# Single-cell characterisation of T lymphocyte immune responses in spondyloarthropathy

Frank Penkava

St. Hilda's college

Thesis submitted to the University of Oxford towards fulfilment of

the degree of DPhil in molecular and cellular medicine

# I. Abstract

The spondyloarthropathies (SpA) are a collection of inflammatory disorders of undetermined etiology affecting the joints and connective tissues, however existing evidence suggests a role for the microbiome, CD4 and CD8 T cells in disease pathogenesis.

A CD154 activation based functional assay was first used to quantify and characterise CD4+ memory T cell populations reactive to a panel of 13 microbes found at mucosal barrier sites, predominantly gut bacteria. Peripheral blood mononuclear cells (PBMC) from 31 SpA, 17 rheumatoid arthritis (RA) and 14 healthy controls, as well as synovial fluid mononuclear cells (SFMC) from 9 SpA patients were separated and stimulated overnight with a panel of microbial lysates. Cells were then stained for CD154, TNFα, IFNγ, IL-17A, GM-CSF, IL-22 and IL-2 and analysed by flow cytometry. Single-cell RNA sequencing (scRNA-seq) was then used to characterise *S. typhimurium*-reactive synovial Th memory cells from a psoriatic arthritis (PsA) patient, in addition to *ex vivo* CD4 and CD8 memory T cells from the peripheral blood and synovial fluid of 6 PsA patients.

The frequency of Th memory cells reactive to microbes capable of inducing a type 17/17.1 response were enriched in SpA synovial fluid relative to blood, and scRNA-seq analysis revealed a trajectory along which Th17.1 cells could differentiate into a "stem-like" memory pool capable of proliferation or into a Th17 regulatory phenotype. Pronounced clonal expansions of *ex vivo* CD8 memory T-cells within the joints of PsA patients were also identified by scRNA-seq. They exhibited distinct gene expression profiles including cycling, activation, tissue homing and tissue residency markers. Pseudotime analysis of these clonal CD8 populations identified trajectories in which tissue residency can represent an intermediate developmental state giving rise to activated, cycling and exhausted CD8 populations. Comparing T cell clonality across patients further revealed specificity convergence of clones against a putative common antigen.

A role for gut microbes in SpA pathogenesis was supported by the higher frequency of Th17 cells reactive to *Salmonella typhimurium* found in SpA synovial fluid compared with SpA peripheral blood, and characterisation of these cells revealed trajectories of differentiation which could potentially be targeted to skew Th17.1 responses towards a regulatory phenotype. The identification of pro-inflammatory and clonally expanded CD8 T cells by scRNA-seq also suggests an antigen driven expansion of pathogenic CD8 T cells in PsA.

# II. Declaration

I declare that this DPhil thesis titled "Single-cell characterisation of T lymphocyte immune responses in spondyloarthropathy" is entirely my own work, and that any contributions made by collaborators has been clearly acknowledged. I have not submitted any part of this work for any other degree or qualification at the University of Oxford or elsewhere.

Frank Penkava

Trinity Term 2019

Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences

Botnar Research Centre

University of Oxford

# III. Publications arising from this thesis

Penkava, Frank, Martin Del Castillo Velasco-Herrera, Matthew D. Young, Nicole Yager, Alicia Lledo Lara, Charlotte Guzzo, Ash Maroof, et al. “Single-Cell Sequencing Reveals a Clonal Expansion of pro-Inflammatory Synovial CD8 T Cells Expressing Tissue Homing Receptors in Psoriatic Arthritis.” *BioRxiv*, July 16, 2019, 704494. https://doi.org/10.1101/704494.[1] (under review at Nature Communications journal)

# IV. Acknowledgements

Firstly I would like to acknowledge my main supervisor Professor Paul Bowness. I could not have hoped to find a better mentor, and I consider myself extremely lucky to have been able to study and research under his guidance. I would also like to thank my co-supervisor Professor Fiona Powrie, whose wealth of knowledge and experience helped guide my research, and Dr. Stephen Sansom, for his support and helpful advice regarding the technically challenging field of single-cell sequencing. I also acknowledge the generous funding of the Kennedy Institute of Rheumatology, and I thank them for providing me with the opportunity to pursue this DPhil at the University of Oxford.

