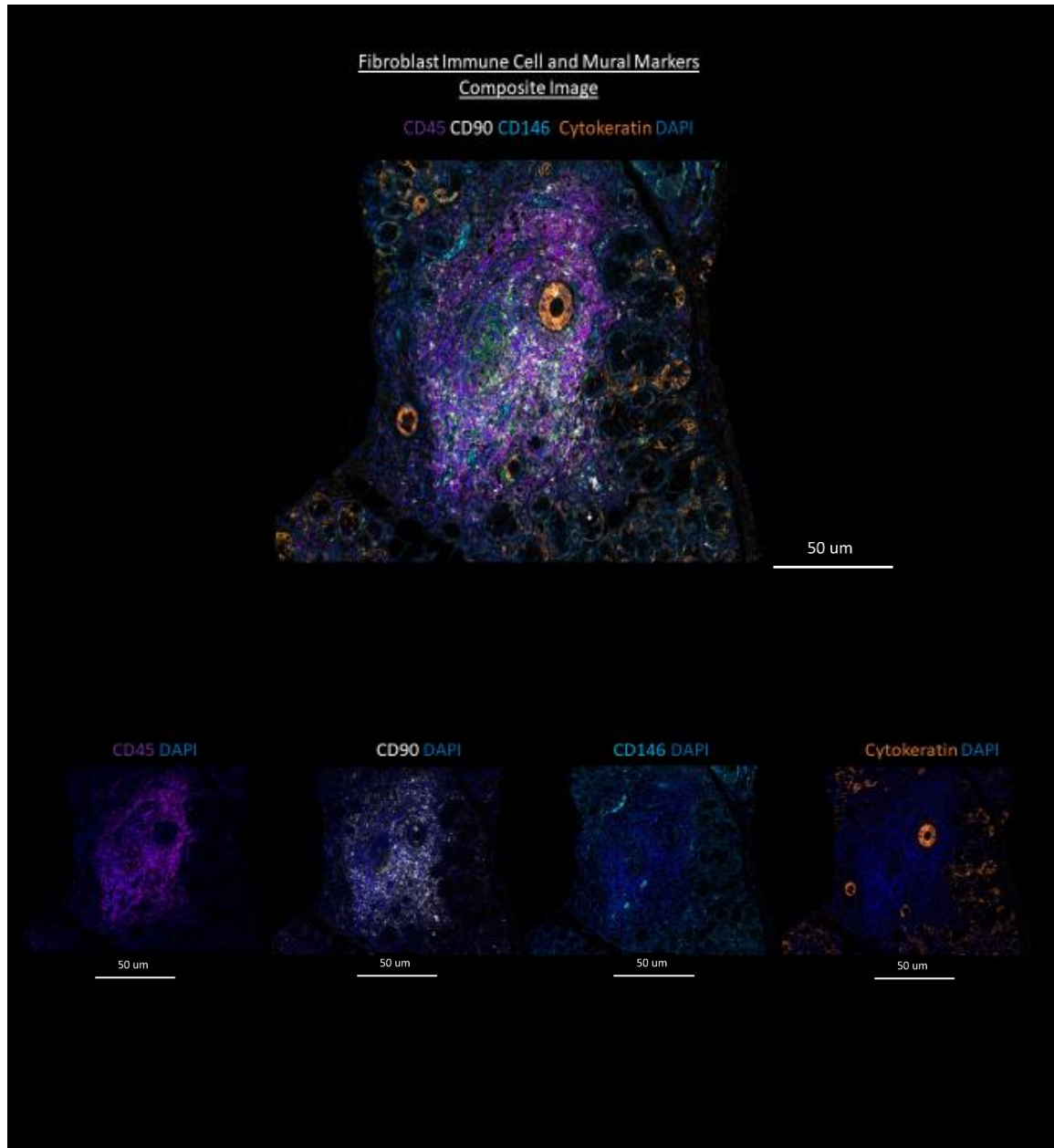


Figure 3C: Histology of the human pSS SG using the markers CD45, CD90, CD146, PDPN, Cytokeratin and the nuclei stain DAPI. This tissue histology demonstrates the expression of the genes used to demarcate the unique fibroblast populations and the *immunopericyte* population within the human pSS scRNAseq dataset relative to the immune cell and stromal cell compartments within the human pSS SG. Image representative of 1 tissue sample.

3.4: GC heterogeneity in pSS SG

The GC reaction is critical in the generation of immunological memory. Within GCs B cells encounter antigens and receive the necessary survival and developmental cues to progress to becoming memory B cells and antibody producing PC [122, 123, 136]. Currently, the exact signalling cascades which result in the formation of TLS able to harbour functional GCs have not yet been completely elucidated. The fibroblast population anchors the formation of TLS and are present in different subpopulations as we demonstrated using scRNAseq of the human pSS SG. We speculate that the heterogeneity of the fibroblast population may contribute to the differential maturation of TLS such that *mature* TLS are defined by the presence of GCs and more evident organisation of B cell and T cell areas compared to *immature* TLS which usually do not form GCs or demonstrate diffuse aggregation of lymphocytes. This definition is not only anatomically relevant; in the context of melanoma different prognosis has been associated to the presence or absence of GC within cancer-associated TLS [19, 66]. Similarly, histological analysis of TLS has unveiled presence of TLS characterised by different degree of organization not only in different patients but also within the same patient and even within the same salivary gland. This suggests asynchrony and differential distribution in the tissue of the signals involved in full maturation of TLS. Interestingly the absence of GC has been thought to be a negative prognostic factor for lymphoma development in pSS, whilst the presence of GC has not been assigned a strong prognostic factor [122, 123]. Considering that MALT lymphoma bears the features of a post-GC B cell malignancy this suggest that the mere formation of GC in the inflammatory SG microenvironment is required but not sufficient to induce lymphomagenesis and that

additional factors or specific features in the GC must be acquired for this to facilitate the expansion of malignant B cells.

To map the cellular heterogeneity that we have identified in the context of the anatomical organization of the TLS we undertook bulk sequencing transcriptomic analysis of 15 GCs which were microdissected from 10 female pSS patient samples. Microdissection under the microscope allowed us not only to spatially segregate TLS from healthy adjacent SG tissue, but also to dissect specific anatomical elements of the TLS. Of note, the GC isolated from SG were all derived from auto-antibody positive patients, suggesting that the differences in the GC profiles were not defined by the Ro/SSA antibody status. In this investigation we also used three GCs micro dissected from tonsillitis derived tonsils of two patients, to compare the GCs isolated from inflamed human pSS SG. In this manner, the tonsils were used as biological calibrators to compare the phenotypes as demonstrated by gene expression patterns within the human pSS GCs (**Fig 4A**).

Exploration of this data using PCA demonstrated the pSS SG GCs to form 2 separate clusters (GC1 and GC2). To explore the potential functional differences between the GC1, GC2 and tonsil derived GC clusters we explored the differentially expressed genes between the different GCs as volcano plots (**Fig 4B**). The differential gene expression patterns identified within these datasets aided us to infer the possible phenotype and functional characteristics of the different GCs we captured. In the first instance, we investigated whether any of the genes associated with the classical lymphoid chemokines and lymphocyte survival factors associated with TLS formation appeared as part of the differentially expressed genes between the GCs (**Fig 4C**). Predominant mRNA expression of the lymphoid chemokines *CCL19*, *CCL21*

and *CXCL13* was detected in a mixture of both TLS and tonsil GCs sampled within our microdissection data (**Fig 4C**). Interestingly, *CCL19* expression was greater in GC14, GC6, GC3, GC7, GC8 and GC5, which were collectively grouped as GC1, as compared to GCs in GC2 and tonsil samples. Conversely *CCL21* expression was greater in tonsil GC (**Fig 4C**). Gene expression for *CXCL12* was enriched in TLS GC as compared to tonsil, suggesting a potential association of this molecule with the highly inflammatory ectopic setting of TLS as it functions to promote the infiltration of CXCR4⁺ immune cells (**Fig 4D**). However, to validate this observation would require quantification of this chemokine at the protein level from such samples. We also looked at the expression of canonical transcription factors which classically demarcate a GC. Intriguingly, the expression of the recombinase *AICDA*, was captured exclusively in the 3 tonsil GCs but not the pSS SG GCs. Previously, it has been demonstrated that *AICDA* expression is low in TLS as compared to SLO, where it functions to confer decreased susceptibility to apoptosis within GC B cells and partake in SMH of B cells [18, 206]. This low level *AICDA* expression supports the observation of irregular B cell survival and the lack of autoreactive B cell clone selection within the ectopic GCs (**Fig 4D**).

The presence of a strong IFN signature has been associated with high lymphocytic infiltration in the pSS SG which warranted its exploration within our dataset [207]. In this manner, we identified an increased mRNA expression of *IFNG Receptor 1* and *2* across all GCs in the microdissection dataset (**Fig 4E**). Additionally, inflammatory genes which participate in the formation of TLS development such as the chemokines *CXCL10*, *CXCL9* and their receptor *CXCR3* were also explored in this data. These additional chemokines were selected for as they are known to be involved in the recruitment and activation of T cells, eosinophils, monocytes, and NK cells [208]. These cells types contribute to the formation of TLS [209]. Interestingly,

these chemokines were detected in TLS GC, but were absent in tonsil GCs (**Fig 4E**), suggesting that TLS GCs are defined by a different pro-inflammatory profile signature from that of tonsil GC.

Exploration of the microdissected GCs in this manner has highlighted that GCs from the same disease may display differing phenotypical characteristics on the basis of the differential mRNA expression patterns of the specific genes which we have explored. The implications of these differences should be considered when evaluating the trajectory development of some of these GC in pre-lymphomatous lesions.

Figure 4: Microdissection analysis of human pSS SG

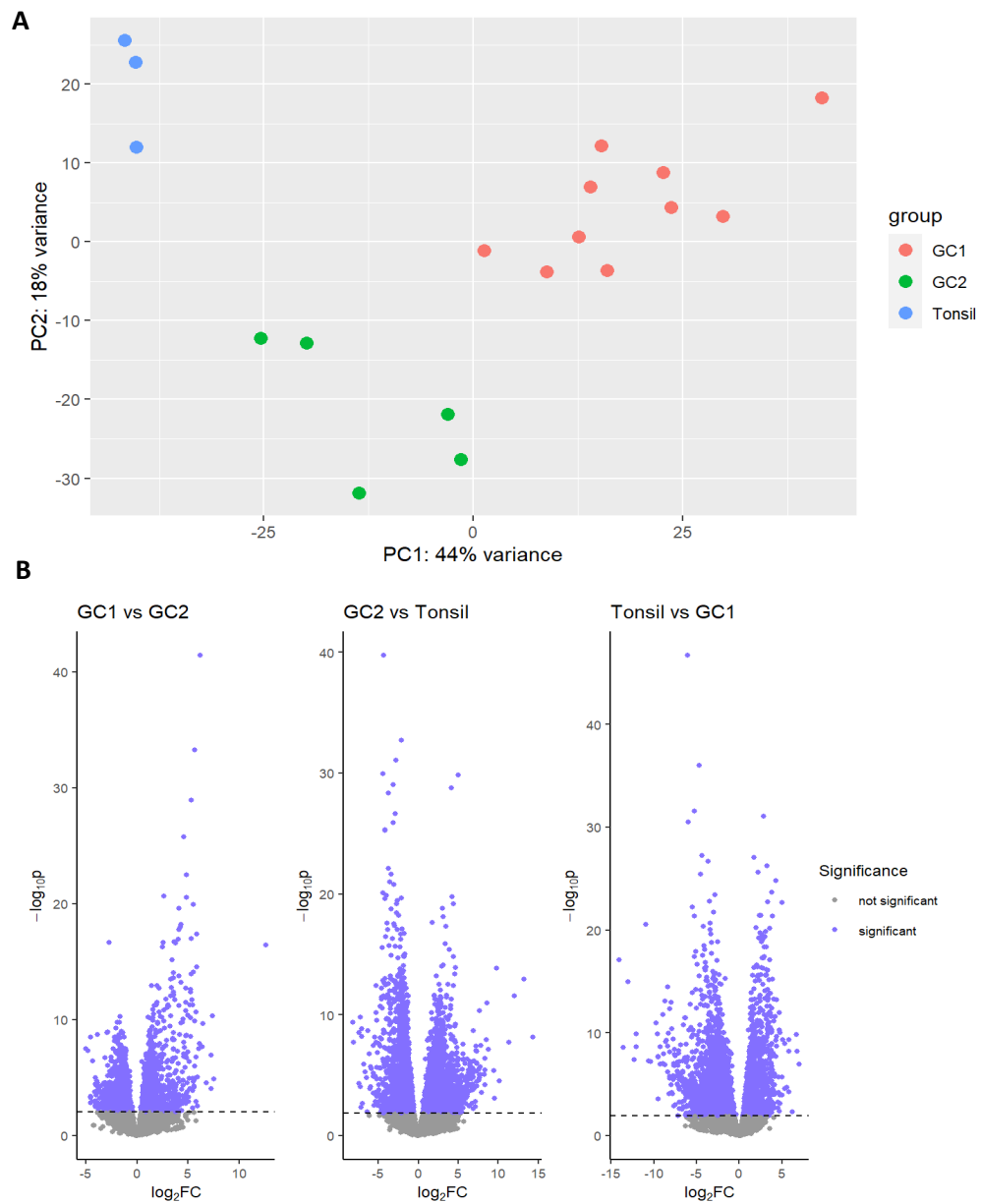


Figure 4: Analysis of microdissected GCs from human pSS SG and human tonsils. A:

Principal component analysis of GCs microdissected from human pSS SG and human tonsils. **B:** Volcano plots of comparison between GC1, GC2 and Tonsil.

Figure 4: Microdissection analysis of human pSS SG with gene expression of key chemokines, interleukins and receptors

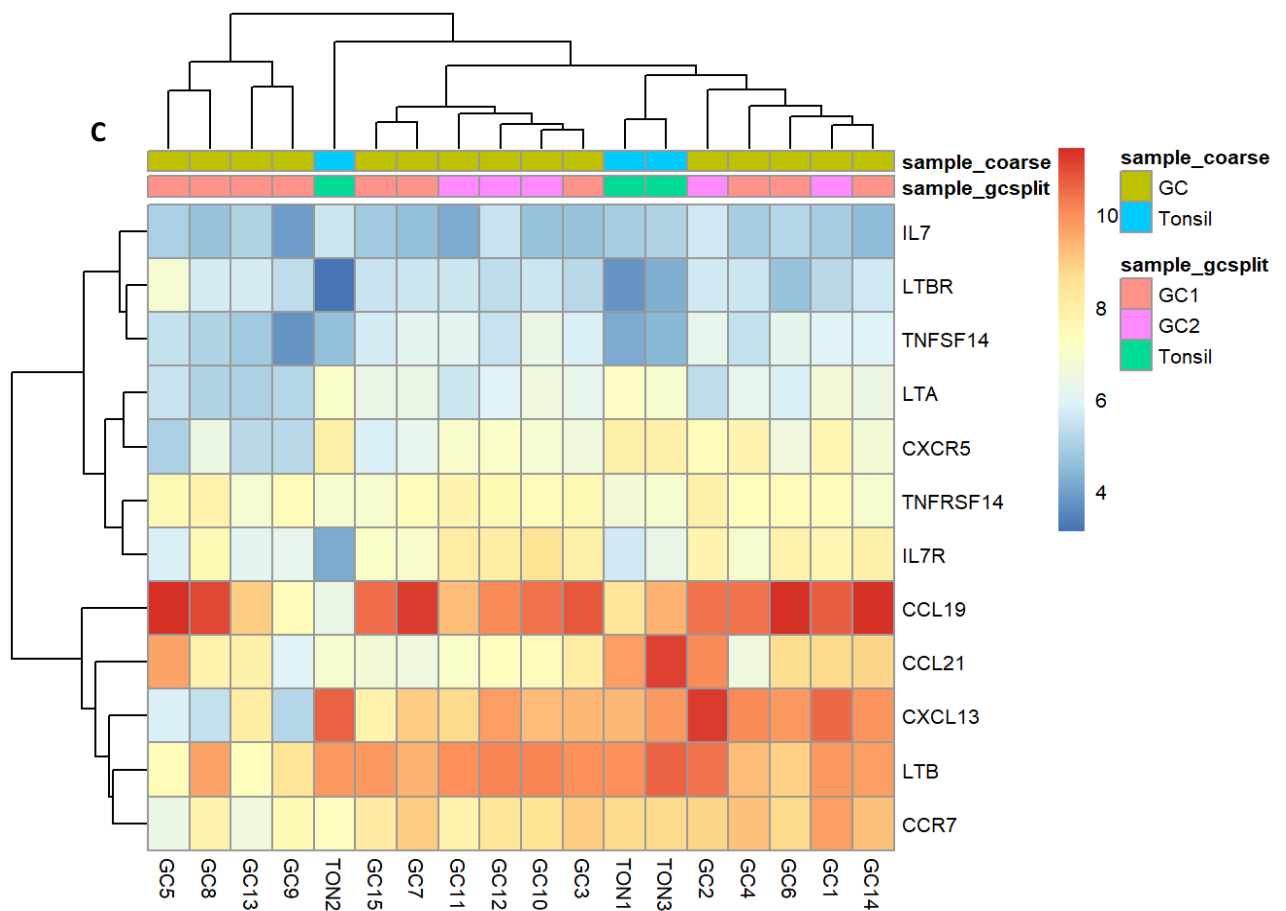


Figure 4: Analysis of microdissected GCs from human pSS SG and human tonsils C:

Heatmap of lymphoneogenesis genes from bulk sequencing of microdissected GCs.

Figure 4: Microdissection analysis of human pSS SG with gene expression of key chemokines, interleukins, receptors and transcription factors

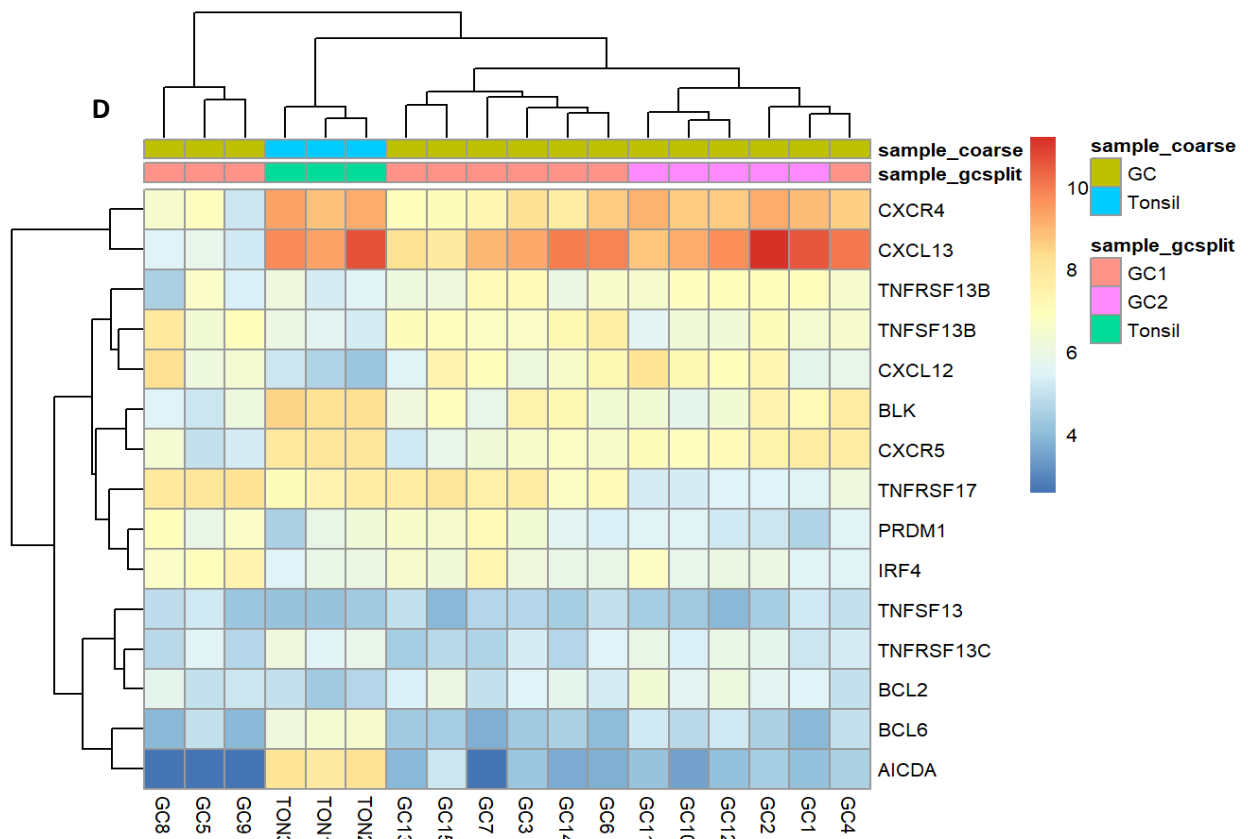


Figure 4: Analysis of microdissected GCs from human pSS SG and human tonsils. D: Heatmap of genes associated with B cell survival and maturation.

Figure 4: Microdissection analysis of human pSS SG with gene expression of key chemokines, receptors and cytokines

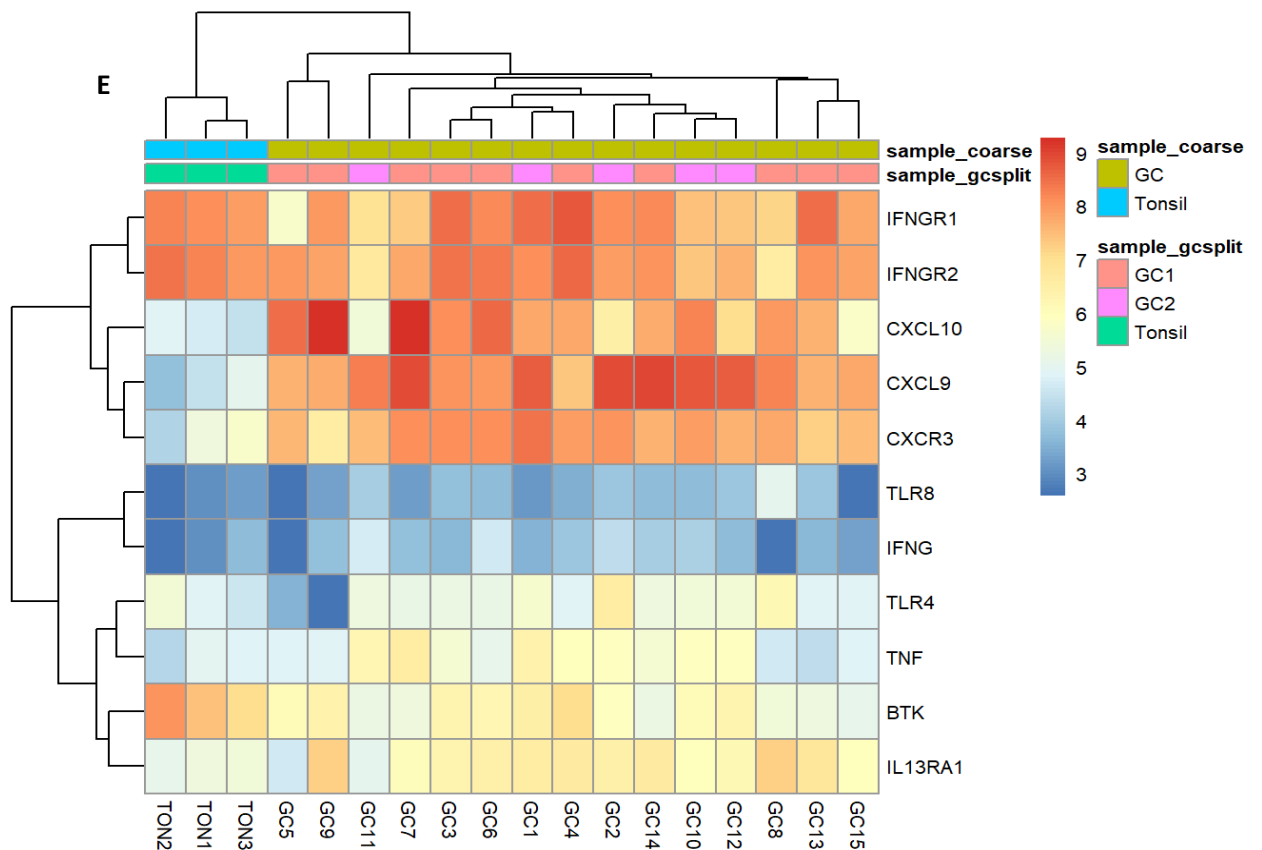


Figure 4: Analysis of microdissected GCs from human pSS SG and human tonsils

E: Heatmap of inflammatory genes from bulk sequencing of microdissected GCs.

3.5: Pathway enrichment analysis of the CD34⁺ CD146⁻ Immunofibroblasts and the CD34⁻ CD146⁺ Pericytes

During SLO development, the compartmentalisation and phenotypic development of the fibroblasts is mainly driven by LT β R signalling [2, 14, 86, 88, 95, 101, 146, 213]. The activation of the canonical and non-canonical NF κ B signalling pathways downstream of the LT β R signalling cascade results in the acquisition of specific lymphoid like characteristics such as the expression of the adhesion molecules ICAM-1, VCAM-1 or the lymphoid chemokines CCL19, CCL21 and CXCL13 [109]. This gives rise to the heterogeneous fibroblast populations characterised within SLO such as the LN [89]. In the context of TLS formation, the contribution of this pathway has also been investigated and identified as key elements for the development of immunofibroblasts in the context of these ectopic organs [109, 128, 133, 213].

In previous work our group has explored TLS formation in LT β R^{-/-} and Rory^{-/-} mice as compared to the TLS development that has been characterised in the wild type (WT) mouse using the inducible TLS murine model [8, 93]. This data reinforced the critical contributions of LT β R signalling, as PDPN⁺ fibroblasts demonstrated reduced chemokine expression which correlated with the observation of less organised TLS in the tissue as compared to WT mice [8]. Exploration of the TLS formation within the Rory^{-/-} mice was warranted as Rory⁺ LTi cells provide the LT $\alpha_1\beta_2$ signalling needed for the embryological development of the SLO mesenchyme [99]. In this manner, TLS development within Rory^{-/-} demonstrated a similar phenotype as the LT β R^{-/-} mice with suboptimal level of chemokine expression therefore resulting in the disorganised formation of TLS [8]. This *in vivo* analysis has provided insight into the contributions of differing signalling cascades on the functional capacity and heterogeneity of stromal populations in the context of TLS.

The individual or synergistic contributions of these signalling pathways, amongst others, could result in the acquisition of the *immunofibroblast* phenotype and different fibroblast identities we observe in our dataset. To gain a bioinformatic insight of the possible different signalling cascades involved within the human pSS SGs in our datasets that could be mediating the formation of pathological cell types, we applied pathway enrichment analysis on our scRNAseq dataset (**Fig 5A**). The odds ratio quantifies the relative abundance of specific genes and whether the expression, the outcome, will occur in a particular exposure i.e. signalling pathway.

In this analysis, we identify the participation of different signalling cascades in the different fibroblast subpopulations. Those include prostaglandin, NFkB, TNF, Wingless/Integrated (Wnt), Bone morphogenic protein (BMP) and a selection of IFN pathways. For each population of cells within the PDPN⁺ CD34⁺ or PDPN⁺ CD34⁻ population we identified a dominating signalling pathway signature. The vascular smooth *SLAMP* population displayed involvement of both the TNF and IFN pathways, whilst the vascular smooth *SLC38A11* predominantly presented involvement of TNF signalling. Intriguingly, the CD34⁻ CD146⁺ CCL19⁺ CCL21⁺ pericyte population showed specific engagement of the prostaglandin pathway compared with the CD34⁻ CD146⁺ non-lymphoid pericyte, which were defined by the participation of genes of the TNF pathway. More specifically within the CD34⁺ CD146⁻ fibroblast populations we observed signatures of different signalling pathways within the different fibroblast subsets. Firstly, the SERPINE2⁺ F2R⁺ population presented an expression signature involving the Wnt and BMP pathways as compared to the other fibroblast subsets, suggesting this population has an element of plasticity as the Wnt and BMP pathways are associated with cellular development and renewal [214-216]. Secondly, the FOSB⁺ cluster demonstrated an

enrichment of the TNF and IFN pathways. The signalling cascades differed in the ACKR3⁺ CD55⁺ fibroblast population had expression of TNF, IFN and BMP pathway related genes suggesting this fibroblast population is responsive to different exogenous signals. Additionally, the ABCC9⁺ fibroblast cluster demonstrated engagement of prostaglandin, Wnt and a strong TNF signature, suggesting that this stromal subset may behave as progenitor population as the association of Wnt signalling has been made with cells which harbour renewal and differentiating capacities [214, 215, 217]. Most intriguingly, the CD34⁺ CCL21⁺ CCL19⁺ *immunofibroblast* population showed specific gene expression of the components of the NFkB pathway, although the expression of LTβR was homogenous across the CD34⁺ CD146⁻ subsets, which was absent in the other CD34⁺ fibroblast populations. This demonstrated the specific involvement of this pathway in the *immunofibroblast* population and is in alignment of the contributions of the LTβR pathway in assigning lymphoid function to mesenchyme cells. Overall, this data captures the different signalling pathways that maybe participating within the CD34⁺ CD146⁻ and CD34⁻ CD146⁺ populations, providing insight into the different mechanisms regulating specific functions within the inflamed human pSS SG microenvironment.

Figure 5: The CD34⁺ CD146⁻ fibroblast and CD34⁻ CD146⁺ mural populations

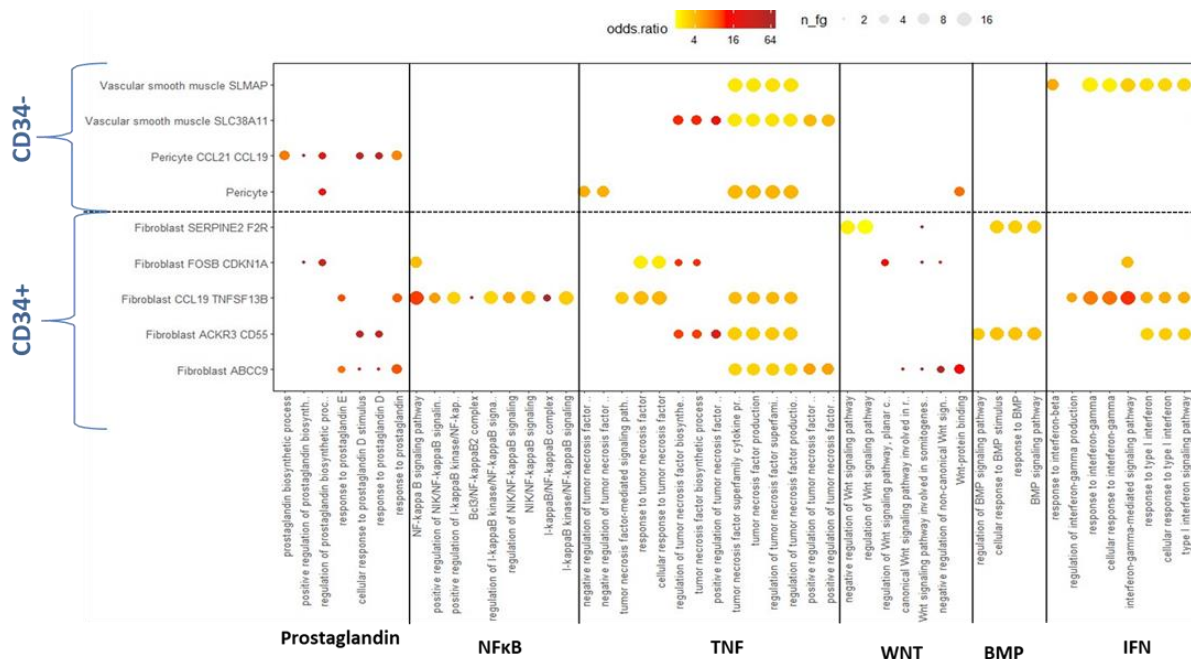


Figure 5: Exploration of different signalling pathways engaged within the human pSS SG in the scRNAseq dataset. A: Pathway enrichment analysis and identification of engagement of different signalling pathways for each CD34⁺ fibroblast population and CD146⁺ pericyte population. Statistical test applied was Fisher's exact test

3.6: The endothelium population

Within the inflamed human pSS SG we identified 4 subclusters of endothelial cells (**Fig 6A**). The 4 endothelial populations were firstly sub-setted into arterial or venous endothelium, with the former exclusively expressing NOTCH4 assigning them an arterial phenotype, while P-selectin (SELP) specifically identified the venous endothelium (**Fig 6B**). Further exploration of gene enrichment within these subsets identified 2 populations of arterial endothelial cells which had unique expression of carbonic anhydrase 4 (CA4) and a regulator of angiogenesis Semaphorin-3G (SEMA3G). The venous endothelial population were, interestingly, was defined by the expression of VCAM-1, whilst we were unable to assign a unique gene identifier for the second venous endothelial population (**Fig 6B**). We also explored whether the relative proportions of these populations differed according to the Ro/SSA status of the patients (**6C**). Intriguingly, this data demonstrated that greater proportion of endothelial venous population within the Ro/SSA⁺ patients as compared to the Ro/SSA⁻. This may suggest that an increased proportion of this cell population exists, and maybe expanded within the Ro/SSA⁺ patients. Furthermore, the endothelial arterial population was present in greater proportion in the Ro/SSA⁺ patients and the endothelial arterial SEMA3G population appeared to be present in a slightly greater proportion in the Ro/SSA⁻ patients (**Fig 6C**). Furthermore, across all the patients evaluated independent of the Ro/SSA status, we demonstrated the presence of this cell type within our analysis (**Fig 6D**).

Figure 6: The endothelium populations in the human pSS SG

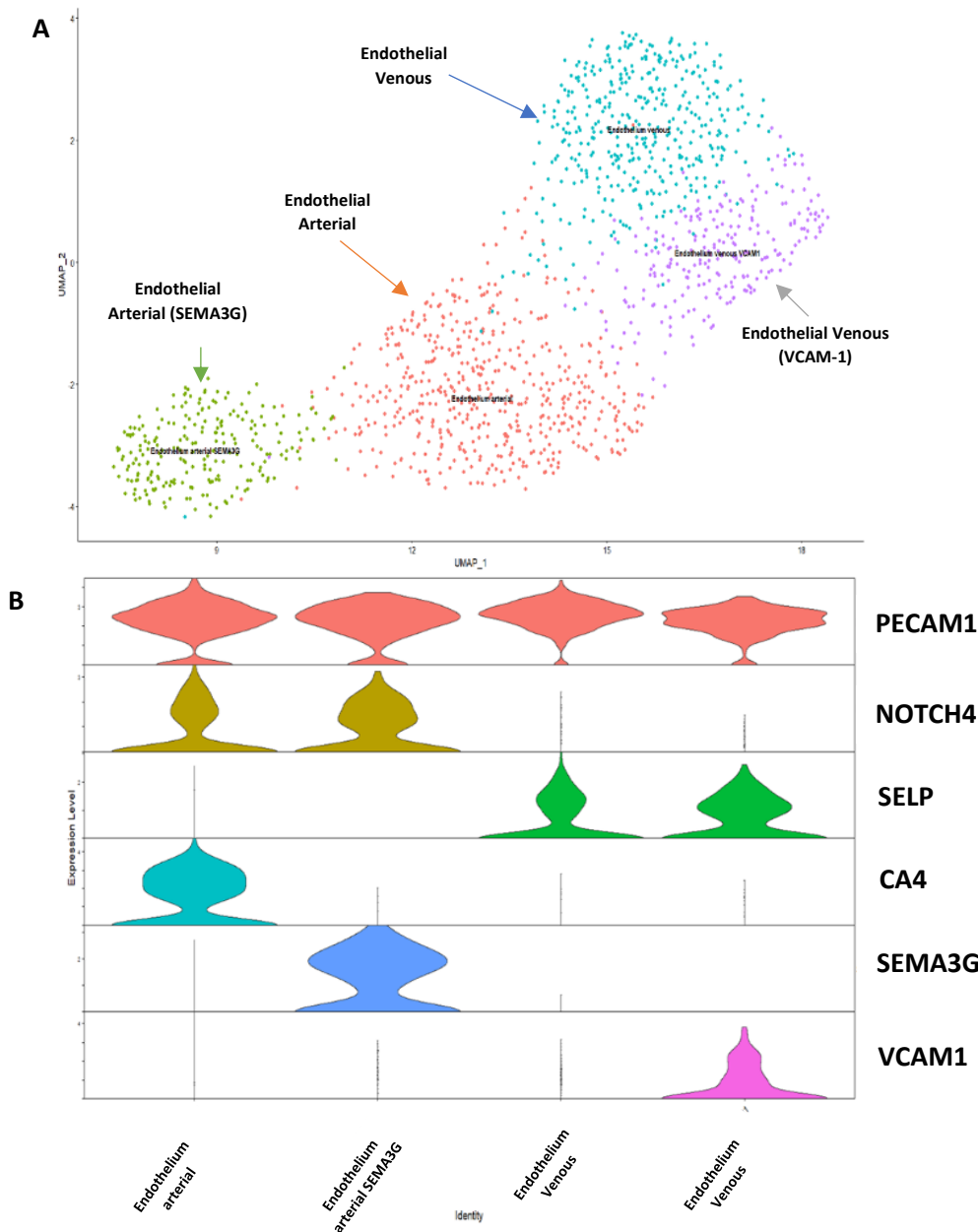


Figure 6: Exploration of the endothelium population in the human pSS scRNAseq dataset. **A:** UMAP of 4 endothelium cell subclusters within the human pSS SG. **B:** Stacked violin plot of unique identifying genes for each endothelium subcluster.

Figure 6: The endothelium populations in the human pSS SG

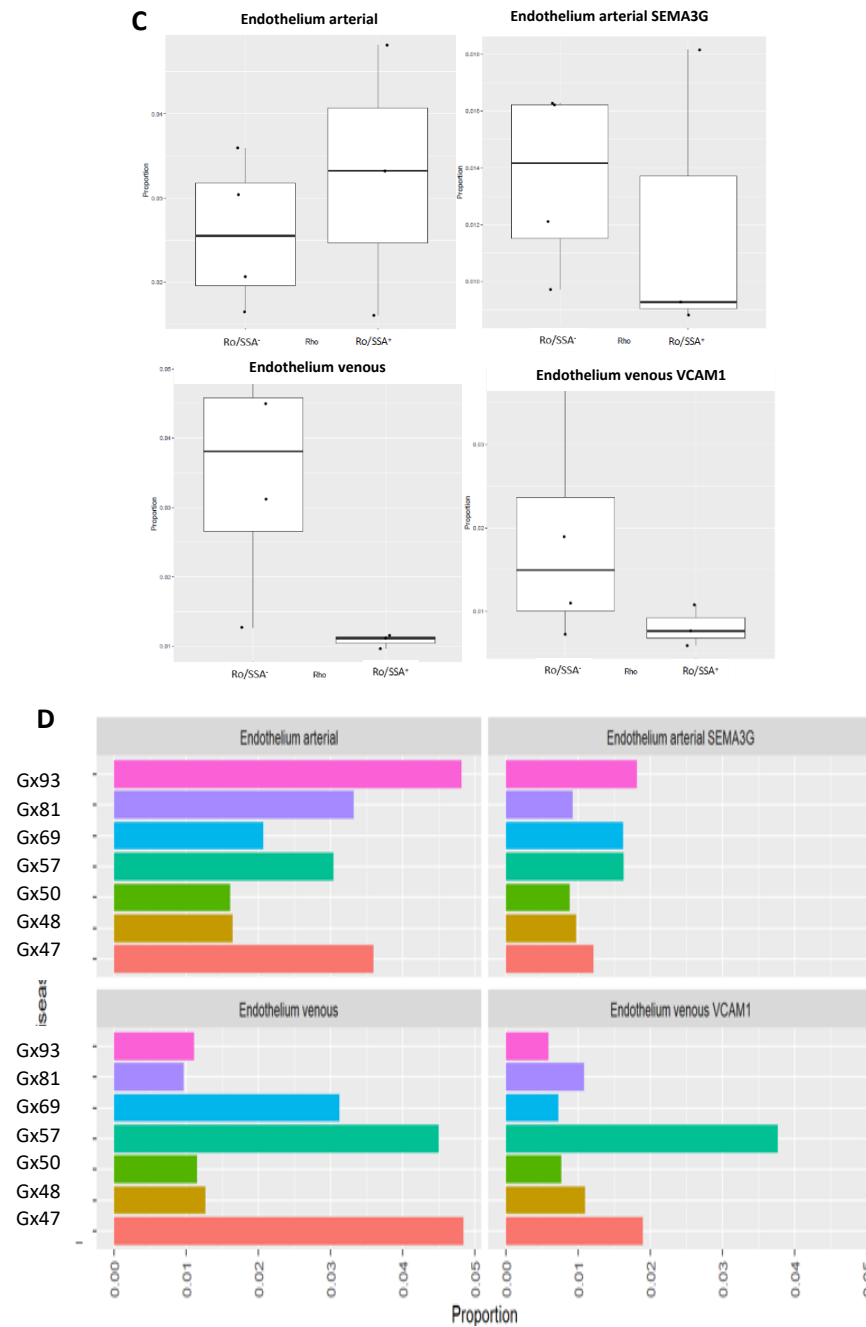


Figure 6: Exploration of the endothelium population in the human pSS scRNAseq dataset. C:

Proportion of selected endothelium subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57).

D: Proportion of each endothelium cell subcluster from specific human pSS patient sample.

3.7: The epithelium population

The epithelium population is categorised as either acinar or ductal and display the primary function of secreting saliva [218]. Many *in vitro* studies have characterised SG epithelial cells (SGECs) from pSS patients as pSS was originally coined as *autoimmune epithelitis* by *Moutsopoulos* who described the presence of epithelial cells with a dysregulated phenotype in close proximity to lymphocytic clusters [219]. In the context of human pSS ductal epithelium has been demonstrated to participate in the pathogenesis of pSS due to the recognition of elevated Ro52 expression within these cells which correlated with increased level of inflammation [220]. These epithelial cells also demonstrated an enhanced expression of MHC-II thus the ability to present antigens, the secretion of pro-inflammatory cytokines and autoantigens within exosomes which are thought to collectively exasperate the inflammatory response within the human pSS SG [221]. In addition, these cells have also been demonstrated to possess the necessary machinery to process PAMPs and DAMPs allowing them to participate in immunomodulatory functions [9, 53]. In our single cell dataset, we identified 7 clusters of epithelial cells within the pSS SG comprising of 6 epithelial populations and 1 myoepithelial cluster (**Fig 7A**). Within this population, 4 of the subclusters expressed Epcam whilst 3 did not. Of these three Epcam⁻ populations, 2 of the subclusters expressed the genes for keratins 7,8 and 14 (*KRT7*, *KRT8* and *KRT14*) and the last of the Epcam⁻ population did not express keratin associated genes (**Fig 7B**). In addition, unique genes which identified each subcluster of Epcam⁺ subclusters comprised of *Ankyrin repeat domain-containing protein 36C* (*ANKRD36C*), *Lysozyme C* (*LYZ*), *Mucin-5B* (*MUC5B*), *Mucin-16* (*MUC16*) whilst the Epcam⁻ populations had unique gene expression of *Keratin type I cytoskeletal 14* (*KRT14*), *Solute carrier family 13 member 2* (*SLC13A2*), and whilst the myoepithelium population

differentially expressed the Notch ligand *Dickkopf-related protein 3 (DKK3)* (**Fig 7B**). The proportions of these epithelium clusters were associated with the antibody status of the patients (**Fig 7C**). Interestingly, the proportion of the epithelial *KRT14* cells and epithelial *MUC16* cluster of cells was greater in Ro/SSA⁻ patients (**Fig 7C**). On the other hand, the proportion of the epithelial *LYZ^{high}* and epithelial *MUC5B* were found in greater abundance in the Ro/SSA⁺ patient subsets (**Fig 7C**). This differing in epithelial population proportions in the different Ro/SSA status patients may be indicative of the expansion, involvement, or loss of specific epithelial populations depending on the level of inflammation present within the pSS SG. Lastly, when we explored the presence of these epithelial populations across our patient samples, we observed that specific populations such as epithelium ANKRD36C, epithelium *KRT14* and epithelium *LYZN* were present in greater abundance than epithelium *MUC5B*, epithelium *MUC16*, epithelium *SLC13A2* and myoepithelium (**Fig 7D**). This maybe owed to the biological contributions and proportions of these populations in the human pSS SG.