Special thanks must go to Dr. Hussein Al-Mossawi, who had the drive and ambition to co-ordinate a large-scale single-cell sequencing project, and had enough faith to give me the opportunity of taking part in it. We battled through much R code together, and it was a pleasure to work with him. I similarly thank others involved in the project, particularly Dr. Nicole Yager and Dr. Davide Simone from our lab, and Dr. Sam Behjati from the Wellcome Sanger Institute along with members of his group Dr. Martin Del Castillo and Dr. Mathew Young. I also thank Prof. Ahmed Hegazy for all his help with setting up the CD154 assay. For making the last 4 years such a wonderful experience, I also thank Anna for showing me the ropes; Liye for our insightful discussions; Hui, Nancy, Laura, Claudia and India for their help in the lab and great all-round teamwork; and Chao for burning the midnight oil with me. I would also like to thank Dr. Andrew Knight, who supervised my MSc degree at Newcastle University and gave me encouragement to pursue a future in immunology, and to my family who have provided excellent feedback and support.

# V. Contents

# VI. List of Tables

# VII. List of figures

# VIII. List of abbreviations

| | |
|---|---|
| AAU | acute anterior uveitis |
| AS | ankylosing spondylitis |
| CD154/CD40LG | cluster of differentiation 154 |
| CD40 | cluster of differentiation 40 |
| EOMES | eomesodermin |
| ERA | juvenile enthesitis-related arthritis |
| GM-CSF | granulocyte macrophage colony-stimulating factor |
| GPR35 | G protein-coupled receptor 35 |
| GPR37 | G protein-coupled receptor 37 |
| GPR65 | G protein-coupled receptor 65 |
| HC | healthy control |
| HLA-B27 | human leukocyte antigen (HLA) B27 |
| IBD | inflammatory bowel disease |
| ICOS | Inducible T-cell costimulator |
| ICOSLG | ICOS ligand |
| IFNγ/IFNg | interferon gamma |
| IL-12B | interleukin 12 subunit beta |
| IL-17A | interleukin 17A |
| IL-2 | interleukin 2 |
| IL-22 | interleukin 22 |
| IL-23 | interleukin 23 |
| IL27 | interleukin 27 |
| IL23R | interleukin 23 receptor |
| JAK2 | janus kinase 2 |
| LOD | limit of detection |
| NKX2-3 | Homeobox protein Nkx-2.3 |
| PB | peripheral blood |
| PBMC | peripheral blood mononuclear cells |
| PsA | psoriatic spondyloarthritis |
| ReA | reactive arthritis |
| RNA | ribonucleic acid |
| RUNX3 | runt-related transcription factor 3 |
| SF | synovial fluid |
| SFMC | synovial fluid mononuclear cells |
| SpA | spondyloarthropathy |
| TBX21 | T-box transcription factor 21 |
| TCR | T-cell receptor |
| Th | T-helper |
| TNFα/TNFa | tumor necrosis factor alpha |
| TYK2 | tyrosine kinase 2 |
| ZMIZ1 | zinc finger MIZ domain-containing protein 1 |

# Chapter 1: Introduction and aims

## 1.1 The spondyloarthropathies (SpA)

The spondyloarthropathies (SpA) are a collection of inflammatory disorders affecting the joints and connective tissues, which share overlapping symptoms and genetic associations, the strongest being with MHC class I (MHC-I) genes. So strong is the association with MHC-I, that "MHC-I-opathy" has been suggested as an umbrella term covering spondyloarthropathies in addition to psoriasis and Behçet's disease.[1] Ankylosing spondylitis (AS) is the prototypic SpA, afflicting 0.3% of the adult population with an estimated heritability of 90% and a suspected ubiquitous environmental trigger[2]. Inflammation primarily affects the axial skeletal system with early sacroiliac (SI) joint involvement, but may also affect peripheral joints, the gut, and eyes (uveitis). Enthesitis, new bone formation and joint fusion are all key features of AS, which can lead to bone fusion of the spine from which the disease derives its name. Psoriatic arthritis (PsA) is also a member of the SpA family, characterised by skin inflammation in the form of psoriasis in addition to joint inflammation. PsA can be further subdivided into subgroups based on the pattern of joint involvement present. For example some patients have more axial involvement manifesting in symptoms closely aligned to AS, while large-joint oligo PsA typically involves inflammation of larger joints such as the knees and has a weaker association with HLA-B*27[3]. Reactive arthritis (ReA) is the only spondyloarthropathy to have been definitively associated with infection. It is known to occur 1 to 4 weeks after infection with bacteria from any of the following genera: *Salmonella*, *Chlamydia*, *Campylobacter*, *Shigella* or *Yersinia*. Symptoms similar to those of other SpA can manifest in ReA, including large joint, axial, skin and eye inflammation, but usually remit after 4 to 6 months. However,

15-30% of ReA patients, particularly those bearing the HLA-B*27 gene, go on to develop a chronic form of the disease, with 12-24% eventually being diagnosed with AS [4].

Recent years have seen advances in our knowledge of SpA associated genetic pathways, SpA microbiome environments and the success of anti IL-17A and anti IL-23 treatment options for SpA. Much of the knowledge we have gained was dependent on advancements and increased affordability of genetic sequencing technology, enabling large scale GWAS and metagenomic microbiome studies. Despite these advances, and although several plausible models of SpA diseases have been suggested, they remain unproven, and the aetiology of SpA remains unknown. The recent association of Type 17 immune pathways, MHC-II genes and the efficacy of anti-IL17A/IL-23 treatment supports a role for CD4+ T helper (Th) cells in SpA. However GWAS studies have also discovered additional genes related to antigen presentation by MHC class I, and the extremely strong association of SpA with HLA-B27 continues to implicate CD8 T cell involvement.

This thesis takes advantage of single-cell sequencing technology as well as multi-colour fluorescence based FACS analysis to investigate at high resolution the role of CD4+ and CD8+ memory T-cells in SpA, with a focus on analysing PsA and AS blood and synovial fluid.

## 1.2 MHC-I, MHC-II and antigen presentation

T cell recognition of antigen is critical to the acquired immune system. Retaining "memory" to previously encountered antigen is only possible if a system exists to accurately identify antigen in a reproducible manner. Presentation of major histocompatibility complex (MHC) bound peptides to T cells with cognate T cell receptors (TCR) forms the basis of this process. This section briefly describes the relationship between MHC peptide presentation and T lymphocyte recognition of antigen.

The MHC gene region is a feature of all vertebrates[5], which in humans contains the human leukocyte antigen (HLA) genes. HLA genes are grouped into MHC class I and class II based on the ability of their encoded MHC molecules to form heterodimers capable of binding either CD8 or CD4 T cell co-receptors respectively. Figure 1-1 illustrates the current naming convention used to denote variants within the HLA coding region. Both MHC class I and class II molecules also contain a peptide binding groove which can have varying binding affinity towards different peptides. The resulting peptide bound MHC complexes (pMHC) provide an MHC "restricted" immune interface on the cell surface available for engagement by the TCR complex present on T cells.