Figure 7: The Epithelial population of the human pSS SG

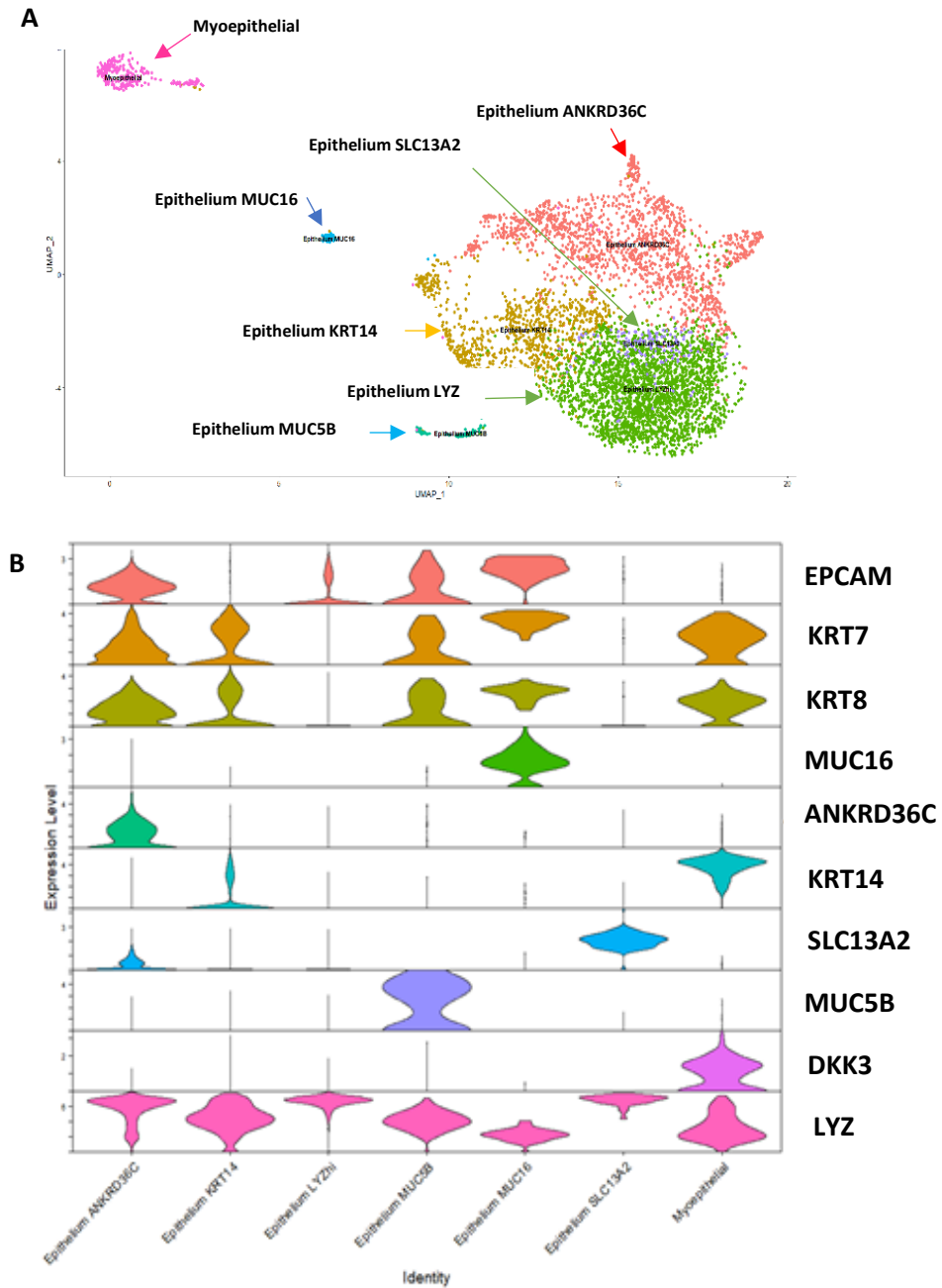


Figure 7: Exploration of the epithelium population in the human pSS scRNAseq dataset. **A:** UMAP of 6 epithelium cell subclusters and 1 myoepithelial cluster within the human pSS SG. **B:** Stacked violin plot of unique identifying genes for each epithelial subcluster. **C:** Proportion of selected epithelial subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57). **D:** Proportion of each epithelial cell subcluster from specific human pSS patient sample.

Figure 7: The Epithelial population of the human pSS SG

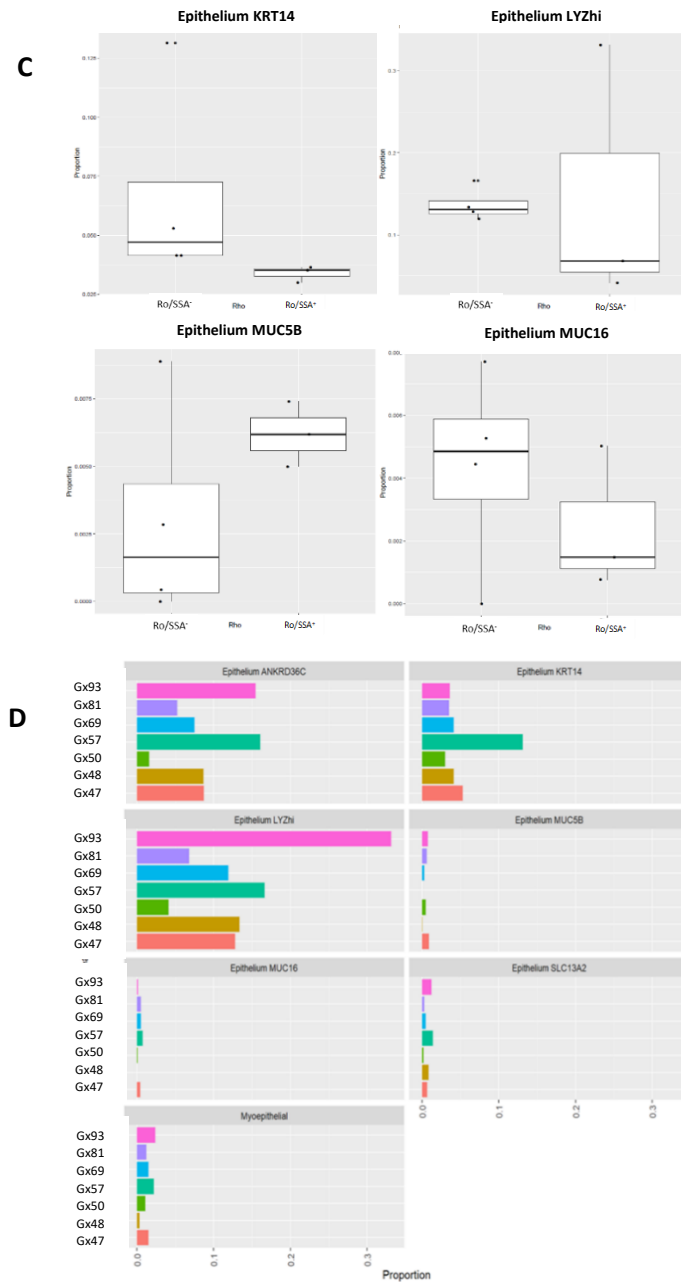


Figure 7: Exploration of the epithelium population in the human pSS scRNAseq dataset. C: Proportion of selected epithelial subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57). **D:** Proportion of each epithelial cell subcluster from specific human pSS patient sample.

3.8: The Glial cell population

Glial cells are a population of cells which function to maintain neurological homeostasis by providing mechanical support, protection, and nourishment to nerves [222]. Within this population of structural cells, Schwann cells and neuroepithelial cells have been identified [222]. The Schwann cell population serves a purpose to insulate the nerve axon as well as nourish neurones with the production of various growth factors whilst the neuroepithelial cell population constitutes the stem cell population of the nervous system [222].

In our dataset we identify a distinct population of neuronal associated cells (**Fig 8A**), which we labelled Schwann cells on the basis of the expression of *Neural Cell Adhesion Molecule 1 (NCAM1)* and *Myelin Protein Zero (MPZ)*. Neuronal epithelial cells were instead identified by the expression of *NCAM1*, *Periostin (POSTN)*, *PLXCIN1*, and *Myelin basic protein (MBP)* (**Fig 8B**). A third cluster was identified closely to the *NCAM1*⁺ *MPZ*⁺ cluster but did not express biologically unique genes thus this additional cluster was not confirmed as a true biological cluster (**Fig 8B**).

Figure 8: The Glial Cell population of the human pSS SG

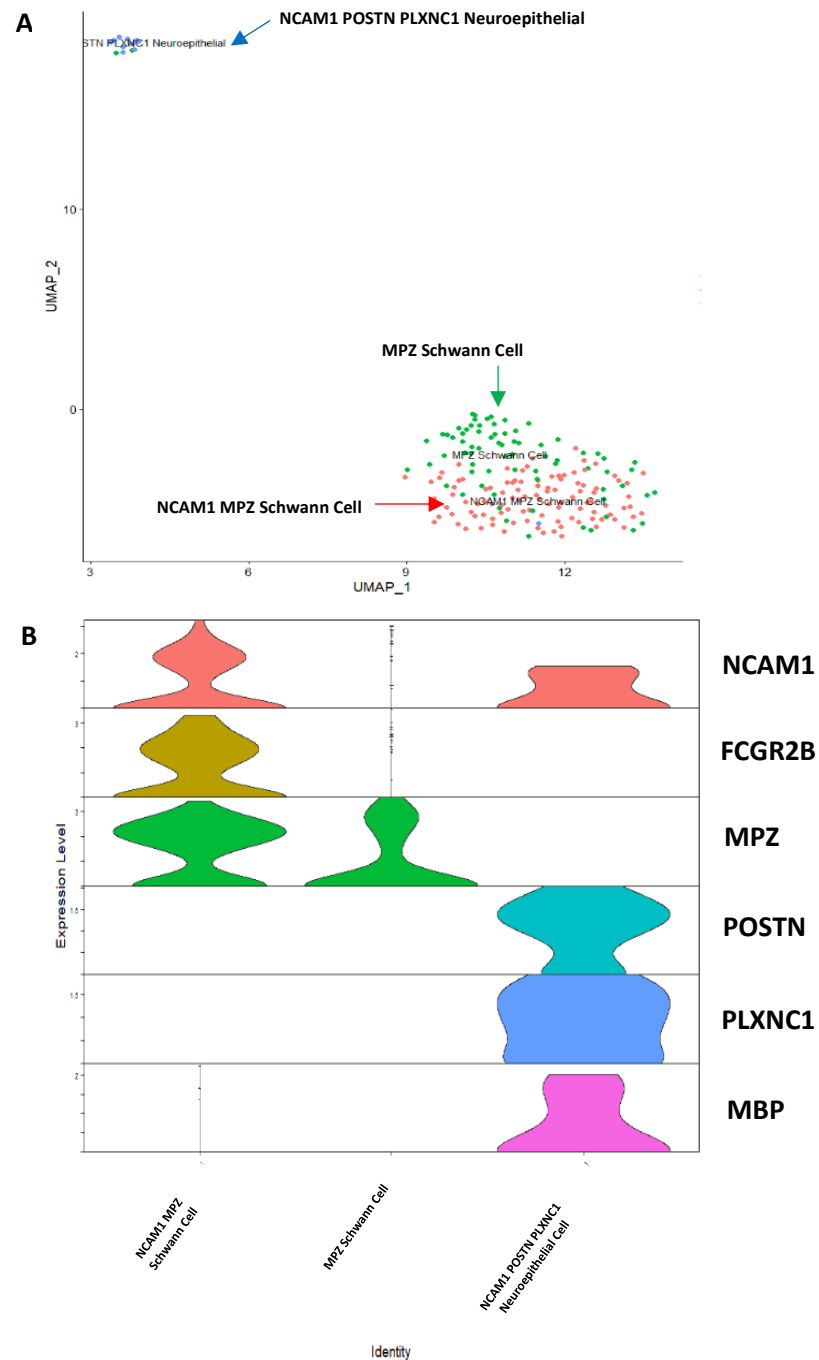


Figure 8: Exploration of the Glial cell populations in the human pSS scRNAseq dataset **A:** UMAP of 3 glial cell clusters. **B:** UMAP visualising the number of features and counts of mRNA to demonstrate the low capture of a subset of glial cells termed MPZ ambient. **C:** Stacked violin plot of unique identifying genes for each subcluster. **C:** UMAP demonstrating the number of UMIs and number of genes expressed within the 3 subclusters to demonstrate sample quality and identify areas of ambience.

3.9: T Cells of the human pSS SG

Contributions of the T cell population in the remodelling and maturation of fibroblasts at different anatomical sites have been explored by various groups [5, 6, 67, 105]. Broadly, exploration of the T cell and fibroblast interaction has demonstrated the ability of these lymphocytes to enhance and prolong the *immunofibroblast* phenotype [8, 172]. This was demonstrated in our work whereby cannulation of the adenovirus in the SG of Rag2^{-/-} mice demonstrated a reduced ability of the fibroblast population to express the lymphoid chemokines [8]. This suggested that whilst chemokine induction can occur in the absence of immune cells, the stabilisation of this lymphoid phenotype is dependent on immune cell interaction. In our dataset, we identified 8 clusters of T cells which comprised of 4 CD4⁺ and 3 CD8⁺ populations (**Fig 9A**). The classical CD4 populations consisted of CD4 Tregs which expressed *FOXP3*, CD4 T helper 1 (Th1) which expressed *TNF*, CD4 Tfh which expressed *CXCL13* and may also be the recently described Tph cells but this needs further exploration [223]. This selection of genes were utilised to assign the identity to the T cells as the classical T cell markers in their mRNA form were not detected in our dataset (**JT**). The absence of expression signature in our data maybe due to the depth of sequencing not being optimal or the sensitivity of the scRNAseq chemistry not being able to capture all mRNA molecules. Lastly, we also identified a population of CD4 T cell central memory (T_{Cem}) which expressed *SELL* and *CCR7* (**Fig 9A**). This data demonstrated that the human pSS SG contains a mixture of various CD4 T cell populations which have contributions to inflammation in the context of pSS.

However, in our dataset we did not find conclusive signatures of Th2 or Th17 cells. This may be due to presence of these cell populations in low abundance within the human pSS SG samples such that they were not captured in order to generate a detectable signature in our dataset.

Additionally, 2 of the 3 CD8 T cell subsets were uniquely identified by the expression of *IFN-γ* and *Granzyme A* (*GZMA*) whilst the third subset did not have expression of these particular genes but rather expressed *IL7R* and *lamin A* (*LMNA*) (**Fig 9B**). This suggests the presence of activated CD8 T cells within the human pSS SG which secrete the pro-inflammatory cytokines demonstrated to have a strong clinical association in pSS [21, 67].

Further exploration of the proportion of the T cell populations demonstrated differences in proportions according to each individual patient (**Fig 9C**). In this instance, patient Gx57 demonstrated an abundance of CD4 Th1 cells and another CD8 population uniquely identified with the expression signature of *IL7R GZMA* cells. Furthermore, a reduced proportion of CD4 Tfh and CD4 T_{Cem} cells suggests that this patient may have had a greater inflammatory signature with the presence of *TNF* and *IFN-γ* as compared to others. Patient Gx81 displayed an increased proportion of CD8 *IL7R GZMA* cells relative to the other T cell populations identified. This trend was also seen in patient Gx57. Contrastingly, Patient Gx69 displayed a greater proportion of CD8 T cell *GZMK* as compared to the other CD4 and CD8 subsets; while patient Gx48 demonstrated to have greater proportions of CD4 Tregs, CD4 Th1 cells, CD4 Tfh cells and CD4 T_{Cem} cells as compared to the CD8 subsets.

Intriguingly, upon analysis of the cell proportions in Ro/SSA⁺ and Ro/SSA⁻ patients, we could detect an increased proportion CD8 *IL7R GZMA* in the Ro⁺ patients whilst CD4 Tregs, CD4 Th1 cells, CD4 Tfh cells and CD4 T_{Cem} cells were present in greater proportions in the Ro/SSA⁻

patients (**Fig 9D**). Interestingly, it would have been expected that the described CD4 populations would be greater in the Ro/SSA⁺ patients which would naturally have greater inflammation than the Ro/SSA⁻ patients. However, this is not the case and could be owed to the low numbers examined within the Ro/SSA⁺ patients or these cells may not reside in the human pSS SG with the same abundance as those within the Ro/SSA⁻ patients.

The pSS SG microenvironment is a potent inflammatory milieu defined by a strong signature of the IFN related family of genes in addition to the TNF family of cytokines [54]. To explore the engagement of this pathway in our dataset, we interrogated our human scRNAseq dataset for the gene expression of the cognate receptor and ligands which mediate the engagement and activation of the TNF- α and IFN- γ signalling cascade.

We explored the expression patterns of TNF, TNFR1A (TNFR1), TNFR1B (TNFR2), IFNGR1, IFNGR2, IFNG, and the classical lymphoid chemokine CCL19 and its cognate receptor CCR7 in the scRNAseq dataset (**Fig 10**). We observed that *TNF* expression was confined to the CD8 and CD4 T cell populations with greatest expression in the CD4 Th1 subset followed by the CD8 *IL7R* *GZMA* subset. Interestingly, *TNFR2* expression was also confined to the T cell subsets, with greatest expression of this receptor being identified within the CD4 T regulatory population. Contrastingly, the mRNA expression of TNFR1 was identified within the CD34⁺ CD146⁻ stromal cell compartment, with greatest expression being demonstrated within the CCL19⁺ pericyte population. This may indicate that the T cells derived TNF could be communicating preferentially via TNFR1 on the stromal cell populations. However, as these observations are mRNA based, we would need to validate these expression patterns at the protein level to confirm the expression of TNFR1 on the specific populations of the CD34⁺ CD146⁻ fibroblasts

and further intervene with the TNF signalling cascade to observe for differences in cell phenotypes. Contrastingly, the mRNA expression of the IFNGR2 was exclusive to the CD34⁺ CD146⁻ fibroblasts as compared to the CD34⁻ CD146⁺ pericyte populations, with increased expression in the CCL19⁺ TNFSF13B⁺ *immunofibroblasts* relative to the other fibroblast populations in our dataset. The mRNA expression of IFNGR1 was also greater than IFNGR2 amongst the CD34⁺ CD146⁻ populations in addition to its detection within the CD34⁻ CD146⁺ pericytes. Moreover, whilst expression of IFNGR1 was homogenous across the CD34⁺ CD146⁻ populations it demonstrated greater expression in the FOSB CDKN1A fibroblast population compared to others. The mRNA expression of the ligand for these receptors, IFNG, was expressed by activated CD8 T cells with no gene signature detected in CD4 T cells.

Lastly, we confirmed the mRNA expression of the T cell lymphoid chemokine CCL19 in the CD34⁺ CD146⁻ *immunofibroblast* population and the CD34⁻ CD146⁺ CCL19⁺ pericyte population. The mRNA of the receptor for this chemokine CCR7 was expressed on the CD4 T cell subset with little expression within the CD8 T cell population. This dataset has demonstrated the gene expression patterns and source of the receptors and ligands within the specific populations we have characterised in our human pSS SG scRNAseq dataset. The observation of these specific receptors emphasises the possibility of therapeutically modulating this interaction such that fibroblast activation is prevented in the context of inflammation or promoted in the context of cancer associated TLS .

Figure 9: T cell populations in the human pSS SG

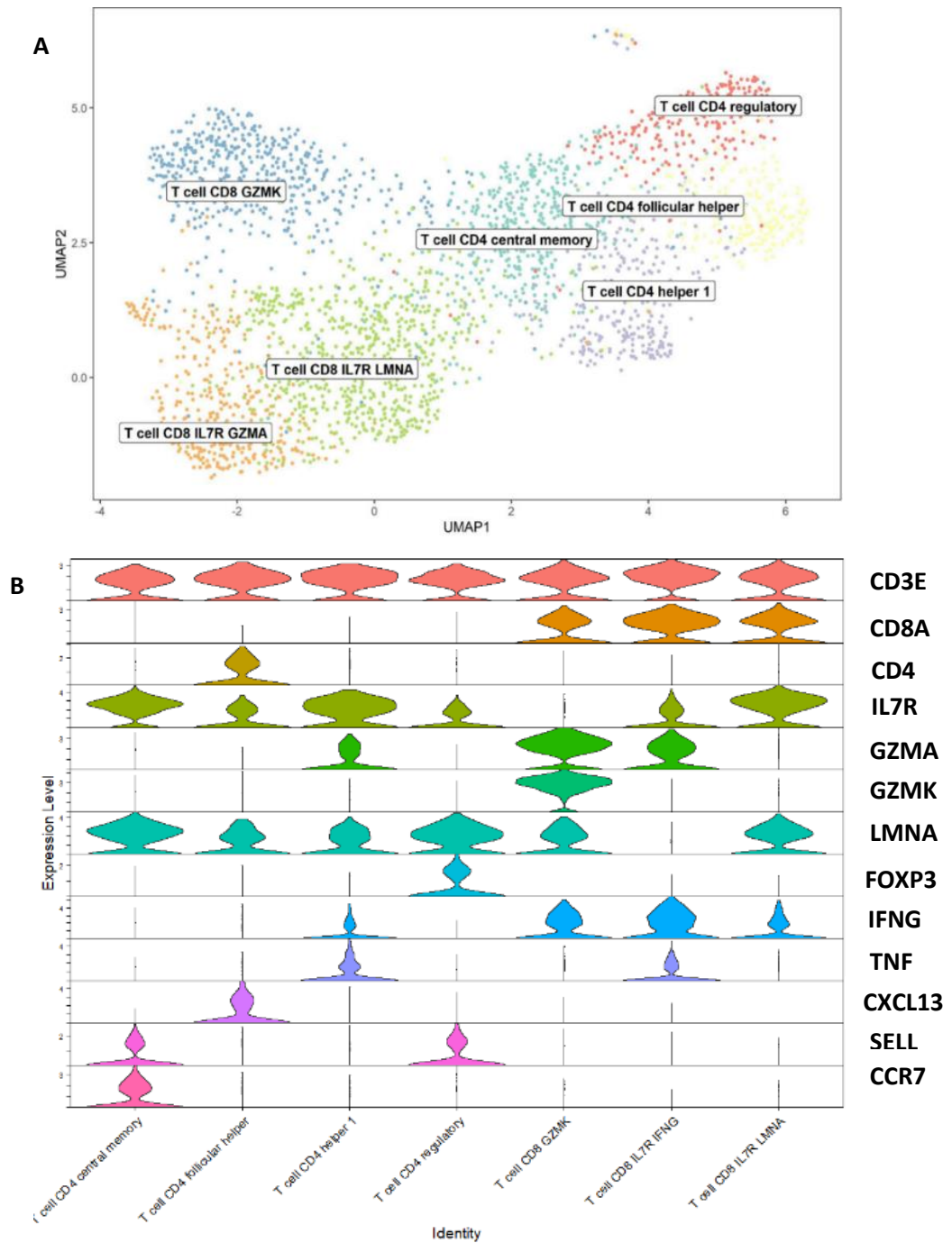


Figure 9: Exploration of the T cell population in the human pSS scRNAseq dataset **A:** UMAP of T cell subclusters within the human pSS SG presents 4 CD4 T cell and 3 CD8 T cell subclusters. **B:** Stacked violin plot of unique identifying genes for each subcluster.

Figure 9: T cell populations in the human pSS SG

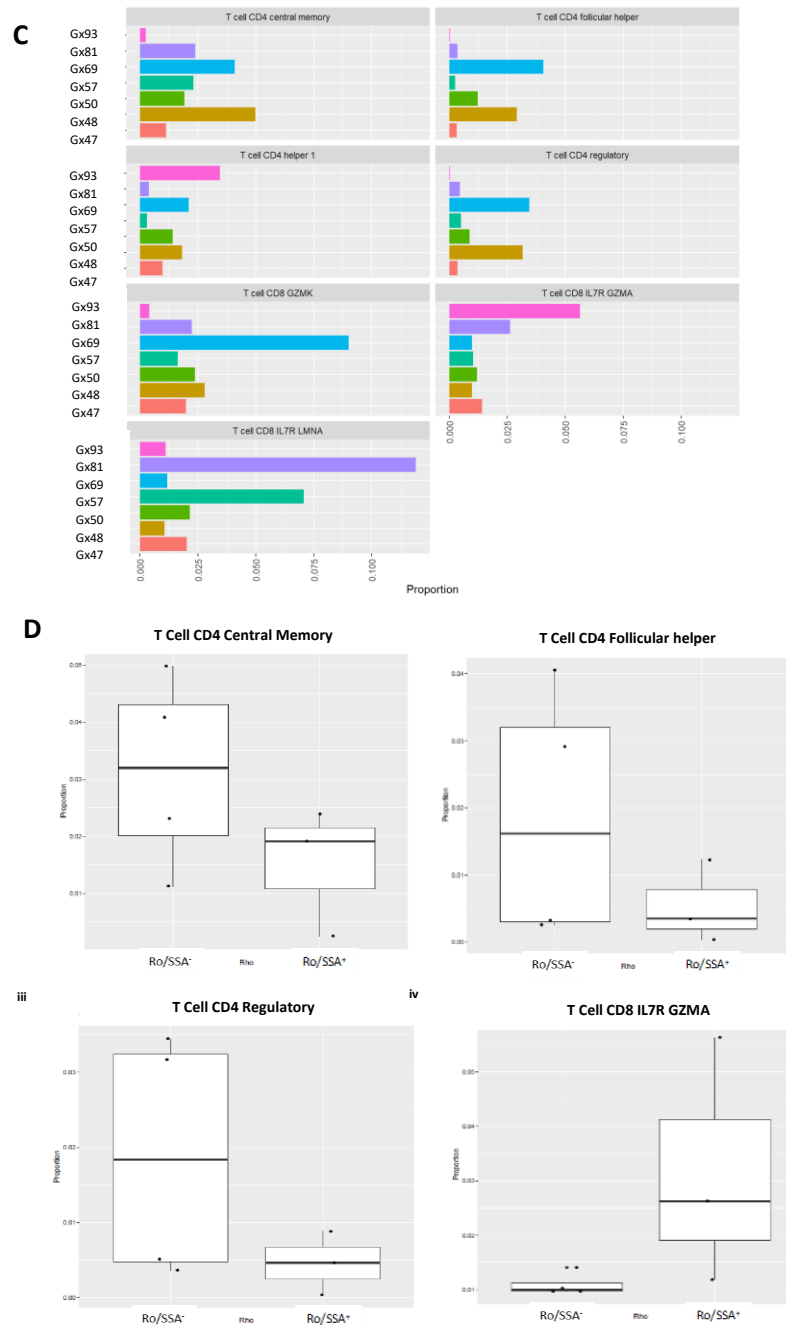


Figure 9: Exploration of the T cell population in the human pSS scRNAseq dataset C: Proportion of each T cell subcluster from human pSS patient sample. D: Proportion of selected T cell subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis Ro/SSA⁺ or Ro/SSA⁻ (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57).

Figure 10: Gene expression of receptor and ligand interactions between T cells and fibroblasts from human pSS scRNAseq dataset

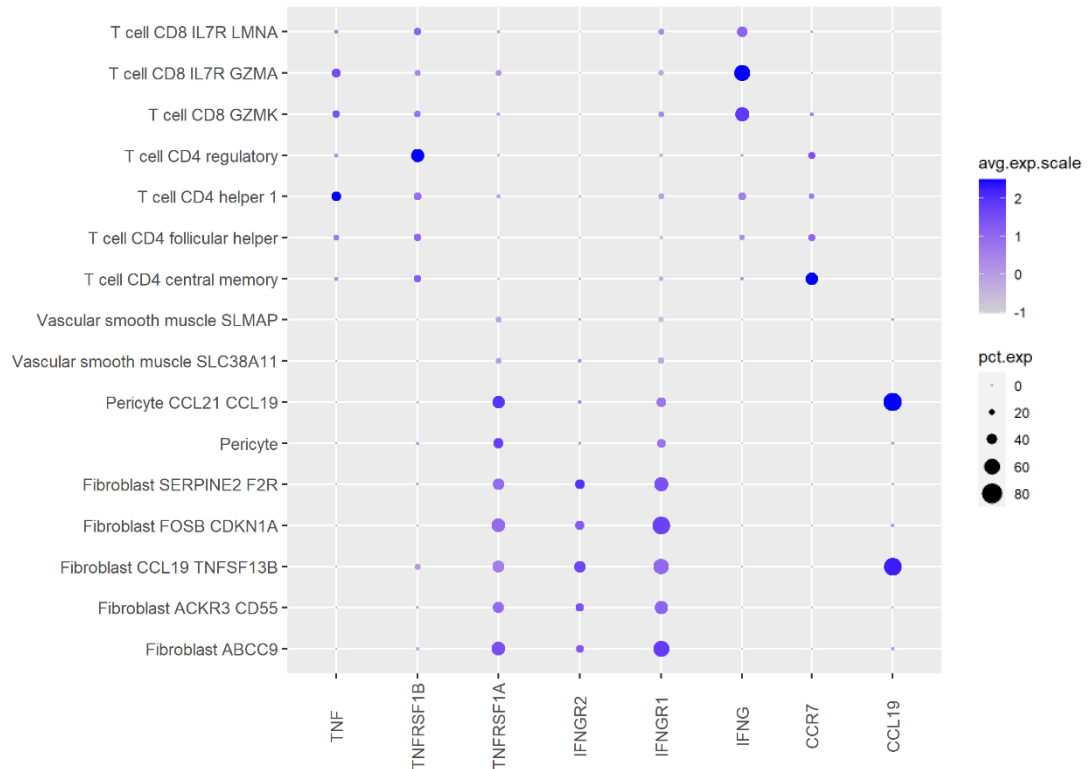


Figure 10: Dot plot of the receptor and ligand expression of the TNF, IFN and lymphoid chemokines family from the human pSS SG scRNAseq dataset. E) The interacting populations were focused on CD8, CD4 T cells and the CD34⁺ CD146⁻ and CD34⁻ CD146⁺ fibroblasts, mural and vascular cells. Larger dots demonstrate large cell proportions whilst darker dots demonstrate enriched expression within the cell groups.

3.10: The B cell population of the human pSS

The presence of B cells within the human pSS SG has been used as a biomarker to define mature TLS and are associated with histological and systemic disease progression [19, 33, 40, 48, 224]. The B cell survival factor BAFF is enriched within the SG of pSS and promotes the accumulation of these lymphocytes [40]. This ectopic accumulation elevates the risk of the lymphoma development within pSS patients as these B cells undergo aberrant SMH and VDJ recombination resulting in the generation of oncogenic B cell clones [50]. In our human pSS dataset, we have identified 3 different B cell subpopulations (**Fig 11A**), determined by the expression of *CD20* (*MS4A1*), *CD27* and the surface immunoglobulin *IgD* (**Fig 11B**). The first cluster of B cells had an expression signature of *IgD*⁺ *CD27*⁻ which resonates as naïve antigen inexperienced cells which also expressed unique genes such as *IL4R* and *FCGR2* (*CD23*), a low affinity receptor for IgE and CD21 [225] (**Fig 11B**). The second cluster of B cells in our dataset were *IgD*⁻ *CD27*⁺ and uniquely expressed the genes for *CCR7*⁺ suggesting this is an activated B cell population [226]. This cluster uniquely expressed the *B cell lymphoma 2-related protein A1* (*BCL2A1*), a pro-survival gene regulated downstream of NfκB activity [227]. The identification of the high expression of this anti-apoptotic gene within this B cell population may confer this population with a form of resistance observed in the context of the use of anti-CD20 depletion therapies in pSS [42,

228]. Lastly, a third cluster of cells were identified as *IgD⁻ CD27⁺ CCR7⁻* a switched memory B cell population, which uniquely expressed the gene *B cell Shisa Family Member 8 (SHISA8)*, previously associated with the regulation of glutamate receptor expression and resistance to apoptosis in the TME [229] **(Fig 11B)**. Interestingly, when we observed the proportion of these three B cell populations according to the Ro/SSA status we observed that the B cell IgD and B cell *SHISA8* populations were present in larger proportions in the Ro/SSA⁻ patients as compared to the Ro/SSA⁺ patients **(Fig 11C)**. Moreover, we observed that the proportion of B cells amongst the individual patient samples was variable **(Fig 11D)**. More specifically, patient Gx48 demonstrated the greatest proportion of B cells of all three B cell populations in contrast to patient Gx93 which had no B cell signature according to our data. This demonstrates that there is variability in the infiltration of B cells within the human pSS SG.

Figure 11: The B cell populations in the human pSS SG

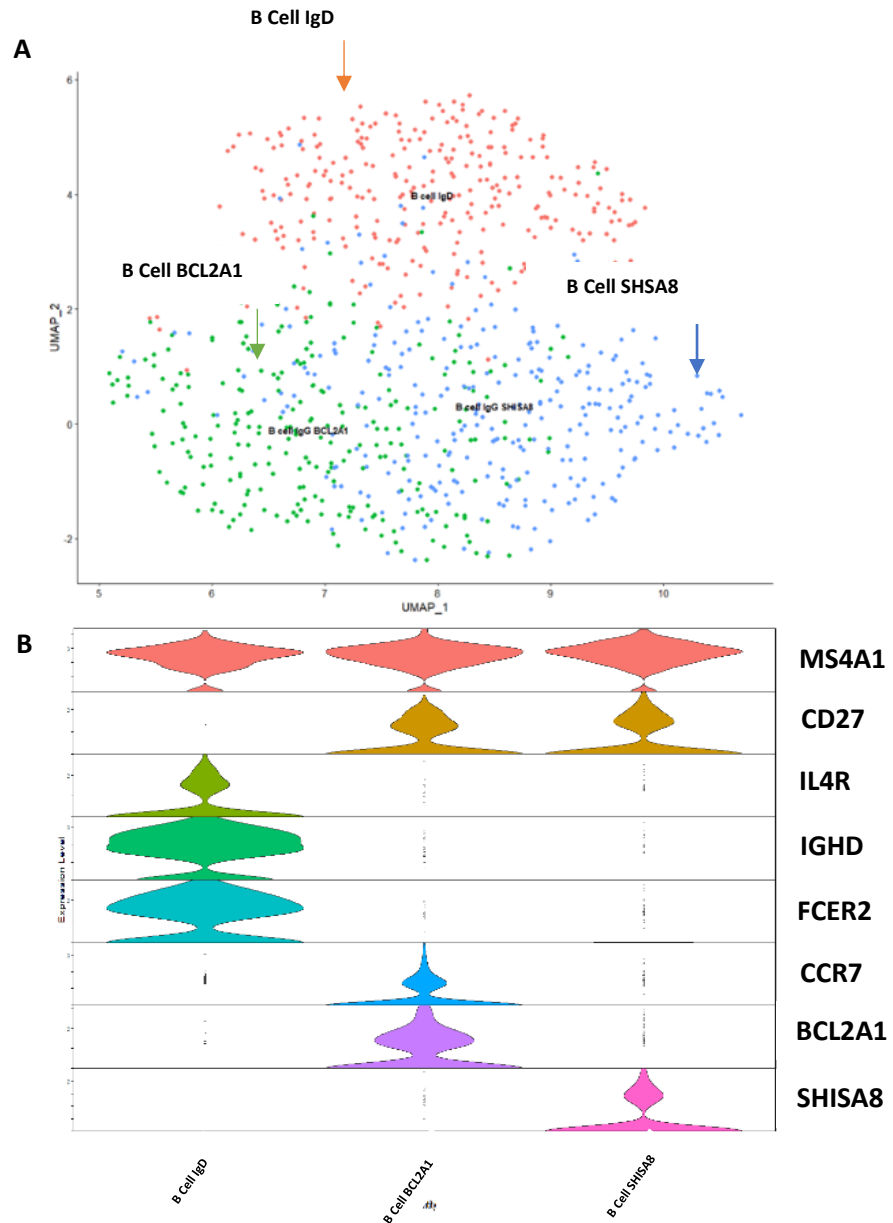


Figure 11: Exploration of the B cell population in the human pSS scRNAseq dataset **A:** UMAP of B cell subclusters within the human pSS SG presents 3 B cell subclusters. **B:** Stacked violin plot of uniquely identifying genes for each subcluster.

Figure 11: The B cell populations in the human pSS SG

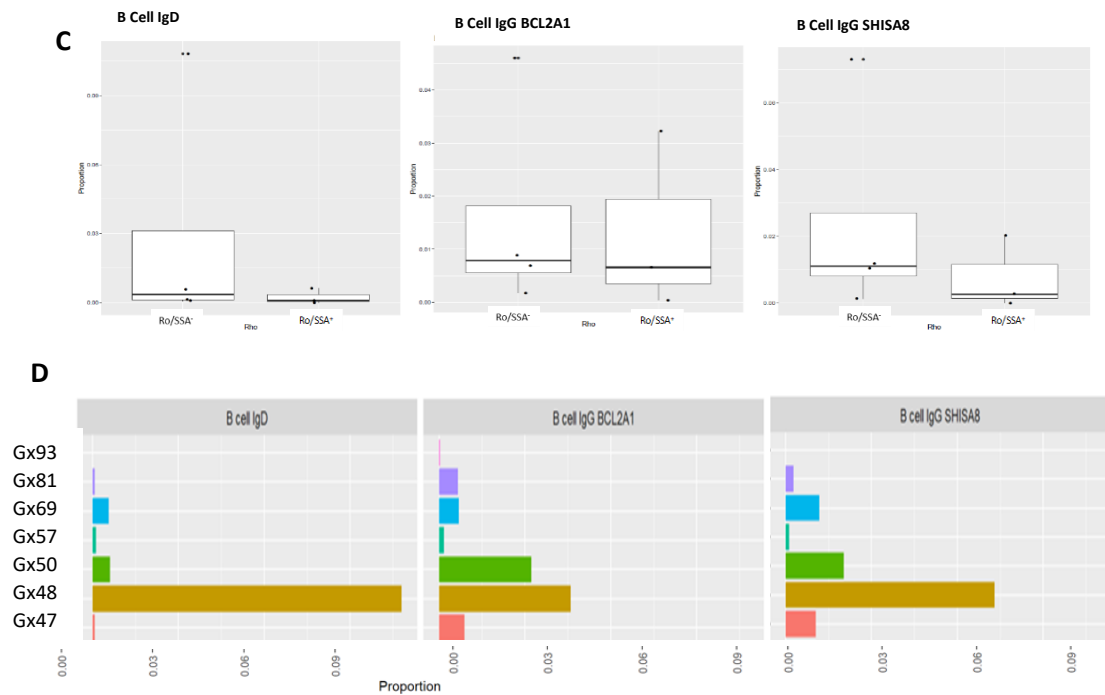


Figure 11: Exploration of the B cell population in the human pSS scRNAseq dataset **C:** Proportion of selected B cell subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57). **D:** Proportion of each B cell subcluster from specific human pSS patient sample.

3.11: Plasma Cells of the pSS SG: the autoantibody factories

It has been demonstrated that antibody producing PC are present within the SG and are responsible for the production of Ro and La antibodies in pSS [39, 230, 231]. Interestingly, the biological heterogeneity of this population has not been deconvoluted to date in the context of the human pSS SG. In our analysis we identified 17 clusters of PC (**Fig 12A**). These clusters were identified using the isotypes discussed below. We also explored if the isotypes were the true drivers of the heterogeneity seen within this cell population. In this manner, we reasoned to remove these genes and observed whether the original clustering pattern persisted to which end we did not maintain the resolution of heterogeneity as described below. This confirmed that the different PC sub-populations were uniquely identified by the expression of the isotype genes.

We first sought to confirm the validity of these cell clusters by visualising the overall gene expression and gene count within this subset (**Fig 12B**). Of these PC clusters, 4 subclusters were uniquely identified due to differential expression of the *immunoglobulin lambda light chain variable region (IGLV)*, 6 subclusters had distinct *immunoglobulin kappa light chain variable regions (IGKV)*, 3 subclusters had *immunoglobulin heavy chain variable regions (IGHV)* and 1 subcluster was identified with the unique expression of the *immunoglobulin heavy chain gamma (IGHG)* (**Fig 12C**). Within this PC population, a CD138^{low} CD38^{low} subcluster was also identified suggesting this cell subcluster to be a possible progenitor PC population. In addition, a single PC subcluster was identified with the *NFKB1A* being the unique identifying gene.

Aside from the differing clusters of PC present within the human pSS dataset, we were also able to observe the different proportions of each subcluster within the patients (**Fig 12D**) to

which end patient Gx50 demonstrated increased PC proportions compared to the other patients. Furthermore, we also explored the proportion of PC subclusters depending on the Ro/SSA status of the patients within our dataset and observed the presence of PC subclusters in greater proportions within the Ro/SSA⁺ subtype than the Ro/SSA⁻ (**Fig 12E**).

Figure 12: The Plasma Cell populations in the human pSS SG

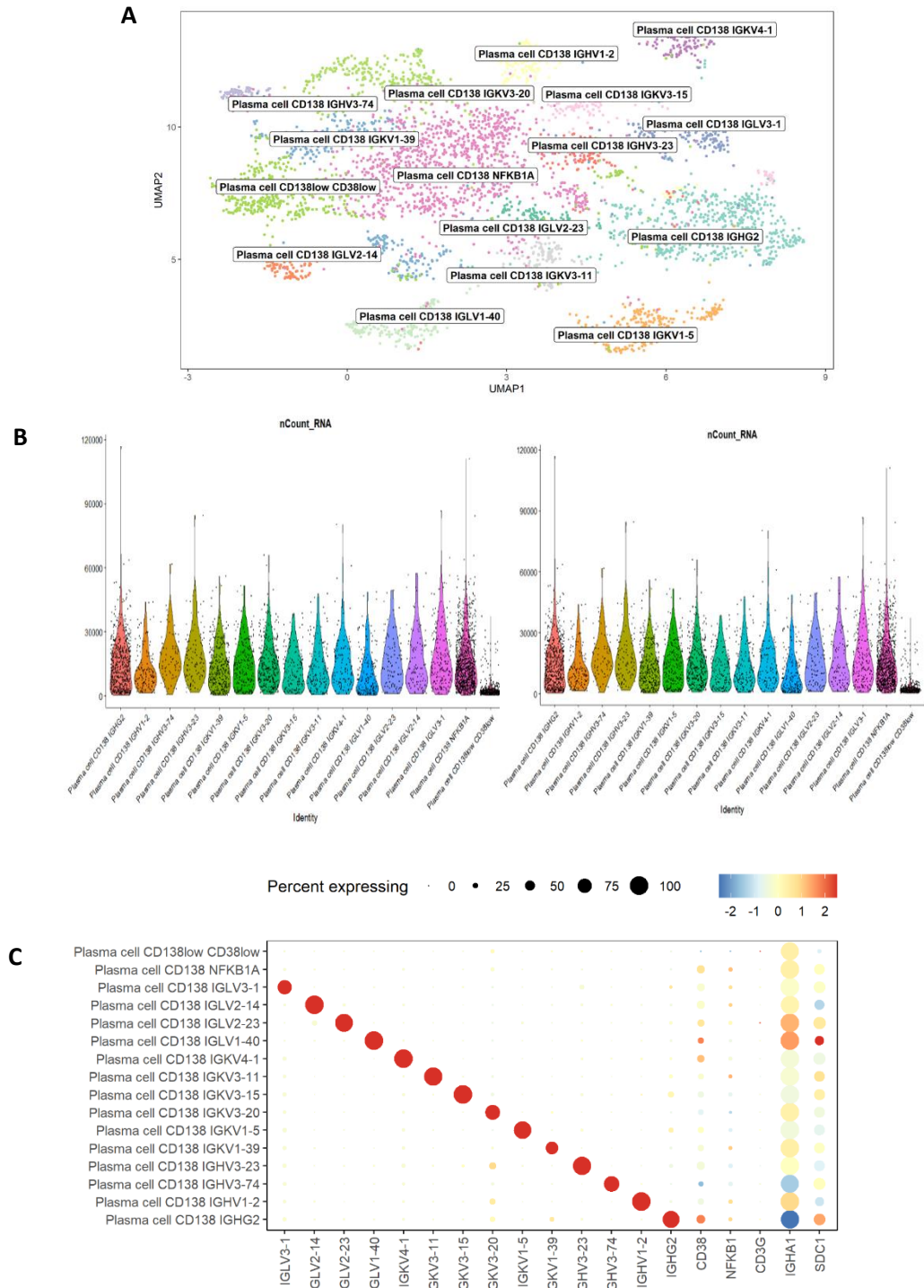


Figure 12: Exploration of the Plasma cell signature in the human pSS scRNAseq dataset. **A:** UMAP of 17 plasma cell subclusters within the human pSS SG. **B:** UMAP of the plasma cell population displaying the expression of features (genes) and counts of mRNA within this cell population. **C:** Dot plot of unique identifying genes for each subcluster

Figure 12: Plasma Cell proportions in the human pSS SG

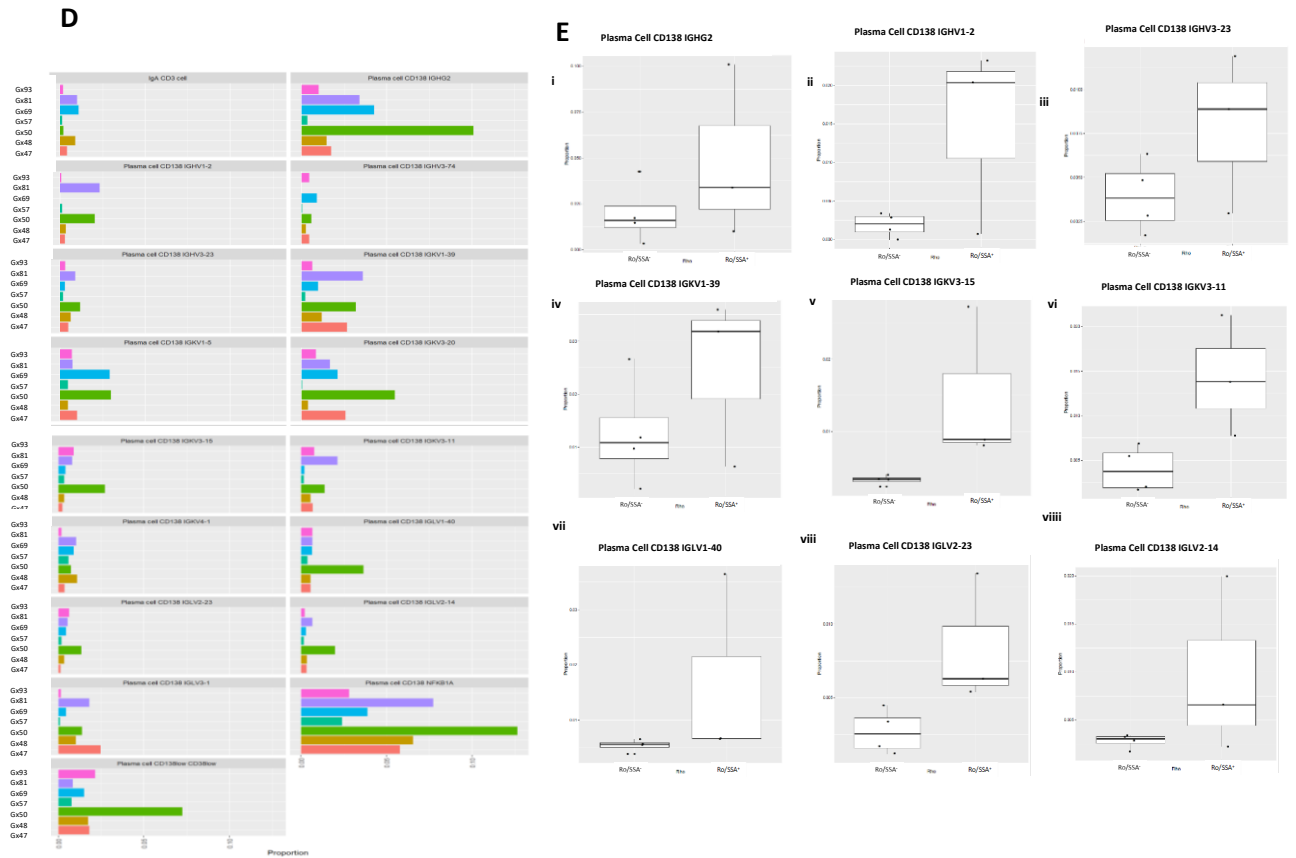


Figure 12: Exploration of the Plasma cell signature in the human pSS scRNAseq dataset. D: Proportion of each plasma cell subcluster from specific human pSS patient sample. **E:** Proportion of selected plasma cell subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57).