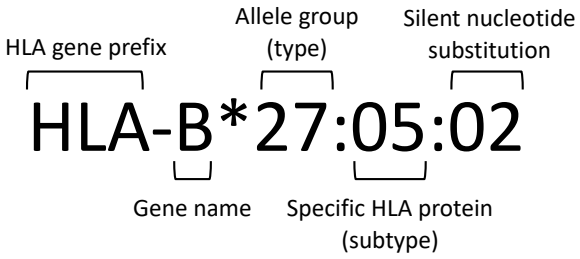


**Figure 1-1 HLA nomenclature**

MHC-I genes are co-dominantly expressed in almost all nucleated cells of the human body, and MHC-I molecules continuously sample intracellular peptides for presentation on the cell surface. Surveillance of MHC-I bound peptides by CD8 T cells provides a mechanism to identify infected or dysfunctional cells for clearance by the immune system. Professional antigen presenting cells (APC) such as dendritic cells (DC), macrophages and B cells additionally have specialized cell processes capable of internalising proteins from exogenous sources for presentation. These proteins are broken down into peptides and loaded onto MHC-II molecules in mature endosomal compartments. pMHC-II complexes are then transported to the cell surface, thus exposing exogenous antigen to immune surveillance by CD4 T cells, also called T "helper" (Th) cells. An additional mechanism, known as cross-presentation, allows DC to present exogenous antigens on MHC-I molecules, and it is also possible for MHC-II to present certain endogenous antigens[5].

T helper cells are named for their ability to support and co-ordinate the immune response of other immune cells. Th cell cytokine signalling and CD154 (CD40L) interaction plays a crucial role in assisting B and T cells to mount robust immune responses and acquire long lived memory phenotypes. Unlike B cells, CD8 T cells are not ordinarily capable of MHC-II antigen presentation, suggesting some other

mechanism exists to hone CD4 help only to CD8 T cells mounting a response to related antigen. It has been shown that DC capable of both MHC-I and MHC-II antigen presentation can to this end be "licensed" by Th cells to provide CD8 T cell help, acting as an intermediary between the two cell types[6]. Recent research has elaborated on the mechanisms underlying this process and suggests a model by which CD8 T cells are provided cognate CD4 T cell help during an infection[7][8]. S. Eickhoff and colleagues showed in mice that on initial infection with a virus, CD4 and CD8 T cells were independently activated by uninfected and infected DC respectively. These CD4 and CD8 T cells would then translocate to a special area of the paracortex and interact with a third uninfected XCR1+ DC subset presenting cognate antigen on both MHC-I and MHC-II, thereby facilitating CD4 T cell help of already activated cognate CD8 T cells. Previous work by researchers from the same team also showed that initial Th licensing of DC results in a high production of CCL3/4, capable of attracting activated CCR5+ CD8 T cells[9].

## 1.3 The T cell receptor (TCR)

In order to recognise the distinct sequences of peptide presented by MHC molecules, T cells utilize a TCR-CD3 complex consisting of a TCR heterodimer capable of ligand recognition, in addition to CD3γε, CD3δε and ζζ signalling modules[10][11]. The CD3 molecule is expressed almost exclusively by mature T cells, making it an ideal surface marker to help identify T cells. This thesis deals primarily with T cells having a TCR constructed from alpha and beta chains, which make up the vast majority of circulating human T cells. Most of these T cells recognise peptide antigen in the context of MHC class I and II presentation discussed previously in this chapter. A notable exception are invariant NK-T cells, more limited in their diversity, which can recognise lipids and glycolipids presented by CD1d molecules rather than MHC. Other T cells possessing a TCR constructed from gamma and delta chains form a very small (~5%) portion of total T cells but are concentrated in larger numbers in the gut mucosa. They also have limited diversity, are not MHC restricted and therefore do not require CD4/CD8 expression to recognise their cognate antigen[12].

In order to properly distinguish immune targets via MHC peptide presentation, a sufficiently diverse array of peptides must be recognizable in the context of an individual's MHC background. The beta chain of a TCR consists of a variable (V), diversity (D), joining (J) and constant (C) region, while the alpha chain has only V, J and C regions. A process known as V(D)J recombination[13] provides the diversity needed for an effective TCR repertoire by selecting only one gene from each pool of available V, (D for beta chain) and J genes, then combining these into a single DNA sequence. The machinery involved in this process cleaves and joins DNA, and during this process additional palindromic (p) and non-germline (n) nucleotides can be

inserted where V, D and J genes are joined, creating a total estimated potential diversity of 10^15 -10^20 TCR clonotypes. However, in reality there are only an estimated 3 x 10^11 naive T cells in the entire human body[14], so any one individual's TCR repertoire will never contain all possible variations.