3.12: The Myeloid population

The myeloid population of cells comprises of DCs and macrophages, which are the classical APCs of the immune system [232]. Classically, these cells have been studied and demonstrated to initiate a plethora of immunomodulatory functions which comprise of the activation of the secretion of inflammatory cytokines and chemokines which results in the activation of resident stromal cells and infiltration of lymphocytes [74]. However, the contribution of the myeloid does not accord them with a major role in the pathogenesis of pSS and are a cell population that warrant further research in this context.

Within the human scRNAseq dataset we identified 3 myeloid populations (**Fig 13A**) determined by the expression of *CD11C (ITGAX)* (**Fig 13B**). Within these 3 populations, 2 subsets expressed *CX3CR1* (**Fig 13B**). Individually, the first myeloid cell cluster was defined by the unique expression of the antigen presenting protein *CD1C* and the receptor for IL-1, *IL1R2* suggesting this population of cells to participate in the activation of neighbouring cells through antigen presentation as well as the canonical production of IL-1. The second cluster of myeloid cells were characterised by the expression of *TGFB1*, the pro-inflammatory cytokine *TNF* and complement component *C3* (**Fig 13B**). This cassette of genes suggests a more “pro-inflammatory” phenotype of this population which may participate in the complement pathway with the detected expression of *C3*. The final population of myeloid cells was defined by the expression of the scavenger receptor *CD163*, the enzyme *Retinal dehydrogenase 1 (ALDH1A1)* and the chemokine *CXCL12* (**Fig 13B**).

The gene signature for each of the myeloid cell clusters was present within each patient sample within our dataset suggesting this myeloid cell population was common amongst all

patients. Furthermore, exploration of these populations within the Ro/SSA⁺ and Ro/SSA⁻ subtypes demonstrated that the myeloid *CD1C* and *C3* populations were present in increased proportions within the Ro/SSA⁻ patient group whilst we observed an inverse trend in the myeloid *CD163* population (**Fig 13C**).

Figure 13: The myeloid populations in the human pSS SG

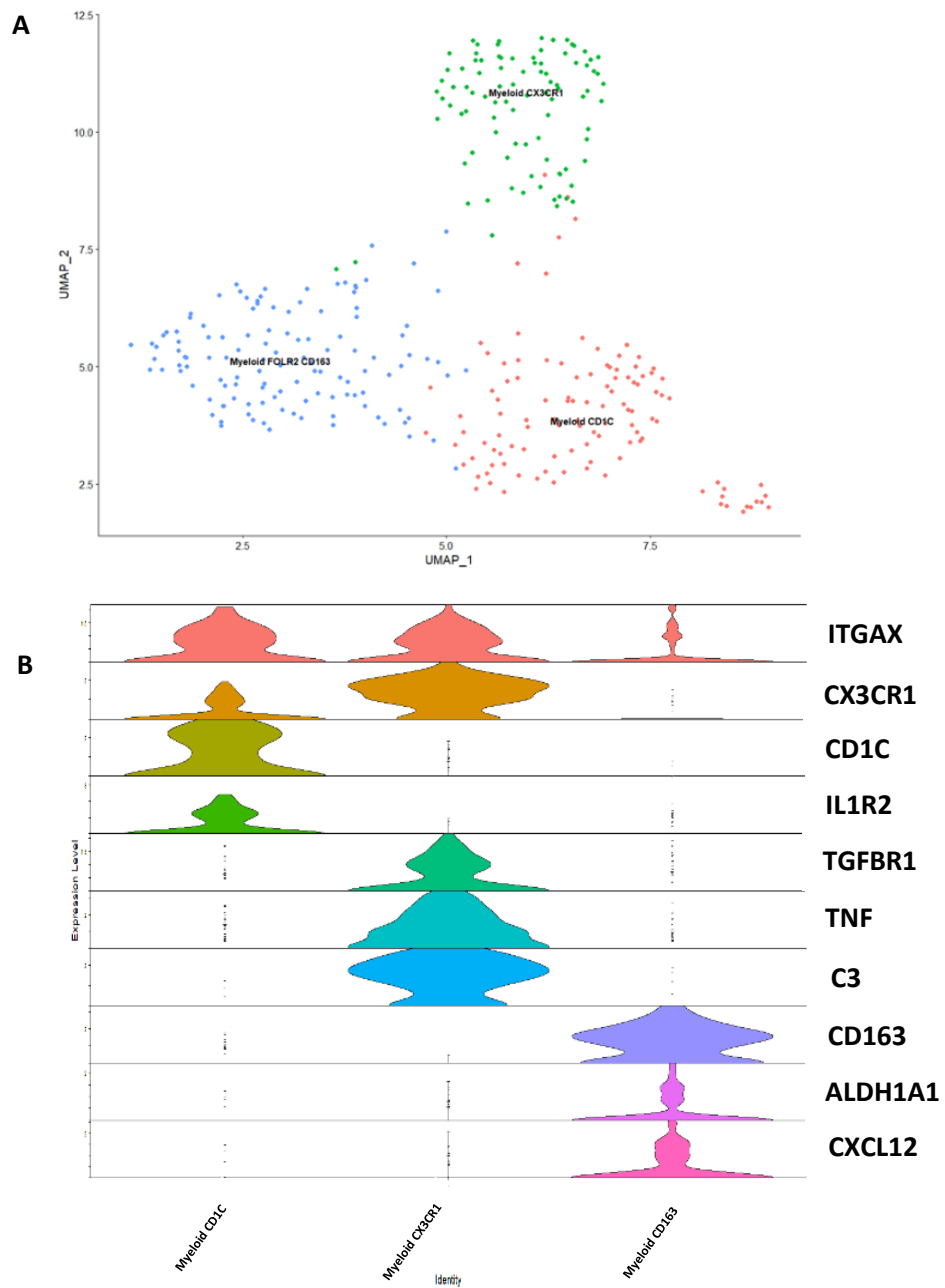


Figure 13: Exploration of the myeloid cell signature in the human pSS scRNAseq dataset. **A:** UMAP of 3 myeloid cell subclusters within the human pSS SG. **B:** Stacked violin plot of unique identifying genes for each myeloid subcluster.

Figure 13: The myeloid populations in the human pSS SG

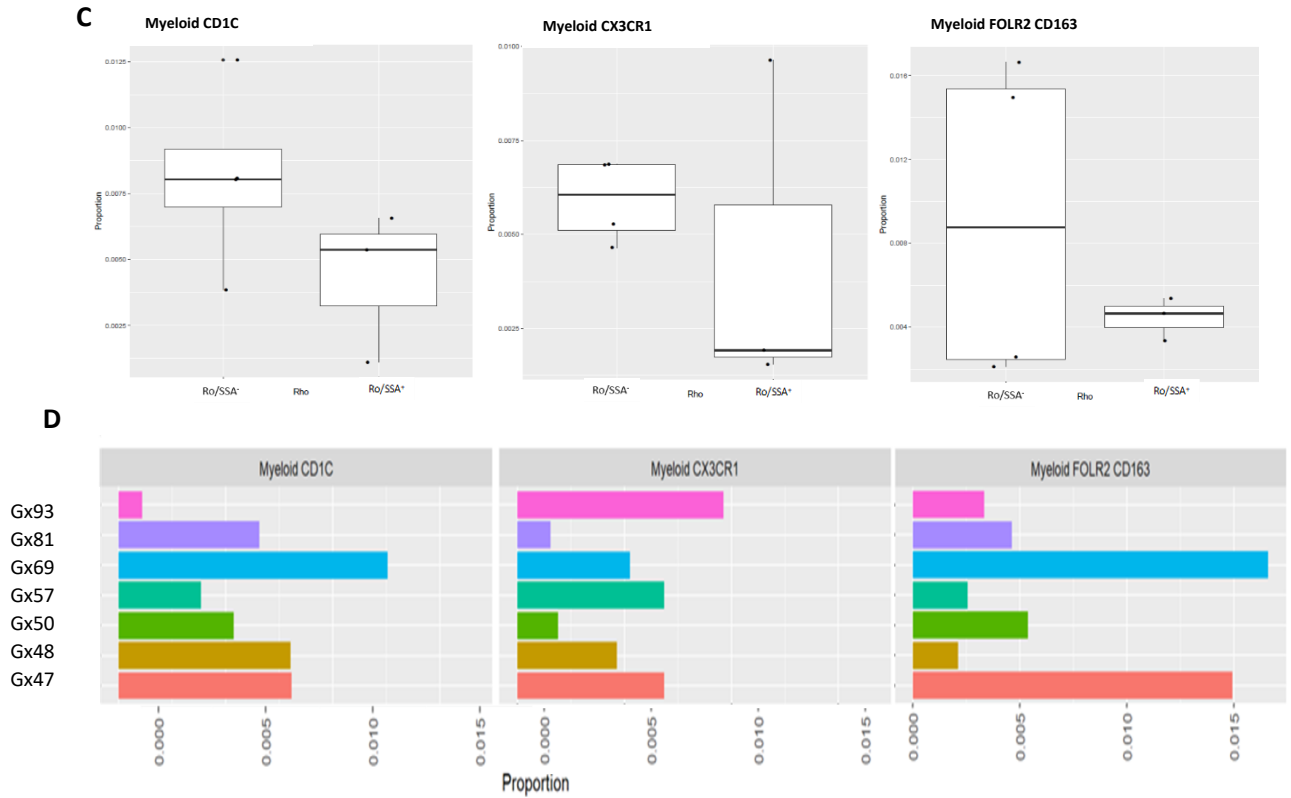


Figure 13: Exploration of the myeloid cell signature in the human pSS scRNAseq dataset. C:

Proportion of selected myeloid subclusters differing based on the Ro⁺ or Ro⁻ subtype of the patients used in the analysis (Ro⁺ patients Gx50, Gx81, Gx93. Ro⁻ patients Gx47, Gx48, Gx69, Gx57).. **D:** Proportion of each myeloid cell subcluster from specific human pSS patient sample.

Chapter 3: Discussion

Fibroblasts in human salivary gland TLS

To date, the human pSS SG microenvironment has not yet been characterised at the single cell level. Here, for the first time, we used single cell RNA sequencing to resolve the different cellular populations residing within human pSS SG. Previous studies have included *in vitro* and *in vivo* approaches to characterise the contributions of the various cell populations which may reside within the pSS SG and modulate the permissive tissue to accommodate TLS initiation and development. The heterogeneity and the complexity of the fibroblast populations inhabiting the TLS supportive SG has not been resolved to our knowledge. Our dataset generated from 7 human pSS SG samples identified 6 immune cell and 5 stromal cell populations. The characterisation of the phenotype of each of these cell types have allowed us to gain an overview of the cellular participants possibly driving TLS formation within the pSS SG.

In the past, the fibroblast compartment was assumed to exert passive functions of structural support to organ parenchyma, mainly through the production of the extra cellular matrix (ECM) [10]. More recently this limited view has expanded, and it is now more widely acknowledged that fibroblasts play a more complex and critical role both in homeostasis and in disease. In cancer, CAFs have garnered attention as critical curators of the TME with their secretion of immunosuppressive and pro-angiogenetic factors [60, 61, 63, 64, 152, 154]. Within the inflammatory microenvironment fibroblasts have been demonstrated to possess immunomodulatory functions and the capability to acquire lymphoid like features [8, 10, 43, 131, 156, 180, 233-236]. Prior to the use of scRNA sequencing approaches *Nayar et al* have

demonstrated that fibroblasts respond to various pro-inflammatory cues, producing lymphoid chemokines and lymphocyte survival factors [8]. We expanded and completed these data, identifying a division of functions within the human pSS SG PDPN⁺ fibroblasts which were driven by observing the differential expression of CD34 within the PDPN⁺ fibroblast populations [8]. Using this data as a guide, we explored the CD34⁺ and CD34⁻ stromal cell populations and identified 5 subsets of CD34⁺ fibroblasts and 2 stromal cell populations in the CD34⁻ cluster which comprised a CD34⁻ CD146⁺ population of pericytes and a population of vascular smooth muscle cells.

The presence of *immunofibroblasts*, previously characterised using *in vivo* and *in vitro* techniques was confirmed at single cell level in our dataset [7, 8]. This pathogenic *immunofibroblast* population in our scRNA data set was identified by the unique mRNA expression of the classical lymphoid genes CCL19, the B cell survival factor BAFF, and the co-stimulatory molecule CD82. This was coupled with the mRNA expression of the lymphocyte adhesion markers ICAM-1 and VCAM-1 relative to the other PDPN⁺ CD34⁺ fibroblasts and CD34⁻ CD146⁺ cells. *Immunofibroblasts* were histologically identified using CD82 and sequential CD90 expression as a large cluster of cells within human pSS SG, confirming their presence at the protein level within the tissue. This population strongly resembles the resident activated mesenchyme of the SLO such as LN, suggesting the participation of this cell type in the initiation and development of TLS. We further sought to identify the possible signalling cascades which could drive the formation of this cell population. Observations of the pathway enrichment analysis from our dataset, suggested that there was clear involvement of NFkB, TNF and IFN signalling according to the expression of mRNA relative to the other CD34⁺ fibroblast populations. This being in accordance with the engagement of the classical LTβR

pathway and NFkB signalling within the *immunofibroblast* [1, 109, 114]. The strong TNF and IFN signalling pathway maybe reflective of the strong inflammatory TNF and IFN signature within the pSS SG microenvironment which has been explored in the clinical context too [49, 176, 180].

The second population of CD34⁺ fibroblasts were defined by the unique mRNA expression of the ACKR3 receptor molecule, which belongs to the family of the atypical chemokine receptors expressed on stromal cells functioning to modulate chemokine gradients by exercising their scavenger function [237, 238]. The ACKR3 molecule scavenges for the chemokines CXCL11 and CXCL12 which bind to their cognate receptors CXCR4, expressed on B cells and T cells and CXCR3 which is expressed on activated T cells and NK cells [238]. This suggests that this fibroblast population may participate in the modulation of the chemokine gradient and lymphocyte infiltration within the human pSS SG. However, we would need to couple these observations at the mRNA level with further proteomic work to confirm if the increased mRNA expression of ACKR3 correlates with high protein abundance. Further validation would include the possible isolation of this fibroblast subset in order to then explore whether the possible function which we are currently inferring for this specific fibroblast population can be validated at the protein level.

In conjunction to ACKR3, this population of fibroblasts also uniquely expresses the mRNA for CD55 – a cell surface complement regulator which prevents the accumulation of the complement component C3b on the surface of self-cells and binds to its receptor CD97, an Adhesion G protein-coupled receptor E5 expressed on the surface of lymphocytes [239, 240]. This suggests that there may be involvement of the complement pathway within the pSS SG

and the engagement of ACKR3 coupled with CD55 might play a role modulating the pattern of lymphocytic infiltration during TLS formation due to the modulation of the chemokine gradient within the human pSS SG.

A third cluster of CD34⁺ fibroblasts identified by the unique expression of ABCC9 may present itself as a possible progenitor population which is resident within the SG and may give rise to the other fibroblast populations, we have identified in our scRNAseq dataset. This possibility was explored in our trajectory analysis we conducted on the CD34⁺ fibroblast populations. This data demonstrated a developmental trajectory in which the ABCC9 fibroblast gave rise to the other fibroblast clusters we had identified. The signalling cascades activated in this population involved genes for prostaglandin signalling, TNF and hints of involvement of the Wnt pathway. Furthermore, this cell cluster had greater expression of autotaxin (ATX). The enzyme autotaxin functions to synthesise the bioactive phospholipid lysophospholipid acid (LPA) which has been demonstrated in maintaining chronic inflammation in autoimmune diseases [241, 242]. Furthermore, the secretion of LPA from cells such as fibroblasts has been demonstrated to protect fibroblasts from undergoing apoptosis and participate in facilitating T cell entry into LNs [174, 243]. This fibroblast subset expressed the autotaxin gene greater than its counterparts, which express the LPA receptor 1 (LPAR1). Intriguingly, the immunofibroblast population had greater expression of LPAR1 suggesting that this signalling cascade maybe engaged greater in this pathological fibroblast subset. Overall, this third fibroblast subcluster may also function to promote survival and persistence of the fibroblasts in a hypoxic pro-inflammatory environment.

The fourth fibroblast cluster assigned with the mRNA expression of FOSB and CDKN1A did not display any unique differential gene expression compared to the other fibroblast populations. This population had a sufficient capture of features and counts of mRNA, suggesting that this population was not a technical artefact of the analysis. Nonetheless, we could not identify unique genes to this cluster, apart from the presence of genes involved in cell proliferation and transformation [244]. These genes included FOSB, which in accordance with CDKN1A regulate cell proliferation in a STAT1 dependant manner [244]. However, we were unable to draw possible functional contributions of this populations from our dataset.

The final cluster of fibroblasts were assigned their unique identity due to the differential expression of the thrombosis and fibrosis associated genes SERPINE 2 and F2R. The SERPINE2, also known as protease nexin -1, molecule belongs to the serpin superfamily, known to regulate the activity of matrix metalloproteinase (MMPs) and participate to the modulation of the extracellular matrix of the organ. This superfamily also appears to be regulated by the expression of pro-inflammatory cytokines such as IL-1 and IL-8 [245, 246] which are expressed by many stromal cells including fibroblasts and are abundant within the inflammatory cytokine milieu. Overexpression of the serpin superfamily has been demonstrated to participate in pathological cardiac fibrosis. In this instance, knockdown of Serpine2 has been shown to reduced cardiac fibrosis whilst knock in of this protein correlates with increased collagen deposition [245]. Fibrosis in the pSS SG has been previously debated in its manifestation – whether it occurs as result of primary disease pathology or is a phenomenon associated with the ageing SG [247]. Increased fibrosis in pSS SG inversely correlated with saliva flow due to the loss of functional secretory tissue present within the SG [247]. The expression of SERPINE could therefore suggest a link between inflammation, fibroblasts, and fibrosis in the context

of a local inflammatory response. Concomitantly with SERPINE2, the expression of F2R further reinforces the pro-fibrotic nature of this subset of fibroblasts. In the context of cardiovascular inflammation, F2R has been demonstrated to be a critical modulator of vascular inflammation and stimulated endothelial cell populations to secrete the pro-inflammatory cytokines such as IL-6, known fibroblast priming cytokine and amplifier of inflammation [15, 248, 249]. Aside from their involvement in fibrotic pathologies, SERPINE2 and F2R have also been explored as emerging targets within tumours [250, 251]. The expression of SERPINE2 has been found elevated in oral carcinoma samples as compared to healthy tissues and has been suggested to promote tumour growth due to its ability to stabilise the TME [250]. Interestingly, a small proportion of pSS patients develop non-haematological malignancies which include oral carcinomas [51], therefore it is possible that this population of fibroblast is also pathogenic in non-lymphoid function associated manner. Similarly, F2R has also been identified to be overly expressed in glioma malignancies, promoting cell proliferation and epithelial to mesenchyme transition (EMT) due to the downstream activation of the Wnt pathway [251]. Intriguingly, pathway enrichment analysis of this fibroblast population demonstrated engagement of the Wnt and BMP pathways relative to the other fibroblast subsets, which supports the idea this fibroblast compartment may be pro-oncogenic and modulate the SG microenvironment to support pSS associated malignancies.

Within the CD34⁻ cell population we identified the expression of the lymphoid chemokines CCL19 within the CD34⁻ CD146⁺ pericytes. The spatial location of TLS has been identified to be in close proximity to vasculature structures which comprise of pericytes, smooth muscle cells or myofibroblasts [10, 14, 92]. We compared the phenotypic and potential functional characteristics of the CD34⁻ CD146⁺ cells which included pericytes and vascular smooth muscle

cells (VSMCs) in order to understand what molecular components are available for TLS formation to preferentially take place near vascular structures. VSMCs are core cells of arteries and primarily function to maintain the structural integrity of the arterial wall [252]. These cell populations are able to respond to inputs from various neighbouring cell populations such as endothelial cells of the tunica intima, circulating immune cells and hormones which allows VSMCs to undergo phenotypic switching in response to injury or inflammatory cues [252]. This plasticity also allows VSMCs to participate in maintaining cardiovascular immunity due to the formation of arterial tertiary lymphoid structures (ATLS) due to their aforementioned immunomodulatory capabilities. In pSS, VSMCs resident within the SG may become activated in a similar manner to those within the circulatory system, facilitating lymphocyte entry and recruitment due to increased flow through of blood [163, 253, 254]. In our dataset, we identified 2 populations of VSMCs uniquely identified by the differential expression of SLMAP and SLC38A11. Firstly, the protein SLMAP has been characterised to participate in the normal development of skeletal muscle development [255]. This suggests that this protein demarcates a VSMC population which part of the normal physiology of the SG. The second population of VSMC were identified by the expression of SLC38A11, which is related to the family of amino acid transporters, previously demonstrated to induce VSMCs proliferation [256, 257]. This suggests that within the pSS SG this population of VSMCs acquire increased proliferative capabilities, possibly to respond to the need of replacement of damaged VSMCs caused by inflammation or to accommodate increased blood flow within the organ.

One of the 2 populations of pericytes identified withing our dataset was distinguished by the ability to secrete the lymphoid chemokines and we defined this population as

“immunopericytes”. Within SLO such as the LN, ITGA7⁺ pericytes have been identified, which facilitate the entry of lymphocytes within the organ[89]. Furthermore, the identification of “pericyte tracks” with the spleen demonstrated that circulating lymphocytes “walked” on these tracks in order to enter the organ [116]. However, the precise signals and mechanisms which induced the formation of these “pericyte tracks” and the expression of the lymphocyte chemokines in this population was not discussed. The identification of CD34⁻ CD146⁺ CCL19⁺ pericytes within the human pSS SG strongly suggests that there may also be a mural component of lymphocyte infiltration and organisation, which contributes to the formation of TLS. We have been able to demonstrate the presence of this pericyte population at protein level, using the markers TNC and CD36 to discriminate between the immunogenic and non-immunopericyte within the human pSS SG. As expected, CD146 pericytes were located close to vascular structures and, intriguingly, at the outskirts of the fibroblast cluster. Here, the CCL19⁺ pericytes may well represent the initial cell population which facilitates the first wave of T cell entry within the SG which then participate in fibroblast priming and activation.

Using pathway enrichment analysis, we sought to identify which signalling mechanism could lead to the formation of this lymphoid pericyte population but were unable to identify a definitive answer. Intriguingly, our data did demonstrate that the expression of CCL19 was independent of the classically associated LTβR pathway as compared to the CD34⁺ CD146⁻ fibroblast populations, which warrants further investigation as to what may drive the acquisition of this phenotype. More intriguingly, our pathway enrichment data demonstrated the increased expression of the prostaglandin cascade in the “immunopericyte” population. The involvement of the prostaglandin signalling cascade has been explored in the context of pericyte and endothelial cell interactions and T cell migration [258, 259]. In a study conducted

by Perrot *et al*, it was observed that engagement of prostaglandin in a Prostaglandin E2 (PGE2) dependant manner in pericytes, altered their stable interaction with endothelial cells to that of non-adhesive and fast migrating cell types [259]. Whilst their functional ability to secrete the lymphoid chemokines has not been explored, this data suggests that PGE2 modulates the overarching phenotype of the pericytes. It would be intriguing to explore the downstream effects of prostaglandin stimulation on human pSS pericytes *in vitro*, or functionally *in vivo*, on the enhancement or induction of the “*immunopericyte*” phenotype. However, the involvement of prostaglandin has been discussed in the context of T cell migration and the surface expression of CCR7, the cognate receptor of CCL19 and CCL21 [259]. This hints that prostaglandin may govern the CCR7:CCL19/CCL21 axis by exerting different effects on the different cell types such as T cells and from our data, the pericytes. Further functional exploration and characterisation is required to explore the possible mechanistic contributions of this signalling cascade in its contributions to promoting chemokine expression from a specific CD34⁻ CD146⁺ mural cell population. However, given our current observations from the scRNAseq dataset and previous functional characterisation of the *immunofibroblast* a possible mechanism combining the lymphoid expression ability of the CD34⁻ CD146⁺ CD36⁺ TNC⁺ CCL19⁺ CCL21⁺ “*immunopericytes*” in the formation of TLS. In this manner, the “*immunopericyte*” population may initiate the extravasation of lymphocytes from the circulation through the endothelium population by establishing a chemokine gradient. Concomitantly, the resident mesenchyme within the human SG maybe remodelled to acquire an *immunofibroblast* phenotype such that the organisation of the infiltrating lymphocytes then becomes that of a TLS. However, this possible cellular mechanism of TLS formation

requires validation at the functional level to confirm whether the cell populations are synchronous in the formation of TLS.

Figure 14: Diagrammatic summary of the possible functional relationship between the *immunofibroblast* and “*immunopericytes*” in TLS formation.

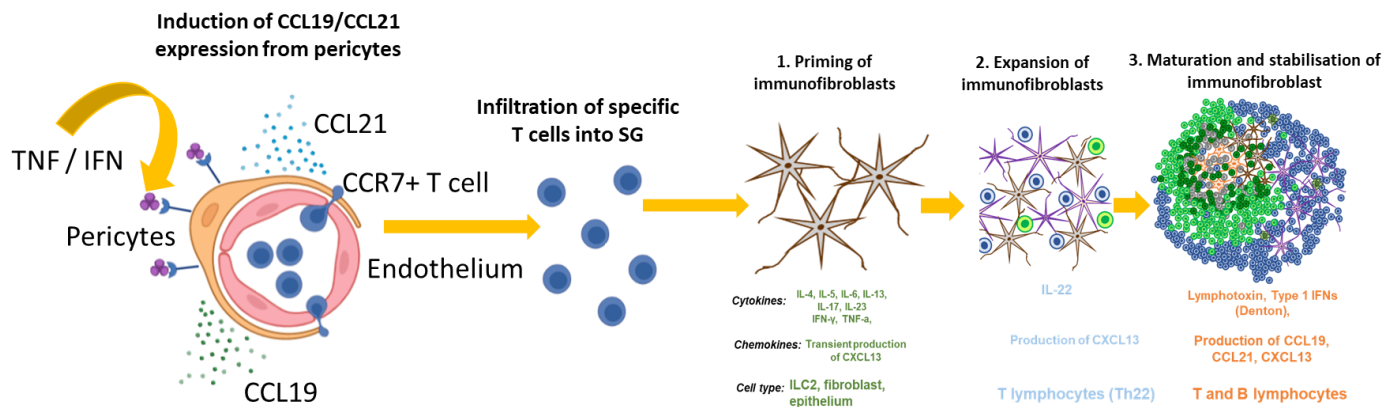


Figure 14: Diagrammatic summary of the possible mechanism of lymphocyte trafficking and TLS formation involving the “*immunopericyte*” and the *immunofibroblast* in TLS formation. The mRNA expression of the lymphoid chemokine *CCL19* and *CCL21* from the CD34⁻CD146⁺CD36⁺TNC⁺ pericyte population from our scRNAseq dataset suggests that initial T cell infiltration into the pSS SG maybe initiated by the “*immunopericyte*” population. Subsequent organisation of infiltrating lymphocytes within the human pSS SG maybe governed by the *immunofibroblasts* and additional fibroblast subsets we have identified in our human pSS scRNAseq dataset,

Using the classical endothelial marker Pecam-1 (CD31), we identified 4 endothelium populations which were divided as either NOTCH4⁺ SELP⁻ arterial or NOTCH4⁻ SELP⁺ venous endothelium. This is in accordance with the presence of these two types of endothelium in normal vasculature architecture [9, 260]. However, the differential expression of CA4 and SEMA3G in arterial endothelium and a VCAM-1 expressing venous endothelium population suggests compartmentalisation of function in these cells. The anhydrase family of enzymes to which CA4 belongs, are involved in the regulation of pH within a cell or tissue by catalysing the conversion of carbon dioxide and water to bicarbonate and a free proton molecule [261]. The cell membrane bound CA4 has been demonstrated in tumour endothelial cells (TECs) to promote TEC survival in lactic acid rich TMEs. The survival of CA4⁺ TECs subsequently allow tumour progression to occur by participating in angiogenesis [262]. Whilst inflammatory microenvironments are characterised to become acidic due to the increased infiltration of activated lymphocytes, the presence of this feature of the human pSS SG has not been formally characterised. The specific unique expression of CA4 on the arterial endothelium in the human pSS SG suggests that this population may participate in the regulation and maintenance of a specific metabolic signature within the SG to allow the progressive infiltration of lymphocytes. Moreover, in a similar context to cancer, this CA4⁺ endothelium population may contribute to the formation of a microenvironment which supports the formation and progression of B cell lymphomas.

The second arterial endothelial population identified presented unique expression of the recently discovered angiogenesis regulator SEMA3G, demonstrated to associate with

angiogenic endothelial cells and governing the patterning of blood vessel development [263, 264]. This class 3 semaphorin is a soluble growth factor that exerts its regulatory functions through its engagement with plexin receptors and neuropilin co-receptors and has been demonstrated in *SEMA3G^{-/-}* mice to repel lymphatic endothelial cells (LECs) from arterial vessels [263]. This suggests the role of this signalling cascade to induce and oversee the correct formation and organisation of the lymphatic vasculature which has been demonstrated to be dynamic and altered during TLS development [100].

Within this endothelial subset, we also identified 2 endothelial population one of which was identified by the unique expression of VCAM-1. The collective expression of VCAM-1 and p-selectin (SELP) indicates a highly activated vascular endothelial population as compared to the SELP⁺ VCAM-1⁻ counterpart. This highly activated venous endothelium has been characterised in other inflammatory contexts to facilitate the transmigration of lymphocytes in response to chemotactic signals [107, 163, 253]. Collectively, the endothelial population we have characterised suggests that the subsets have signatures of genes which may govern the maintenance of the metabolic environment, facilitate the formation of *de novo* lymphatic vessels and lymphocyte transmigration into the human pSS SG.

Another stromal cell population which was predominantly associated with pSS was the SGEC population [9, 53]. The involvement of the epithelium population in this disease has been demonstrated in the spatial organisation of infiltrating immune cells near the SG ducts as a result of the expression of the lymphoid chemokines CCL19, CCL21, CXCL13 and CXCL12 [12]. The activated epithelia also express the surface adhesion molecules ICAM-1 and VCAM-1 which further facilitates immune cell positioning. In our dataset we identified 7 SGEC

populations, 4 of which were positive for the traditional epithelial marker Epcam. Within this population, 2 subsets individually expressed the genes Krt14 and SLC13A2. The expression of Krt14 has been used to identify the basal epithelium of the SG in mice which has been characterised to be responsive to RA signalling [110]. This study identified that inhibition of RA signalling specifically in the basal epithelium resulted in the down regulation of Krt14 and cell proliferation [110]. This may infer that, within the human pSS SG, we have captured a basal epithelium population which may also be responsive to RA signalling.

The second epithelium population demonstrated mRNA expression of the solute carrier SLC13A2, one of many solute carrier proteins characterised within epithelium cells [265]. To function correctly, transport of Cl^- and HCO_3^- is essential in epithelial cells as these ions act as buffers that maintains the correct acid-base balance in the tissue to prevent acidosis [265]. Additionally, HCO_3^- functions to retain the secretory molecules such as mucins and other digestive enzymes in their soluble form [265]. This suggests that this specific population of epithelia within the SG maybe involved in the maintenance of the correct pH balance within the organ, however further exploration and validation at the protein level is required to confirm this function. Four Epcam⁺ populations were identified by the unique mRNA expression of the ion channel inhibitor ANKRD36C [266]. Mutations in this ion channel have been demonstrated in the context of gastric cancers [267]. A different Epcam⁺ cluster uniquely expressed the mRNA for MUC5B which is forms a component of saliva [268]. Interestingly, the lack of MUC5B identified in the saliva of patients which had undergone radio therapy for head and neck cancers correlates with increased oral dryness [268]. Whilst the specific lack of this mucin has not been explored in the context of pSS, the unique expression of this mucin by a specific population of epithelial cells suggests this mucin may form a high proportion of the

saliva content. MUC16 was another mucin mRNA expressed uniquely by another epithelial cell cluster which also forms composition of the saliva [269]. The last population identified was the myoepithelium which was differentially expressed by the Notch ligand DKK3. The myoepithelium is situated at the basement layer of the glandular epithelium and due to their contractile nature, function to expel saliva from the SG [270]. In the case of pSS, damage to this cell population as a result of inflammation results in poor saliva secretion in combination with other anatomical features. This DKK3 expressing myoepithelium population has also been explored in solid adenoid carcinomas and shown to correlate with poor myoepithelium differentiation [271]. This may suggest that the loss of secretory function observed in the pSS patients may be partly owed to increased Notch pathway engagement in the human SG within the myoepithelial population.

Within the human pSS SG we also identified a population of glial cells. Glial cells provide functional and structural support to neurones [272]. In our dataset we identify a neuronal associated population of cells due to the mRNA expression of NCAM-1. We captured a NCAM-1⁺ CD32⁺ MPZ⁺ Schwann cell population and a NCAM-1⁺ POSTN⁺ PLXNC1⁺ MBP⁺ neuroepithelial cell population within the human pSS SG scRNAseq dataset. The Schwann cell population displays gene expression related to its function of insulating neurones such that increased transmission of electrical impulses occurs. Secondly, the neuroepithelial population we identified expresses genes which participate in the cell attachment, cell adhesion in addition to MBP which is associated with early neuronal development [272]. These genes suggest that this small cell population within the SG may be able to participate in the replenishment and generation of new neuronal cells either with age or in response to damage due to the inflammatory cytokine milieu present within the pSS SG.

In the context of pSS, the infiltrating lymphocyte population responsible for the modelling of the resident mesenchyme within the SG has been characterised as CD4 or CD8 T cells [21, 274]. Within the T cell CD4 subset, T helper cells 1 (Th1) and T helper cells 2 (Th2) have been associated with the development and progression of pSS [275]. Classically, CD4 Th1 cells have been well characterised to produce inflammatory cytokines such as TNF- α and IFN- γ . These cytokines have been associated with the priming and activation of the stromal cells which contributes in the amplification of disease [183]. This was demonstrated with the deletion of IFN- γ and IFNR in non-obese diabetic (NOD) mouse models with resulted in the reduction of the clinical signs of pSS as compared to wild type mice [276]. Moreover, the Th2 cytokine profile of IL-4, IL-6 and IL-13 have been demonstrated to contribute to the priming and activation of the fibroblast compartment prior to TLS formation and establishment [8]. In our datasets, we captured four CD4 populations and three CD8 populations within the human pSS SG. Within the CD4 subset we identified a subset of, Th1, Tfh, Treg and T central memory cells. These identities being conferred by the observed mRNA expression patterns of specific genes unique to each cell population as per previously established literature. The contributions of the Th1 population may comprise the priming of the resident fibroblast population with the expression of the pro-inflammatory cytokines IFN- γ and TNF- α which have been demonstrated to mediate expression of adhesion markers in gingival fibroblasts [183]. Therefore, this pro-inflammatory CD4 subset may recapitulate the same functions observed on gingival fibroblasts but on the immunofibroblast population within the inflamed SG by mediating their priming and activation. Similarly, the presence of the Tfh population of cells has been characterised in the context of the DZ of the GC and their contribution to exasperating pSS

with the expression of the B cell chemokine CXCL13 has previously been explored too [277]. In this study, the Tfh population was identified to be abundant in TLS⁺ SG and were responsible for the inflammatory milieu due to the expression of the B cell activating and survival factor IL-21 [277]. Contrastingly, the contributions of Tregs and T central memory cells in the context of pSS has been less explored. The classical function of Tregs is to mitigate the consequences of an immune response mainly by the production of the immunosuppressive interleukin IL-10 [67, 69]. The infiltration of Tregs within the inflamed pSS SG suggests that whilst they are present within the inflammatory microenvironment, they may be unable to execute their functions effectively [278]. This may occur due to the overriding effects of the inflammatory process which prevents Tregs from containing the damage of bystander tissue within the SG. Additionally, the presence of T central memory cells, which function to survey the peripheral tissue and circulation [279], within the SG hints that the antigen or inflammation inducing insult is one recognised by this cell population. The proportion of the CD4 Treg, Tfh and T central memory was seen to be greater in Ro/SSA⁻ patients as compared to Ro/SSA⁺ patients suggesting that in the absence of autoantibody producing cells the SG is occupied predominantly by these T cell subsets prior to established TLS mediated pathology. This also suggests that these cells may be forming the T cell aggregate which is observed at initial TLS formation as the underlying stromal cell compartment undergoes remodelling and maturation. On the other hand, the CD8 subset of cells identified within this dataset by the expression of IFN- γ , granzyme A and granzyme K suggest the occurrence of an anti-viral response within the SG, as the onset of pSS is associated with the presence of an EBV signature [9, 39, 207]. The three uniquely defined CD8 populations collectively demonstrate a pro-inflammatory response which participates in the clearance of viral antigen, promotes targeted

cell death but can also result in the augmentation of the pro-inflammatory cytokine milieu within the SG. Interestingly, an increased proportion of CD8 GZK cell population in the Rho/SSA⁺ subset of patients suggests that these patients have an elevated level of inflammation with the reduced proportions of Treg cells.

Collectively, the presence of the different T cell populations within the pSS SG reinforces the idea that T cells may contribute to the remodelling and priming of the underlying stroma prior to TLS formation. We also explored our scRNAseq dataset to explore the expression pattern of the receptor and ligands between the IFN and TNF family. Firstly, this data demonstrated that the mRNA expression of the receptors TNFR2, IFNGR1 and IFNGR1 were expressed amongst the human CD34⁺ CD146⁻ fibroblasts and the CD34⁻ CD146⁺ pericytes. More specifically, the mRNA expression of the IFNGR1 was present in the fibroblasts and pericytes but not in the vascular smooth muscle cells. Moreover, IFNGR2 gene expression was specific to the CD34⁺ CD146⁻ fibroblast population suggesting the involvement of this receptor is critical in transmitting the downstream signalling cascades which ultimately participate in assigning the fibroblasts a primed and activated phenotype. In this dataset, we also observe the enriched gene expression of IFNG within the CD8 T cell population suggesting that this IFNG signal may be mediated by CD8 T cells in particular, amongst other immune cell present within the human pSS SG during inflammation. The mRNA expression of TNF was enriched within the CD4 T cell subset, more precisely the CD4 Th1 subset of cells, whilst IFNG expression was preferential within the CD8 T cell populations. Thus far, this data enlighten the local crosstalk between T cells and stromal cells occurring in TLS. We further determined the gene expression pattern of the classical chemokine and receptors facilitating the navigation of T cells into the pSS SG. We confirmed the mRNA expression of CCL19 has been described in both the CD34⁺

CD146⁻ CCL19⁺ TNFSF13B⁺ fibroblast population and the CD34⁻ CD146⁺ CCL19⁺ pericyte population in our scRNAseq dataset [8]. The expression of the cognate receptor for CCL19, CCR7, was preferentially expressed within the CD4 subset.

The classical hallmark of TLS formation is the organisation of separate T and B cell compartments in non-immunological organs [9, 10, 14, 34, 61, 92, 93]. For B cells to be accommodated in their allocated spatial location within a TLS the appropriate chemokines, CXCL13 and survival factor, BAFF must be present [8, 136]. Using scRNAseq we identified 3 populations of B cells within the human pSS SG. The phenotype of these cells was explored using the mRNA expression patterns of 2 markers in tandem: IgD and CD27. The three populations of B cells identified comprised of an IgD⁺ CD27⁻ naïve B cell population, IgD⁻ CD27⁻ activated B cell population, and a possible IgD⁻ CD27⁺ switched memory B cells. The presence of these populations within our scRNAseq dataset suggests that a variety of B cells, characterised by different function and phenotype, reside within the inflamed SG. The unique expression of specific genes concomitant with their function allows us to begin to understand the possible functions these populations may have. Firstly, the IgD⁺ CD27⁻ naïve B cell uniquely expressed IL4R, which has been demonstrated to participate in B cell activation in combination with CD40 signalling, to produce the mucosal associated immunoglobulin IgE [280]. Additionally, the expression of FCER2, or CD23, is a low affinity Fc receptor IgE and has been associated with the regulation of the production of this immunoglobulin [225].

This suggests that this could be an IgE secreting B cell population. Moreover, IgE titre has been observed to be higher in pSS patients as compared to healthy controls. The specific contributions of IgE in pSS have not been explored extensively but suggests that this protein

may facilitate glandular inflammation as it is involved in the degranulation of resident mast cells and eosinophils which exacerbate the disease [280].

The second population of B cells suggested an activated population which expressed the mRNA gene expression signature of CD27⁺ CCR7⁺ and BCL2A1⁺ as its unique identifier. The overexpression of this BCL2 related protein has been associated with haematological cancers [227] which may infer that this is a hyperactive population of B cells which be at a predisposition to become oncogenic and give rise to B cell lymphomas which are associated as an outcome of pSS [198]. However, further exploration of this population at the gene and protein level is required to confirm whether this is a pathogenic population of B cells in the pSS SG. Lastly, the IgD⁻ CD27⁺ CCR7⁻ and SHISA8⁺ B cells may identify as switched memory B cells which is indicative of infiltration of B cells in response to inflammatory cues. Different proportion of these B cell subsets were identified in the Ro/SSA⁻ subset of patients as compared to the Ro/SSA⁺ subset, suggesting the possibility that Ro/SSA⁺ patients TLS are inhabited by fewer B cells and greater PC component, in agreement with their higher autoantibody.

Analysis of PC cells localization has been previously performed by Szyszko *et al* who demonstrate that pSS SG PC are localised close to cells producing CXCL12 and IL-6, such as pro-inflammatory fibroblasts [55]. Given the ability of SG TLS to support both CXCL12 and IL-6, it is possible to speculate that those sites might provide protective niches to resident PC against immunosuppressive therapies [42]. Within our dataset, we captured a population of syndecan-1 (CD138)^{low} CD38^{low} plasma cells which suggests this may be a precursor population, or mature centrocytes, which will progressively upregulate the expression of

CD138 and become bona fide PC [281]. In addition to this we have been able to identify 14 different plasma cell subsets which were assigned unique identities based on the expression of differing variable regions. We have confirmed that the variable regions were the differentially expressed genes driving the formation of the independent CD138⁺ PC cell clusters, as removal of these genes resulted in the collapse of the PC populations in a single population of syndecan-1 (CD138)^{low} CD38^{low} cells. The differential expression of these variable region genes suggests that a selection of plasma cells are generated within the SG, emphasising that this tissue is an active site of ectopic B cell maturation and DNA modulation. Intriguingly, the proportion of this cell population is greater in the Ro/SSA⁺ patient subset than in the Ro/SSA⁻. This correlates with the idea that these PC populations are the producers of the autoantibody in pSS patients. Additionally, when exploring the PC signature per patient sample, we demonstrated that some patients have a specific signature of select populations of each cell type whilst others do not. For instance, patient Gx50 had the greatest proportion of the majority of PC as compared to other patient samples whilst patient Gx69 did not display any CD138 IGH1-2 cells and patient G did not display any capture of the CD138 IGHV3-74 population. This suggests that the PC signature is heterogeneous and unique to the patient and may be reflective of the GC heterogeneity we have identified within pSS patients. Aside from the presence of PC with unique variable regions a cluster of cells assigned as IgA⁺ CD3⁺ were also found. Dual expressers cells with a similar profile have been described in the context of diabetes, as a novel pathogenic cell population which may drive this autoimmune disease [282]. However, when we looked to assess the robustness of this cell cluster, we identified that this cell population presented low feature counts, suggesting it may have not been

sampled correctly. For this reason, this cell cluster warrants further analysis before conclusions can be drawn.

The myeloid cell population consists of dendritic cells (DCs), macrophages and monocytes which serve the purpose of transporting and present antigen to the cells within the neighbouring SLO in conjunction with the release of pro-inflammatory cytokines to initiate an inflammatory response [232, 283]. The contributions of the myeloid cell population in the formation of TLS has not been extensively explored. In our dataset, we captured 3 populations of myeloid cells which collectively expressed the mRNA for CD11C as per our scRNAseq dataset. Individually, within our dataset we identified a myeloid DC population with the expression pattern of CD11C, CD1C and ILR2 [284]. The second population of myeloid cells carried a more inflammatory signature with genes such as TGFBR1, TNF and C3, often detected on suppressor cells (MDSCs) [283]. This population of cells have been demonstrated to become expanded in response to inflammatory cues and in the context of cancer, suggesting this population of cells may drive the pSS pathology [283]. The final cluster comprises of the mRNA expression of CD163, ALDH1A1 and CXCL12 suggesting that this myeloid population may participate in the generation of Vit A which may contribute to the post-natal modification of stromal cells such that they mirror B cell supportive stroma development as observed in the SLO mesenchyme [111, 112].