The region of a TCR that comes into most contact with actual antigen and is responsible for the majority of TCR diversity is called complementarity determining region 3 (CDR3). This hypervariable region includes part of the V and J genes and encompasses the entire region they flank. In particular, all of the highly diverse joining regions resulting from V(D)J recombination are part of the CDR3 region (Figure 1-2). Most peptide contact can in fact be attributed to an even smaller sequence of approximately 6 amino acids in length within the CDR3 region of both alpha and beta chains[15]. It is based on this principle that the Grouping of Lymphocyte Interactions by Paratope Hotspots (GLIPH) algorithm was developed to identify clonotypes specific for a common antigen (discussed further in chapter 4). The next most variable regions, CDR1 and CDR2, primarily come into contact with the MHC molecule rather than the peptide it holds.

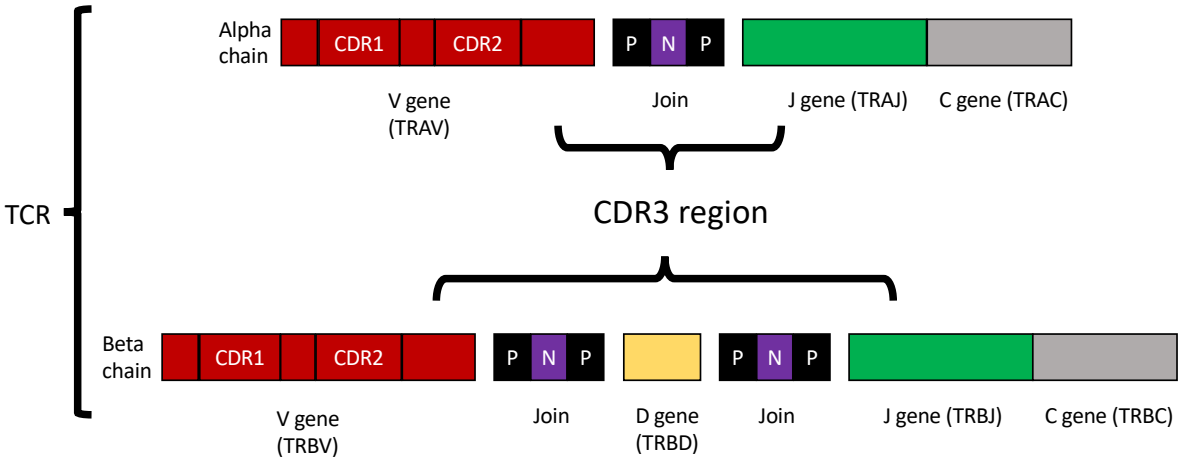


**Figure 1-2 Complementarity determining region 3**

TCR alpha and beta chains, showing the location of complementarity determining regions 1,2 and 3 after V(D)J recombination. Red: V gene; yellow: D gene; green: J gene; grey: constant gene; black: palindromic nucleotides; N: non-germline nucleotides.

## 1.4 Genetics of SpA

**Table 1-1 MHC association with disease in SpA**

| Disease | HLA-B*27 frequency[16] | Other MHC-I | MHC-II |
|---|---|---|---|
| Ankylosing spondylitis (AS) | >90 | Risk:<br>HLA-B*13:02,<br>HLA-B*40:01,<br>HLA-B*40:02,<br>HLA-B*47:01,<br>HLA-B*51:01,<br>HLA-A*02:01[17]<br>Protective:<br>HLA-B*07:02,<br>HLA B*08[17]<br>HLA-B*57:01[17] | Risk: HLA-DRB1*01:03, rs1126513 in HLA-DPB1[17] |
| Juvenile enthesitis-related arthritis (ERA) | 76 | | Risk:<br>HLA-DQA1:01:01<br>HLA-DQB1:05<br>HLA-DRB1:01[18]<br><br>Protective:<br>HLA-DPB1:02:01,<br>HLA-DRB1:07 |
| Psoriatic spondyloarthritis (PsA) | 40-50% | HLA-B*38[19],<br>HLA-A*02:01[20],<br>HLA-Cw*06:02[20],<br>HLA-B*15[21] | Risk:<br>Juvenile PsA:<br>HLA-DQA1:01:01,<br>HLA-DRB1:01[18] |
| Reactive arthritis (ReA) | 30-75 | | |
| Colitis-associated spondyloarthritis | 33-75 | | |
| Acute anterior uveitis (AAU) | 50 | | |