Collectively, the lymphocyte populations we have captured within this single cell data set demonstrate to have inflammatory signatures which may contribute to the remodelling of the underlying stromal cell compartment and in the perpetuation of pSS as a disease. The stromal cell compartment has emerged as having critical immunomodulatory properties, being able

to receive modulatory signals from the infiltrating lymphocytes and to reciprocates signals of activation, proliferation, and survival signals to the immune cells. Within the scRNA landscape we captured the classical populations of endothelial cells, epithelial cells, fibroblasts, pericytes and vascular smooth muscle cells. Each of these stromal cell populations demonstrate to express gene signatures which may contribute to the initiation, formation, and persistence of TLS.

In this thesis, we harvested GCs microdissected from 10 female pSS SG samples to complement our single cell work to provide molecular signatures of specified areas in the context of TLS. Our preliminary analysis on this small dataset suggest that the isolated GC interestingly compartmentalised into 2 distinct groups, suggesting a diversification of phenotype and function of GCs derived from the inflamed tissue. The diversification of the GCs in these two groups and their clustering from GC derived from inflamed human tonsils was driven by the differential expression of specific genes including chemokines, cytokines, and pro-inflammatory molecules.

The lymphoid chemokines CCL19, CCL21 and CXCL13 which orchestrate the infiltration and organisation of T and B cells demonstrated differential expression patterns with GC1 showing increased CCL19 expression than GC2 and tonsil derived GCs. CCL21 expression was, on the contrary higher in GC1 and tonsil GC than in GC2. The inverse expression of these 2 T cell chemo attractants according to which GC group they derived from suggests that a mechanistic switch leads to the expression of one T cell lymphoid chemokine as compared to the other. The reason for this differential gene expression is not clear and may have occurred due to a sampling artefact which resulted in the inclusion, or exclusion, of certain cell aggregates which may have had expression of CCL21.

The mRNA expression of the B cell lymphoid chemokine CXCL13 was also diverse among GC1 and GC2 and tonsil GC. Few GC samples showed low expression of this chemokine. This may indicate that these GCs have not matured to express CXCL13 which can be derived from resident GC FDCs and Tfh cells [133, 135, 136, 285]. In addition to this low level of CXCL13

expression, the B cell survival factor, TNFSF13B (BAFF) and its receptor TNFSFR13B (BAFFR) were also demonstrated to have low expression, suggesting that these survival factors were not abundant within the GCs or rather were rapidly utilised for their function and were not captured in the dataset.

An additional chemokine signature identified within TLS is that of CXCL12 [12, 135, 136, 286]. The canonical receptor for this chemokine is CXCR4, however this ligand can also engage with ACKR3 which modulates the chemokine gradient within tissue [286]. The CXCL12: CXCR4 axis facilitates the migration of cell populations such as B cells, T cells within SLO, but also exasperates the migration of cancer cells during metastasis [287]. In the context of pSS, CXCL12 expression pattern was observed to be predominant within the heavily infiltrated SG and overlapped with promoting the survival of malignant B cell clones present within the tissue [12]. In our dataset the CXCL12 expression demonstrated greater expression levels in GC2 as compared to tonsil derived GCs and GC1 clusters. The expression pattern of CXCR4 overlapped the expression of CXCL12 which was expressed in a mixture of GC1, GC2 and tonsil GC samples suggesting that there was engagement of this receptor ligand axis which promotes survival of B cells within the GCs. Additionally, BLK – a tyrosine protein kinase – which is involved in the development and differentiation of B cells and the CXCL13 receptor CXCR5 had low expression which was present in GCs from tonsils but had dispersed expression in GCs from both the GC1 and GC2 clusters [288]. This suggests that this tyrosine kinase was more engaged in GCs derived from the human pSS SG, as compared to the tonsil GCs in accordance with expected elevated ectopic B cell proliferation and differentiation in the SG inflammatory setting. This exploratory data in our hands has begun to give hints of the additional overarching heterogeneity encompassed between GCs of the same tissue and disease. This

raises great interest, as understanding the cellular and molecular contributions which give rise to the TLS, and GC heterogeneity we observe, would allow us to harness the potential to inhibit or induce the specific signals or cell populations which contribute to this phenomenon.

In this work, we have identified precise fibroblast compartments which reside within the pSS SG in addition to the detection of the immunofibroblast population using transcriptomic analysis. This data has allowed for the first time the transcriptional characterisation of the inflamed SG which opens avenues of further research and investigation. The contributions of the mesenchyme in TLS formation have been explored previously but the immunogenic pericyte population is a novel finding within the pSS SG which presents us with perhaps two key cellular targets which maybe therapeutically manipulated to either dissolve pathogenic TLS in autoimmunity or promote TLS formation in the context of cancer. However, what we still do not understand are the exact signals that specifically induce the formation of the specific fibroblast populations and the $CD34^- CD146^+ CCL21^+$ pericytes which may play a crucial role in the recruitment of lymphocytes into the human SG. We have also explored the GC heterogeneity which exists between patients of pSS in comparison to tonsil derived GCs and observed differential pattern expression of the classical lymphoid and TLS associated genes accordingly. This suggests that GCs within ectopic tissues have different functional and phenotypical characteristics as compared to the SLO GC. Further exploration of this observation may reveal key molecular findings which could be utilised as novel biomarkers determining whether a patient has a functional or dormant GC within their tissue thus predicting clinical severity of the disease and response to therapy in particular patients.

Chapter 4: Transcriptomic characterisation of the murine pSS SG

To validate our findings *in vivo* and address functional relationship between the different cellular populations described in human SG, we took advantage of a well-established model of SS, already exploited by our group to establish the role of fibroblasts in TLS development [7, 8, 93]. This mouse model mimics the initiation, formation, and maturation of TLS in response to virally induced inflammation using the replication deficient adenovirus. Mature TLS form in the murine SG within 15 days (D15) post cannulation (p.c.) and appear to resolve between 23 days and 30 days after infection [93]. Stromal activation markers such as PDPN, FAP, ICAM-1, VCAM-1 on *immunofibroblasts*, and the expression of the lymphoid chemokines CCL19, CCL21, and CXCL13 have been described in this model, in association with the formation of segregated T and B cell areas within the TLS [7, 8, 93].

We sought to complement our functional characterisation of the *immunofibroblasts* by subjecting 3 mice to SG inflammation (D15 p.c.), by cannulation of a replication deficient adenovirus, and 3 resting murine SG (D0) to scRNAseq analysis. Our selection for these time points were undertaken as characterisation of this mouse model at D15 p.c. as demonstrated to be the peak of TLS formation, such that it resonates as close as possible to the TLS signature seen in the human pSS SG [93]. This model of inducible TLS demonstrated loss of SG function comprising of reduced saliva flow after delivery of the replication deficient adenovirus (Adv5) in the submandibular glands of C57BL/6 mice **Ref.** Inflammation induced in this manner leads to the initial early formation of diffuse lymphocytic aggregates towards a more organised and functional TLS which is accompanied with the expression of the lymphoid chemokines CXCL13, CCL19 and CCL19 [93]. This was confirmed from the upregulation of the

mRNA transcripts of CXCL13, CXCR5 and CCL19 in mice treated with Adv5. Cell populations were not sorted, and non-viable cells were removed to yield high quality cells for downstream analysis. In this chapter I will discuss and describe the different stromal cell compartments this dataset has captured in a non-inflammatory vs. inflammatory SG microenvironment. Moreover, I will describe the validation work that I have been able to undertake to establish the functional role of identified signalling cascades in the biology of the selected fibroblast and stromal cell populations.

4.1: Exploring the cellular landscape of the murine SG

In our work we have been able to compare similarities and differences between resting and inflamed SG (**Fig 1 & 2**). Following the removal of ambient gene signatures in our datasets, we overlaid the differing samples within each time point to assess whether the visualisation of the data was driven homogeneously by all samples (**Fig 1A & 2A**). In this manner, we demonstrated that a particular cluster did not derive from a particular sample, suggesting that our data were representative of the tissues that had been sampled. We then proceeded to assign cellular identities to the clusters using feature plots tool within Seurat, using the canonical markers associated with each biological population (**Fig 1B & 2B**). This demonstrated the location of particular cell types in the context of UMAP dimensions which formed the global cluster identity annotation (**Fig 1C & 2C**). This led to the identification of 24 cell clusters at Day 0 comprising of 10 clusters originating from the CD45⁺ compartment. The immune cell clusters comprised of *Cd3e*⁺ *Cd8a*⁺, but no *Cd4*⁺ T cells, *Klrb1c*⁺ (NK1.1) NK cells, *Itgae*⁺ and *Cx3cr1*⁺ (CD103) myeloid cells, *Ms4a1*⁺ (CD20) B cells and Syndecan-1 (*Sdc1*⁺) PC (**Fig 1C**). This is in accordance with the absence of inflammatory cues within the resting murine SG which would facilitate the infiltration of lymphocytes within the tissue. Contrastingly, we identified 14 clusters in the CD45⁻ stromal cell populations which were assigned their overall identities using the expression pattern of *Epcam*⁺ epithelial cells, *Pecam-1*⁺ endothelium, *Pdpr*⁺ fibroblasts, and *Mcam*⁺ (CD146) and *Regulator of G-protein signalling 5* (*Rgs5*)⁺ pericytes (**Fig 1C**).

Within the inflamed Day 15 p.c. SG we observed an increased number of clusters with 12 stromal cell 2 clusters and 35 clusters of lymphocytic cells as compared to our D0 dataset (**Fig**

2A & 2C). The vast expansion of the immune cell population within the SG aligns with previously published data which demonstrates the increased influx of lymphocytes in response to the increased expression of chemokines upon virus mediated inflammation [8]. Assignment of the cellular identities was completed in an identical manner to that of Day 0 cluster analysis.

As we are focusing on exploring and characterising the heterogeneity of the stromal cell compartment in our current work, I have focused on key cell populations which include, fibroblast, pericytes, epithelial cells and endothelial cells within this chapter.

Figure 1: Global Cellular landscape of the Day 0 pSS murine model SG

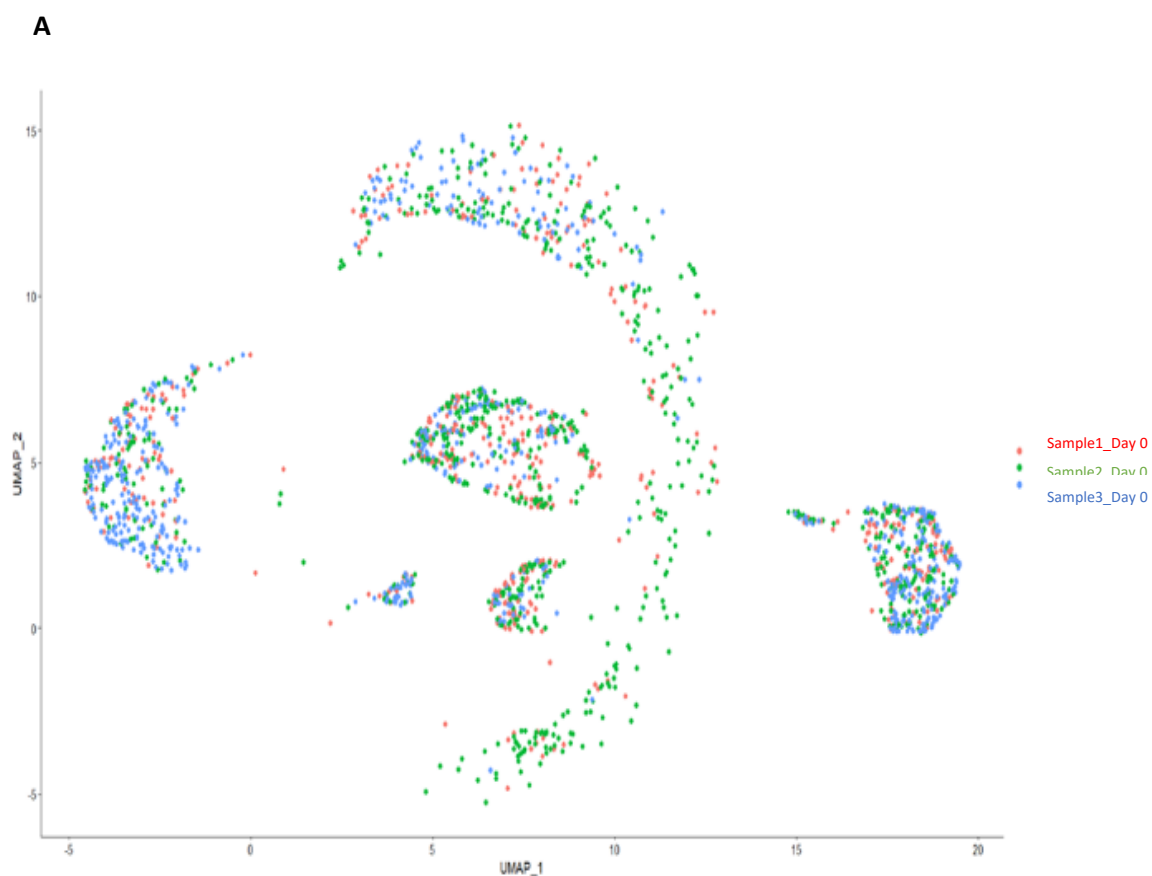


Figure 1: **Course global clustering of the D0 murine salivary gland.** A) Overlay of the 3 samples from 3 individual resting murine SG in UMAP space.

Figure 1: Global Cellular landscape of the Day 0 pSS murine model SG

B

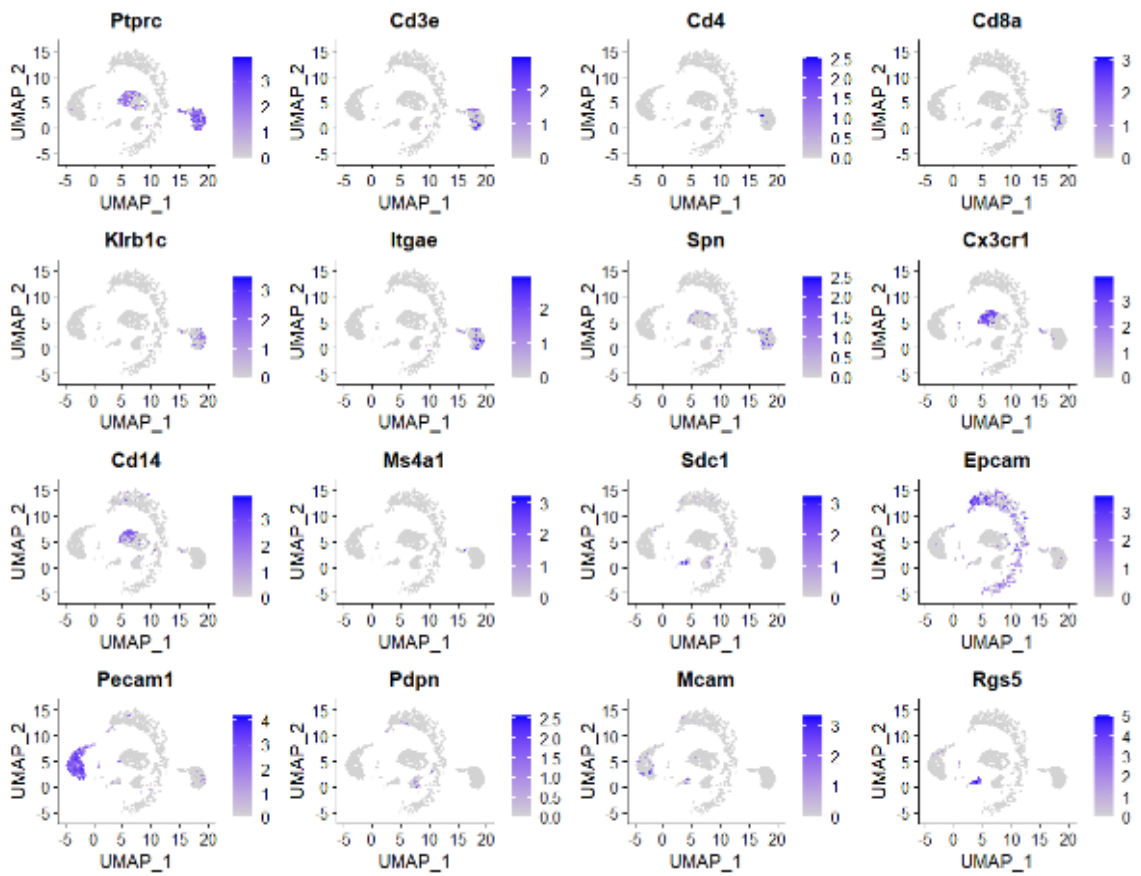


Figure 1: Course global clustering of the D0 murine salivary gland. B) UMAP of Day 0

data after removal of soup signature.

Figure 1: Global Cellular landscape of the Day 0 pSS murine model SG

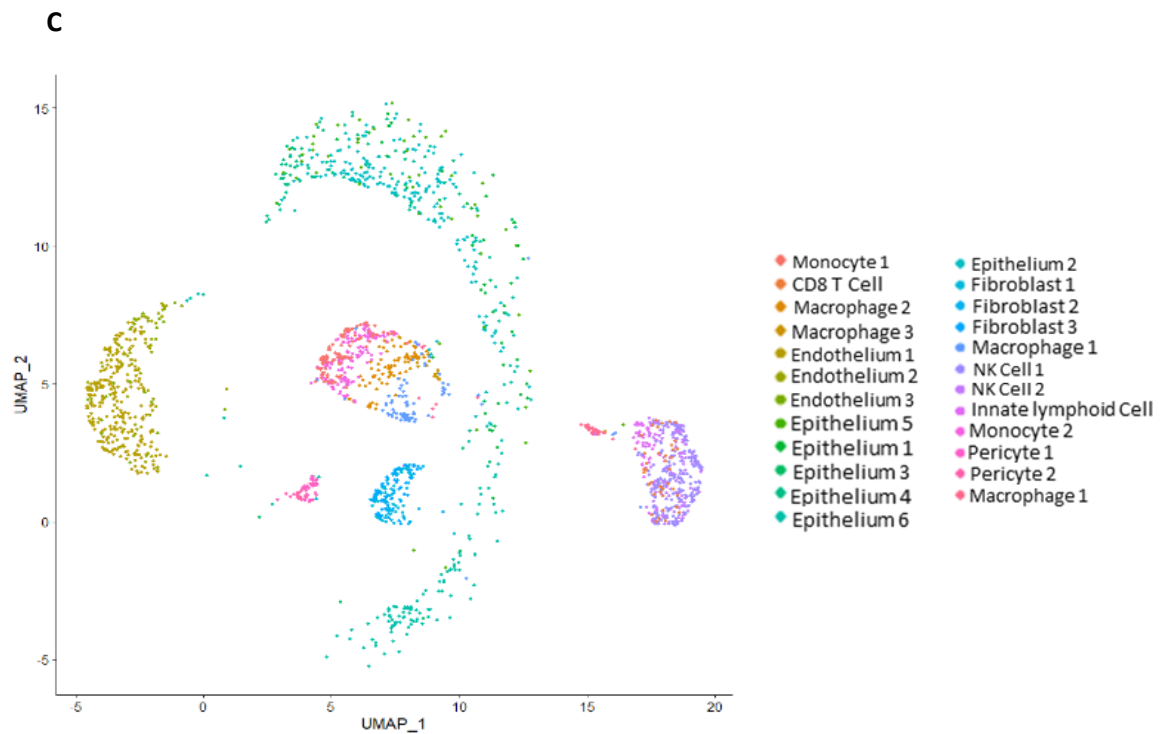


Figure 1: **Course global clustering of the D0 murine salivary gland.** C) Feature plots of key gene markers identifying the classical cellular components of the hematopoietic and the non- hematopoietic compartments and their expression in UMAP space.

Figure 2: Global Cellular landscape of the Day 15 pSS murine model SG

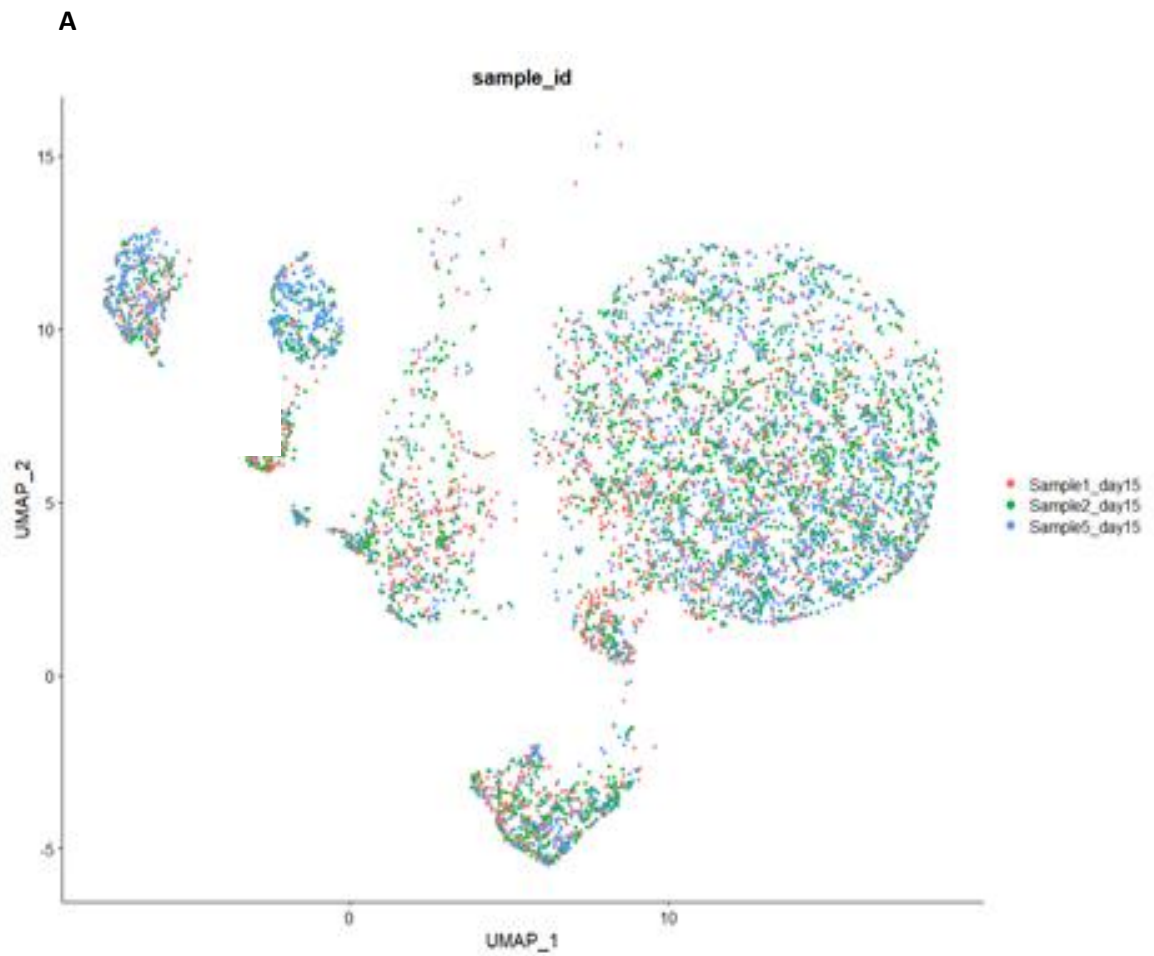


Figure 2: **Course global clustering of D15 p.c. murine salivary gland.** A) Overlay of the 3 samples from 3 individual murine SGs at D15 p.c.

Figure 2: Global Cellular landscape of the Day 15 pSS murine model SG

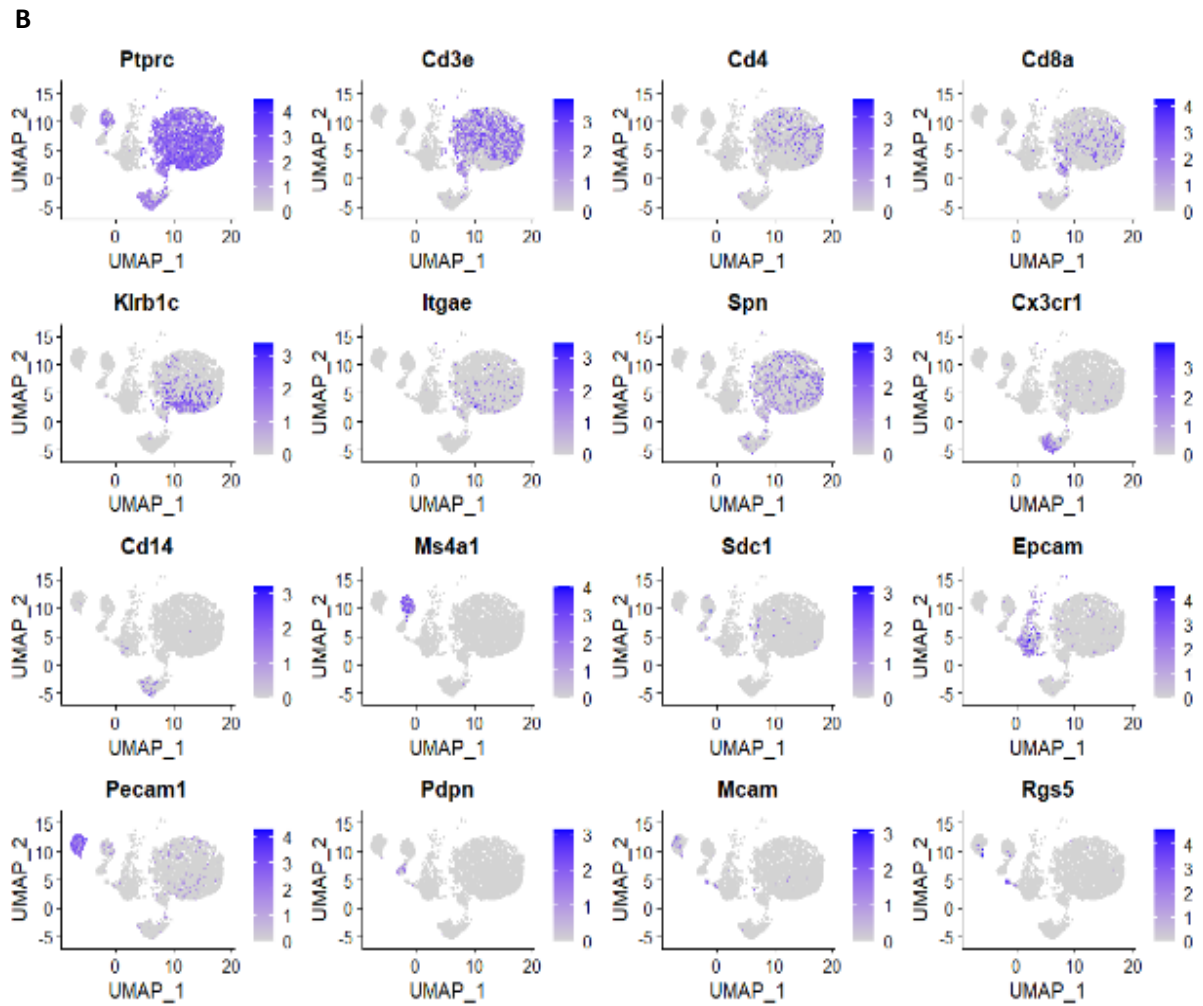


Figure 2: Course global clustering of D15 p.c. murine salivary gland. B) Feature plots of key gene markers identifying the classical cellular components of the haematopoietic and the non- haemopoietic compartments and their expression in UMAP space.

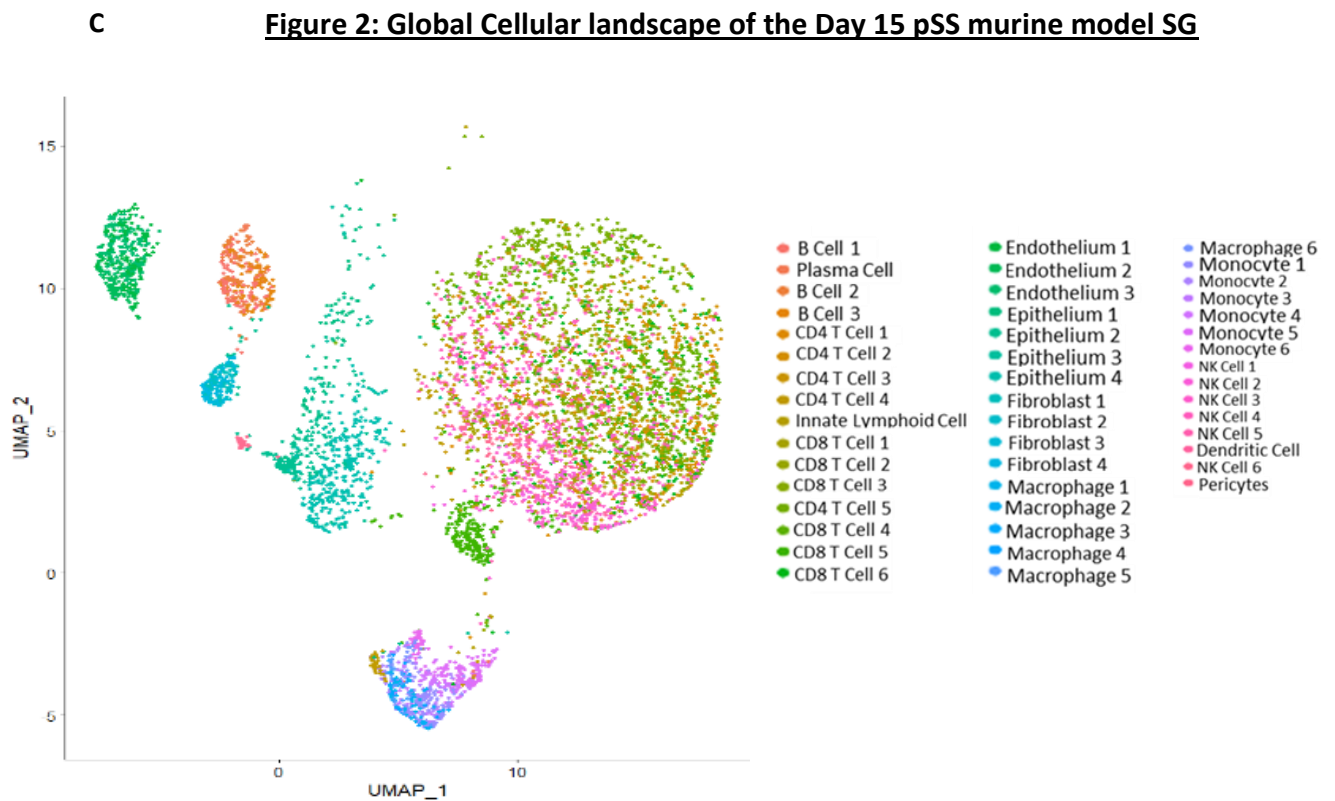


Figure 2: **Course global clustering of D15 p.c. murine salivary gland.** . C) UMAP of Day 15p.c. data after removal of soup signature. H) Stacked violin plots of classical gene markers according to cluster identity.

4.2: The single cell landscape: the stromal cell compartment

Our group has previously demonstrated that the underlying stromal compartment of the murine SG can be induced to express the lymphoid chemokines CCL19, CCL21 and CXCL13 in response to viral induced inflammation [93]. Exploration of the ability of *immunofibroblasts* to acquire lymphoid like features included investigating the effects of various pro-inflammatory cytokines such as IL-13, IL-4 and IL-22 and their contributions to the modulation of the fibroblast compartment [7, 8]. The effects of IL-13 and IL-4R were explored in mice deficient for these genes in which SG inflammation had been induced [8]. These data demonstrated the inability of fibroblasts to upregulate *activation* markers in the early phases of TLS establishment, in absence of IL-13. IL-13 expression in the SG was supported by innate lymphoid cells (ILCs) and the resident fibroblast and epithelial cells [8]. We further explored the requirement for lymphocytes to form the *immunofibroblast* network and demonstrated that, in Rag2^{-/-} mice the expansion phase of the network is impaired, while the early acquisition of activation markers by the resident fibroblast is maintained. Importantly, in Rag2^{-/-} mice the expression of lymphoid chemokines is also abrogated. This suggest that lymphocytes are dispensable for the early priming of fibroblasts but required for their proliferation and full maturation. Accordingly, lymphoid chemokine expression was also reduced in the SG of LTβR^{-/-} mice [8]. Our group had also uncovered an element of organisation within the heterogenous PDPN⁺ fibroblasts using the marker CD34 independently of scRNAseq [8]. In this manner, we demonstrated PDPN⁺ CD34⁻ expressed the lymphocyte survival factors BAFF and IL-7, whilst PDPN⁺ CD34⁺ expressed the lymphoid chemokines CCL19, CCL21 and CXCL13 [8]. This data suggested the presence of a potential functional compartmentalisation within the heterogeneous fibroblast population of the SG.

Whilst the current literature has been able to characterise the stromal heterogeneity in the context of murine LN, we will for the first time explore the stromal heterogeneity present within a TLS permissive inflamed murine SG.

4.3: Fibroblasts of the resting and inflamed murine SG

In our dataset we have identified 3 populations of fibroblasts at Day 0, and 4 populations of fibroblasts at D15 p.c.. One of the fibroblast populations identified within the D15 p.c. dataset were aligned with the *immunofibroblast* population Nayar *et al* have previously characterised using a combination of functional *in vivo* and *in vitro* experiments [8]. Overall, fibroblast populations at both D0 and D15p.c. had expression of the canonical genes *Thy1* (*Cd90*), *Pdpn*, *Cd34* and *Pdgfra* (**Fig 3 A-D**). More concisely, the resting murine SG fibroblast populations were identified by differential expression of various genes associated with a possible specific function (**Fig 3 A & B**). The first fibroblast cluster had specific expression of the Cx3R1 ligand *Cx3cl1*, *Integrin alpha-1* (*Itga1*), *Notch2* and *Retinoic acid receptor gamma* (*Rarg*) but also expressed Indolamine *N*-methyltransferase (*Inmt*) which was observed across all 3 fibroblast clusters. The second cluster had specific expression of genes associated with the Wnt pathway such as *Microfibril-associated glycoprotein 4* (*Mfap4*), *Wnt5a*, *Secreted frizzled-related protein 2* (*Sfrp2*), suggesting this subcluster could function as a progenitor population which is reinforced by the unique expression of *Peptidase inhibitor 16* (*Pi16*), a marker for resting universal fibroblasts [162] (**Fig 3 B**). This expression was concomitant with that of *Dermatopontin* (*Dpt*), which was used to pinpoint perturbed fibroblasts across different murine tissues [162]. The third resting mouse SG fibroblast cluster did not present any unique gene expression however resembled the *Mfap4*⁺ cluster by the lack of Wnt associated genes and *Pi16* suggesting that this population might derive from the *Mfap4*⁺ cluster.

Contrastingly, 4 PDPN⁺ CD34⁺ fibroblast populations were captured in the inflamed murine SG (**Fig 3 D**). The *Ackr3*⁺ and the *Complement decay-accelerating factor* (*Daf*/*Cd55*) fibroblast

population presented an overlapping gene signature with that of the D0 *Itga1*⁺ fibroblast, thus suggesting that this population maybe a resident population within the murine SG which acquires additional unique genes such as *Ackr3* and participation in the complement cascade. The second cluster of D15 p.c. fibroblasts had specific expression of *Fcgr2* (*Cd32*) but overlapped expression of *Fibroblast growth factor 7* (*Fgf7*), *Cx3cl1*, and *Notch2* with the third cluster of fibroblasts at D15 p.c. (**Fig 3 D**). The third cluster as identified by the unique expression of the chemokine *Cxcl9* and the activator of the inflammasome *Guanylate-binding protein 5* (*Gbp5*) which aligns itself to the *Cxcl9* TRC population as characterised by *Rodda et al* [89]. Unfortunately, we were unsuccessful in capturing expression of lymphoid chemokines and other pro-inflammatory cytokines, in the fibroblasts identified in the mouse scRNAseq dataset, which would have allowed us to confirm whether any population of the fibroblasts presented a unique lymphoid stromal cell like signature. However, the upregulated expression of inflammatory associated genes suggested that this fibroblast population maybe the *immunofibroblast* population which arises in response to inflammation. This is confirmed by the absence of these signature genes in the D0 tissue. The final cluster of fibroblasts was uniquely identified by the expression of the gene *Plexin-A2* (*Plxna2*), the co-receptor for the semaphorin *Sema3a* [289]. Although this unique gene identified this cluster, many of the genes expressed overlapped with that of the *Ackr3*⁺ *Cd55*⁺ *Pi16*⁺ fibroblast cluster suggesting that either this is a transitional population from the *Ackr3*⁺ *Cd55*⁺ *Pi16*⁺ or could be a separate population with a similar gene expression pattern.

Collectively, the scRNAseq data suggests that at both time points in the resting and perturbed mouse SG a *Pi16*⁺ population is present which may give rise to the other resident fibroblast populations characterised at D0 and D15 p.c. Intriguingly, at D15 p.c. we observe the

emergence of a pro-inflammatory fibroblast cluster identified with the expression of *Cxcl10* and *Gbp5* which is absent at D0 and expresses an inflammation associated gene pattern within the D15 p.c. tissue, suggesting that it arises from an existing population in response to inflammatory cues. As we had the opportunity to explore both resting and perturbed murine tissue, we applied a pseudo bulk correlation approach to our scRNAseq dataset of the murine fibroblasts and explored whether we could align fibroblasts from resting tissue with fibroblast populations we characterised in the inflamed mouse gland (**Fig 4**). This analysis inferred that the D0 *Itga1*⁺ fibroblast population had a gene signature which correlated with the D15 p.c. *Cd32*⁺ fibroblast population, therefore this may be a resident population within the murine SG which does not appear as a consequence of inflammation. The second D0 fibroblast *Mfap*⁺ population mapped across to the *Ackr3*⁺ *Cd55*⁺ D15p.c. fibroblast population with minimal correlation to the other fibroblast populations in the D15p.c. mouse dataset. Lastly, the D0 murine *Vcam-1*⁺ population presented minimal correlation with the *Ackr3*⁺ *Cd55*⁺ D15p.c. fibroblasts; however, the *Vcam-1*⁺ population shared overlap with three of the 4 populations within the D15p.c. fibroblast populations. This suggests that this population may expand to form a part of the additional populations seen at D15p.c.

Following on from the observed fibroblast data in the scRNAseq dataset, we then went on to explore whether the population of CD34⁺ Pdpn⁺ fibroblasts could be identified in the wild type (Wt) murine SG by flow cytometry (**Fig 5 A-C**). We firstly confirmed that we were demonstrating true CD34 expression by using the mouse colon as a positive control (**Fig 5 A**), as clear segregated CD34 expression was demonstrated and investigated by *Stzepourginski et al* in the context of intestinal stromal cells in the presence and absence of colonic inflammation [203]. Following this, we were able to successfully identify the expression of

PDPN⁺ CD34⁺ fibroblasts which were increased in proportion in the D15 p.c. murine SG suggesting this population was expanded in response to inflammatory cues. This confirmed our groups previous findings of the presence of PDPN⁺ CD34⁺ cells within the human pSS SG suggesting that the expression of CD34 in the fibroblast compartment overlapped in both human and murine SG tissue [8].

Having identified the subsets of murine fibroblasts in both conditions of the murine SG, we explored the potential overlap with the populations we have described in the human SG, a concept which has also been explored by *Beuchler et al* [162]. To address this, we used a pseudo bulk correlation approach of our human pSS CD34⁺ CD146⁺ cells and the inflamed D15 p.c. fibroblast scRNAseq dataset and as summarised in the table (**Fig 6**). Our dataset demonstrated certain alignment and similarities among the human pSS SG fibroblasts and the fibroblasts isolated at D15 p.c. from the murine SG. The first human pSS SG population of ABCC9 fibroblasts demonstrated an overlap with the murine D15p.c. *Cxcl10*⁺ and *Plxna2*⁺ fibroblast, while not presenting similarities in terms of gene expression with the murine *Cd32*⁺ fibroblast. The human pSS FOSB⁺ CDKN1A⁺ subset, demonstrated similarities with the murine D15p.c. *Ackr3*⁺ *Cd55*⁺ and the *Plxna2*⁺ fibroblasts, and a weaker similarity with the murine *Cxcl10*⁺ inflammatory fibroblast population. Contrastingly, the human pSS *ACKR3*⁺ *CD55*⁺ and murine D15 p.c. *Ackr3*⁺ *Cd55*⁺ populations showed the greatest overlap in gene signature, suggesting that these two fibroblast populations might play similar roles across the two species. Similarly, the human pSS *CCL19*⁺ *TNFSF13B*⁺ fibroblast population showed similar gene signature to that of the murine *Cxcl10*⁺ fibroblast population also suggesting conservation of the *immunofibroblast* population across different species at similar permissive sites of TLS formation. Lastly, the human pSS *Serpine2*⁺ *F2R*⁺ fibroblast population

demonstrated strong correlation with all 4 murine D15 p.c. populations but had the greatest correlation with the *Cxcl10*⁺ population. This observation implies that there may be overlap in function between fibroblast populations amongst species.

Altogether, these data demonstrate the presence of 3 populations of fibroblasts within the resting murine SG, which at D15 p.c. expands to 4 populations and one of which aligns with the *immunofibroblast* phenotype. Moreover, features of specific populations of fibroblasts within the human pSS SG can be identified within the inflamed SG of the murine TLS model. This suggests the presence of an overlap of the role of fibroblasts in the pathophysiology of disease at these anatomical sites that also cross the species boundary too.

Figure 3: The Fibroblast Cell Populations at D0

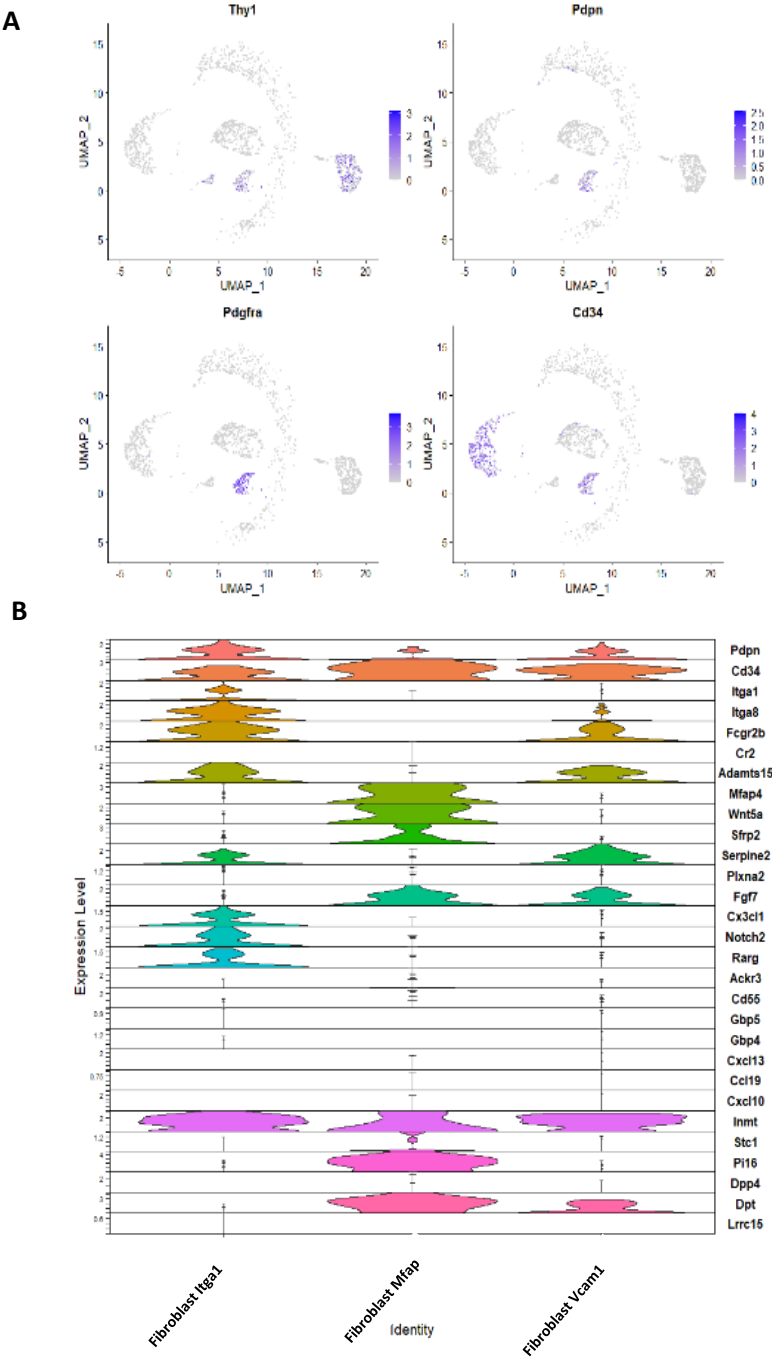


Figure 3 A-D: Characterisation of fibroblasts at Day 0 and Day 15 p.c. in the murine SG. A) Feature plot identifying the fibroblast population using *Thy1*, *Pdpn*, *Pdgfra* and *Cd34* within the global UMAP. **B)** Stacked violin plots of unique identifying genes of the fibroblast populations.

Figure 3: Fibroblasts Cells Day 15 p.c.

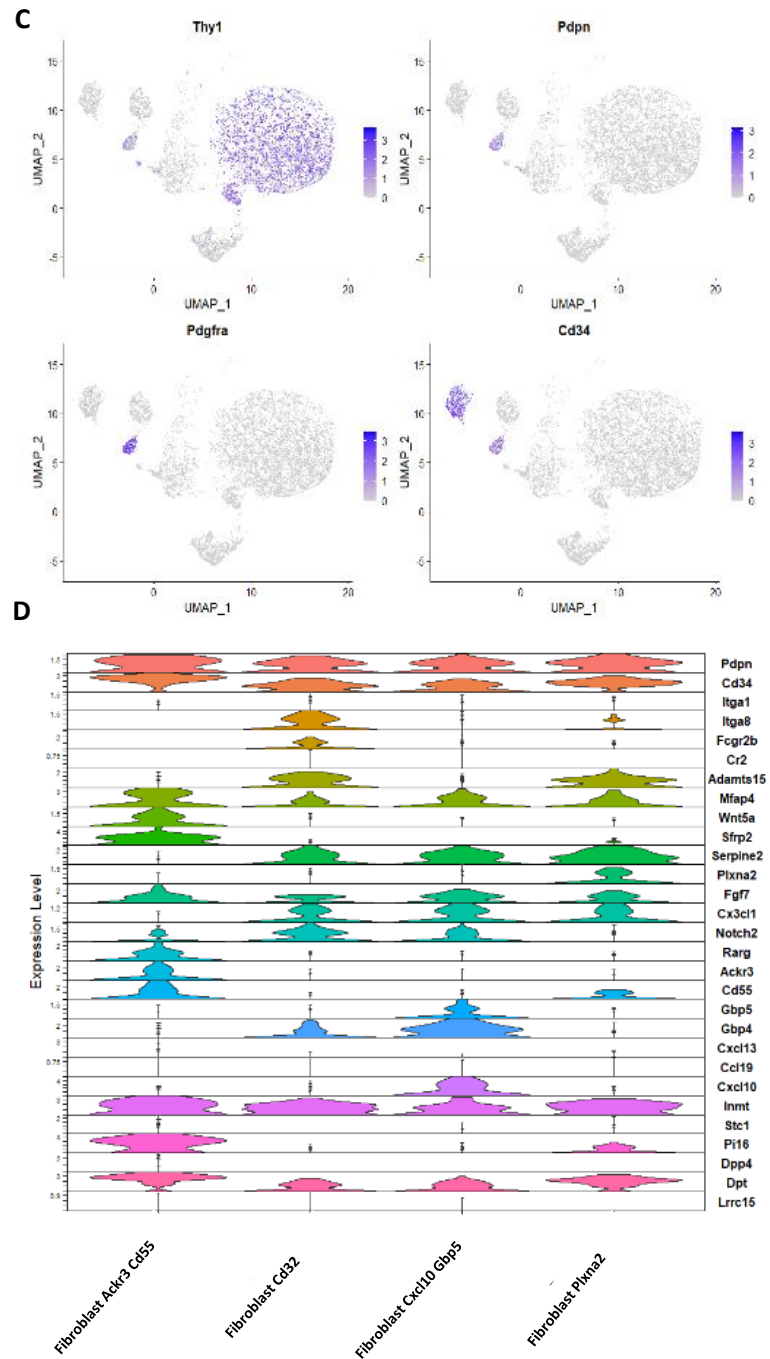


Figure 3 A-D: Characterisation of fibroblasts at Day 0 and Day 15 p.c. in the murine

SG. C) Feature plot of *Thy1*, *Pdpn*, *Pdgfra* and *Cd34* expression in day 15 p.c. murine

SG. D) UMAP of 3 different subclusters of fibroblasts populations at Day 15 p.c.

Figure 4: Comparison of the fibroblast population between D0 and D15 p.c.

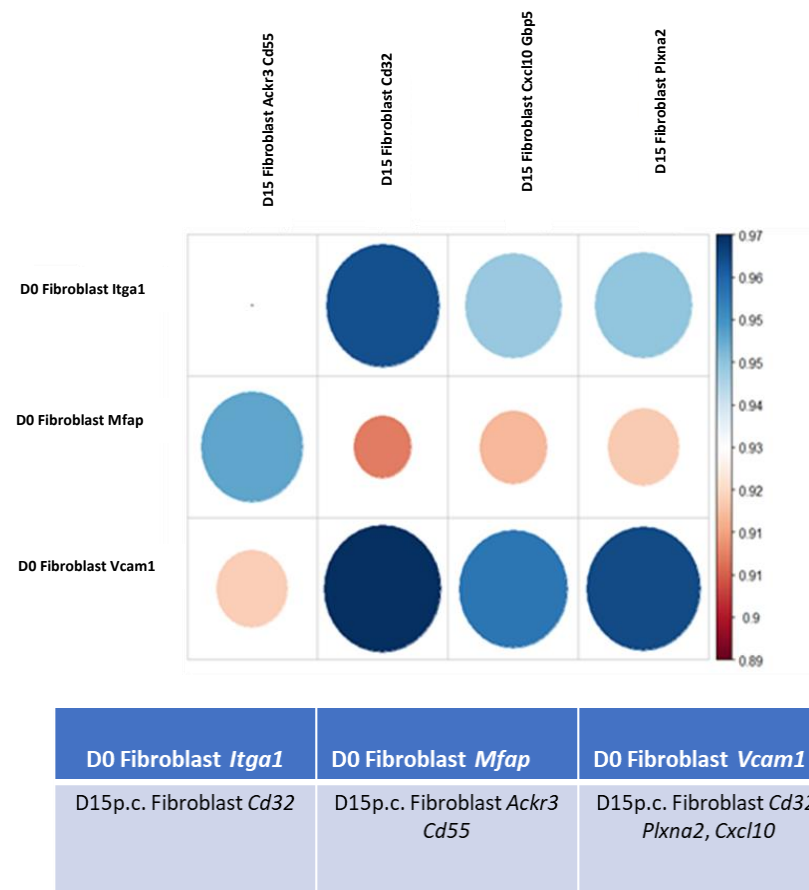
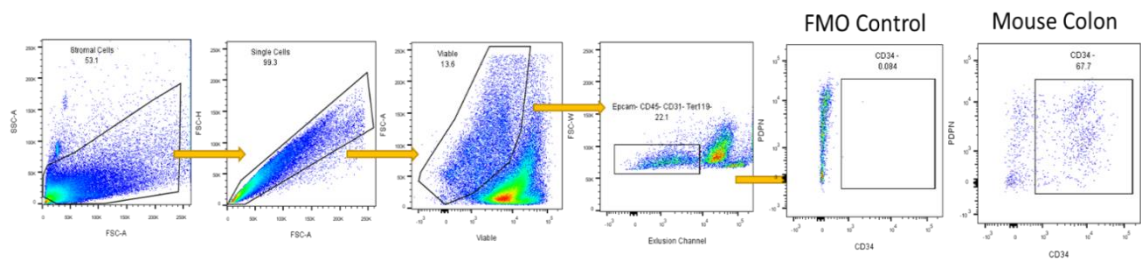


Figure 4: Pseudo bulk correlation of D0 and D15 p.c. fibroblast subsets captured within the scRNAseq dataset. Exploration of the fibroblast population overlap in differing experimental conditions using Pearson’s correlation. Scale bar indicates strength of correlation: 0.89 indicates a weak correlation between populations whilst 0.97 indicates a strong correlation between populations. Table summarises D0 murine fibroblast populations with the corresponding aligning D15p.c. fibroblast populations

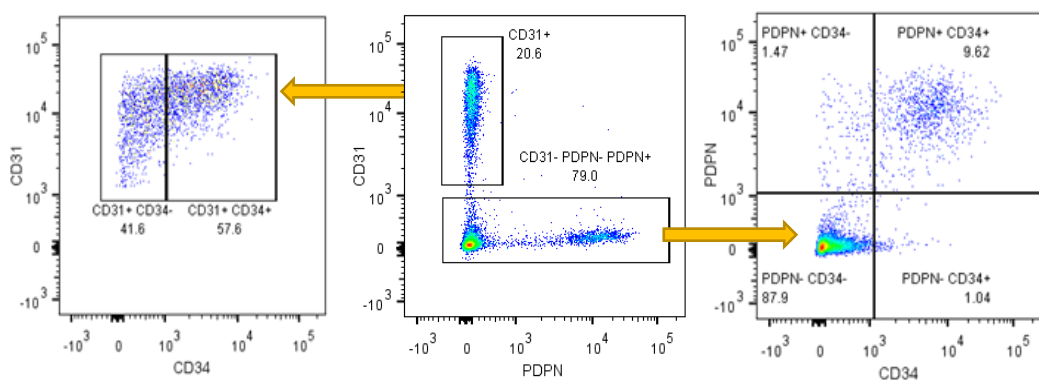
Figure 5: *In vivo* validation of PDPN⁺ CD34⁺ Fibroblasts in the mouse SG

A



Day 0 resting WT murine SG

B



Day 15 p.c. inflamed WT murine SG

C

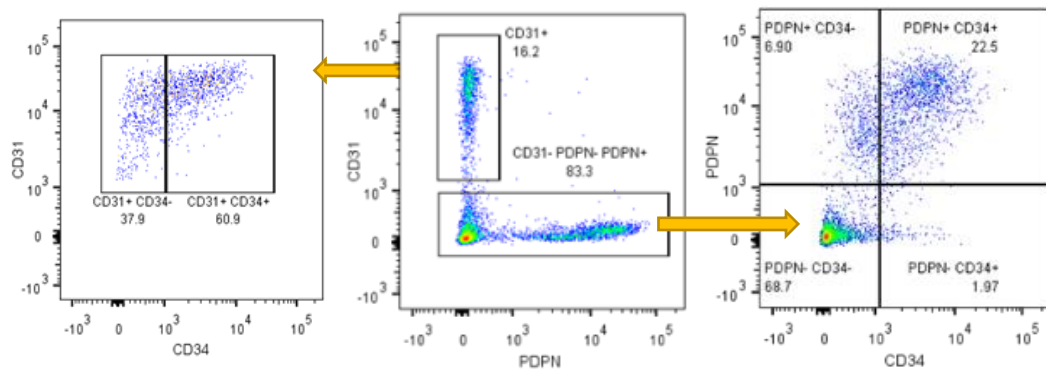
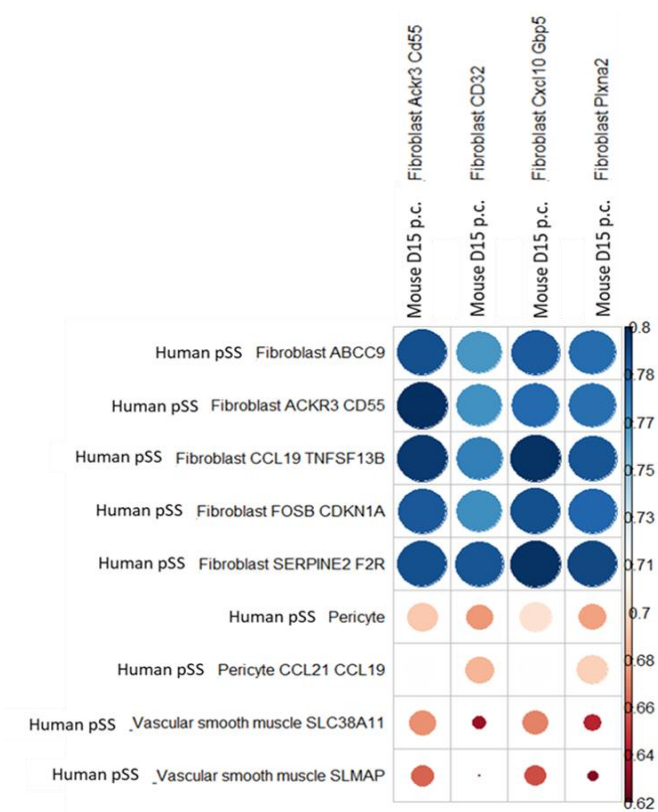


Figure 5: A-C: Representative flow cytometry plots of PDPN⁺ CD34⁺ fibroblasts in the mouse SG at D0 and D15p.c.

A) Mouse colon was used as a positive control to validate the expression of CD34 using the Ram34 clone APC antibody. **B)** Dot plots of the PDPN⁺ CD34⁺ populations in the **D0** mouse SG within the PDPN⁻ CD31⁺ endothelial population and the PDPN⁺ CD31⁻ fibroblast population. **C)** Dot plots of the PDPN⁺ CD34⁺ populations in the **D15 p.c.** mouse SG within the PDPN⁻ CD31⁺ endothelial population and the PDPN⁺ CD31⁻ fibroblast population

Figure 6: Comparison of the fibroblast population between human pSS SG and D15 p.c mouse SG



Human Fibroblast Populations	Mouse D15p.c. Fibroblast Populations
Fibroblast <i>ACKR3 CD55</i>	Fibroblast <i>Ackr3</i>
Fibroblast <i>CCL19 TNFSF13B</i>	Fibroblast <i>Cxcl10 Gbp5</i>
Fibroblast <i>SERPINE2 F2R</i>	Fibroblast <i>Cxcl10 Gbp5</i>

Figure 6: Pseudo bulk correlation of human pSS CD34+ CD146- fibroblasts and murine D15 p.c. fibroblast subsets captured within the scRNAseq dataset. Exploration of the fibroblast population overlap in differing experimental conditions using Pearson’s correlation. Scale bar indicates strength of correlation: 0.62 indicates a weak correlation between populations whilst 0.8 indicates a strong correlation between populations. Table summarises the fibroblast populations which map across from the human and mouse fibroblast dataset.

4.4: Pericytes of the murine SG

Within the human pSS SG the pericytes clustered in two discreet populations, one of which was demonstrated to be able to produce the lymphoid chemokines CCL19 and CCL21. Due to the small number of cells recovered from murine SG and the limited ability of our system to capture lymphoid chemokine expression, we were unable to capture the CD34⁻ CD146⁺ CCL21⁺ population in the murine single cell dataset. However, we were able to identify CD34⁻ CD146⁺ CCL21⁺ population at the protein level within the inflamed murine D15 p.c. SG.

Although we did not specifically capture the *immunopericytes* in the murine scRNAseq dataset, we did identify pericyte populations both within the resting and the inflamed murine SG (**Fig 7 A-D**). At D0 we identified one population of pericytes defined by the expression of the *Mcam* (*Cd146*), *Rgs5*, *Alpha actin 2* (*Acta2*), and *Platelet-derived growth factor receptor beta* (*Pdgfrb*) which did not have expression of any lymphoid chemokines or pro-inflammatory cytokines as previously disclosed (**Fig 7 A & B**). Contrastingly, at D15 p.c. we captured 2 population of pericytes within the murine SG using the same canonical markers as at D0 but with unique differential expression of specific genes (**Fig 7 C & D**). The two populations of pericytes at D15p.c. firstly differed on their expression of *Cd146*, as the first population did not express this cell adhesion marker but instead expressed *Gamma-adducin* (*Add3*), whilst the second population did. This suggests that the second pericyte population may have acquired *Cd146* expression and may arise as an independent entity from the *Cd146*⁻ population in response to inflammatory cues (**Fig 7 D**). Additional genes which differentiated the two populations included *Endothelial PAS domain-containing protein 1* (*Epas1*) and *Notch1* which were both present in the CD146⁺ population but not in the CD146⁻ pericyte cluster (**Fig**

7 D). Interestingly, *Notch2* was not expressed by either population, demonstrating a preferential expression for a particular Notch gene.

Whilst the scRNAseq data was unable to capture the pericyte population to the same transcriptomic depth as in the human pSS SG, the presence of CD146⁺ cells has been demonstrated by other groups [290, 291]. Given our previous data in the human pSS SG scRNAseq dataset, pointed to the presence of a pericyte population able to express chemokines, in absence of a clear NFκB mediated engagement, we aimed to explore the features of these pericyte populations *in vivo* using our murine TLS model (**Fig 8 A-E**).

Firstly, we explored the presence of CD146⁺ pericytes at the protein level within the resting murine SG using flow cytometric analysis (**Fig 8A**) before proceeding with the exploration of this population at D15 p.c. (**Fig 8 A-E**). We identified the presence of CD146⁺ CCL21⁺ pericytes within the SG of Wt mice but were unable to use the markers TNC or CD36, to identify distinct populations in a manner identical to that of the pericytes within the human dataset. We also aimed to evaluate the presence and functional ability of these pericytes to express CCL21 in LTβR^{-/-} mice, as our human dataset suggested that the ability of this set of pericytes to express chemokines was independent from this pathway. Similarly, whilst were unable to capture the expression of *Ccl21* in the murine scRNAseq data set, previous studies have demonstrated that murine SG fibroblasts are able to upregulate the expression of this lymphoid chemokine in response to inflammation [8]. We firstly aimed to confirm the presence of the PDPN⁺ CD34⁺ fibroblasts previously described to produce CCL21 and CCL19 in murine TLS (**Fig 8 A-E**). As expected, we observed the presence of CD34⁺ CCL21⁺ fibroblasts in WT mice. This population was significantly reduced in LTβR^{-/-} suggesting either the reliance of CCL21

production on LT β R or the need for this signalling to drive the formation of this population **(Fig 8C)**. Intriguingly, we observed an increased proportion of CD146⁺ CCL21⁺ pericytes in the LT β R^{-/-} mice compared to the WT mouse, suggesting that the expression of this chemokine and/or the development of this population was governed by an alternative upstream signalling cascade independent of LT β R signalling **(Fig 8 D-E)**. Overall, our data demonstrates that despite the limitation of our mouse RNAseq dataset, we were able to identify this pericyte population *in vivo* using the pSS mouse model. This confirms that this population is present at D15p.c. within the murine SG and appears to be differentially regulated by an alternative signalling cascade independent of the LT β R which is yet to be explored.

Figure 7: The pericyte Cell Populations at D0 in the murine SG

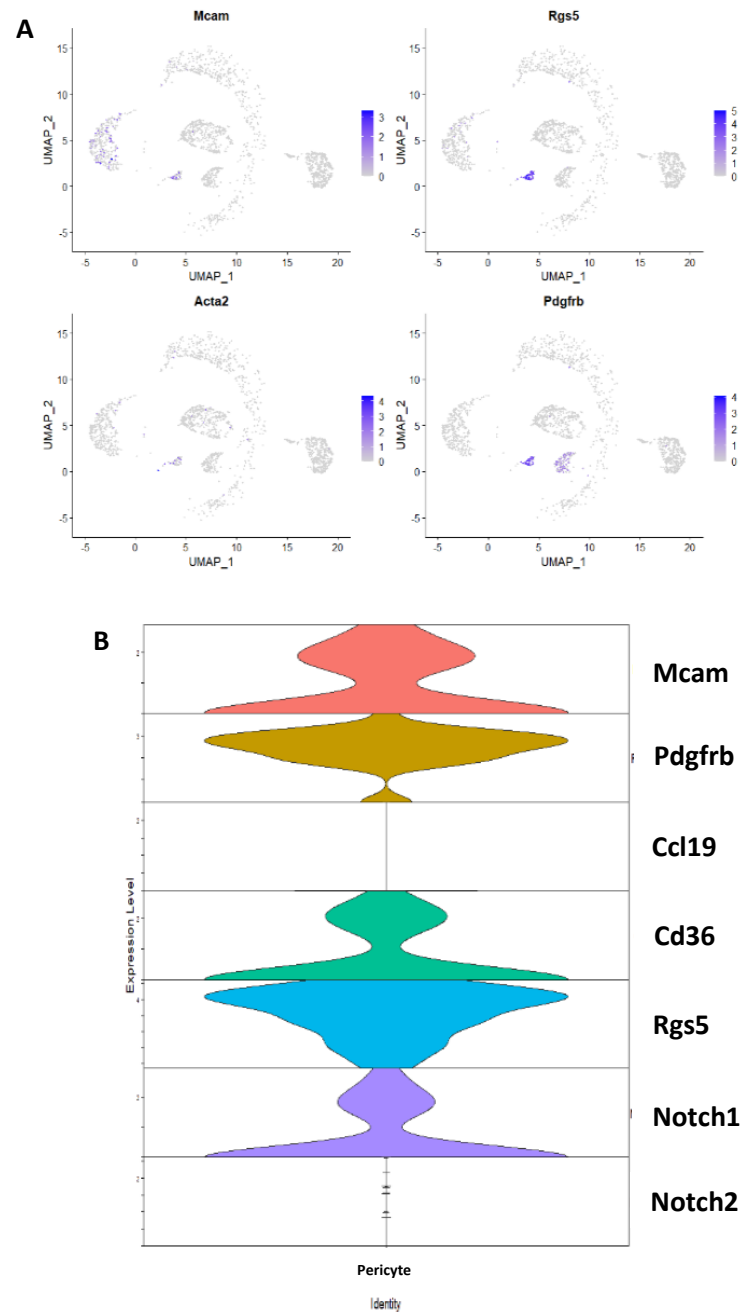


Figure 7: Characterisation of fibroblasts at Day 0 and Day 15 p.c. in the murine SG. A) Feature plot identifying the fibroblast population using *Thy1*, *Pdpn*, *Pdgfra* and *Cd34* within the global UMAP. **B)** Stacked violin plots of unique identifying genes of the fibroblast populations at Day 0.

Figure 7: The pericyte Cell Populations at D15 p.c in the murine SG

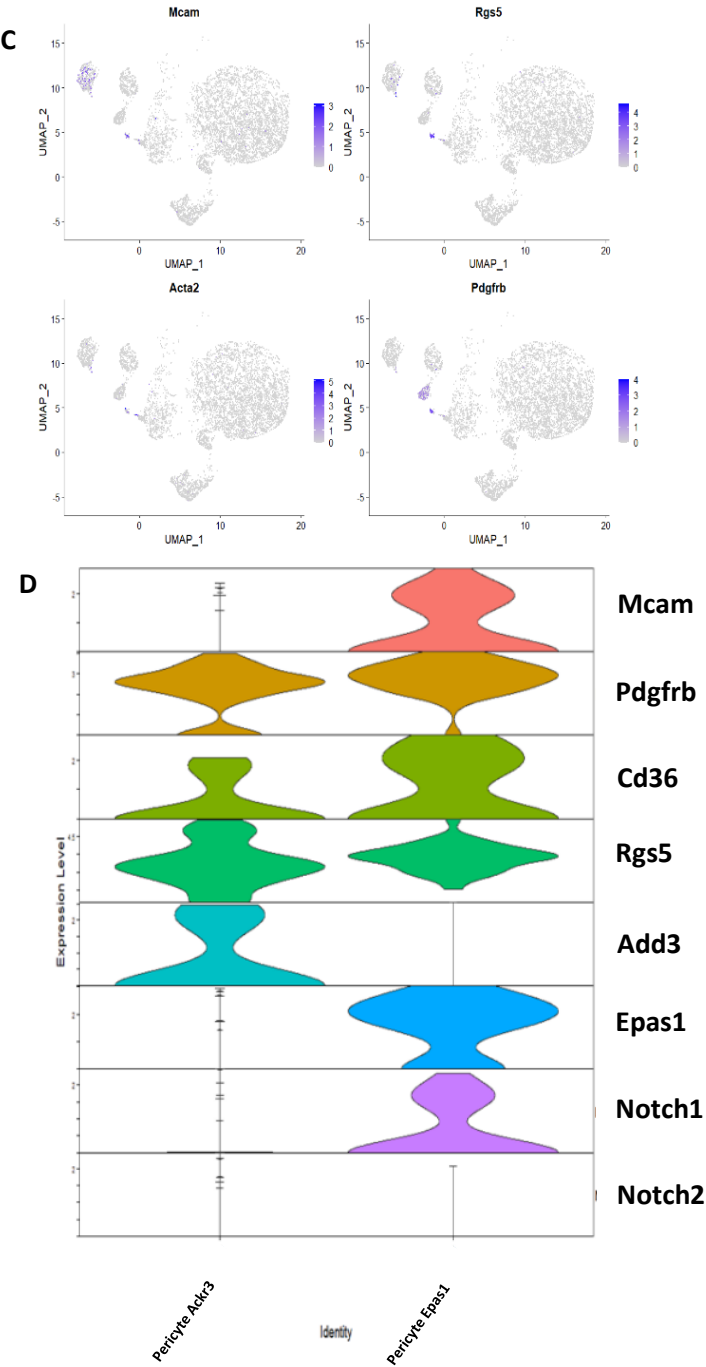


Figure 7: Characterisation of fibroblasts at Day 0 and Day 15 p.c. in the murine SG. C) Feature plot of *Thy1*, *Pdpn*, *Pdgfra* and *Cd34* expression in day 15 p.c. murine SG. *D)* UMAP of 3 different subclusters of fibroblasts populations at Day 15 p.c.

Figure 8: Validation of CD34⁺ CD146⁺ CCL21⁺ pericytes in the D15 p.c. murine SG *in vivo*

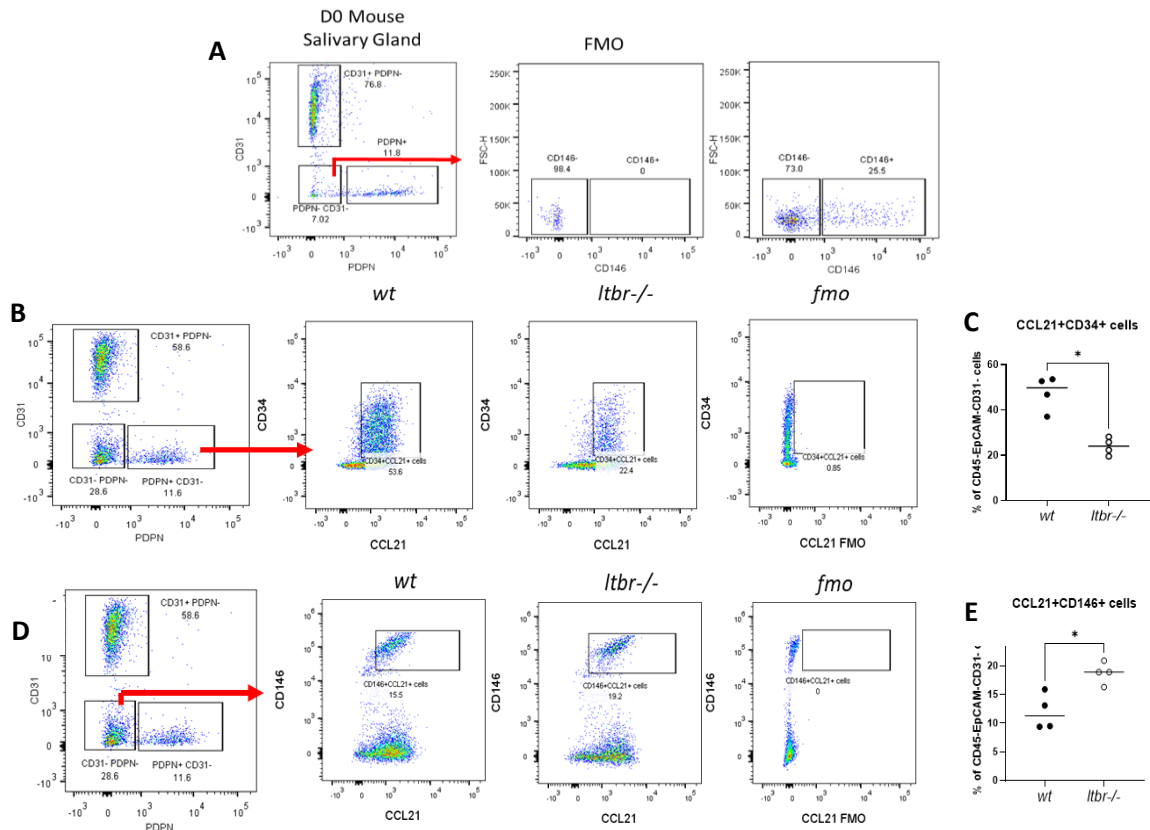


Figure 8: Characterisation of CD34⁺ CD146⁺ CCL21⁺ pericytes at Day 0 and Day 15 p.c. in the murine SG. A) Validation of CD34⁺ CD146⁺ pericytes in the mouse SG at D0 using flow cytometry analysis. **B)** Representative flow cytometry dot plots of D15p.c. CD34⁺ CD146⁺ CCL21⁺ in the murine SGs of both WT and *Ltbr*^{-/-} mice. **C)** The proportion of D15p.c. CD34⁺ CD146⁺ CCL21⁺ cells in WT compared with *Ltbr*^{-/-} demonstrated a significantly increased proportion of CCL21⁺ pericytes in the *Ltbr*^{-/-} mice compared to WT mice (*n*=4 for each time point, each dot is a biological repeat, * = *p* ≤ 0.05, ** = *p* ≤ 0.01, *** = *p* ≤ 0.001, ns = *p* > 0.05. Statistical test was a student's T test). **D)** Representative flow cytometry dot plots of D15p.c. CD34⁺ CD146⁺ CCL21⁺ in the murine SGs of both WT and *Ltbr*^{-/-} mice. **E)** The proportion of D15p.c. CD34⁺ CD146⁺ CCL21⁺ cells in WT compared with *Ltbr*^{-/-} demonstrated a significantly increased proportion of CCL21⁺ pericytes in the *Ltbr*^{-/-} mice compared to WT mice (*n*=4 for each time point, each dot is a biological repeat, * = *p* ≤ 0.05, ** = *p* ≤ 0.01, *** = *p* ≤ 0.001, ns = *p* > 0.05. Statistical test was a student's T test).

4.5: Endothelium Population of the murine SG

Aside from the contributions of the mural cells, endothelial cells form another population of stromal cells, which may contribute to TLS heterogeneity with their functions of facilitating lymphocyte migration within tissue. Traditionally, the endothelium within organs has varied functions comprising of regulation of vascular permeability, angiogenesis, leucocyte trafficking and immunity. The contributions of the endothelium in inflammation are governed by the presence of different signalling factors such as G protein coupled receptors (GPCRs) and pro-inflammatory cytokines TNF and interleukin 1 (IL-1) [253, 292]. Moreover, in response to LT β R signalling, resident endothelial cells have been demonstrated to acquire the phenotype of high endothelial venules (HEVs) which consequently support the formation of TLS within the inflamed tissue by facilitating lymphocyte trafficking into the tissue, thus also warranting their exploration in this context [100].

Within the mouse SG, we identified endothelial cells in both the resting and inflamed gland using the classical marker *Pecam-1* (**Fig 9 A & C**). In the resting gland, 3 populations of endothelial cells were characterised (**Fig 9 B**). These 3 clusters had unique expression of a cassette of genes which allowed us to infer their possible function. The first cluster of cells had expression of *Carbonic anhydrase 4* (*Car4*). The second cluster expressed *Semaphorin-3G* (*Sema3g*), *Transforming growth factor beta-2* (*Tgfb2*), *Claudin 6* (*Cldn6*) and *Adhesion G protein-coupled receptor E5* (*Agdre5/ Cd97*) suggesting this endothelium population to have role in facilitating lymphocyte infiltration and maintaining vascular homeostasis. The last cluster of cells expressed genes *von Willebrand factor* (*Vwf*), *Vcam-1* and *Ackr1* suggesting this population to also participate in the modulation of the chemokine gradients within the SG

microenvironment which consequently governs the migration of lymphocytes within the murine SG (**Fig 9 B**).

Similarly, at D15 p.c. we also identify 3 populations of endothelial cells within the murine SG (**Fig 9 D**). In this instance, some gene markers overlapped with populations identified at D0 such as *Sema3B* & *Tgfb2* (**Fig 9D**) which suggests that this population is resident within the SG independent of inflammation. Of the remaining 2 endothelial populations at D15 p.c. one was specifically identified with a pro-inflammatory signature of *MHC class II-associated invariant chain* (*Cd74*), *Cxcl9*, *Cxcl10* and *Vcam-1*, suggesting this population is immunomodulatory and activated with the expression of components of the antigen presentation cascade, lymphocyte chemokines and adhesion molecules. The remaining population at D15 p.c. had unique expression of *Car4* and *Ackr3* suggesting this population may also be involved in governing the modulation of the chemotactic gradient and the metabolic signature within the SG (**Fig 9 D**).

Overall, our data demonstrates that in perhaps a similar manner to the activation of resident fibroblast populations, a specific immunomodulatory endothelial population arises, or becomes activated from a resident endothelial cell population. This may facilitate the infiltration of lymphocytes from the circulation due to the expression of chemo attractants and cell adhesion markers. Aside from fibroblasts, mural cells and endothelial cells, the epithelial cell population has also been described to participate modulating the SG in pSS for the formation of TLS.

Figure 9: The Endothelium Cell Populations in D0 murine pSS SG

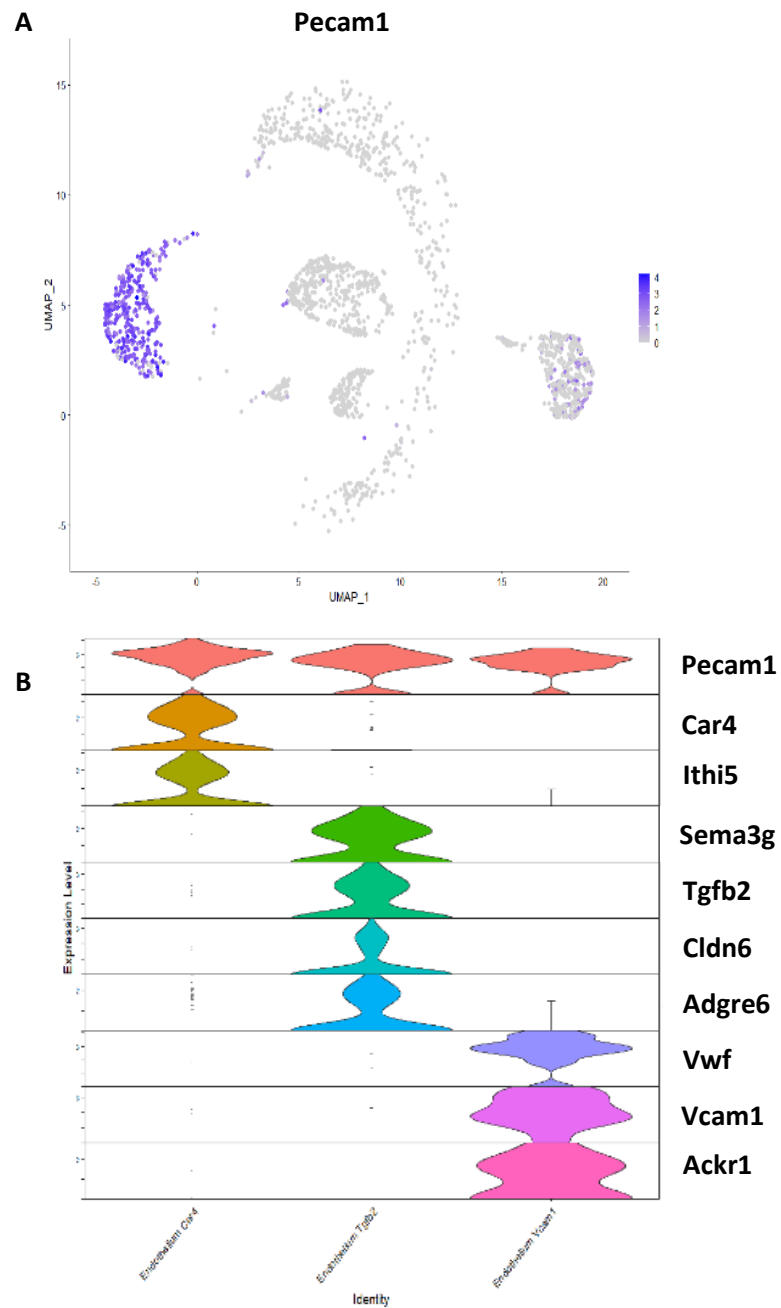


Figure 9: Characterisation of endothelial cells at Day 0 and Day 15 in the murine SG. A) Feature plot identifying the endothelial population using *Pecam-1* within the global UMAP. **B)** Stacked violin plot of unique identifying genes of the endothelial populations. **C)** Feature plot of *Pecam-1* expression in day 15 p.c. murine SG. **D)** Stacked violin plot of unique identifying genes of the endothelial subsets at Day 15 p.c.

Figure 9: The Endothelium Cell Populations in D15p.c. murine pSS SG

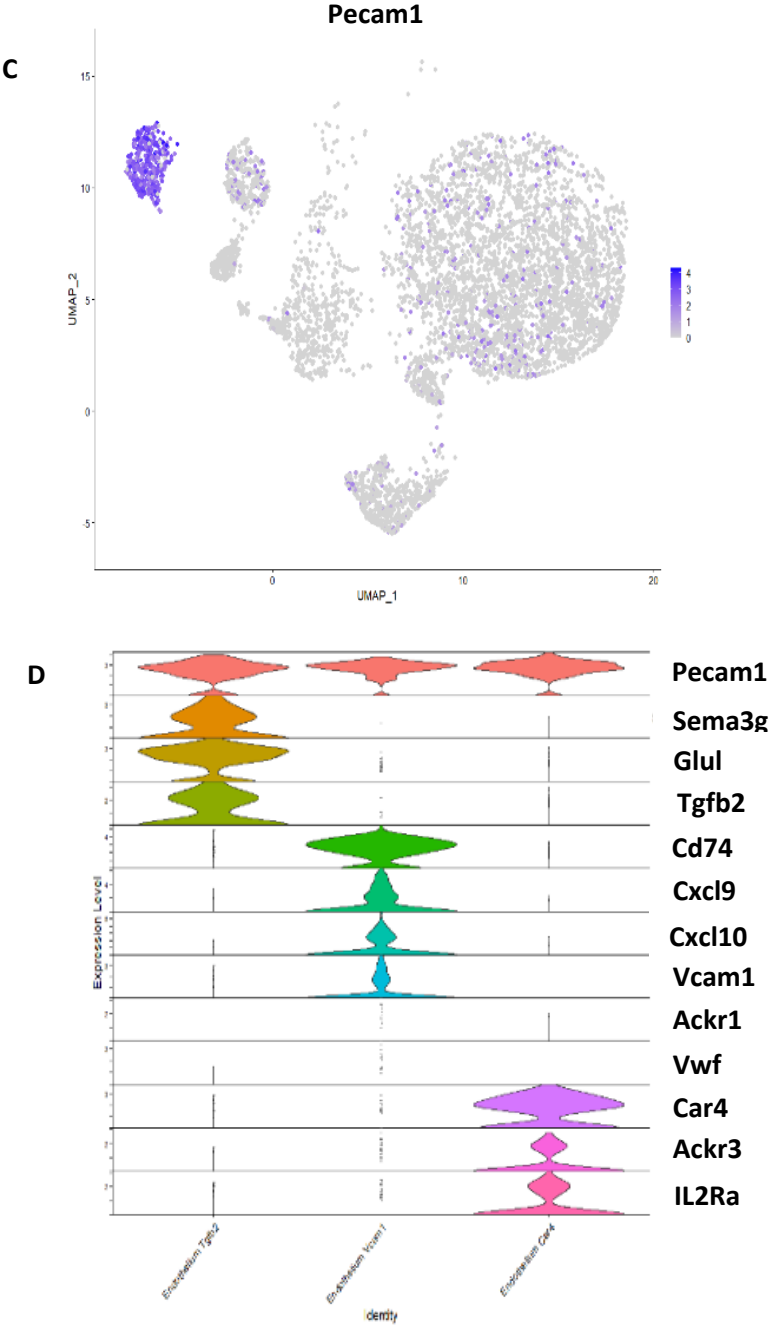


Figure 9: Characterisation of endothelial cells at Day 0 and Day 15 in the murine SG. A) Feature plot identifying the endothelial population using *Pecam-1* within the global UMAP. **B)** Stacked violin plot of unique identifying genes of the endothelial populations. **C)** Feature plot of *Pecam-1* expression in day 15 p.c. murine SG. **D)** Stacked violin plot of unique identifying genes of the endothelial subsets at Day 15 p.c.

4.6: Epithelium Population of the murine SG

In our murine SG scRNAseq dataset, we identified 6 populations of epithelial cells at Day 0, and 4 populations within the inflamed murine SG (**Fig 10**). All of the epithelial cell populations expressed the epithelial marker *Epcam* in combination with the expression of unique set of markers. At D0 the epithelial clusters uniquely expressed a selection of genes assigning them potentially different functions (**Fig 10 B**). The first cluster of epithelial cells expressed the transcriptional repressor *Achaete-scute homolog 3 (Ascl3)* and *Asialoglycoprotein receptor 1 (Asgr1)* which has been associated with the endocytosis of plasma glycoproteins [293]. The second cluster was identified with the expression of *Cell growth regulator with EF hand domain protein 1 (Cgref1)* which participates in regulating cell-to-cell adhesion [294]. The third cluster was identified by the expression of *Meg3* – a currently uncharacterised gene and protein, and the pyruvate dehydrogenase *Pdp2*. The fourth cluster expressed the gene *Exocrine gland secreted peptide 4 (Esp4)* and *Doublecortin-like kinase 1 (Dclk1)* which has been attributed to participate in the maturation of the nervous system. A fifth cluster of epithelial cells had unique expression of *Potassium voltage-gated channel subfamily E member 1 (Kcne1)* which in the context of the epithelial aligns with their function to maintain the optimum pH and osmotic balance within the tissue. The final cluster expressed *Beta-nerve growth factor (Ngf)*, and *Wnt4* which suggests this population to have a stem-cell, plasticity like signature.

Contrastingly at D15 p.c. we identified 4 clusters of epithelial cells one of which was lacking the expression of *Epcam* (**Fig 10 D**). In this instance, the unique genes identifying the first *Epcam*⁺ cluster included *Major urinary protein 6 (Mup6)* and *Olfactomedin-like protein 3 (Olfml3)* which has been associated to have a developmental role [295, 296]. The second

cluster of epithelial cells included the expression of *Esp4* which we had also identified within an epithelial cell cluster at Day 0, and *Desmoyokin (Ahnak)* which has been found to be expressed in the plasma membrane of epithelium [297]. The third Epcam⁻ epithelium expressed the gene for *CDK5 regulatory subunit-associated protein 1-like 1 (Cdkal1)* which has been suggested to participate in EMT in the context of cancer [298] The final Epcam⁺ cluster was identified by the expression of *FXVD domain-containing ion transport regulator 2 (Fxyd2)* which functions to support ion transportation within the epithelium.

This data within the murine SG has identified the presence of epithelial populations which serve their canonical functions based on the gene expression profiles they have demonstrated.

Figure 10: The Epithelium Cell Populations in D0 murine SG

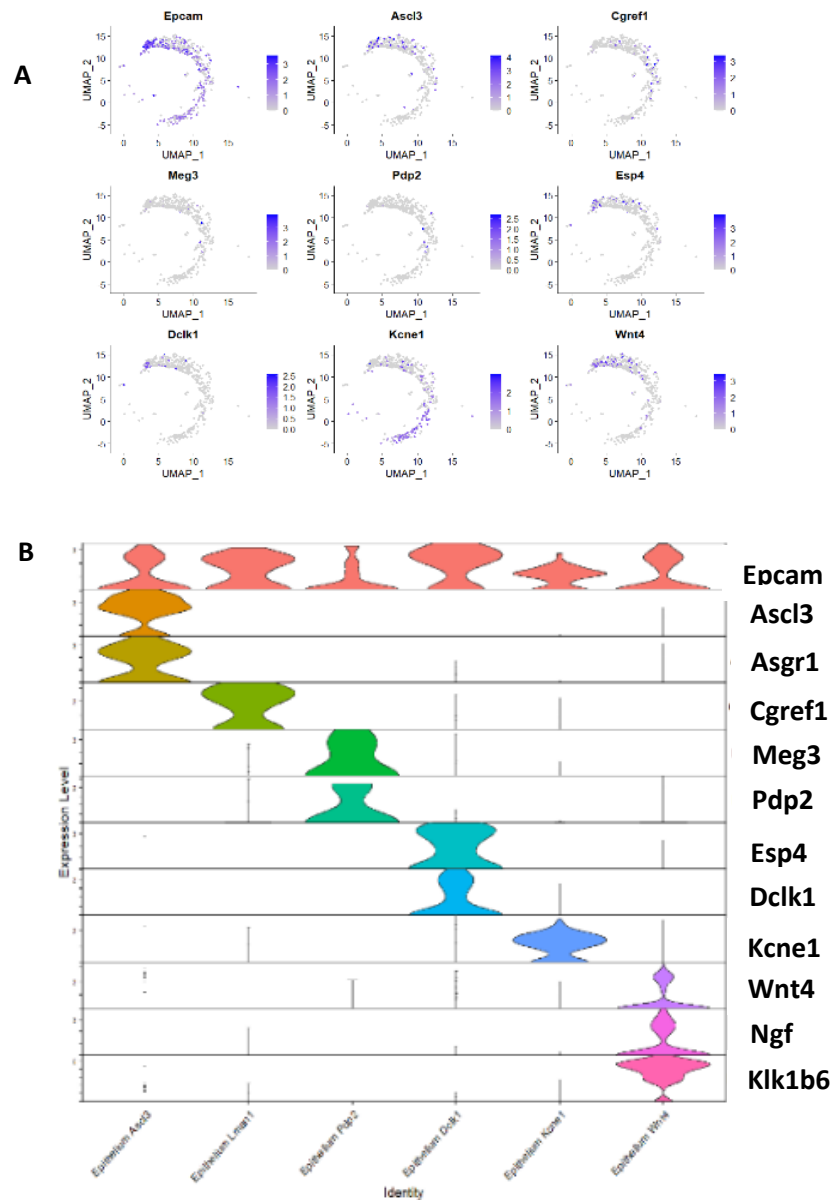


Figure 10: Characterisation of epithelial cells at Day 0 and Day 15 p.c. in the murine SG. A) Feature plot identifying the monocyte population using Epcam gene expression within the global UMAP. **B)** UMAP of the 6 subclusters of the epithelium populations at Day 0.