The MHC-I gene HLA-B*27 is the strongest single gene associated with risk of developing AS, and is also associated with developing other forms of SpA as outlined in Table 1-1[16]. Since the discovery of its association with AS more than 40 years ago[22][23], HLA typing has been refined and much larger genome wide association studies (GWAS) have been carried out. This has led to the identification of additional genes associated with SpA and a better appreciation for how variations within the HLA-B*27 gene itself can affect its ability to confer disease risk. The genetics surrounding AS was recently reviewed by Matthew Brown and colleagues[24] which the reader is urged to consult for further information on this topic.

MHC-I genes other than HLA-B*27 associated with increased risk of AS include HLA-B*13:02, HLA-B*40:01, HLA-B*40:02, HLA-B*47:01, HLA-B*51:01 (also associated with Becet's disease) and HLA-A*02:01. Conversely HLA B*07:02, HLA-B*08 and HLA B*57:01 are protective. An association with increased risk of developing PsA has been found for HLA-Cw*06:02 (also independently associated with developing Psoriasis only), HLA-B*38[19], HLA-A*02:01[20] and HLA-B*15[21].

What is particularly striking is that most HLA-B gene alleles conferring both risk and protection in AS involve modification of the amino acid residue at position 97 of the HLA-B molecule[17,24]. This position is located within the C/F pocket of the peptide binding groove and is therefore likely to influence the type of antigens able to bind HLA-B. The next strongest gene associated with AS risk is ERAP1, encoding an enzyme which trims peptides before they are loaded onto MHC-I molecules (typically able to accommodate peptides of 9-10 amino acids in length). ERAP1 is also strongly associated with uveitis[25]. Epistasis has been demonstrated between alleles of ERAP1 and both HLA-B*27 and HLA-B*40 in AS[17]; ERAP1 and HLA-Cw6 in psoriasis; and ERAP1 and HLA-B*51 in Behçet disease[24]. This strongly supports antigen processing and presentation as critical to disease pathogenesis in AS, and directly associates these 4 MHC-I-opathies[1,24]. In addition ERAP2, also encoding an aminopeptidase, has been associated with AS, Crohn's disease, psoriasis and birdshot chorioretinopathy (a form of uveitis)[24].

One possible implication of these genetic associations is that the peptides presented to CD8 T-cells in the context of an individual's MHC-I background affects the risk of developing disease. This line of reasoning would support CD8 T cells playing a key

role in SpA pathogenesis. In contrast to the risk attributed to HLA-B*27 in SpA, a protective association exists for HIV patients. HLA-B*27 positive HIV patients benefit from a higher chance of being "elite controllers", with slow disease progression and lower viral loads, sometimes remaining asymptomatic for decades[26]. The reasons behind this protective effect remain unclear. One hypothesis is that it is difficult for HIV to escape immune control of the HLA-B*27 restricted immunodominant KK10 peptide "KRWIILGLNK". The number and complexity of mutations required to do so would take a long time, therefore extending the asymptomatic period of HIV infection[26][27]. Separate evidence has shown that in contrast to other MHC class I restricted CD8 T cell responses towards HIV, HLA-B*27 (and HLA-B*57) restricted responses are characterised by an extended proliferative capacity, upregulation of TIM-3, and resistance to Treg suppression[28]. It has therefore been suggested that a similarly robust CD8 T cell response, resistant to regulation and resolution, could take place in HLA-B*27 positive SpA patients targeting an arthritogenic peptide[26]. The concept of an arthritogenic HLA-B*27 restricted "self" peptide being cross-reactive with HLA-B*27 restricted peptide from an infectious agent is the earliest attempt at explaining the association of HLA-B*27 with AS[29]. However, to date no definitive arthritogenic peptide, or auto-antigen, has been identified for any SpA. An alternative, but not incompatible theory, is based on the ability of HLA-B*27 molecules to form $Cys^{67}$-dependent homodimers ($B27_2$) in the absence of $\beta_2$-microglobulin[30]. Experiments have shown that cell surface $B27_2$ and HLA-B*27 heavy chains free of $\beta_2$-microglobulin (FHC) are able to bind KIR3DL2 receptors on T-cells with high affinity[31]. This can act as a "volume knob", enhancing the pro-inflammatory capacity of T cells[32], increasing their resistance to apoptosis and encouraging a Th17 phenotype in AS[33][16].