Figure 10: The Epithelium Cell Populations in D15p.c. murine SG

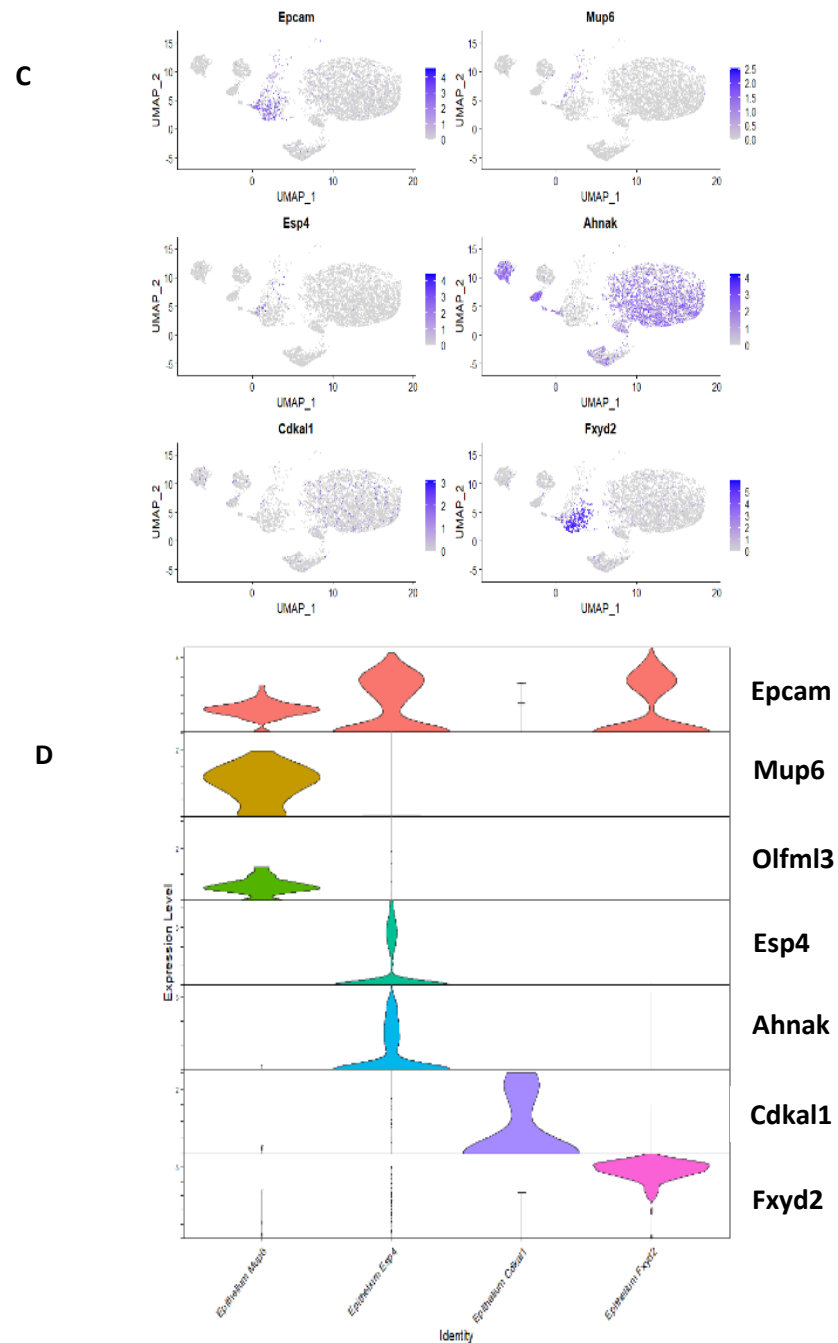


Figure 10: Characterisation of epithelial cells at Day 0 and Day 15 p.c. in the murine

SG.. C) Stacked violin plot of unique identifying genes of the epithelial population.

D) Feature plot of *Epcam* expression in day 15 p.c. murine SG.

Chapter 4: Discussion

The murine mouse model of pSS has previously enabled us to investigate TLS formation and development in a functional manner to characterise the various different pro-inflammatory cytokines produced at the different developmental stages. This has allowed us to profile the phenotypical and functional elements of the underlying *immunofibroblast* compartment [8]. However, our *in vivo* approaches have not enabled us to completely characterise the heterogeneity underpinning a fully formed TLS. To this end, we have performed single cell sequencing on digested SG isolated from three resting (D0) and three mice cannulated with an AdV5 adenovirus at day 15 post cannulation (D15p.c.). This later time point was selected on the basis of previous observations pointing to the TLS formed by D15p.c. in murine SG [8]. This material generated a dataset comprising of tissues of two different states which allowed us to compare differences and similarities in terms of cellular distribution in resting and inflamed conditions.

We have been able to identify immune cell and stromal cell populations at both time points using established canonical markers. As expected, the inflamed SG demonstrated increased presence of immune cell populations such as myeloid cells, NK cells, ILCs, B cells, PC, and T cells which is expected in response to inflammatory cues. The expected trends in response to inflammation that we have seen have included the expansion of specific populations which acquire the required functional phenotypes in response to virally induced inflammation. For the purposes of this work, we have not discussed the lymphoid populations I have and kept the focus on the heterogeneity of the fibroblasts, pericytes, endothelial and epithelial cells. This allowed us to not only identify specific fibroblast subsets, but also to compare fibroblast

populations between our human pSS SG dataset and inflamed murine D15 p.c. SG. Naturally, we identified the classical stromal cells such as endothelial cells, epithelium cells, fibroblasts and pericytes. However, we were not able to identify Schwann cells or vascular smooth muscle populations in the murine dataset, this may be due to a processing error which led to the inability to capture this cell population in this dataset. We were also unsuccessful in the capture of the lymphoid chemokine genes which we have previously demonstrated to be expressed by the fibroblasts in our *in vivo* work [8].

The fibroblast compartment of our mouse scRNAseq dataset comprised of 3 populations within the resting gland as compared to 4 populations within the D15p.c. inflamed gland [8]. Within our fibroblast populations and, in particular in the CD34⁺ subset, we were unable to detect the expression of lymphoid chemokines, this was attributed to the technical limitation of our assay. Nonetheless, we were able to assign some functional identity to the fibroblast populations by applying the gene signatures demonstrated by *Rodda et al* in their characterisation of the stromal cell heterogeneity within the murine LN [89]. Furthermore, given the available scRNAseq data from both resting and inflamed tissue, we also had the possibility to explore the gene cassette demonstrated by *Beuchler et al*, in which a universal fibroblast population identified by *Pi16* and additional genes such as *Dpp4*, *Dpt* and *Lrrc15* was described [162]. Within our resting tissue we identified a unique cluster expressing *Pi16* alongside genes associated with the Wnt pathway such as *Mfap4*, *Sfrp2* and *Wnt5a*, suggesting a potential identity of this population as resident progenitor, pro-repair population. The Wnt pathway is well known to associate with developmental and cellular plasticity; moreover, studies of murine SG have previously identified the involvement of Wnt pathways in the recovery of SG damage[215-217, 299-302]. The two remaining populations

identified at D0 demonstrated similar gene signatures which included the expression of the integrins *Itga1* and *Itga8*, the low affinity receptor *Fcgr2b*, the metalloproteinase *Adamts15* and the methyl transferase *Inmt*. Collectively, this gene signature suggests the possibility that these fibroblasts may participate in the maintenance of the ECM, a classical function associated with fibroblasts prior to the unveiling of their immunological competency. The distinguishing genes between these fibroblast subsets included the expression of the genes *Cx3c1*, *Notch2* and *Rarg*. The chemokine *Cx3c1* binds to its cognate receptor *Cx3cr1* which are expressed on immune cells such as NK cells suggesting this fibroblast population may engage with NK cells, detected in the SG. Whilst the expression of *Notch2* is traditionally associated to occur in mural cells, it has been demonstrated to be upregulated in the presence of retinoic acid [303]. In particular it was observed that murine mesenchymal cells of the dental cavity treated with retinoic acid, as compared with fibroblast growth factors, significantly upregulated the expression of *Notch2* [303]. This was associated with the appropriate patterning of the mesenchyme within the cavity. In murine peripheral neuropathy [304], the expression of *Cx3c1* has been shown to be driven by Notch, and a similar process might happen in the salivary glands, supporting the co-expression of these two genes in the same fibroblast population. Overall, the 3 fibroblast subsets in the non-inflamed mouse gland associate mainly with genes associated with ECM modulation and maintenance.

Four populations of fibroblasts were identified in the samples collected at D15 p.c. two of which showed similar gene pool to two populations identified at D0. Whilst the D15p.c. *Ackr3⁺ Cd55⁺* fibroblast population uniquely expressed the aforementioned genes, this population correlated in its gene signature with the D0 *Mfap⁺* population. This population had upregulated the expression of its atypical chemokine receptors and complement pathway

components. The murine *Ackr3⁺ Cd55⁺* also correlated with the *ACKR3⁺ CD55⁺* population within the human pSS SG, demonstrating the presence of this fibroblast population in the presence of an inflammatory microenvironment with a potential function to modulate the chemokine gradient of its cognate ligand CXCL11 and CXCL12 [305]. The second D15p.c. murine fibroblast population showed a significant overlap with the D0 resident *Vcam-1⁺* population and differentially expressed *Fcrg2b*. The function of this FCR2 receptor has been associated with the engulfment of IgE coated molecules [81].

The most interesting observation we found within the D15p.c. murine SG consisted of the identification of an inflammatory fibroblast population identified by the chemokine gene *Cxcl10* and the inflammasome activators *Gbp5* and *Gbp4*. These genes were also used to characterise the active TRC population of the murine LN, suggesting that this fibroblast population possess the genetic signature aligning it to a lymphoid stromal cell population [89]. Following this observation, the *Cxcl10⁺* fibroblast population appears as an independent population arising in response to inflammatory cues. From our pseudo bulk correlations we observed this *Cxcl10⁺* fibroblast population to overlap with the *Vcam-1⁺* population at D0, which suggests gene signature overlap with three of the 4 D15p.c. populations. Alignment between the human pSS *CCL19⁺ TNFSF13B⁺* population and the murine D15p.c. *Cxcl10⁺* population was also observed, suggesting that these inflammatory populations can be characterised in the same tissue of both species. The human pSS *SERPINE2⁺ F2R⁺* fibroblast population also demonstrated strong correlation to the murine D15p.c. *Cxcl10⁺* population. The last population of fibroblast we identified within the inflamed SG differentially expressed the co-receptor for the semaphorins, *Plxna2*, shown to participate in the regulation of the vascular architecture within tissue as well as governing cell migration and cell growth.

Intriguingly, this cluster also shared overlap with the D0 *Vcam-1*⁺ population and showed modest correlation with the human pSS *SERPINE2*⁺ *F2R*⁺ and *CCL19*⁺ *TNFSF13B*⁺ fibroblasts. Overall, this murine population encompassed gene features which were similar to other D15p.c. fibroblast subsets which may hint at this cell population being a transitional or highly plastic cell population.

Collectively, our data infers that there are conserved microenvironmental cues in diverse species, which induce the acquisition of the *immunofibroblasts* within murine SG harbouring TLS, supporting our effort aimed at using the murine system to address functional properties of the immunofibroblast.

Exploiting the same experimental conditions as those used to generate our transcriptomic dataset, we identified by flow cytometry on digested tissue a CD34⁺PDPN⁺ double positive fibroblast population and a CD34⁻ CD146⁺ pericyte population. As previously described [8], this population expanded in the D15p.c. SG as compared to the resting, un-cannulated, SG. Whilst our scRNAseq dataset did not capture expression of the lymphoid chemokines, previous studies exploring the lymphoid characteristics of murine fibroblasts have demonstrated the ability of the mesenchyme to upregulate gene expression of *Ccl19*, *Ccl21*, and *Cxcl13* [7, 8, 26]. We also identified a CD34⁻ CD146⁺ CCL21⁺ pericyte population. Contributions of the pericyte population have not extensively been explored in the murine TLS model, however in the context of neuronal inflammation, pericytes have been demonstrated to have immunological features such as production of pro-inflammatory cytokines IL-6, IL-1 α and chemokines such as CCL2 and CX3C1 which mediates the influx of monocytes and macrophages within the meninges [306]. Furthermore, CD146⁺ α -SMA⁺ CD36⁺ cells have been previously identified in

the PDPN⁻ CD31⁻ population of cells within tonsil lymphoid stromal cells and identified to express CCL19 and CCL21 [307]. This suggests that the pericyte population has lymphoid features as identified in the tonsil but can possibly acquire this phenotype also in non-immunological tissues, in the context of inflammation. The pericyte population in the murine TLS model have not been extensively explored, but previous data has demonstrated CCL21 expression surrounding vascular structures in LTβR^{-/-} SG histology. Due to the small cell number captured within the murine dataset and lack of overall lymphoid chemokine signature, we identified CD34⁻ CD146⁺ pericytes at both time points but were unable to align them with the human population of CD34⁻ CD146⁺ CCL21⁺. Interestingly, within the D0 resting gland we observed only one population of CD146⁺ pericytes whilst within the inflamed gland we captured two population of pericytes, as confirmed by the expression of genes such as *Rgs5* and *Pdgfrβ* but differed on the basis of *Cd146* gene expression. This demonstrates that two population of pericytes may reside within the inflamed SG and perhaps the expression of *Cd146* is acquired or lost depending on the functional or maturation state of the particular population. Aside from being differentiated by the gene expression pattern of *Cd146*, these two populations also expressed unique gene expression of the cell matrix gene *Add3* or the regulator of VEGF expression *Epas1* in combination with *Notch1*. The latter two genes were expressed concomitantly with *Cd146* and demonstrated a classical pericyte population able to engage in the regulation of vasculature development and associate with the Notch signalling cascade [308]. As previously mentioned, whilst we could not detect *Ccl19*, *Ccl21* or *Cxcl13* expression in our transcriptomic data, we explore the CD146 pericyte population *in vivo* and the expression of lymphoid chemokines in this population.

Within our human transcriptomic dataset, we observed that whilst the CD34⁺ CD146⁻ *immunofibroblast* demonstrated engagement of LTβR signalling, the CD34⁻ CD146⁺ pericyte population did not. In this manner, we hypothesised that we would observe a reduction in CCL21 expression in CD34⁺ CD146⁻ fibroblasts within LTβR^{-/-} mice and no change in chemokine expression within the CD34⁻ CD146⁺ pericyte. To functionally explore this hypothesis at protein level, we firstly identified expression of CCL21 within the PDPN⁺ CD34⁺ fibroblast population in WT mice and identified a significantly reduced expression of CCL21 expression in the PDPN⁺ CD34⁺ of LTβR^{-/-} mice. This demonstrated that CCL21 expression in the CD34⁺ fibroblast compartment was governed by the classical engagement of the LTβR signalling cascade which has been demonstrated to be critical in establishing the lymphoid phenotype during SLO formation [13, 14, 16, 87, 98, 101, 104, 109, 129, 131, 145, 213, 285]. Intriguingly, in the CD146⁺ population the CCL21 expression was greater in the LTβR^{-/-} suggesting that this chemokine expression was independent of LTβR signalling and was regulated in a different manner. This suggests that it is possible the first influx of lymphocyte infiltration maybe mediated by the pericyte derived CCL21 expression which would then allow infiltrating cells to begin remodelling the resident dormant mesenchyme such that the *immunofibroblast* phenotype can be acquired. This first phase appears to occur independently of LTβR signalling. However, successful stromal cell maturation is dependent on this signalling pathway as supported by the evidence that TLS are profoundly disorganised in LTβR^{-/-} mice. We have yet to uncover the exact signalling cascade governing the CCL21 expression pattern specifically in the CD146⁺ pericytes.

Aside from exploring the heterogeneity of the fibroblast and pericyte populations we also investigated distribution of additional populations present in the SG. During inflammation,

endothelial cells have been described to upregulate relevant lymphocyte adhesion markers and chemokines to facilitate the extravasation of immune cells from the circulation [9]. Within the mouse model, we identified the presence of 3 *Pecam-1*⁺ endothelial cells in both the resting and inflamed setting suggesting that there is an alteration of function rather than the emergence of novel populations in response to inflammatory cues. The three populations identified within the resting SG expressed genes supporting their function as explored by other groups. The first cluster differentially expressed *Car4* which associates with their function to regulate the pH within their corresponding tissue in a similar manner to the *CA4* expressing endothelium in the human scRNAseq dataset. The second D0 endothelium population expressed genes associated with governing blood vessel development due to the expression of *Sema3g*. Moreover, this subset also exclusively expressed *Tgfb2* which has been demonstrated to participate in the development and differentiation of smooth muscle vasculature which may function synergistically with *Cldn6* (*Claudin 6*) which has been observed to modulate tight junctions between cells in the vasculature [309] [310]. Intriguingly, this population also demonstrated expression of lymphocyte activating genes such as *Adgre5* (*Cd97*) which when engaged with *Cd55* results in the proliferation and differentiation of endothelial cells [311]. This suggests this population may be more responsive to external signalling cues allowing it to differentiate in order to accommodate for increased blood flow in response to inflammation. The final population of endothelial cells had a gene expression signature resonating with arterial endothelial cells. This gene set included *Vwf*, *Vcam-1* and *Ackr1* which has been associated with arterial function.

Moreover, at D15p.c. the *Sema3g* and *Tgfb2* expressing endothelium aligns with the similar gene expressing population at D0 suggesting this is a resident endothelial population present

in resting and inflammatory conditions. Intriguingly, we identified an inflammatory endothelium due to the expression of genes comprising of antigen presentation such as *Cd74*, the chemokines *Cxcl9* binding to CXCR3 and *Cxcl16* binding to CXCR6 and the lymphocyte adhesion marker *Vcam-1*. This suggests that this is an *activated* endothelium which possess the capabilities to secrete chemokines which may drive the influx of immune cells such as T cells and B cells and facilitate their activation with CD74 mediated interactions [208]. Lastly, at D15p.c. expressed *Car4* and *Ackr3* which resembled the arterial endothelial population identified at within the resting SG suggesting that this is another resident population which does not change its phenotype extensively in the presence or absence of inflammation. Overall, the endothelial populations fulfil their classical functions of maintaining and governing the SG vasculature in the presence and absence of inflammation. However, the emergence of a pro-inflammatory endothelial population with the ability to produce chemokines in response to inflammation reinforces the contributions of this population to facilitate lymphocyte entry.

Aside from the endothelial population, the epithelial cell population has been demonstrated to participate in the inflammatory response with their ability to secrete pro-inflammatory cytokines such as IL-1 α in combination with NF κ B mediated signalling [57]. The contribution of the epithelial cell population in the SG provides architecture and mechanical structure to the tissue and support non-classical APCs function in the tissue [9, 53, 57, 293]. In our dataset, we identified six Epcam⁺ populations within the resting SG and three clusters within the inflamed SG. Within the resting SG the overall functions of the epithelial cells consisted of regulating the cell-to-cell adhesion through the expression of *Asc/3* and the metabolism of glucose and fatty acids *Pdp2*. Moreover, the resting SG contained a population of epithelia

expressing *Kcne1* which has been previously associated with the transport of water and salts across the epithelium and form the beta subunit for the voltage gated potassium channels [312]. This aligns with the natural function of these cell types to govern the ionic homeostasis of the SG. The sixth subset of epithelial cells were associated with a progenitor cell type signature with the expression of genes such as *Wnt4*, *Ngf* and the peptidase *Klk1b5* and is in accordance with the identification of intestinal epithelial stem cells (iESCs) in the murine colon which also demonstrated a regenerative gene signature which participated in the maintenance of the resident mesenchyme in response to inflammatory injury [203]. Contrastingly, within the D15 p.c. we identified three populations of epithelial cells one of which did not express the Epcam marker. Broadly, the two Epcam⁺ populations expressed genes associated with the transport of the sodium potassium ions [312]. The final cluster which did not express Epcam instead uniquely expressed the gene *Ahnak*. The protein expression of this gene was spatially identified within the squamous epithelium in the respiratory tract and associated with promoting EMT downstream of interaction with TGFβ [313]. The progression of EMT has been associated with the downregulation of Epcam which corresponds with the gene signature of our Epcam⁺ subset which expresses the EMT pro-oncogene *Ahnak*. We further explored the expression pattern of the mucin family of genes which are derived from epithelial cells and constitute a large proportion of the saliva component. We were unable to capture genes associated to these proteins.

Overall, the murine scRNAseq dataset has provided us an overview with the various cell populations which are present in the resting versus the inflamed gland in the context of TLS. This provides a foundation to explore and validate the presence of other cell populations

independent of the fibroblasts which may have critical contributions in TLS formation. This dataset has allowed for the transcriptomic characterisation of this mouse model and also allowed us to explore similarities and differences of the heterogeneous fibroblast populations between the human and mouse dataset. Further exploration of this dataset would involve characterising the cellular interactions which may be occurring using transcriptomic analysis packages and furthering our validation studies with spatial analysis to understand where key populations are positioned within the inflamed tissue. This would not only validate the presence of these populations, but also allow us to infer the overarching communication taking place amongst cell populations within the SG tissue in order to understand their contributions to the formation of a TLS supportive microenvironment.

Chapter 5: Final Thesis Discussion

5.1: Overall Discussion

TLS are lymphocytic aggregates which develop within microenvironments harbouring molecular and supportive stromal populations functionally similar those present within lymphoid organs [1, 3, 7, 8, 11, 92, 94, 149, 170]. The presence of TLS has been identified in the context of different physiological and pathological states such as inflammation, cancer, tissue rejection and anti-viral and bacterial responses [3, 8, 19, 33, 61, 66, 80, 148, 314]. The presence of TLS in these differing contexts has been described and our group has specifically focused on the functional contributions of the stromal cell compartment in the formation of TLS [3, 7, 8, 10, 11, 19, 33, 61, 66, 80, 148, 314]. At present, in this thesis, we have furthered our understanding of the contributions of fibroblasts and pericytes, amongst other stromal and immune cells, in the context of TLS biology, and the heterogeneity which exists within these cell populations using single cell sequencing analysis on human pSS SG and inflamed murine SG.

Previously, we have investigated and demonstrated the contributions of various signals and cellular interactions which induce the formation of *immunofibroblasts* in non-lymphoid tissues such as the SG [8]. The exploration of TLS in the context of pSS using primary human pSS fibroblasts in conjunction with the inducible TLS mouse model has generated a 3-step model of *immunofibroblast* generation which includes stromal cell *priming*, *expansion*, and *maturation* [93] [7, 8]. Our previous work has functionally described and characterised the emerging immunomodulatory roles of fibroblasts in non-lymphoid organs and their ability to express the lymphoid chemokines and necessary lymphocyte survival factors in response to proinflammatory cues [8]. We have also made evident that pro-inflammatory cytokines such

as IL-13, TNF- α and LT $\alpha_1\beta_2$ can induce the expression and upregulation of markers such as ICAM-1, VCAM-1 and PDPN, indicating the *priming* and *activation* of stromal cells [8]. This upregulation facilitates the stabilisation of the cellular interaction with LFA-1 and VLA-4 expressing lymphocytes and stromal cells allowing for the exchange of molecular cues resulting in the *activation* and *maturation* of each cell type. We further demonstrated that the *expansion* of these activated cells was driven by the effects of the cytokine IL-22 [7]. This was followed by the stabilisation and further *maturation* of the *immunofibroblasts* which occurs in an LT β R dependant manner. Further exploration of the maintenance of the *immunofibroblast* phenotype has also been demonstrated to involve dependency on infiltrating lymphocytes to maintain the engagement of the LT β R signalling cascade within the *immunofibroblast* [8].

The heterogeneity of the fibroblast population in the human pSS SG, amongst other stromal cells, has not been, to date, characterised in its entirety. However, with the emergence of novel transcriptomic tools allowing scientists to tease apart this cellular cocktail, various sub populations of fibroblasts in the context of autoimmunity and cancer have now been identified [43, 154, 156, 160, 202, 315]. Specific to TLS biology, this fibroblast heterogeneity may underpin the overall TLS heterogeneity which is observed in patient biopsies in cancer and inflammation [19, 33, 66]. In this regard, to understand the underlying fibroblast compartment within TLS and the potential mechanisms leading to their formation would allow for novel therapeutic strategies to either induce or dissolve TLS in a disease specific context.

In this thesis we have leveraged our previously published work and explored the fibroblast heterogeneity within human pSS SG by exploring stromal cells compartmentalising as

CD34⁺ CD146⁻ and CD34⁻ CD146⁺ populations (Chapter 3). The five fibroblast populations we uncovered in this dataset demonstrated individual gene expression patterns assigning them unique identities to which we could begin inferring some functionality. Within these subpopulations we identified the *immunofibroblast* characterised by the gene expression of *CCL19* and *BAFF*; this population was enriched for NfKB mediated signalling. These findings were in accordance with the activation of the NfKB mediated signalling downstream of LTβR in the context of lymphoid tissue differentiation and specialisation [2, 98, 104, 109, 149]. Further subsets we identified exhibited stem cell like features which suggests the presence of a resident pool of fibroblasts which have the capabilities of acquiring *immunomodulatory* features in response to inflammation or modulatory cytokines present within the SG microenvironment. Additional fibroblast subsets demonstrated the expression of certain genes which indicated their possible participation in specific biological functions. For example, the expression of *ACKR3* indicates the possible participation of specific fibroblast populations in the modulation of the chemokine gradient within the human pSS SG. Contrastingly, other fibroblast populations demonstrated participation in vascular clotting functions with signature expression of *SERPINE2* and *F2R* genes, which I have discussed in depth in Chapter 1. The diverging functions and differential engagement of NfKB, TNF-α, and IFN-γ signalling pathways, inferred from our data, suggests that a diversity of fibroblast functions is perhaps necessary in order to establish a microenvironment able to harbour TLS. On the contrary, the induction of specific fibroblast populations, for example the *immunofibroblast*, in immunologically suppressed environments such as cancer, may provide a novel manner of “reactivating” the immune signature to generate an anti-tumoral response.

[152, 154, 202, 316][152][154][153]. Aside from cancer, the deconvolution of the cellular landscape within inflammatory disease such as RA and UC also found specific populations of fibroblasts which were observed to harbour specific gene signatures which assigned them specific functional and phenotypical characteristics [43, 155-157]. Our work has now shed light on the heterogeneity and functional differentiation of fibroblasts in the context of TLS. We further contribute by providing important data on the “*immunopericyte*” population in the context of TLS, data which warrants further exploration and investigation in the context of other TLS harbouring disease such as cancer and permissive tissues.

This CD34⁻ CD146⁺ “*immunopericyte*” population subset which expressed *CCL19* and *CCL21* was also distinguished by the gene, and protein, expression of *CD36* and *TNC*, which are both absent within the non-lymphoid like pericyte subset. This finding suggests that within the human pSS SG there are two sources of cells which produce *CCL19* and *CCL21* and mediate T cell infiltration and positioning in SG TLS. We hypothesise that the “*immunopericytes*” could initially facilitate the extravasation of CXCR7⁺ lymphocytes from the circulation into the SG itself as immune cell aggregates are identified in close proximity to SG vasculature from previous histological analysis [100]. The mechanism by which T cells are trafficked via resident pericyte populations within the spleen were explored by *Chauveau et al* [116]. Their work visualised the movement of T cells within the spleen in which they identified T cell trafficking occurred in a *CCL21* dependant manner which was associated with the presence of closely resident fibroblasts. In their work, *Chauveau et al* did not explore, or identify, the exact source of the *CCL21*, whereas we have identified at the transcriptomic level and at the protein level in human pSS SG and murine SG tissue that *CCL21* expression is specific to the CD34⁻ CD146⁺ population of pericyte, not just the CD34⁺ CD146⁻ fibroblasts. Intriguingly, from our

pathway enrichment analysis we observed that the prostaglandin signalling cascade was engaged within the CD34⁻ CD146⁺ pericyte population as compared to the other CD34⁻ CD146⁺ populations and the CD34⁺ CD146⁻ fibroblast populations.

The contributions of the prostaglandin pathway in migration through pericytes have been explored in the context of cancer and migration of CCR7⁺ T cells [258] [259]. Previous investigation by *Ahmed et al* observed initial rapid T cell migration towards endothelial cells activated with TNF- α and IFN- γ that was CCL19 and CCL21 dependant, but it was the role of the prostaglandin PGD₂ which facilitated trans endothelial migration [258]. These observations, coupled with our findings of CCL21 expressing pericytes with involvement of the prostaglandin pathway, may suggest that these *immunopericyte* possess the machinery to facilitate T cell migration from the circulation into the inflamed SG, and the *immunofibroblasts*, amongst the other fibroblast populations, control their navigation to the designated site of TLS formation within the tissue.

Moreover, we began our exploration of possible signals driving the expression of this lymphoid chemokine in *LT β R*^{-/-} mice as this pathway is critical for CCL21 expression in our immunofibroblast population. Intriguingly, CCL21 protein expression was preserved within the CD34⁻ CD146⁺ population suggesting that functional ability of this population to produce lymphoid chemokines is independent of *LT β R* engagement. Importantly our scRNAseq data of the human pSS SG also indicated the mRNA expression of IFN and TNF receptors within the CCL21⁺ CD34⁻ CD146⁺ pericyte population suggesting that there may be a dependency on these signalling cascades in the differentiation of the “*immunopericyte*”. The engagement of the IFN and TNF pathways was also emphasised within specific CD34⁺ CD146⁻ fibroblast

populations at the gene expression level which demonstrated not only the preferential engagement of certain signalling cascades in particular fibroblast subsets but also reinforces the engagement of IFN and TNF, as clinically these cytokines have been considered as biomarkers in pSS [176, 177].

In addition, we have undertaken transcriptomic analysis of the murine SG at D0 and D15 p.c. These two time points allowed us to capture a transcriptomic snapshot of the resting and inflamed murine SG allowing us to compare and contrast different cell populations at these time points and physiological states (Chapter 4). In this manner, we identified several subsets of fibroblasts each with unique mRNA expression of particular genes. We were successfully able to identify *immunofibroblasts* within the D15p.c. murine SG which we were not able to identify at D0. This data suggests that this pro-inflammatory population arises in the presence of inflammatory cues and interestingly did not demonstrate overlap with populations which we characterised at D0 using our pseudobulk correlation analysis. This suggests that this *immunofibroblast* population arises as an independent entity, or it may be the contributions of different fibroblast populations which have intrinsic capabilities that allow them to respond to inflammatory signals and become *immunofibroblasts*. However, further investigation at the protein level remains to be undertaken to confirm this hypothesis. Additionally, we were not able to identify the “*immunopericyte*” in the mouse dataset at the mRNA level but took advantage of the inducible TLS mouse model to identify this “*immunopericyte*” population at the protein level.

In addition to using scRNAseq to deconvolute the cellular heterogeneity within the human pSS and murine SG to understand the contributions of fibroblasts in the formation of TLS [162,

195][150, 184]. The overarching TLS heterogeneity has also been observed in histological analysis of human pSS SG biopsies. In these data, varying TLS structures had also been identified within the human pSS SG. This variability of TLS may be driven by the stromal heterogeneity that we have uncovered in our human pSS scRNAseq data. We furthered this exploration by conducting bulk sequencing of the microdissected lymphocyte aggregates ranging from small foci to TLS with GC from human pSS SG tissue and human tonsillitis tissue too. PCA analysis from this data gave us insights, requiring further formal characterisation and investigation, which demonstrated compartmentalisation of specific GC samples into 3 distinct groups comprising of tonsil GCs, SG GC1 and SG GC2. Preliminary analysis of this dataset appears to suggest that GC heterogeneity also exists within TLS and encompasses perhaps distinct similarities and differences depending on whether the GCs derive from SLO or inflamed SG tissue.

Intriguingly, our current dataset, which still requires additional samples to strengthen the statistical significance, identifies the 2 human pSS derived GC groups to either align closer to the tonsil derived GC or demonstrate a more pro-inflammatory and potentially pathological GC phenotype. These observations were coupled with gene expression patterns of the lymphoid chemokines and receptors associated with the inflammatory process which also teased the two GC subgroups apart. These data suggest that one set of GCs may be more pro-inflammatory than the other. The detection of GC heterogeneity may also provide insight into the functional and phenotypical signature of the overall TLS it sits within and the signalling cascades which may distinguish GC with different pathogenic profiles within TLS. Interestingly, the absence of GCs in the SG of patients with SS represent a negative prognostic factor for lymphoma development. Conversely the presence of GC failed to significantly correlate with

lymphoma formation, suggesting either that additional factors to GC formation would contribute to the process of lymphomagenesis or that there is variability in this process in patients harbouring GCs. In this context, the heterogeneity that we have identified in our data set warrant further exploration.

In conclusion, in the work underpinning this research project, we have utilised the contributions of scRNAseq to deconvolute the stromal cell heterogeneity which resides within the inflamed SG in both human and mice, and further explored the consequences of this cellular heterogeneity in the context of the differing GCs present within the pSS SG. We have also begun to unravel the possible unique contributions that specific signalling pathways may have on the formation of the different fibroblast and pericyte populations that we have unveiled within these datasets. I believe that this work has contributed to characterising the cellular landscape in the human pSS SG, providing an additional context to the development of novel therapeutic targets in the context of TLS associated diseases.

5.2: Limitations within this work

The use of scRNAseq has allowed scientists to explore and interrogate the genome within single cells of specific tissues in varying physiological contexts. This approach has allowed for the characterisation of the heterogeneity of cell populations which can drive the exploration of precise scientific questions at the lab bench. Whilst this technique is powerful there are limitations which we have encountered in the process of analysing and interpreting this data which hinder the total exploratory potential of our samples of interest at the single cell level.

Primarily, the interpretation of such data needs to consider carefully prior to inferring cellular identity and function as this mRNA data does not confirm or imply the correlation of the mRNA to the protein expression which may occurring *in vivo*. Therefore, it is critical that further and exploration of this scRNAseq data at the protein level is undertaken to demonstrate and confirm that the data seen at the mRNA level translated at the protein level in the tissue of interest. Furthermore, we observed that classic canonical cell populations were either not identified in the scRNAseq dataset or hallmark mRNA transcripts were not captured within certain cell populations. For example, key transcription factors identifying T cell subsets demonstrated no gene expression in our data, whilst it is established that these genes and proteins are expressed within differentiated cells. This demonstrates that the scRNAseq chemistry may not be sensitive enough to capture all the genes of the cell or rather the depth of sequencing may not be sufficient for which reason we have not captured all the mRNA we would expect to observe. As this is one example of many, this demonstrates that this dataset may not completely and accurately capture the heterogeneity and the mRNA profile of the cells.

This was additionally demonstrated with our inability to capture the CD34⁻CD146⁺ “immunopericyte” population in the murine scRNAseq dataset due to the inability to capture the expression of the lymphoid chemokines in our dataset. However, at the protein level using primary inflamed murine tissue we demonstrated the presence of this “CD34⁻CD146⁺CCL21⁺ immunopericyte” population within WT and LTβR^{-/-} murine SG at the protein level which confirmed that although we were unable to identify these cells in our scRNAseq dataset they were present *in vivo*.

Therefore, our scRNAseq datasets may not be capturing the entirety of the cell populations present within a tissue. This can result due to the caveats associated with the processing of the tissue and the 35-50% capture efficiency of the 10x chromium platform. To circumvent these issues, libraries for this data maybe sequenced at a greater depth in order increase the capture of mRNA which we may not be demonstrating at present or additional biological replicates may need to be processed in order to capture enough cells within the suspension to be replicated within the scRNAseq dataset.

5.3: Future experimental work

The work I have currently explored and discussed in this thesis has utilised novel genomic technologies which can profile the gene signature of single cells at the mRNA level. In the previous section I have discussed and described the limitations of such a technique, however gene expression datasets offer researchers a “roadmap” to better navigate and explore currently known knowledge and scientific trends. In this manner, the work I have currently explored requires validation at the protein and functional level to demonstrate whether our observations have true translational significance and can be explored further to understand the contributions of specific cell populations in TLS formation and heterogeneity. This includes the remaining stromal cell populations of endothelial, epithelial, and neuronal populations as well as further validation and exploration of the immune cells such as T cells, B cells, NK cells, DCs, monocytes and macrophages.

Primarily, with access to primary human tissue for pSS SGs, we can complement our human scRNAseq dataset by using flow cytometric or imaging analysis to confirm the presence of the proteins from the mRNA expression patterns we have identified. This would not only validate the mRNA expression we have observed in our datasets but also provide the basis on which specific cell populations can be further functionally characterised and explored in their contribution to the onset and maintenance of pSS. For example, the unique mRNA signature we have identified for the *immunofibroblast* with the expression of the lymphoid chemokines CCL19 and TNFSF13B, future work could include specifically isolating this subset of fibroblasts from primary tissues. For example, exploring these populations in T cell knockout and B cell knockout mice would be interesting to explore as the involvement of these lymphocytes,

amongst others, has been demonstrated to contribute to the remodelling of the resident mesenchyme such that it adopts immunomodulatory properties. With this novel data, and access to human pSS fibroblasts, it would also be interesting to explore whether specific fibroblast phenotypes can be induced *in vitro* with stimulation of the appropriate cytokines we have observed from our pathway enrichment dataset. This would not only validate our transcriptomic work but also allow us to dissect the possible contributions of the different signals that are present within the inflamed pSS SG which generate the fibroblast heterogeneity within the microenvironment.

Moreover, as our histological analysis of the data is conducted on one tissue sample it is necessary to use the same marker panel for additional pSS SG samples to observe for any distinct trends and patterns for the protein expression within the tissue. This would allow us to truly confirm the mRNA expression trends we observe in our dataset within the tissue. Additionally, we may be able to observe specific trends in the expression of protein according to the pathology and clinical profile of the tissue being used.

On the other hand, to understand the cellular landscape better within the human pSS SG, it would be intriguing to explore the cellular landscape of the sicca SG. The pathology of sicca overlaps with that of pSS as it is an inflammatory condition of the SG amongst other mucosal tissues, however the cellular basis of these two diseases differs. In this instance, whilst pSS has the hallmark presence of TLS, the histological characterisation of sicca tissue demonstrates the presence of diffuse immune cell infiltration. It would be of interest to compare and contrast the scRNAseq datasets of these two diseases in order to observe whether the presence or absence of specific cell populations, or mRNA expression profiles,

have specific contributions in TLS formation, maintenance and organisation in the context of pSS.

More specifically, it would be insightful to explore whether these fibroblast populations are also present in sicca tissue and additionally in the scRNAseq datasets of other cancer tissues which may harbour TLS. This would inform us of whether TLS associated fibroblasts are present in other tissues in which case these findings would support the use of therapies to perhaps induce an *immunofibroblast* phenotype which would hypothetically support the formation of TLS within the TME. Furthermore, exploration of whether the CD34⁻ CD146⁺ CCL21⁺ CCL19⁺ CD36⁺ "*immunopericyte*" population is present and functionally relevant in other permissive tissue such as within TLS harbouring cancers or sicca syndrome, would also be an additional interesting avenue to pursue as this would indicate how unique or redundant this cell population is in the context of disease and more specifically in the formation of TLS. Further exploration of our dataset would include pinpointing specific cellular interactions, using tools such as Cell Phone DB, between stromal cells and lymphocytes which could be contributing to the modulation and formation of the underlying TLS supportive mesenchyme. Further exploration of our dataset would include using tools such as Cell Phone DB which pinpoints specific cellular interactions which are inferred from the gene expression dataset. However, as with the gene expression data, this would require protein validation using histological techniques such as immunohistochemistry to determine if the inferred cellular populations are spatially located in proximity and can support the observations made from the scRNA Seq dataset.

Furthermore, whilst this work would still explore this “*immunopericyte*” population, in my thesis we have demonstrated the presence of this population at the protein level in human pSS SG and the murine inducible pSS model SG. We have also conducted initial functional exploration using the murine inducible mouse model to explore the possible signalling mechanism which results in the expression of the CCL21 in the CD34⁻ CD146⁺ CCL21⁺ pericytes. Future functional investigations of these cell types could include exploring the CCL21 expression pattern of these cells in different murine knockout models such as *IFN* γ ^{-/-} and *TNF*^{-/-} mice in order to deduce the specific signals potentially involved in their development and function. This opens up potential avenues to functionally explore the signalling cascades which may govern the phenotype of this pericyte population as we have observed that the lymphoid chemokine expression is independent of LT β R signalling. In addition, it would also be insightful to further explore whether any functional changes are observed in the context of T cell knockout or B cell knockout mice and whether the absence of these lymphocytes impacts the lymphoid phenotype and function of the “*immunopericyte*”.

These additional validation and functional studies would strengthen the observations made from the scRNAseq database and confirm the trends and possible functions I have inferred in my thesis at the protein level.

References

1. Asam, S., et al., *Stromal cells in tertiary lymphoid structures: Architects of autoimmunity*. Immunological Reviews. **n/a**(n/a).
2. Kratz, A., et al., *Chronic inflammation caused by lymphotoxin is lymphoid neogenesis*. J Exp Med, 1996. **183**(4): p. 1461-72.
3. Neyt, K., et al., *Tertiary lymphoid organs in infection and autoimmunity*. Trends in Immunology, 2012. **33**(6): p. 297-305.
4. Rangel-Moreno, J., et al., *The development of inducible bronchus-associated lymphoid tissue depends on IL-17*. Nat Immunol, 2011. **12**(7): p. 639-46.
5. Grogan, J.L. and W. Ouyang, *A role for Th17 cells in the regulation of tertiary lymphoid follicles*. Eur J Immunol, 2012. **42**(9): p. 2255-62.
6. Peters, A., et al., *Th17 cells induce ectopic lymphoid follicles in central nervous system tissue inflammation*. Immunity, 2011. **35**(6): p. 986-96.
7. Barone, F., et al., *IL-22 regulates lymphoid chemokine production and assembly of tertiary lymphoid organs*. Proceedings of the National Academy of Sciences, 2015. **112**(35): p. 11024-11029.
8. Nayar, S., et al., *Immunofibroblasts are pivotal drivers of tertiary lymphoid structure formation and local pathology*. Proceedings of the National Academy of Sciences, 2019. **116**(27): p. 13490-13497.
9. Asam, S., et al., *The role of stroma and epithelial cells in primary Sjögren's syndrome*. Rheumatology, 2019.
10. Barone, F., et al., *Stromal Fibroblasts in Tertiary Lymphoid Structures: A Novel Target in Chronic Inflammation*. Frontiers in Immunology, 2016. **7**(477).
11. Pipi, E., et al., *Tertiary Lymphoid Structures: Autoimmunity Goes Local*. Frontiers in Immunology, 2018. **9**(1952).
12. Barone, F., et al., *CXCL13, CCL21, and CXCL12 Expression in Salivary Glands of Patients with Sjögren's Syndrome and MALT Lymphoma: Association with Reactive and Malignant Areas of Lymphoid Organization*. The Journal of Immunology, 2008. **180**(7): p. 5130-5140.
13. Ngo, V.N., et al., *Lymphotoxin alpha/beta and tumor necrosis factor are required for stromal cell expression of homing chemokines in B and T cell areas of the spleen*. J Exp Med, 1999. **189**(2): p. 403-12.
14. Cupedo, T. and R.E. Mebius, *Role of chemokines in the development of secondary and tertiary lymphoid tissues*. Semin Immunol, 2003. **15**(5): p. 243-8.
15. Romano, M., et al., *Role of IL-6 and its soluble receptor in induction of chemokines and leukocyte recruitment*. Immunity, 1997. **6**(3): p. 315-25.
16. Rangel-Moreno, J., D. Carragher, and T.D. Randall, *Role of lymphotoxin and homeostatic chemokines in the development and function of local lymphoid tissues in the respiratory tract*. Immunologia, 2007. **26**(1): p. 13-28.
17. Wiede, F., et al., *TNF-dependent overexpression of CCL21 is an underlying cause of progressive lymphoaccumulation in generalized lymphoproliferative disorder*. Eur J Immunol, 2007. **37**(2): p. 351-7.

18. Bombardieri, M., et al., *Activation-induced cytidine deaminase expression in follicular dendritic cell networks and interfollicular large B cells supports functionality of ectopic lymphoid neogenesis in autoimmune sialoadenitis and MALT lymphoma in Sjögren's syndrome*. J Immunol, 2007. **179**(7): p. 4929-38.
19. Helmink, B.A., et al., *B cells and tertiary lymphoid structures promote immunotherapy response*. Nature, 2020. **577**(7791): p. 549-555.
20. Kang, S., et al., *BAFF Induces Tertiary Lymphoid Structures and Positions T Cells within the Glomeruli during Lupus Nephritis*. The Journal of Immunology, 2017: p. 1600281.
21. Zhou, H., et al., *CD8+ T Lymphocytes: Crucial Players in Sjögren's Syndrome*. Frontiers in Immunology, 2021. **11**(3632).
22. Rizzo, C., et al., *Primary Sjogren Syndrome: Focus on Innate Immune Cells and Inflammation*. Vaccines, 2020. **8**(2): p. 272.
23. Roozendaal, R. and R.E. Mebius, *Stromal cell-immune cell interactions*. Annu Rev Immunol, 2011. **29**: p. 23-43.
24. Noort, A.R., et al., *Tertiary Lymphoid Structures in Rheumatoid Arthritis: NF- κ B-Inducing Kinase-Positive Endothelial Cells as Central Players*. The American Journal of Pathology, 2015. **185**(7): p. 1935-1943.
25. Liu, C., et al., *Characteristics of primary Sjögren's syndrome patients with IgG4 positive plasma cells infiltration in the labial salivary glands*. Clin Rheumatol, 2017. **36**(1): p. 83-88.
26. Barone, F., et al., *Association of CXCL13 and CCL21 expression with the progressive organization of lymphoid-like structures in Sjögren's syndrome*. Arthritis Rheum, 2005. **52**(6): p. 1773-84.
27. Barone, F., et al., *IgA-Producing Plasma Cells Originate From Germinal Centers That Are Induced by B-Cell Receptor Engagement in Humans*. Gastroenterology, 2011. **140**(3): p. 947-956.
28. Weinstein, A.M., et al., *Association of IL-36 γ with tertiary lymphoid structures and inflammatory immune infiltrates in human colorectal cancer*. Cancer immunology, immunotherapy : CII, 2019. **68**(1): p. 109-120.
29. Buettner, M. and M. Lochner, *Development and Function of Secondary and Tertiary Lymphoid Organs in the Small Intestine and the Colon*. Front Immunol, 2016. **7**: p. 342.
30. Prados, A., et al., *Fibroblastic reticular cell lineage convergence in Peyer's patches governs intestinal immunity*. Nature Immunology, 2021.
31. Guedj, K., et al., *Adipocytes orchestrate the formation of tertiary lymphoid organs in the creeping fat of Crohn's disease affected mesentery*. J Autoimmun, 2019. **103**: p. 102281.
32. Sautès-Fridman, C., et al., *Tertiary lymphoid structures in the era of cancer immunotherapy*. Nature Reviews Cancer, 2019. **19**(6): p. 307-325.
33. Petitprez, F., et al., *B cells are associated with survival and immunotherapy response in sarcoma*. Nature, 2020. **577**(7791): p. 556-560.
34. Gago da Graça, C., L.G.M. van Baarsen, and R.E. Mebius, *Tertiary Lymphoid Structures: Diversity in Their Development, Composition, and Role*. The Journal of Immunology, 2021. **206**(2): p. 273-281.
35. Lee, H.J., et al., *Tertiary lymphoid structures: prognostic significance and relationship with tumour-infiltrating lymphocytes in triple-negative breast cancer*. Journal of Clinical Pathology, 2016. **69**(5): p. 422-430.

36. Marinkovic, T., et al., *Interaction of mature CD3+CD4+ T cells with dendritic cells triggers the development of tertiary lymphoid structures in the thyroid*. The Journal of Clinical Investigation, 2006. **116**(10): p. 2622-2632.
37. Humby, F., et al., *Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium*. PLoS Med, 2009. **6**(1): p. e1.
38. Bombardieri, M., M. Lewis, and C. Pitzalis, *Ectopic lymphoid neogenesis in rheumatic autoimmune diseases*. Nat Rev Rheumatol, 2017. **13**(3): p. 141-154.
39. Fox, R.I., et al., *Primary Sjogren syndrome: clinical and immunopathologic features*. Semin Arthritis Rheum, 1984. **14**(2): p. 77-105.
40. Mackay, F. and J.L. Browning, *BAFF: a fundamental survival factor for B cells*. Nat Rev Immunol, 2002. **2**(7): p. 465-75.
41. Mueller, S.N. and R.N. Germain, *Stromal cell contributions to the homeostasis and functionality of the immune system*. Nat Rev Immunol, 2009. **9**(9): p. 618-29.
42. Alcântara, C., M.J. Gomes, and C. Ferreira, *Rituximab therapy in primary Sjögren's syndrome*. Ann N Y Acad Sci, 2009. **1173**: p. 701-5.
43. Croft, A.P., et al., *Distinct fibroblast subsets drive inflammation and damage in arthritis*. Nature, 2019. **570**(7760): p. 246-251.
44. Ropes, M.W., et al., *Proposed diagnostic criteria for rheumatoid arthritis*. Ann Rheum Dis, 1957. **16**(1): p. 118-25.
45. Munthe, E. and J.B. Natvig, *Immunoglobulin classes, subclasses and complexes of IgG rheumatoid factor in rheumatoid plasma cells*. Clin Exp Immunol, 1972. **12**(1): p. 55-70.
46. Boling, E.P., et al., *Primary Sjogren's syndrome and autoimmune hemolytic anemia in sisters. A family study*. Am J Med, 1983. **74**(6): p. 1066-71.
47. Salomonsson, S., et al., *Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with Sjögren's syndrome*. Arthritis Rheum, 2003. **48**(11): p. 3187-201.
48. Nocturne, G. and X. Mariette, *Sjögren Syndrome-associated lymphomas: an update on pathogenesis and management*. British Journal of Haematology, 2015. **168**(3): p. 317-327.
49. Imgenberg-Kreuz, J., J.K. Sandling, and G. Nordmark, *Epigenetic alterations in primary Sjögren's syndrome - an overview*. Clin Immunol, 2018. **196**: p. 12-20.
50. Ibrahem, H.M., *B cell dysregulation in primary Sjögren's syndrome: A review*. Jpn Dent Sci Rev, 2019. **55**(1): p. 139-144.
51. Brito-Zerón, P., et al., *Characterization and risk estimate of cancer in patients with primary Sjögren syndrome*. Journal of hematology & oncology, 2017. **10**(1): p. 90-90.
52. Pırlıdar, T., et al., *Endothelial dysfunction in patients with primary Sjögren's syndrome*. Rheumatology International, 2005. **25**(7): p. 536-539.
53. Verstappen, G.M., et al., *Epithelial-immune cell interplay in primary Sjögren syndrome salivary gland pathogenesis*. Nature Reviews Rheumatology, 2021.
54. Nordmark, G., M.L. Eloranta, and L. Ronnblom, *Primary Sjögren's syndrome and the type I interferon system*. Curr Pharm Biotechnol, 2012. **13**(10): p. 2054-62.
55. Szyszko, E.A., et al., *Salivary glands of primary Sjögren's syndrome patients express factors vital for plasma cell survival*. Arthritis Research & Therapy, 2011. **13**(1): p. R2.

56. Bahler, D.W. and S.H. Swerdlow, *Clonal Salivary Gland Infiltrates Associated With Myoepithelial Sialadenitis (Sjögren's Syndrome) Begin as Nonmalignant Antigen-Selected Expansions*. Blood, 1998. **91**(6): p. 1864-1872.
57. Cheng, D.-s., et al., *Airway Epithelium Controls Lung Inflammation and Injury through the NF- κ B Pathway*. The Journal of Immunology, 2007. **178**(10): p. 6504-6513.
58. Voulgarelis, M. and A.G. Tzioufas, *Pathogenetic mechanisms in the initiation and perpetuation of Sjögren's syndrome*. Nature Reviews Rheumatology, 2010. **6**: p. 529.
59. Wahren, M., et al., *Autoantibody repertoire to Ro/SSA and La/SSB antigens in patients with primary and secondary Sjögren's syndrome*. J Autoimmun, 1996. **9**(4): p. 537-44.
60. Lim, B., et al., *Inflammatory breast cancer biology: the tumour microenvironment is key*. Nature Reviews Cancer, 2018. **18**(8): p. 485-499.
61. Colbeck, E.J., et al., *Tertiary Lymphoid Structures in Cancer: Drivers of Antitumor Immunity, Immunosuppression, or Bystander Sentinels in Disease?* Frontiers in immunology, 2017. **8**: p. 1830-1830.
62. Avagliano, A., et al., *Metabolic Reprogramming of Cancer Associated Fibroblasts: The Slavery of Stromal Fibroblasts*. Biomed Res Int, 2018. **2018**: p. 6075403.
63. Elmusrati, A.A., et al., *Cancer-associated fibroblasts promote bone invasion in oral squamous cell carcinoma*. British journal of cancer, 2017. **117**(6): p. 867-875.
64. Karagiannis, G.S., et al., *Cancer-associated fibroblasts drive the progression of metastasis through both paracrine and mechanical pressure on cancer tissue*. Mol Cancer Res, 2012. **10**(11): p. 1403-18.
65. Ziani, L., S. Chouaib, and J. Thiery, *Alteration of the Antitumor Immune Response by Cancer-Associated Fibroblasts*. Front Immunol, 2018. **9**: p. 414.
66. Cabrita, R., et al., *Tertiary lymphoid structures improve immunotherapy and survival in melanoma*. Nature, 2020. **577**(7791): p. 561-565.
67. Mu, L., et al., *CD4+ and Perivascular Foxp3+ T Cells in Glioma Correlate with Angiogenesis and Tumor Progression*. Frontiers in Immunology, 2017. **8**(1451).
68. Kumagai, S., et al., *The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies*. Nature Immunology, 2020. **21**(11): p. 1346-1358.
69. Patel, S.A., et al., *Treg/Th17 polarization by distinct subsets of breast cancer cells is dictated by the interaction with mesenchymal stem cells*. J Cancer Stem Cell Res, 2014. **2014**(2).
70. de Chaisemartin, L., et al., *Characterization of Chemokines and Adhesion Molecules Associated with T cell Presence in Tertiary Lymphoid Structures in Human Lung Cancer*. Cancer Research, 2011. **71**(20): p. 6391-6399.
71. Posch, F., et al., *Maturation of tertiary lymphoid structures and recurrence of stage II and III colorectal cancer*. Oncoimmunology, 2018. **7**(2): p. e1378844.
72. Kroeger, D.R., K. Milne, and B.H. Nelson, *Tumor-Infiltrating Plasma Cells Are Associated with Tertiary Lymphoid Structures, Cytolytic T-Cell Responses, and Superior Prognosis in Ovarian Cancer*. Clinical Cancer Research, 2016. **22**(12): p. 3005-3015.
73. GeurtsvanKessel, C.H., et al., *Dendritic cells are crucial for maintenance of tertiary lymphoid structures in the lung of influenza virus-infected mice*. J Exp Med, 2009. **206**(11): p. 2339-49.

74. Wang, J.P., et al., *Toll-like receptor-mediated activation of neutrophils by influenza A virus*. Blood, 2008. **112**(5): p. 2028-34.
75. Ramakrishnan, L., *Revisiting the role of the granuloma in tuberculosis*. Nature Reviews Immunology, 2012. **12**(5): p. 352-366.
76. Teillaud, J.L., et al., *Exploring the Role of Tertiary Lymphoid Structures Using a Mouse Model of Bacteria-Infected Lungs*. Methods Mol Biol, 2018. **1845**: p. 223-239.
77. Koscsó, B., et al., *Gut-resident CX3CR1(hi) macrophages induce tertiary lymphoid structures and IgA response in situ*. Sci Immunol, 2020. **5**(46).
78. Magro, D.O., et al., *Remission in Crohn's disease is accompanied by alterations in the gut microbiota and mucins production*. Sci Rep, 2019. **9**(1): p. 13263.
79. Thaunat, O., et al., *Chronic rejection triggers the development of an aggressive intragraft immune response through recapitulation of lymphoid organogenesis*. J Immunol, 2010. **185**(1): p. 717-28.
80. Huibers, M.M., et al., *The composition of ectopic lymphoid structures suggests involvement of a local immune response in cardiac allograft vasculopathy*. J Heart Lung Transplant, 2015. **34**(5): p. 734-45.
81. Koenig, A. and O. Thaunat, *Lymphoid Neogenesis and Tertiary Lymphoid Organs in Transplanted Organs*. Front Immunol, 2016. **7**: p. 646.
82. Porcheray, F., et al., *Chronic humoral rejection of human kidney allografts associates with broad autoantibody responses*. Transplantation, 2010. **89**(10): p. 1239-46.
83. Dieudé, M., et al., *Extracellular vesicles derived from injured vascular tissue promote the formation of tertiary lymphoid structures in vascular allografts*. American Journal of Transplantation, 2020. **20**(3): p. 726-738.
84. Stuhrt, S., et al., *Lymphatic Neoangiogenesis in Human Renal Allografts: Results from Sequential Protocol Biopsies*. American Journal of Transplantation, 2007. **7**(2): p. 377-384.
85. Dieudé, M., et al., *Extracellular vesicles derived from injured vascular tissue promote the formation of tertiary lymphoid structures in vascular allografts*. Am J Transplant, 2020. **20**(3): p. 726-738.
86. Krishnamurty, A.T. and S.J. Turley, *Lymph node stromal cells: cartographers of the immune system*. Nature Immunology, 2020. **21**(4): p. 369-380.
87. Koning, J.J. and R.E. Mebius, *Interdependence of stromal and immune cells for lymph node function*. Trends Immunol, 2012. **33**(6): p. 264-70.
88. Koning, J.J. and R.E. Mebius, *Stromal cells and immune cells involved in formation of lymph nodes and their niches*. Curr Opin Immunol, 2020. **64**: p. 20-25.
89. Rodda, L.B., et al., *Single-Cell RNA Sequencing of Lymph Node Stromal Cells Reveals Niche-Associated Heterogeneity*. Immunity, 2018. **48**(5): p. 1014-1028.e6.
90. Randall, T.D., D.M. Carragher, and J. Rangel-Moreno, *Development of secondary lymphoid organs*. Annu Rev Immunol, 2008. **26**: p. 627-50.
91. van de Pavert, S.A. and R.E. Mebius, *Development of secondary lymphoid organs in relation to lymphatic vasculature*. Adv Anat Embryol Cell Biol, 2014. **214**: p. 81-91.
92. Carragher, D.M., J. Rangel-Moreno, and T.D. Randall, *Ectopic lymphoid tissues and local immunity*. Semin Immunol, 2008. **20**(1): p. 26-42.

93. Bombardieri, M., et al., *Inducible Tertiary Lymphoid Structures, Autoimmunity, and Exocrine Dysfunction in a Novel Model of Salivary Gland Inflammation in C57BL/6 Mice*. The Journal of Immunology, 2012. **189**(7): p. 3767-3776.
94. van de Pavert, S.A. and R.E. Mebius, *New insights into the development of lymphoid tissues*. Nat Rev Immunol, 2010. **10**(9): p. 664-74.
95. Bénézech, C., et al., *Ontogeny of Stromal Organizer Cells during Lymph Node Development*. The Journal of Immunology, 2010. **184**(8): p. 4521-4530.
96. Mebius, R.E., *Organogenesis of lymphoid tissues*. Nat Rev Immunol, 2003. **3**(4): p. 292-303.
97. Bénézech, C. and L.H. Jackson-Jones, *ILC2 Orchestration of Local Immune Function in Adipose Tissue*. Front Immunol, 2019. **10**: p. 171.
98. Vondenhoff, M.F., et al., *LTbetaR signaling induces cytokine expression and up-regulates lymphangiogenic factors in lymph node anlagen*. J Immunol, 2009. **182**(9): p. 5439-45.
99. Kim, M.Y., et al., *Lymphoid tissue inducer cells: architects of CD4 immune responses in mice and men*. Clinical and experimental immunology, 2009. **157**(1): p. 20-26.
100. Nayar, S., et al., *Bimodal Expansion of the Lymphatic Vessels Is Regulated by the Sequential Expression of IL-7 and Lymphotoxin α ₁ β ₂ in Newly Formed Tertiary Lymphoid Structures*. The Journal of Immunology, 2016. **197**(5): p. 1957-1967.
101. Cupedo, T. and R.E. Mebius, *Cellular interactions in lymph node development*. J Immunol, 2005. **174**(1): p. 21-5.
102. Ansel, K.M., et al., *A chemokine-driven positive feedback loop organizes lymphoid follicles*. Nature, 2000. **406**(6793): p. 309-14.
103. Ruddle, N.H. and E.M. Akirav, *Secondary lymphoid organs: responding to genetic and environmental cues in ontogeny and the immune response*. J Immunol, 2009. **183**(4): p. 2205-12.
104. White, A., et al., *Lymphotoxin α -dependent and -independent signals regulate stromal organizer cell homeostasis during lymph node organogenesis*. Blood, 2007. **110**(6): p. 1950-1959.
105. Cupedo, T., G. Kraal, and R.E. Mebius, *The role of CD45⁺CD4⁺CD3⁻ cells in lymphoid organ development*. Immunological Reviews, 2002. **189**(1): p. 41-50.
106. Sabin, F.R., *On the origin of the lymphatic system from the veins and the development of the lymph hearts and thoracic duct in the pig*. American Journal of Anatomy, 1902. **1**(3): p. 367-389.
107. Jussila, L. and K. Alitalo, *Vascular growth factors and lymphangiogenesis*. Physiol Rev, 2002. **82**(3): p. 673-700.
108. Ansel, K.M., R.B. Harris, and J.G. Cyster, *CXCL13 is required for B1 cell homing, natural antibody production, and body cavity immunity*. Immunity, 2002. **16**(1): p. 67-76.
109. Bénézech, C., et al., *Lymphotoxin- β receptor signaling through NF- κ B2-RelB pathway reprograms adipocyte precursors as lymph node stromal cells*. Immunity, 2012. **37**(4): p. 721-734.
110. Abashev, T.M., et al., *Retinoic acid signaling regulates Krt5 and Krt14 independently of stem cell markers in submandibular salivary gland epithelium*. Dev Dyn, 2017. **246**(2): p. 135-147.

111. van de Pavert, S.A., et al., *Chemokine CXCL13 is essential for lymph node initiation and is induced by retinoic acid and neuronal stimulation*. Nat Immunol, 2009. **10**(11): p. 1193-9.
112. Koning, J.J., et al., *Development of follicular dendritic cells in lymph nodes depends on retinoic acid mediated signaling*. bioRxiv, 2020: p. 2020.05.20.106385.
113. den Haan, J.M., R.E. Mebius, and G. Kraal, *Stromal cells of the mouse spleen*. Frontiers in immunology, 2012. **3**: p. 201-201.
114. Aloisi, F. and R. Pujol-Borrell, *Lymphoid neogenesis in chronic inflammatory diseases*. Nat Rev Immunol, 2006. **6**(3): p. 205-17.
115. Mueller, C.G. and E. Hess, *Emerging Functions of RANKL in Lymphoid Tissues*. Front Immunol, 2012. **3**: p. 261.
116. Chauveau, A., et al., *Visualization of T Cell Migration in the Spleen Reveals a Network of Perivascular Pathways that Guide Entry into T Zones*. Immunity, 2020. **52**(5): p. 794-807.e7.
117. Coles, M.C., et al., *Role of T and NK cells and IL7/IL7r interactions during neonatal maturation of lymph nodes*. Proc Natl Acad Sci U S A, 2006. **103**(36): p. 13457-62.
118. Gardner, J.M., et al., *Extrathymic Aire-expressing cells are a distinct bone marrow-derived population that induce functional inactivation of CD4⁺ T cells*. Immunity, 2013. **39**(3): p. 560-72.
119. Wculek, S.K., et al., *Dendritic cells in cancer immunology and immunotherapy*. Nature Reviews Immunology, 2020. **20**(1): p. 7-24.
120. Subramanian, V., et al., *Immune response to tissue-restricted self-antigens induces airway inflammation and fibrosis following murine lung transplantation*. Am J Transplant, 2014. **14**(10): p. 2359-66.
121. Schwartz, R.H., *T cell anergy*. Annu Rev Immunol, 2003. **21**: p. 305-34.
122. Victora, G.D. and M.C. Nussenzweig, *Germinal Centers*. Annual Review of Immunology, 2012. **30**(1): p. 429-457.
123. Allen, C.D., T. Okada, and J.G. Cyster, *Germinal-center organization and cellular dynamics*. Immunity, 2007. **27**(2): p. 190-202.
124. Link, A., et al., *Association of T-zone reticular networks and conduits with ectopic lymphoid tissues in mice and humans*. Am J Pathol, 2011. **178**(4): p. 1662-75.
125. Baccin, C., et al., *Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization*. Nature Cell Biology, 2020. **22**(1): p. 38-48.
126. Acton, S.E., et al., *Dendritic cells control fibroblastic reticular network tension and lymph node expansion*. Nature, 2014. **514**(7523): p. 498-502.
127. Perez-Shibayama, C., C. Gil-Cruz, and B. Ludewig, *Fibroblastic reticular cells at the nexus of innate and adaptive immune responses*. Immunological reviews, 2019. **289**(1): p. 31-41.
128. Brown, F.D. and S.J. Turley, *Fibroblastic reticular cells: organization and regulation of the T lymphocyte life cycle*. J Immunol, 2015. **194**(4): p. 1389-94.
129. Fletcher, A.L., et al., *Lymph node fibroblastic reticular cells directly present peripheral tissue antigen under steady-state and inflammatory conditions*. The Journal of experimental medicine, 2010. **207**(4): p. 689-697.

130. Fletcher, A.L., S.E. Acton, and K. Knoblich, *Lymph node fibroblastic reticular cells in health and disease*. Nature Reviews Immunology, 2015. **15**(6): p. 350-361.
131. Choi, S.Y., et al., *YAP/TAZ direct commitment and maturation of lymph node fibroblastic reticular cells*. Nat Commun, 2020. **11**(1): p. 519.
132. Mionnet, C., et al., *Identification of a new stromal cell type involved in the regulation of inflamed B cell follicles*. PLoS Biol, 2013. **11**(10): p. e1001672.
133. Cyster, J.G., et al., *Follicular stromal cells and lymphocyte homing to follicles*. Immunol Rev, 2000. **176**: p. 181-93.
134. Krautler, N.J., et al., *Follicular dendritic cells emerge from ubiquitous perivascular precursors*. Cell, 2012. **150**(1): p. 194-206.
135. Allen, C.D.C., et al., *Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5*. Nature Immunology, 2004. **5**(9): p. 943-952.
136. Pikor, N.B., et al., *Remodeling of light and dark zone follicular dendritic cells governs germinal center responses*. Nat Immunol, 2020. **21**(6): p. 649-659.
137. Gustafson, C.E., C.M. Weyand, and J.J. Goronzy, *T follicular helper cell development and functionality in immune ageing*. Clin Sci (Lond), 2018. **132**(17): p. 1925-1935.
138. Amé-Thomas, P., et al., *CD10 delineates a subset of human IL-4 producing follicular helper T cells involved in the survival of follicular lymphoma B cells*. Blood, 2015. **125**(15): p. 2381-2385.
139. Linterman, M.A., et al., *Foxp3+ follicular regulatory T cells control the germinal center response*. Nat Med, 2011. **17**(8): p. 975-82.
140. Picker, L.J. and E.C. Butcher, *Physiological and molecular mechanisms of lymphocyte homing*. Annu Rev Immunol, 1992. **10**: p. 561-91.
141. Moyron-Quiroz, J.E., et al., *Role of inducible bronchus associated lymphoid tissue (iBALT) in respiratory immunity*. Nat Med, 2004. **10**(9): p. 927-34.
142. Bénézech, C., et al., *Inflammation-induced formation of fat-associated lymphoid clusters*. Nat Immunol, 2015. **16**(8): p. 819-828.
143. Jackson-Jones, L.H., et al., *Stromal Cells Covering Omental Fat-Associated Lymphoid Clusters Trigger Formation of Neutrophil Aggregates to Capture Peritoneal Contaminants*. Immunity, 2020. **52**(4): p. 700-715.e6.
144. Navarro, R., et al., *Immune Regulation by Pericytes: Modulating Innate and Adaptive Immunity*. Frontiers in immunology, 2016. **7**: p. 480-480.
145. Cupedo, T., et al., *Induction of Secondary and Tertiary Lymphoid Structures in the Skin*. Immunity, 2004. **21**(5): p. 655-667.
146. Brendolan, A. and J. Caamano, *Mesenchymal cell differentiation during lymph node organogenesis*. Frontiers in Immunology, 2012. **3**(381).
147. Cheng, H.-W., et al., *Origin and differentiation trajectories of fibroblastic reticular cells in the splenic white pulp*. Nature Communications, 2019. **10**(1): p. 1739.
148. Goc, J., et al., *Characteristics of tertiary lymphoid structures in primary cancers*. Oncoimmunology, 2013. **2**(12): p. e26836-e26836.
149. Tang, H., et al., *Lymphotoxin signalling in tertiary lymphoid structures and immunotherapy*. Cellular & molecular immunology, 2017. **14**(10): p. 809-818.
150. Germain, C., S. Gnjjatic, and M.-C. Dieu-Nosjean, *Tertiary Lymphoid Structure-Associated B Cells are Key Players in Anti-Tumor Immunity*. Frontiers in Immunology, 2015. **6**(67).

151. Krausgruber, T., et al., *Structural cells are key regulators of organ-specific immune responses*. Nature, 2020.
152. Davidson, S., et al., *Single-Cell RNA Sequencing Reveals a Dynamic Stromal Niche That Supports Tumor Growth*. Cell Rep, 2020. **31**(7): p. 107628.
153. de Jong, M.M.E., et al., *The multiple myeloma microenvironment is defined by an inflammatory stromal cell landscape*. Nature Immunology, 2021. **22**(6): p. 769-780.
154. Chen, Z., et al., *Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma*. Nature Communications, 2020. **11**(1): p. 5077.
155. Parikh, K., et al., *Colonic epithelial cell diversity in health and inflammatory bowel disease*. Nature, 2019. **567**(7746): p. 49-55.
156. Kinchen, J., et al., *Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease*. Cell, 2018. **175**(2): p. 372-386.e17.
157. Smillie, C.S., et al., *Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis*. Cell, 2019. **178**(3): p. 714-730.e22.
158. Friščić, J., et al., *The complement system drives local inflammatory tissue priming by metabolic reprogramming of synovial fibroblasts*. Immunity.
159. Martin, J.C., et al., *Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy*. Cell, 2019. **178**(6): p. 1493-1508.e20.
160. Solé-Boldo, L., et al., *Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming*. Communications Biology, 2020. **3**(1): p. 188.
161. Korsunsky, I., et al., *Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases*. bioRxiv, 2021: p. 2021.01.11.426253.
162. Buechler, M.B., et al., *Cross-tissue organization of the fibroblast lineage*. Nature, 2021. **593**(7860): p. 575-579.
163. Weyl, A., et al., *Expression of the adhesion molecules ICAM-1, VCAM-1, and E-selectin and their ligands VLA-4 and LFA-1 in chronic venous leg ulcers*. J Am Acad Dermatol, 1996. **34**(3): p. 418-23.
164. Ekwall, A.K., et al., *The tumour-associated glycoprotein podoplanin is expressed in fibroblast-like synoviocytes of the hyperplastic synovial lining layer in rheumatoid arthritis*. Arthritis Res Ther, 2011. **13**(2): p. R40.
165. Honma, M., et al., *Podoplanin expression in wound and hyperproliferative psoriatic epidermis: regulation by TGF- β and STAT-3 activating cytokines, IFN- γ , IL-6, and IL-22*. J Dermatol Sci, 2012. **65**(2): p. 134-40.
166. Astarita, J., S. Acton, and S. Turley, *Podoplanin: emerging functions in development, the immune system, and cancer*. Frontiers in Immunology, 2012. **3**(283).
167. Kumar, V., et al., *A Dendritic-Cell-Stromal Axis Maintains Immune Responses in Lymph Nodes*. Immunity, 2015. **42**(4): p. 719-730.
168. Jensen, S.S. and M. Gad, *Differential induction of inflammatory cytokines by dendritic cells treated with novel TLR-agonist and cytokine based cocktails: targeting dendritic cells in autoimmunity*. Journal of Inflammation, 2010. **7**(1): p. 37.
169. Kocks, J.R., et al., *Regulatory T cells interfere with the development of bronchus-associated lymphoid tissue*. J Exp Med, 2007. **204**(4): p. 723-34.

170. Mueller, C.G., et al., *Cellular and Vascular Components of Tertiary Lymphoid Structures*. Methods Mol Biol, 2018. **1845**: p. 17-30.
171. Wiede, F., et al., *TNF-dependent overexpression of CCL21 is an underlying cause of progressive lymphoaccumulation in generalized lymphoproliferative disorder*. European journal of immunology, 2007. **37**: p. 351-7.
172. Pikor, N.B., et al., *Integration of Th17- and Lymphotoxin-Derived Signals Initiates Meningeal-Resident Stromal Cell Remodeling to Propagate Neuroinflammation*. Immunity, 2015. **43**(6): p. 1160-73.
173. Lavoie, T.N., et al., *Expression of interleukin-22 in Sjögren's syndrome: significant correlation with disease parameters*. Scandinavian journal of immunology, 2011. **74**(4): p. 377-382.
174. Zhang, Y., et al., *Autotaxin through lysophosphatidic acid stimulates polarization, motility, and transendothelial migration of naive T cells*. J Immunol, 2012. **189**(8): p. 3914-24.
175. van Hamburg, J.P. and S.W. Tas, *Molecular mechanisms underpinning T helper 17 cell heterogeneity and functions in rheumatoid arthritis*. Journal of Autoimmunity, 2018. **87**: p. 69-81.
176. Kopitar-Jerala, N., *The Role of Interferons in Inflammation and Inflammasome Activation*. Front Immunol, 2017. **8**: p. 873.
177. McNab, F., et al., *Type I interferons in infectious disease*. Nat Rev Immunol, 2015. **15**(2): p. 87-103.
178. Hosokawa, Y., et al., *Cytokines differentially regulate ICAM-1 and VCAM-1 expression on human gingival fibroblasts*. Clin Exp Immunol, 2006. **144**(3): p. 494-502.
179. Spoelstra, F.M., et al., *Interferon-gamma and interleukin-4 differentially regulate ICAM-1 and VCAM-1 expression on human lung fibroblasts*. Eur Respir J, 1999. **14**(4): p. 759-66.
180. Denton, A.E., et al., *Type I interferon induces CXCL13 to support ectopic germinal center formation*. J Exp Med, 2019. **216**(3): p. 621-637.
181. Sonar, S. and G. Lal, *Role of Tumor Necrosis Factor Superfamily in Neuroinflammation and Autoimmunity*. Front Immunol, 2015. **6**: p. 364.
182. Monaco, C., et al., *Anti-TNF therapy: past, present and future*. Int Immunol, 2015. **27**(1): p. 55-62.
183. Hosokawa, Y., et al., *Cytokines differentially regulate ICAM-1 and VCAM-1 expression on human gingival fibroblasts*. Clinical and experimental immunology, 2006. **144**(3): p. 494-502.
184. Machowicz, A., et al., *Mediterranean diet and risk of Sjögren's syndrome*. Clin Exp Rheumatol, 2020. **38 Suppl 126**(4): p. 216-221.
185. Angerer, P., et al., *Single cells make big data: New challenges and opportunities in transcriptomics*. Current Opinion in Systems Biology, 2017. **4**: p. 85-91.
186. Luecken, M.D. and F.J. Theis, *Current best practices in single-cell RNA-seq analysis: a tutorial*. Molecular Systems Biology, 2019. **15**(6): p. e8746.
187. Gehring, J., et al., *Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins*. Nature Biotechnology, 2020. **38**(1): p. 35-38.
188. Kobschull, J.M. and A.M. Zador, *Cellular barcoding: lineage tracing, screening and beyond*. Nature Methods, 2018. **15**(11): p. 871-879.

189. Chen, W., et al., *UMI-count modeling and differential expression analysis for single-cell RNA sequencing*. Genome Biology, 2018. **19**(1): p. 70.
190. Brennecke, P., et al., *Accounting for technical noise in single-cell RNA-seq experiments*. Nature Methods, 2013. **10**(11): p. 1093-1095.
191. Bellman, R., *Dynamic Programming*. Science, 1966. **153**(3731): p. 34-37.
192. Alvarez, M., et al., *Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM*. Scientific Reports, 2020. **10**(1): p. 11019.
193. Young, M.D. and S. Behjati, *SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data*. GigaScience, 2020. **9**(12).
194. Wolock, S.L., R. Lopez, and A.M. Klein, *Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data*. Cell Systems, 2019. **8**(4): p. 281-291.e9.
195. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony*. Nature Methods, 2019. **16**(12): p. 1289-1296.
196. Bergen, V., et al., *Generalizing RNA velocity to transient cell states through dynamical modeling*. bioRxiv, 2019: p. 820936.
197. Fisher, B.A., et al., *Standardisation of labial salivary gland histopathology in clinical trials in primary Sjögren's syndrome*. Annals of the Rheumatic Diseases, 2017. **76**(7): p. 1161-1168.
198. Alunno, A., et al., *Lymphoma and Lymphomagenesis in Primary Sjögren's Syndrome*. Frontiers in medicine, 2018. **5**: p. 102-102.
199. Yoshimi, R., et al., *Clinical and pathological roles of Ro/SSA autoantibody system*. Clinical & developmental immunology, 2012. **2012**: p. 606195-606195.
200. Fragkioudaki, S., C.P. Mavragani, and H.M. Moutsopoulos, *Predicting the risk for lymphoma development in Sjogren syndrome: An easy tool for clinical use*. Medicine, 2016. **95**(25): p. e3766-e3766.
201. Efremova, M., et al., *CellPhoneDB v2.0: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes*. bioRxiv, 2019: p. 680926.
202. Bartoschek, M., et al., *Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing*. Nature Communications, 2018. **9**(1): p. 5150.
203. Stzepourginski, I., et al., *CD34⁺ mesenchymal cells are a major component of the intestinal stem cells niche at homeostasis and after injury*. Proceedings of the National Academy of Sciences, 2017. **114**(4): p. E506-E513.
204. Sidney, L.E., et al., *Concise review: evidence for CD34 as a common marker for diverse progenitors*. Stem Cells, 2014. **32**(6): p. 1380-9.
205. Sitnik, Katarzyna M., et al., *Context-Dependent Development of Lymphoid Stroma from Adult CD34⁺ Adventitial Progenitors*. Cell Reports, 2016. **14**(10): p. 2375-2388.
206. Colafrancesco, S., *"Role of Germinal Centers in Sjögren's Syndrome: a descriptive and functional study"*. 2018.
207. Cobb, B.L., et al., *Genes and Sjögren's syndrome*. Rheum Dis Clin North Am, 2008. **34**(4): p. 847-68, vii.
208. Tokunaga, R., et al., *CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy*. Cancer treatment reviews, 2018. **63**: p. 40-47.

209. Hosoya, S., et al., *Innate immune responses involving natural killer and natural killer T cells promote liver regeneration after partial hepatectomy in mice*. American Journal of Physiology-Gastrointestinal and Liver Physiology, 2013. **304**(3): p. G293-G299.
210. Herold, T., et al., *High expression of MZB1 predicts adverse prognosis in chronic lymphocytic leukemia, follicular lymphoma and diffuse large B-cell lymphoma and is associated with a unique gene expression signature*. Leuk Lymphoma, 2013. **54**(8): p. 1652-7.
211. Weissmiller, A.M., et al., *Inhibition of MYC by the SMARCB1 tumor suppressor*. Nature Communications, 2019. **10**(1): p. 2014.
212. Kohashi, K. and Y. Oda, *Oncogenic roles of SMARCB1/INI1 and its deficient tumors*. Cancer science, 2017. **108**(4): p. 547-552.
213. Gonzalez, M., et al., *The sequential role of lymphotoxin and B cells in the development of splenic follicles*. J Exp Med, 1998. **187**(7): p. 997-1007.
214. Catalan, V., et al., *Activation of noncanonical Wnt signaling through WNT5A in visceral adipose tissue of obese subjects is related to inflammation*. J Clin Endocrinol Metab, 2014. **99**(8): p. E1407-17.
215. Hai, B., et al., *Wnt/beta-catenin signaling regulates postnatal development and regeneration of the salivary gland*. Stem Cells Dev, 2010. **19**(11): p. 1793-801.
216. Mikels, A.J. and R. Nusse, *Wnts as ligands: processing, secretion and reception*. Oncogene, 2006. **25**(57): p. 7461-7468.
217. Fernandez-Torres, J., et al., *Risk of Wnt/beta-catenin signalling pathway gene polymorphisms in primary Sjogren's syndrome*. Rheumatology (Oxford), 2020. **59**(2): p. 418-425.
218. Perez, P., et al., *Salivary epithelial cells: an unassuming target site for gene therapeutics*. Int J Biochem Cell Biol, 2010. **42**(6): p. 773-7.
219. Moutsopoulos, H.M., *Sjögren's Syndrome: Autoimmune Epithelitis*. Clinical Immunology and Immunopathology, 1994. **72**(2): p. 162-165.
220. Aqrabi, L.A., et al., *Ductal epithelial expression of Ro52 correlates with inflammation in salivary glands of patients with primary Sjögren's syndrome*. Clin Exp Immunol, 2014. **177**(1): p. 244-52.
221. Katsiogiannis, S., R. Tenta, and F.N. Skopouli, *Autoimmune epithelitis (Sjögren's syndrome); the impact of metabolic status of glandular epithelial cells on auto-immunogenicity*. Journal of Autoimmunity, 2019. **104**: p. 102335.
222. Sheridan, C., *Glial cells on the radar*. Nature Biotechnology, 2009. **27**(2): p. 114-116.
223. Rao, D.A., et al., *Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis*. Nature, 2017. **542**(7639): p. 110-114.
224. Ibrahim, H.M., *B cell dysregulation in primary Sjögren's syndrome: A review*. Japanese Dental Science Review, 2019. **55**(1): p. 139-144.
225. Chan, M.A., et al., *FCER2 (CD23) Asthma-Related Single Nucleotide Polymorphisms Yields Increased IgE Binding and Egr-1 Expression in Human B Cells*. American Journal of Respiratory Cell and Molecular Biology, 2014. **50**(2): p. 263-269.
226. Agematsu, K., et al., *B cell subpopulations separated by CD27 and crucial collaboration of CD27+ B cells and helper T cells in immunoglobulin production*. European Journal of Immunology, 1997. **27**(8): p. 2073-2079.

227. Vogler, M., *BCL2A1: the underdog in the BCL2 family*. Cell Death & Differentiation, 2012. **19**(1): p. 67-74.
228. Gottenberg, J.E., et al., *Efficacy of rituximab in systemic manifestations of primary Sjogren's syndrome: results in 78 patients of the AutoImmune and Rituximab registry*. Ann Rheum Dis, 2013. **72**(6): p. 1026-31.
229. Liu, X., et al., *Metabotropic glutamate receptor 3 is involved in B-cell-related tumor apoptosis*. Int J Oncol, 2016. **49**(4): p. 1469-78.
230. Leah, E., *Long-lived plasma cells in Sjögren's syndrome—a salivary survival niche?* Nature Reviews Rheumatology, 2011. **7**(3): p. 130-130.
231. Szyszko, E.A., et al., *Phenotypic diversity of peripheral blood plasma cells in primary Sjögren's syndrome*. Scand J Immunol, 2011. **73**(1): p. 18-28.
232. Stegelmeier, A.A., et al., *Myeloid Cells during Viral Infections and Inflammation*. Viruses, 2019. **11**(2).
233. Crowley, T., et al., *Priming in response to pro-inflammatory cytokines is a feature of adult synovial but not dermal fibroblasts*. Arthritis Res Ther, 2017. **19**(1): p. 35.
234. Kündig, T.M., et al., *Fibroblasts as efficient antigen-presenting cells in lymphoid organs*. Science, 1995. **268**(5215): p. 1343-7.
235. Öhlund, D., et al., *Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer*. J Exp Med, 2017. **214**(3): p. 579-596.
236. Peduto, L., et al., *Inflammation recapitulates the ontogeny of lymphoid stromal cells*. J Immunol, 2009. **182**(9): p. 5789-99.
237. Lokeshwar, B.L., G. Kallifatidis, and J.J. Hoy, *Chapter One - Atypical chemokine receptors in tumor cell growth and metastasis*, in *Advances in Cancer Research*, A.K. Shukla, Editor. 2020, Academic Press. p. 1-27.
238. García-Cuesta, E.M., et al., *The Role of the CXCL12/CXCR4/ACKR3 Axis in Autoimmune Diseases*. Front Endocrinol (Lausanne), 2019. **10**: p. 585.
239. Cocuzzi, E.T., et al., *Upregulation of DAF (CD55) on orbital fibroblasts by cytokines. Differential effects of TNF-β and TNF-α*. Current Eye Research, 2001. **23**(2): p. 86-92.
240. Hoek, R.M., et al., *Deletion of either CD55 or CD97 ameliorates arthritis in mouse models*. Arthritis & Rheumatism, 2010. **62**(4): p. 1036-1042.
241. Aoki, J., *Mechanisms of lysophosphatidic acid production*. Semin Cell Dev Biol, 2004. **15**(5): p. 477-89.
242. Hwang, Y.S., et al., *Secretion of IL-6 and IL-8 from lysophosphatidic acid-stimulated oral squamous cell carcinoma promotes osteoclastogenesis and bone resorption*. Oral Oncol, 2012. **48**(1): p. 40-8.
243. Song, J., et al., *Autotaxin (lysoPLD/NPP2) protects fibroblasts from apoptosis through its enzymatic product, lysophosphatidic acid, utilizing albumin-bound substrate*. Biochemical and Biophysical Research Communications, 2005. **337**(3): p. 967-975.
244. Ji, Z., et al., *Inflammatory regulatory network mediated by the joint action of NF-κB, STAT3, and AP-1 factors is involved in many human cancers*. Proceedings of the National Academy of Sciences, 2019. **116**(19): p. 9453-9462.
245. Li, X., et al., *Overexpression of SerpinE2/protease nexin-1 Contribute to Pathological Cardiac Fibrosis via increasing Collagen Deposition*. Scientific Reports, 2016. **6**(1): p. 37635.

246. Santoro, A., et al., *SERPINE2 Inhibits IL-1 α -Induced MMP-13 Expression in Human Chondrocytes: Involvement of ERK/NF- κ B/AP-1 Pathways*. PloS one, 2015. **10**(8): p. e0135979-e0135979.
247. Leehan, K.M., et al., *Minor salivary gland fibrosis in Sjögren's syndrome is elevated, associated with focus score and not solely a consequence of aging*. Clinical and experimental rheumatology, 2018. **36 Suppl 112**(3): p. 80-88.
248. Gigante, B., et al., *The interaction between coagulation factor 2 receptor and interleukin 6 haplotypes increases the risk of myocardial infarction in men*. PloS one, 2010. **5**(6): p. e11300-e11300.
249. Valin, A., et al., *IL6/sIL6R regulates TNF α -inflammatory response in synovial fibroblasts through modulation of transcriptional and post-transcriptional mechanisms*. BMC Molecular and Cell Biology, 2020. **21**(1): p. 74.
250. Gao, S., et al., *Overexpression of protease nexin-1 mRNA and protein in oral squamous cell carcinomas*. Oral Oncology, 2008. **44**(3): p. 309-313.
251. Gao, G., et al., *Coagulation factor 2 thrombin receptor promotes malignancy in glioma under SOX2 regulation*. Aging, 2020. **12**(11): p. 10594-10613.
252. Hu, D., et al., *Vascular Smooth Muscle Cells Contribute to Atherosclerosis Immunity*. Frontiers in immunology, 2019. **10**: p. 1101-1101.
253. Hsu, T., H.-H. Nguyen-Tran, and M. Trojanowska, *Active roles of dysfunctional vascular endothelium in fibrosis and cancer*. Journal of Biomedical Science, 2019. **26**(1): p. 86.
254. Pober, J.S. and W.C. Sessa, *Evolving functions of endothelial cells in inflammation*. Nat Rev Immunol, 2007. **7**(10): p. 803-15.
255. Guzzo, R.M., et al., *Regulated expression and temporal induction of the tail-anchored sarcolemmal-membrane-associated protein is critical for myoblast fusion*. Biochem J, 2004. **381**(Pt 3): p. 599-608.
256. Osman, I., et al., *TEAD1 (TEA Domain Transcription Factor 1) Promotes Smooth Muscle Cell Proliferation Through Upregulating SLC1A5 (Solute Carrier Family 1 Member 5)-Mediated Glutamine Uptake*. Circulation Research, 2019. **124**(9): p. 1309-1322.
257. Grossi, M., et al., *Vascular smooth muscle cell proliferation depends on caveolin-1-regulated polyamine uptake*. Biosci Rep, 2014. **34**(6): p. e00153.
258. Ahmed, S.R., et al., *Prostaglandin D₂ Regulates CD4⁺ Memory T Cell Trafficking across Blood Vascular Endothelium and Primes These Cells for Clearance across Lymphatic Endothelium*. The Journal of Immunology, 2011. **187**(3): p. 1432-1439.
259. Perrot, C.Y., et al., *Prostaglandin E2 breaks down pericyte-endothelial cell interaction via EP1 and EP4-dependent downregulation of pericyte N-cadherin, connexin-43, and R-Ras*. Sci Rep, 2020. **10**(1): p. 11186.
260. Pober, J.S. and W.C. Sessa, *Evolving functions of endothelial cells in inflammation*. Nature Reviews Immunology, 2007. **7**(10): p. 803-815.
261. Koltai, T., S.J. Reshkin, and S. Harguindey, *Chapter 7 - Carbonic anhydrases*, in *An Innovative Approach to Understanding and Treating Cancer: Targeting pH*, T. Koltai, S.J. Reshkin, and S. Harguindey, Editors. 2020, Academic Press. p. 157-176.
262. Annan, D.A., et al., *Carbonic anhydrase 2 (CAII) supports tumor blood endothelial cell survival under lactic acidosis in the tumor microenvironment*. Cell Communication and Signaling, 2019. **17**(1): p. 169.

263. Liu, X., et al., *Semaphorin 3G Provides a Repulsive Guidance Cue to Lymphatic Endothelial Cells via Neuropilin-2/PlexinD1*. Cell Rep, 2016. **17**(9): p. 2299-2311.
264. Kutschera, S., et al., *Differential Endothelial Transcriptomics Identifies Semaphorin 3G as a Vascular Class 3 Semaphorin*. Arteriosclerosis, Thrombosis, and Vascular Biology, 2011. **31**(1): p. 151-159.
265. Dorwart, M.R., et al., *The solute carrier 26 family of proteins in epithelial ion transport*. Physiology (Bethesda), 2008. **23**: p. 104-14.
266. Basu, M.K., et al., *Exome Sequencing Identifies Abnormalities in Glycosylation and ANKRD36C in Patients with Immune-Mediated Thrombotic Thrombocytopenic Purpura*. Thromb Haemost, 2021. **121**(4): p. 506-517.
267. Cong, T., et al., *[Exome sequencing of gastric cancers screened the differences of clinicopathological phenotypes between the mutant and the wide-type of frequently mutated genes]*. Zhonghua Yi Xue Za Zhi, 2018. **98**(28): p. 2242-2245.
268. Dijkema, T., et al., *MUC5B levels in submandibular gland saliva of patients treated with radiotherapy for head-and-neck cancer: a pilot study*. Radiation oncology (London, England), 2012. **7**: p. 91-91.
269. Teshima, T.H.N., et al., *Development of human minor salivary glands: expression of mucins according to stage of morphogenesis*. Journal of Anatomy, 2011. **219**(3): p. 410-417.
270. Redman, R.S., *Myoepithelium of salivary glands*. Microscopy Research and Technique, 1994. **27**(1): p. 25-45.
271. Zhang, Y., et al., *Notch activation leads to loss of myoepithelial differentiation and poor outcome in solid adenoid cystic carcinoma*. Oral Diseases, 2020. **26**(8): p. 1677-1686.
272. Jessen, K.R. and R. Mirsky, *The origin and development of glial cells in peripheral nerves*. Nature Reviews Neuroscience, 2005. **6**(9): p. 671-682.
273. Sommakia, S. and O.J. Baker, *Neurons Self-Organize Around Salivary Epithelial Cells in Novel Co-Culture Model*. J Stem Cell Regen Biol, 2016. **2**(2).
274. Karabiyik, A., A.B. Peck, and C.Q. Nguyen, *The important role of T cells and receptor expression in Sjögren's syndrome*. Scandinavian journal of immunology, 2013. **78**(2): p. 157-166.
275. Ramos-Casals, M., et al., *Th1/Th2 cytokine imbalance in patients with Sjögren syndrome secondary to hepatitis C virus infection*. Semin Arthritis Rheum, 2002. **32**(1): p. 56-63.
276. Nguyen, C.Q., et al., *Differential gene expression in the salivary gland during development and onset of xerostomia in Sjögren's syndrome-like disease of the C57BL/6.NOD-Aec1Aec2 mouse*. Arthritis Res Ther, 2009. **11**(2): p. R56.
277. Pontarini, E., et al., *Unique expansion of IL-21+ Tfh and Tph cells under control of ICOS identifies Sjögren's syndrome with ectopic germinal centres and MALT lymphoma*. Ann Rheum Dis, 2020. **79**(12): p. 1588-1599.
278. Christodoulou, M.I., et al., *Foxp3+ T-regulatory cells in Sjogren's syndrome: correlation with the grade of the autoimmune lesion and certain adverse prognostic factors*. Am J Pathol, 2008. **173**(5): p. 1389-96.
279. Raphael, I., R.R. Joern, and T.G. Forsthuber, *Memory CD4(+) T Cells in Immunity and Autoimmune Diseases*. Cells, 2020. **9**(3).
280. Sanjuan, M.A., D. Sagar, and R. Kolbeck, *Role of IgE in autoimmunity*. J Allergy Clin Immunol, 2016. **137**(6): p. 1651-1661.

281. Arce, S., et al., *CD38 low IgG-secreting cells are precursors of various CD38 high-expressing plasma cell populations*. J Leukoc Biol, 2004. **75**(6): p. 1022-8.
282. Ahmed, R., et al., *A Public BCR Present in a Unique Dual-Receptor-Expressing Lymphocyte from Type 1 Diabetes Patients Encodes a Potent T Cell Autoantigen*. Cell, 2019. **177**(6): p. 1583-1599.e16.
283. Gabrilovich, D.I. and S. Nagaraj, *Myeloid-derived suppressor cells as regulators of the immune system*. Nature reviews. Immunology, 2009. **9**(3): p. 162-174.
284. Collin, M., N. McGovern, and M. Haniffa, *Human dendritic cell subsets*. Immunology, 2013. **140**(1): p. 22-30.
285. Jarjour, M., et al., *Fate mapping reveals origin and dynamics of lymph node follicular dendritic cells*. The Journal of experimental medicine, 2014. **211**(6): p. 1109-1122.
286. García-Cuesta, E.M., et al., *The Role of the CXCL12/CXCR4/ACKR3 Axis in Autoimmune Diseases*. Frontiers in Endocrinology, 2019. **10**(585).
287. Kim, S.Y., et al., *Inhibition of the CXCR4/CXCL12 chemokine pathway reduces the development of murine pulmonary metastases*. Clinical & Experimental Metastasis, 2008. **25**(3): p. 201-211.
288. Gauld, S.B. and J.C. Cambier, *Src-family kinases in B-cell development and signaling*. Oncogene, 2004. **23**(48): p. 8001-8006.
289. Mah, S., et al., *Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia*. Mol Psychiatry, 2006. **11**(5): p. 471-8.
290. Hardy, W.R., et al., *Transcriptional Networks in Single Perivascular Cells Sorted from Human Adipose Tissue Reveal a Hierarchy of Mesenchymal Stem Cells*. Stem Cells, 2017. **35**(5): p. 1273-1289.
291. Corselli, M., et al., *The tunica adventitia of human arteries and veins as a source of mesenchymal stem cells*. Stem Cells Dev, 2012. **21**(8): p. 1299-308.
292. Kadl, A. and N. Leitinger, *The role of endothelial cells in the resolution of acute inflammation*. Antioxid Redox Signal, 2005. **7**(11-12): p. 1744-54.
293. Polito, A.J. and D. Proud, *Epithelial cells as regulators of airway inflammation*. Journal of Allergy and Clinical Immunology, 1998. **102**(5): p. 714-718.
294. He, M., et al., *Chloride channels regulate differentiation and barrier functions of the mammalian airway*. Elife, 2020. **9**.
295. Perkins, T.N., et al., *Differences in gene expression and cytokine production by crystalline vs. amorphous silica in human lung epithelial cells*. Part Fibre Toxicol, 2012. **9**(1): p. 6.
296. Miljkovic-Licina, M., et al., *Targeting Olfactomedin-like 3 Inhibits Tumor Growth by Impairing Angiogenesis and Pericyte Coverage*. Molecular Cancer Therapeutics, 2012. **11**(12): p. 2588-2599.
297. Sussman, J., et al., *Protein kinase B phosphorylates AHNK and regulates its subcellular localization*. J Cell Biol, 2001. **154**(5): p. 1019-30.
298. Gaczowska, A., et al., *Association of CDKAL1 nucleotide variants with the risk of non-syndromic cleft lip with or without cleft palate*. Journal of Human Genetics, 2018. **63**(4): p. 397-406.
299. Cadigan, K.M. and M.L. Waterman, *TCF/LEFs and Wnt signaling in the nucleus*. Cold Spring Harbor perspectives in biology, 2012. **4**(11): p. a007906.
300. Cao, H.Z., et al., *LGR5 promotes cancer stem cell traits and chemoresistance in cervical cancer*. Cell Death Dis, 2017. **8**(9): p. e3039.

301. Hai, B., et al., *Concurrent Transient Activation of Wnt/ β -Catenin Pathway Prevents Radiation Damage to Salivary Glands*. International Journal of Radiation Oncology*Biophysics, 2012. **83**(1): p. e109-e116.
302. van Dijk, E.M., et al., *Noncanonical WNT-5B signaling induces inflammatory responses in human lung fibroblasts*. Am J Physiol Lung Cell Mol Physiol, 2016. **310**(11): p. L1166-76.
303. Mitsiadis, T.A., et al., *Expression of Notch 1, 2 and 3 is regulated by epithelial-mesenchymal interactions and retinoic acid in the developing mouse tooth and associated with determination of ameloblast cell fate*. J Cell Biol, 1995. **130**(2): p. 407-18.
304. Qin, B., et al., *Notch activation enhances microglial CX3CR1/P38 MAPK pathway in rats model of vincristine-induced peripheral neuropathy*. Neuroscience Letters, 2020. **715**: p. 134624.
305. Gustavsson, M., et al., *Structural basis of ligand interaction with atypical chemokine receptor 3*. Nature Communications, 2017. **8**(1): p. 14135.
306. Rustenhoven, J., et al., *TGF-beta1 regulates human brain pericyte inflammatory processes involved in neurovasculature function*. J Neuroinflammation, 2016. **13**: p. 37.
307. Mourcin, F., et al., *Follicular lymphoma triggers phenotypic and functional remodeling of the human lymphoid stromal cell landscape*. Immunity.
308. Wei, K., et al., *Notch signalling drives synovial fibroblast identity and arthritis pathology*. Nature, 2020.
309. Tsukita, S. and M. Furuse, *Pores in the wall: claudins constitute tight junction strands containing aqueous pores*. J Cell Biol, 2000. **149**(1): p. 13-6.
310. Perrella, M.A., M.K. Jain, and M.E. Lee, *Role of TGF-beta in vascular development and vascular reactivity*. Miner Electrolyte Metab, 1998. **24**(2-3): p. 136-43.
311. Wang, T., et al., *CD97, an adhesion receptor on inflammatory cells, stimulates angiogenesis through binding integrin counterreceptors on endothelial cells*. Blood, 2005. **105**(7): p. 2836-2844.
312. Barro-Soria, R., et al., *KCNE1 and KCNE3 modulate KCNQ1 channels by affecting different gating transitions*. Proceedings of the National Academy of Sciences, 2017. **114**(35): p. E7367-E7376.
313. Sohn, M., et al., *Ahnak promotes tumor metastasis through transforming growth factor- β -mediated epithelial-mesenchymal transition*. Sci Rep, 2018. **8**(1): p. 14379.
314. Lucchesi, D. and M. Bombardieri, *The role of viruses in autoreactive B cell activation within tertiary lymphoid structures in autoimmune diseases*. Journal of leukocyte biology, 2013. **94**.
315. Liu, Y., et al., *Single-Cell Profiling Reveals Divergent, Globally Patterned Immune Responses in Murine Skin Inflammation*. iScience, 2020. **23**(10).
316. Zhang, L., et al., *Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer*. Cell, 2020. **181**(2): p. 442-459.e29.
317. Regard, L., et al., *Effective control of *Staphylococcus aureus* lung infection despite tertiary lymphoid structure disorganisation*. European Respiratory Journal, 2021. **57**(4): p. 2000768.

Appendix

1.0: Seurat Packages required for scRNAseq analysis

library (Seurat)

library(plyr)

library(dplyr)

library (Matrix)

library(tenxutils)

library(reshape2)

library(harmony)

library (SingleR)

library(ggplot2)

library(slingshot)

library (RColorBrewer)

library(gsfisher)

library(clustree)

library(pheatmap)

1.1: Set variables used throughout

```
tenxdir <- "//its-rds.bham.ac.uk/rdsprojects/b/baronef-  
dtaa/Saba_sjogrens_10x/FASTQ/aggr/all_batches/scrub.agg.processed.dir"  
  
outdir <- "//its-rds.bham.ac.uk/rdsprojects/b/baronef-  
dtaa/Saba_sjogrens_10x/FASTQ/aggr/all_batches/scrub.agg.processed.dir/seurat_jdt"  
  
if(!dir.exists(outdir)){  
  dir.create(outdir)}  
  
project <- "PSSvSICCA_10xAll_scrubbed"  
  
mincells <- 3  
  
mingenes <- 200  
  
meta <- file.path(tenxdir, "metadata.tsv")  
  
groupby <- "sample_id"  
  
qcmingenes <- 200  
  
qcmaxpercentmito <- 0.20  
  
latentvars <- "nUMI,percent.mito"  
  
latent.vars <- strsplit(latentvars, ",")[[1]]  
  
modeluse <- "linear" #Regress out variation
```


1.2: Declare functions

```
getCellNumbers <- function(s, cell_numbers="none", stage="input",
groupby="sample_id") {
  _cell_info <- getCellInfo(s)
  counts <- as.data.frame(table(s@meta.data[[groupby]]))
  colnames(counts) <- c(groupby, stage)
  rownames(counts) <- counts[[groupby]]
  counts[[groupby]] <- NULL
  if ( identical(cell_numbers, "none") ) {
    result <- counts
    colnames(counts) <- stage} else {
    cnames <- c(colnames(cell_numbers), stage)
    result <- cbind(cell_numbers, counts[[stage]])
    colnames(result) <- cnames}
  return(result)}

getExpressedGenesFromSeuratObject <- function(seurat_object,
clusters,
min.pct=0.1)
{ expressed <- c()
  for(cluster in clusters)}
```

[1.3: Get genes detected in the cluster](#)

```
cluster_cells <- names(seurat_object@ident[seurat_object@ident==cluster])  
  
clust_pcts <- apply(seurat_object@data[,cluster_cells],  
1, function(x) sum(x>0)/length(x))  
  
detected_in_clust <- names(clust_pcts[clust_pcts>min.pct])
```

[1.4: Get genes detected in the other cells](#)

```
other_cells <- names(seurat_object@ident[seurat_object@ident!=cluster])  
  
other_pcts <- apply(seurat_object@data[,other_cells],  
1, function(x) sum(x>0)/length(x))  
  
detected_in_other_cells <- names(other_pcts[other_pcts>min.pct])  
  
expressed <- c(expressed, detected_in_clust, detected_in_other_cells)}  
  
expressed <- unique(expressed)  
  
expressed}
```

[1.5: Read in data](#)

```
stats <- list()  
  
cat("Importing matrix from: ", tenxdir, " ... ")  
  
data <- Read10X(tenxdir)  
  
cat("Done.\n")
```

[1.6: Retaining ensembl IDs and identifying hsapiens orthologs](#)

```
inFile <- file.path(tenxdir, "genes.tsv")

cat("Importing gene information from: ", inFile, " ... ")

genes <- read.table(inFile, as.is=TRUE)

cat("Done.\n")

colnames(genes) <- c("gene_id", "gene_name")

genes$seurat_id <- as.character(make.unique(genes$gene_name))

rownames(genes) <- genes$seurat_id

write.table(

genes, file.path(outdir, "annotation.txt"),

sep="\t", col.names=TRUE, row.names=FALSE, quote=FALSE)
```

[1.7: Create seurat object](#)

```
cat("Creating Seurat object ... ")

s <- CreateSeuratObject(

raw.data=data, min.cells=mincells, min.genes=mingenes,

## do.scale=F,

## do.center=F,

#save.raw=T,

project=project)

cat("Done.\n")

s@misc <- genes
```

1.8: Read in the metadata

```
if ( identical(meta, "none") ) {  
  stop("No metadata file given")}  
  
cat("Importing metadata ... ")  
  
metadata <- read.table(meta, sep="\t", header=TRUE, as.is=TRUE)  
  
cat("Done.\n")  
  
if (!"barcode" %in% colnames(metadata)) {  
  stop('Mandatory "barcode" column missing from the metadata')  
  
rownames(metadata) <- metadata$barcode  
  
metadata$barcode <- NULL  
  
metadata <- metadata[s@cell.names, ]  
  
if (!identical(rownames(metadata), s@cell.names)) {  
  ## print out some debugging information  
  
  cat("Count of rownames in metadata: ", length(rownames(metadata)), "\n")  
  
  cat("Count of cell.names in Seurat object: ", length(s@cell.names), "\n")  
  
  cat("--\n")  
  
  cat("First rownames in metadata:\n")  
  print(head(rownames(metadata)))  
  
  cat("First cell.names in Seurat object:\n")  
  print(head(s@cell.names))  
  
  cat("--\n")  
  
  cat("Last rownames in metadata:\n")  
  print(tail(rownames(metadata)))  
  
  cat("Last cell.names in Seurat object:\n")  
  print(tail(s@cell.names))  
  
  stop("Metadata barcode field does not match cell.names")  
  
  cat("Adding meta data ... ")  
  
  s <- AddMetaData(s, metadata)  
  
  cat("Done.\n")  
  
  if (!groupby %in% colnames(s@meta.data)) {
```

```
stop(paste("The specified grouping factor:",  
groupby,  
"was not found in the metadata",  
sep=" "))
```

[1.9: QC filtering](#)

```
mito.genes <- grep("^MT-", rownames(s@data), value=TRUE, ignore.case=TRUE)

percent.mito <- Matrix::colSums(s@data[mito.genes, ]) / Matrix::colSums(s@data)

s <- AddMetaData(s, percent.mito, "percent.mito")

gp <- VlnPlot(
  s, c("nGene", "nUMI", "percent.mito"), size.title.use=14, size.x.use=12,
  group.by=groupby, x.lab.rot=TRUE, point.size.use=0.1, nCol=3)

gp

plot_fn <- function() {
  par(mfrow=c(1, 2))
  GenePlot(s, "nUMI", "percent.mito", cex.use=1)
  GenePlot(s, "nUMI", "nGene", cex.use=1)
  par(mfrow=c(1, 1))}

plot_fn()

cell_numbers <- getCellNumbers(s, groupby=groupby)
stats$no_cells <- ncol(s@data)
stats$qc_min_gene_threshold <- qcmingenes
stats$qc_max_percent_mito_threshold <- qcmaxpercentmito
s <- FilterCells(
  s, subset.names=c("nGene", "percent.mito"),
  low.thresholds=c(qcmingenes, -Inf),
  high.thresholds=c(Inf, qcmaxpercentmito))
stats$no_cells_after_qc <- ncol(s@data)
cell_numbers <- getCellNumbers(s, cell_numbers=cell_numbers, stage="after_qc_filters",
  groupby=groupby)
cat("Data dimensions after subsetting:\n")
print(dim(s@data))
```

[1.10: Subsetting data](#)

```
s <- SetIdent(s, ident.use = s@meta.data$sample_group)

pss <- SubsetData(s, ident.use = c("pss"))

sic <- SubsetData(s, ident.use = c("sicca"))

##### PSS pipeline #####
#####

#Normalise data

pss <- NormalizeData(
  object=pss, normalization.method="LogNormalize", scale.factor=10E3)
```

[1.11: Perform regression without cell cycle correction](#)

```
latent.vars <- strsplit(latentvars, ", ")[[1]]

print(latent.vars)

pss <- ScaleData(object=pss, vars.to.regress=latent.vars,
  model.use=modeluse)
```

[1.12: Downsample to lowest cell number](#)

```
par(mar=c(6,6,3,3))

barplot(table(pss@meta.data$sample_id),
  las=2, ylab = "Cell counts", main="Cell# before downsampling")

pss <- SetIdent(object = pss, ident.use = pss@meta.data$sample_id)

min.cells <- min(cell_numbers <- getCellNumbers(pss, groupby='sample_id'))

pss <- SubsetData(pss, ident.use = levels(pss@ident), max.cells.per.ident = min.cells)

par(mar=c(6,6,3,3))

barplot(table(pss@ident), las=2, ylab = "Cell counts", main="Cell# after downsampling")

stats$no_cells_after_downsample_pss <- ncol(pss@data)
```

[1.13: Identify variable genes](#)

```
pss <- FindVariableGenes(pss, y.cutoff = 0.5)

xthreshold <- 0.1

xx <- pss@hvg.info

xx$var.gene = FALSE

xx$var.gene[rownames(xx) %in% pss@var.genes] <- TRUE

xsm <- melt(xx, id.vars=c("var.gene", "gene.mean"))

gp <- ggplot(xsm, aes(gene.mean, value, color=var.gene))

gp <- gp + scale_color_manual(values=c("black", "red"))

gp <- gp + geom_point(alpha = 1, size=0.5)

gp <- gp + facet_wrap(~variable, scales="free")

gp <- gp + theme_classic()

if(xthreshold > 0){

  gp <- gp + geom_vline(xintercept=xthreshold, linetype="dashed", color="blue")}

gp

cat("no. variable genes: ", length(pss@var.genes), "\n")

stats$no_variable_genes_pss <- length(pss@var.genes)

print(stats)
```

[1.14: Dimensionality reduction](#)

```
# perform PCA using the variable genes

pss <- RunPCA(pss, pc.genes=pss@var.genes, pcs.compute=50, do.print=FALSE)

n_cells_pca <- min(1000, length(pss@cell.names))

DimPlot(pss, reduction.use = "pca", group.by = "sample_id")
```


1.15: Write out heatmaps showing the genes loading the first 12 components

```
plot_fn <- function() {  
  PCHeatmap(  
    pss, pc.use=1:12, cells.use=n_cells_pca, do.balanced=TRUE,  
    label.columns=FALSE, cexRow=0.8, use.full=FALSE )}  
plot_fn()  
nPCs <- min(dim(pss@dr$pca@cell.embeddings)[2],50)  
PCElbowPlot(pss, num.pc=nPCs)  
pss <- JackStraw(pss, num.replicate=100,  
  num.pc = nPCs)  
pss <- JackStrawPlot(pss, PCs=1:nPCs)  
gp <- pss@dr$pca@misc$jackstraw.plot  
gp
```

1.16: Harmony for Seurat v2

```
harmonyEmbed <- HarmonyMatrix(pss@dr$pca@cell.embeddings,  
pss@meta.data,  
c("seq_id"),  
do_pca = FALSE,0,  
theta = c(2),  
max.iter.harmony = 20,  
max.iter.cluster = 100,  
plot_convergence = T,  
return_object = FALSE)  
rownames(harmonyEmbed) <- row.names(pss@dr$pca@cell.embeddings)  
colnames(harmonyEmbed) <- paste0("Harmony", seq_len(ncol(harmonyEmbed)))  
pss <- Seurat::SetDimReduction(object = pss,  
reduction.type = "harmony",  
slot = "cell.embeddings",  
new.data = harmonyEmbed)  
pss <- Seurat::SetDimReduction(object = pss,  
reduction.type = "harmony",  
slot = "key",  
new.data = "harmony")  
pss <- Seurat::ProjectDim(pss, reduction.type = "harmony", dims.print = 1:5,  
dims.store = 30, genes.print = 30, replace.dim = TRUE,  
do.center = FALSE, do.print = TRUE, assay.type = "RNA")
```

[1.17: Generating TSNE](#)

```
pss <- RunTSNE(pss, reduction.use = "harmony", dims.use = 1:30,  
do.fast = T, perplexity = 100)
```

[1.18: Generating UMAP](#)

```
pss <- RunUMAP(pss, reduction.use = "harmony", dims.use = 1:30)  
DimPlot(pss, reduction.use = "umap")
```

[1.19: Find clusters](#)

```
pss <- FindClusters(pss, reduction.type = "harmony",  
resolution = 0.57, dims.use = 1:30, force.recalc = TRUE,  
print.output = FALSE)
```

```
DimPlot(pss, reduction.use = "umap", do.label = T)
```

[1.20: Plotting key markers using Violin \(Vln\) plots](#)

```
VlnPlot(pss, c("PDGFRB", "PDGFRA", "NOTCH3", "THY1", "VCAM1", "CD248"))
```

```
#Stromal
```

```
VlnPlot(pss, c("MT-CO3", "MT-ND3", "MT-ND4L", "MT-ND4", "MT-ND5", "MT-ND6"), x.lab.rot = T)
```

```
pss <- RenamIdent(pss, old.ident.name = 0,  
new.ident.name = "Epithelium1")
```

```
DimPlot(pss, reduction.use = "umap", do.label = T)
```

```
All.markers.pss <- FindAllMarkers(object = pss, only.pos = TRUE, min.pct = 0.25,  
thresh.use = 0.25)
```

[1.21: Generating Subclusters](#)

```
fls.pss <- SubsetData(pss, ident.use = c("Fibroblast"))

fls.pss <- FindClusters(fls.pss, reduction.type = "harmony",
resolution = 0.35,dims.use = 1:30, force.recalc = TRUE,
print.output = FALSE)

DimPlot(fls.pss, reduction.use = "umap")

All.markers.fls.pss <- FindAllMarkers(object = fls.pss, only.pos = TRUE, min.pct = 0.25, thresh.use =
0.25)

write.csv(All.markers.fls.pss, file.path(outdir, "All.markers.fls.pss.05.all.csv"))

index <- match(names(fls.pss@ident),names(pss@ident))

ident <- as.character(pss@ident)

ident[index] <- as.character(fls.pss@ident)

pss <- SetIdent(pss, ident.use = ident)

DimPlot(pss, reduction.use = "umap", do.label = T)

rm(fls.pss)
```

[1.22: Build clustree and validate clusters](#)

```
pss <- BuildClusterTree(pss)

pss@meta.data$set_idents <- pss@ident

pss <- BuildSNN(pss, reduction.type = "harmony", dims.use = 1:30, plot.SNN = T)

pss <- ValidateClusters(pss, pc.use=1:30)
```

[1.23: Saving seurat objects](#)

```
saveRDS(pss, "pss_scrub_subcluster_preAmbientClass_200315.RDS")
```

1.24: Fibroblast/mural enrichment rerun post fine clustering

#Ensure to run DEG for only pos markers

```
pss.mural.fibro <- SubsetData(pss, ident.use = c("Fibroblast SERPINE2+F2R+",  
Fibroblast ACKR3+CLEC3B+CD55+",  
"Fibroblast C7+APOC1+",  
"Fibroblast CCL19+",  
"Pericyte",  
"Vascular smooth muscle"))  
DimPlot(pss.mural.fibro, reduction.use = "umap")  
All.markers.mural.fibro.pss <- FindAllMarkers(pss.mural.fibro,  
test.use = "negbinom",  
only.pos = T,  
latent.vars = "soup_prob")  
All.markers.mural.fibro.pss$cluster <- factor(All.markers.mural.fibro.pss$cluster)  
levels(All.markers.mural.fibro.pss$cluster)  
levels(pss.mural.fibro@ident)  
GSout <- paste(outdir, "Fibroblast_GSFisher", sep = " ")  
if (dir.exists(GSout)){  
  print("Directory already exists")  
} else{  
  dir.create(GSout)}  
index <- match(All.markers.mural.fibro.pss$gene, pss@misc$seurat_id)  
All.markers.mural.fibro.pss$ensembl <- pss@misc$gene_id[index]  
FilteredGeneID <- as.data.frame(rownames(pss.mural.fibro@data))  
index <- match(FilteredGeneID$`rownames(pss.mural.fibro@data)` , pss@misc$seurat_id)  
FilteredGeneID$ensemblUni <- pss@misc$gene_id[index]  
for(cluster in levels(pss.mural.fibro@ident)){  
  cat("Running go enrichment for cluster", cluster, "...")  
  Go <- runGO(foreground_ids =  
All.markers.mural.fibro.pss$ensembl[All.markers.mural.fibro.pss$cluster == cluster],
```

```

background_ids = FilteredGenelD$ensemblUni,
gene_id_type = "ensembl", species = "hs")
cat("Running Kegg enrichment for cluster", cluster, "...")

Kegg <- runKEGG(foreground_ids =
All.markers.mural.fibro.pss$ensembl[All.markers.mural.fibro.pss$cluster == cluster],
background_ids = FilteredGenelD$ensemblUni,
gene_id_type = "ensembl", species = "hs")
cat("Joining results tables for cluster", cluster, "...")
results <- full_join(Go, Kegg)
results$p.adj <- p.adjust(results$p.val, method = "BH")
cat("Filtering genesets for ", cluster, "...")
filtered_genesets <- filterGenesets(results,
min_foreground_genes = 2,
max_genes_geneset = 500,
min_odds_ratio = 1.5,
padjust_method="BH",
use_adjusted_pvalues=TRUE,
pvalue_threshold=0.05)
filtered_genesets$cluster <- cluster
cat("Writing out results table for ", cluster, "...")
write.csv(filtered_genesets, file.path(GSout, paste("Geneset Enrichment pss",
cluster,"csv", sep=".")))}

All_enrichment_path = list.files(path=GSout, pattern="Geneset Enrichment pss*", full.names=TRUE)
All_enrichments.pss = ldply(All_enrichment_path, read.csv)
All_enrichments.pss <- All_enrichments.pss[,-1]
write.csv(All_enrichments.pss, file.path(GSout, paste("All.enrichments.pss.fibro.mural", "csv",
sep=".")))

```

1.25: Plotting cell populations

```
All_enrichments.bp <- All_enrichments.pss[grep("BP", All_enrichments.pss$ontology),]

All_enrichments.bp$cluster <- factor(All_enrichments.bp$cluster, levels=c("Fibroblast
SERPINE2+F2R+",
"Fibroblast ACKR3+CLEC3B+CD55+",
"Fibroblast C7+APOC1+",
"Fibroblast CCL19+",
"Pericyte",
"Vascular smooth muscle"))

# Top.Enrich <- All_enrichments.bp %>% group_by(cluster) %>% top_n(10, p.adj)

# Top.Enrich <- Top.Enrich[!duplicated(Top.Enrich$geneset_id),]

# levels(Top.Enrich$cluster)

gp <- sampleEnrichmentDotplot(All_enrichments.bp,

selected_genesets =
unique(All_enrichments.bp$geneset_id[grep("*wnt*|*Wnt*|*WNT*|*notch*|*Notch*|*NOTCH*",
All_enrichments.bp$description)]),

selection_col = "geneset_id",

sample_id_col = "cluster",

sample_levels = levels(All_enrichments.bp$cluster),

min_dot_size = 1, max_dot_size = 6,

maxl = 45,

pvalue_threshold = 0.05,

title = "geneset",

rotate_sample_labels = T

gp

gp <- sampleEnrichmentDotplot(All_enrichments.bp,

selected_genesets = unique(All_enrichments.bp$geneset_id[grep("*kappa*|*tumor necrosis*",
All_enrichments.bp$description)]),

selection_col = "geneset_id",

sample_id_col = "cluster",

sample_levels = levels(All_enrichments.bp$cluster),
```

```

min_dot_size = 1, max_dot_size = 6,

maxl = 45,

pvalue_threshold = 0.05,

title = "geneset",

rotate_sample_labels = T)

gp

Kegg.enrichment <- All_enrichments.pss[is.na(All_enrichments.pss$ontology),]

Top.Enrich <- Kegg.enrichment %>% group_by(cluster) %>% top_n(10, p.adj)

Top.Enrich <- Top.Enrich[!duplicated(Top.Enrich$geneset_id),]

Top.Enrich$cluster <- factor(Top.Enrich$cluster)

levels(Top.Enrich$cluster)

gp <- sampleEnrichmentDotplot(Kegg.enrichment,

selected_genesets = Top.Enrich$geneset_id,

selection_col = "geneset_id",

sample_id_col = "cluster",

sample_levels = levels(Kegg.enrichment$cluster),

min_dot_size = 1, max_dot_size = 6,

maxl = 45,

pvalue_threshold = 0.05,

title = "geneset"

gp

```


[1.26: Cluster identity assignments](#)

```
cluster_assignments <- data.frame(barcodes=as.character(names(pss@ident)),  
cluster_id=as.character(pss@ident))  
write.table(cluster_assignments,  
gzfile(file.path(outdir, "pss_cluster_assignments.txt.gz")),  
sep="\t", col.names=T, row.names=F, quote=F)
```

[1.27: Selecting cluster colours](#)

```
cluster_colors <- gg_color_hue(length(levels(pss@ident)))  
write.table(cluster_colors, file=file.path(outdir, "pss_cluster_colors.txt"),  
quote=FALSE, col.names = FALSE, row.names = FALSE)
```

1.28: Exporting data for Cibersortx

```
pss_data <- pss@data  
  
colnames(pss_data) <- pss@ident  
  
#####Check for duplicated rownames and remove duplicates based simply on order of occurrence  
  
annot <- read.table(paste(outdir,"annotation.txt", sep = "/"),header = T)  
  
index <- match(rownames(pss_data), annot$seurat_id)  
  
genes.inc <- annot[index,]  
  
table(duplicated(genes.inc$gene_id))  
  
index <- match(genes.inc$seurat_id, rownames(pss_data))  
  
pss_data <- pss_data[index,]  
  
dim(pss_data)  
  
dim(genes.inc)  
  
table(rownames(pss_data) == genes.inc$seurat_id)  
  
rownames(pss_data) <- genes.inc$gene_id  
  
#Subset data to hvg due to size of output  
  
index <- match(pss@var.genes, genes.inc$seurat_id)  
  
ens.use <- genes.inc$gene_id[index]  
  
pss_data <- pss_data[ens.use,]  
  
pss_data <- cbind(c(rownames(pss_data)), as.matrix(pss_data))  
  
colnames(pss_data) <- c("GeneSymbol", colnames(pss_data)[-1])  
  
write.table(as.matrix(pss_data),  
  
file.path(outdir, "PSS_exprsFinal_Cibersortx.txt"),  
  
sep = "\t",row.names = F)
```

[1.29: GSEA plots](#)

```
All_enrichments_pss_fibro_mural <- read_csv("//its-rds.bham.ac.uk/nr-castles/projects/b/baronef-dtaa/Saba_sjogrens_10x/FASTQ/aggr/all_batches/scrub.agg.processed.dir/seurat_jdt_Fibroblast_GS_Fisher/All.enrichments.pss.fibro.mural.csv")
```

```
All_enrichments.pss <- All_enrichments_pss_fibro_mural
```

```
All_enrichments.pss$cluster <- factor(All_enrichments.pss$cluster, levels=c("Fibroblast  
SERPINE2+F2R+",
```

```
"Fibroblast ACKR3+CLEC3B+CD55+",
```

```
"Fibroblast C7+APOC1+",
```

```
"Fibroblast CCL19+",
```

```
"Pericyte",
```

```
"Vascular smooth muscle"))
```

```
gp <- sampleEnrichmentDotplot(All_enrichments.pss,
```

```
selected_genesets =
```

```
unique(All_enrichments.pss$geneset_id[grepl("*Wnt*|*Notch*|*BMP*|*TGF*|*transforming  
growth factor*|*Apelin*|*Hippo*", All_enrichments.pss$description)]),
```

```
selection_col = "geneset_id",
```

```
sample_id_col = "cluster",
```

```
sample_levels = levels(All_enrichments.pss$cluster),
```

```
min_dot_size = 1, max_dot_size = 6,
```

```
maxl = 45,
```

```
pvalue_threshold = 0.05,
```

```
title="geneset",
```

```
rotate_sample_labels = T)
```

```
gp
```

```
gp <- sampleEnrichmentDotplot(All_enrichments.pss,
```

```
selected_genesets =
```

```
unique(All_enrichments.pss$geneset_id[grepl("*interferon*|*MHC*|*interleukin*",  
All_enrichments.pss$description)]),
```

```
selection_col = "geneset_id",
```

```
sample_id_col = "cluster",
```

```
sample_levels = levels(All_enrichments.pss$cluster),
```

```
min_dot_size = 1, max_dot_size = 6,
```

```

maxl = 45,

pvalue_threshold = 0.05,

title="geneset",

rotate_sample_labels = T)

gp

gp <- sampleEnrichmentDotplot(All_enrichments.pss,

selected_genesets = unique(All_enrichments.pss$geneset_id[grep("*TNF*|*tumor necrosis*|*NF\kappaB*", All_enrichments.pss$description)])),

selection_col = "geneset_id",

sample_id_col = "cluster",

sample_levels = levels(All_enrichments.pss$cluster),

min_dot_size =1, max_dot_size = 6,

maxl = 45,

pvalue_threshold = 0.05,

title="geneset",

rotate_sample_labels = T)

gp

gp <- sampleEnrichmentDotplot(All_enrichments.pss,

selected_genesets = unique(All_enrichments.pss$geneset_id[grep("*extracellular matrix*", All_enrichments.pss$description)])),

selection_col = "geneset_id",

sample_id_col = "cluster",

sample_levels = levels(All_enrichments.pss$cluster),

min_dot_size =1, max_dot_size = 6,

maxl = 45,

pvalue_threshold = 0.05,

title="geneset",

rotate_sample_labels = T)

gp

gp <- sampleEnrichmentDotplot(All_enrichments.pss,

```

```

selected_genesets = unique(All_enrichments.pss$geneset_id[grepl("*complement activation*",
All_enrichments.pss$description)]),
selection_col = "geneset_id",
sample_id_col = "cluster",
sample_levels = levels(All_enrichments.pss$cluster),
min_dot_size = 1, max_dot_size = 6,
maxl = 45,
pvalue_threshold = 0.05,
title="geneset",
rotate_sample_labels = T)

```

[1.30: Removal of ambient signatures](#)

```

pss.clean <- SubsetData(pss, ident.remove = grep("Ambient", levels(pss@ident), value = T))

DimPlot(pss.clean, reduction.use = "umap", do.label=TRUE, no.legend=TRUE)

#Identify variable genes

pss.clean <- FindVariableGenes(pss.clean, y.cutoff = 0.5)

#Dimension reduction

perform PCA using the variable genes

pss.clean <- RunPCA(pss.clean, pc.genes=pss.clean@var.genes, pcs.compute=50, do.print=FALSE)

```

[1.31: Calculating Relative Cell Proportions](#)

```

fibro.numbers <- clus.cell.numbers[grepl("Fibroblast", rownames(clus.cell.numbers)),]

fibro.props <- prop.table(as.matrix(fibro.numbers), margin=2)

```

[1.32: Proportion per cluster by Rho pos or neg](#)

```
pss@meta.data$Rho <- pss@meta.data$sample_id

pss@meta.data$Rho <- gsub("PSS_A", "Rho_n", pss@meta.data$Rho)
pss@meta.data$Rho <- gsub("PSS_B", "Rho_n", pss@meta.data$Rho)
pss@meta.data$Rho <- gsub("PSS_D", "Rho_p", pss@meta.data$Rho)
pss@meta.data$Rho <- gsub("PSS_E", "Rho_n", pss@meta.data$Rho)
pss@meta.data$Rho <- gsub("PSS_F", "Rho_n", pss@meta.data$Rho)
pss@meta.data$Rho <- gsub("PSS_G", "Rho_p", pss@meta.data$Rho)
pss@meta.data$Rho <- gsub("PSS_H", "Rho_p", pss@meta.data$Rho)

clus.cell.numbers <- as.data.frame.matrix(table(pss@ident, pss@meta.data$sample_id))
clus.props <- prop.table(as.matrix(clus.cell.numbers), margin=2)
colnames(clus.props) <- c("Rho_n", "Rho_n", "Rho_p", "Rho_n", "Rho_n", "Rho_p", "Rho_p")
long.clus.props <- melt(clus.props)
colnames(long.clus.props) <- c("Cluster", "Rho", "Proportion")
ggplot(dat=long.clus.props, aes(fill=Rho,y=Proportion,x = Cluster)) +
  geom_bar(position="dodge", stat="identity") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
ggplot(dat=long.clus.props, aes(y=Proportion,x = Rho)) +
  geom_boxplot() +
  geom_jitter(height=0, width=0.1) +
  facet_wrap(~ Cluster)
pdf("Rho_proportions.pdf")
for(cluster in unique(long.clus.props$Cluster)){
  plotting <- long.clus.props[long.clus.props$Cluster %in% cluster,]
  p <- ggplot(dat=plotting, aes(y=Proportion,x = Rho)) +
    geom_boxplot() +
    geom_jitter(height=0, width=0.1)+
    ggtitle(cluster)
  print(p)}
dev.off()
```

```

long.plasma <- long.clus.props[grepl("Plasma", long.clus.props$Cluster),]
ggplot(long.plasma, aes(y=Proportion,x = Rho)) +
geom_boxplot(fill=NA, outlier.alpha = 0) +
geom_jitter(dat=long.plasma, aes(colour=Cluster),height=0, width=0.1)
ggplot(long.plasma, aes(y=Proportion,x = Rho, fill=Cluster)) +
geom_boxplot(outlier.alpha = 0) +
geom_point(position=position_jitterdodge())
clus.cell.numbers <- as.data.frame.matrix(table(pss@ident, pss@meta.data$sample_id))
clus.props <- prop.table(as.matrix(clus.cell.numbers), margin=2)
long.clus.props <- melt(clus.props)
colnames(long.clus.props) <- c("Cluster", "Sample", "Proportion")
long.plasma <- long.clus.props[grepl("Plasma", long.clus.props$Cluster),]
ggplot(long.plasma, aes(x=Sample, y=Proportion))+
geom_bar(stat="identity") +
facet_wrap(~ Cluster) +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

```

1.33: Pseudobulk and correlate between mouse and human pss cell clusters

#Generate pseudobulk per cluster for the pss fibros

```
clusterIds <- pss.fibro@ident
matrixUMI <- pss.fibro@data
stopifnot(is.factor(clusterIds))
stopifnot(!is.null(names(clusterIds)))
stopifnot(
  all(names(clusterIds) %in% colnames(matrixUMI)))
clusterCountMatrix <- matrix(
  nrow = nrow(matrixUMI),
  ncol = nlevels(clusterIds),
  dimnames = list(
    gene_ids = rownames(matrixUMI),
    cluster = paste("cluster", levels(clusterIds), sep="_")))
for (clusterId in levels(clusterIds)){
  cat("Processing cluster:", clusterId, "...")
  cellsInCluster <- names(clusterIds)[which(clusterIds == clusterId)]
  pseudobulkCount <- rowSums(matrixUMI[, cellsInCluster])
  clusterCountMatrix[, paste("cluster", clusterId, sep = "_")] <-
    pseudobulkCount[rownames(clusterCountMatrix)]
  cat("Done.\n")}
pss.pseudo <- clusterCountMatrix
```


1.34: Generate pseudobulk for the murine day15 fibroblasts

```
clusterIds <- Day15_fibro@iden
matrixUMI <- Day15_fibro@data
stopifnot(is.factor(clusterIds))
stopifnot(!is.null(names(clusterIds)))
stopifnot(
  all(names(clusterIds) %in% colnames(matrixUMI)))
clusterCountMatrix <- matrix(
  nrow = nrow(matrixUMI),
  ncol = nlevels(clusterIds),
  dimnames = list(
    gene_ids = rownames(matrixUMI),
    cluster = paste("cluster", levels(clusterIds), sep="_")
  ))
for (clusterId in levels(clusterIds)){
  cat("Processing cluster:", clusterId, "...")
  cellsInCluster <- names(clusterIds)[which(clusterIds == clusterId)]
  pseudobulkCount <- rowSums(matrixUMI[, cellsInCluster])
  clusterCountMatrix[, paste("cluster", clusterId, sep = "_")] <-
  pseudobulkCount[rownames(clusterCountMatrix)]
  cat("Done.\n")
}
day15.pseudo <- clusterCountMatrix
```

1.35: Finding orthologs and matching identities

```
ensembl = useMart("ensembl",dataset="hsapiens_gene_ensembl")

searchAttributes(mart = ensembl, pattern = "ortho")

attributes = listAttributes(ensembl)

human_genes_ens <- pss.fibro@misc$gene_id[pss.fibro@misc$seurat_id %in%
rownames(pss.fibro@data)]

annot <- getBM(attributes = c("ensembl_gene_id",
"mmusculus_homolog_associated_gene_name",
"mmusculus_homolog_orthology_type"),
filters = "ensembl_gene_id",
values = human_genes_ens,
mart = ensembl)

annot <- annot[annot$mmusculus_homolog_orthology_type=="ortholog_one2one",]

index <- match(annot$ensembl_gene_id, pss.fibro@misc$gene_id)

merged <- pss.fibro@misc
merged$ens <- NA
merged$mm_gene <- NA
merged$seurat[index] <- annot$ensembl_gene_id
merged$mm_gene[index] <- annot$mmusculus_homolog_associated_gene_name
merged <- merged[!is.na(merged$mm_gene),]
merged$mm_seurat <- Day15_fibro@misc$seurat_id[match(merged$mm_gene,
Day15_fibro@misc$gene_name)]
merged <- merged[!is.na(merged$mm_seurat),]
merged <- merged[merged$seurat_id %in% rownames(pss.pseudo),]
merged <- merged[merged$mm_seurat %in% rownames(day15.pseudo),]

pss.pseudo.sub <- pss.pseudo[rownames(pss.pseudo) %in% merged$seurat_id,]
day15.pseudo.sub <- day15.pseudo[rownames(day15.pseudo) %in% merged$mm_seurat,]
indexH <- match(merged$seurat_id, rownames(pss.pseudo.sub))
pss.pseudo.sub <- pss.pseudo.sub[indexH,]
indexM <- match(merged$mm_seurat,rownames(day15.pseudo.sub))
```

```
day15.pseudo.sub <- day15.pseudo.sub[indexM,]  
rownames(day15.pseudo.sub) <- rownames(pss.pseudo.sub)  
summary(rownames(pss.pseudo.sub) == rownames(day15.pseudo.sub))  
corRes <- cor(pss.pseudo.sub, day15.pseudo.sub)  
tiff(filename = "PssFibroDay15FibroOtholog.cor.tiff", width = 1028,height = 1028)  
corrplot::corrplot(corRes,  
  tl.cex = 1,  
  cl.cex = 1,  
  tl.col = "black",  
  is.corr = F)  
dev.off()
```

Publications:

1. **Asam S**, Nayar S, Gardner D, Barone F. Stromal cells in tertiary lymphoid structures: Architects of autoimmunity. *Immunol Rev.* 2021 Jul;302(1):184-195. doi: 10.1111/imr.12987. Epub 2021 May 31. PMID: 34060101.
2. **Asam S**, Neag G, Berardicurti O, Gardner D, Barone F. The role of stroma and epithelial cells in primary Sjögren's syndrome. *Rheumatology (Oxford).* 2019 Mar 3;kez050. doi: 10.1093/rheumatology/kez050. Epub ahead of print. PMID: 30945742.
3. Nayar S, Campos J, Smith CG, Iannizzotto V, Gardner DH, Mourcin F, Roulois D, Turner J, Sylvestre M, **Asam S**, Glaysher B, Bowman SJ, Fearon DT, Filer A, Tarte K, Luther SA, Fisher BA, Buckley CD, Coles MC, Barone F. Immunofibroblasts are pivotal drivers of tertiary lymphoid structure formation and local pathology. *Proc Natl Acad Sci U S A.* 2019 Jul 2;116(27):13490-13497. doi: 10.1073/pnas.1905301116. Epub 2019 Jun 18. PMID: 31213547; PMCID: PMC6613169.