MHC-II genetic associations have also been discovered for AS, juvenile ERA, and a juvenile form of PsA (Table 1-1), potentially implicating CD4 T cells. AS risk associated with HLA-DRB1 and HLA-DPB1 correlates with altered amino acid residues in the peptide binding groove for these MHC-II molecules (positions 70 and 11 respectively)[17]. Thus SpA genetic associations potentially implicate antigen presentation to both CD8 and CD4 T cells in disease pathogenesis.

When comparing AS with IBD, 20 additional loci are shared with the same direction of effect, including *IL23R*, *IL12B*, *TYK2*, *JAK2* and *IL27* which point to a role for IL-23 signalling; *TBX21*, *RUNX3*, *EOMES*, *ICOSLG*, *NKX2-3* and *ZMIZ1* involved in mucosal immunity; and *GPR35*, *GPR37* and *GPR65* involved in bacterial sensing[24]. Co-familiality studies further support overlap between IBD and AS, indicating that first degree relatives of AS patients have a 3 fold higher chance of developing IBD and implying a degree of shared aetiology between these diseases[24,34]. As IBD is associated with intestinal dysbiosis, these findings suggest that the gut environment may similarly be dysregulated in SpA.

An association with IL-23 is particularly interesting as this cytokine has been found necessary for the expansion and survival of pathogenic T-helper 17 (Th17) cells [35], and both serum levels of IL-17[36] and Th17 cells[37][38] are increased in AS patients. In addition the monoclonal antibodies Secukinumab (binding IL-17A) and Ustekinumab (binding the p40 subunit of IL-23) have both proved efficacious in treating AS[39][40]. This highlights a clear role for both IL-17A and IL-23 in SpA.

## 1.5 Evidence of gut microbiome as a driver of disease in SpA

The gut has a surface area of approximately 32m$^2$ [41] exposed to a highly diverse [42][43] microbial community, initially inherited at birth from the mother and later influenced by the environment and genes of an individual [44]. It forms a critical interface between self and non-self, needing to allow the passage of nutrients while maintaining a barrier to harmful pathogens. Constant immune monitoring of this interface to foster a diverse set of beneficial microbes while limiting the expansion of harmful microbes is critical in maintaining homeostasis [45]. Differences in the relative abundance of various microbial species and taxa compared to healthy controls has now been observed for a large number of autoimmune and inflammatory diseases summarized in Table 1-2. This leads to the key question of whether gut microbial composition might be associated with pathogenesis in these diseases, and whether interventions intended to modify this composition (such as diet, probiotics, prebiotics, antibiotics, bacteriophage therapy and faecal microbiome transplant) could be a useful approach to treatment[46].

## Table 1-2 Relative abundance studies in inflammatory and autoimmune diseases

| | | | | | |
|---|---|---|---|---|---|
| AS ankylosing spondylitis | JIA-P juvenile idiopathic arthritis - polyarticular | Increase in diversity | △ | Increase in abundance | ▲ |
| PsA psoriatic arthritis | IBD inflammatory bowel disease | Decrease in diversity | ▽ | Decrease in abundance | ▼ |
| Ps psoriasis | CD-I Crohn's disease - Ileitis | No diversity difference | ◁▷ | No difference in abundance | ◀▶ |
| ReA reactive arthritis | T1D type I diabetes | | | | |
| JIA-E juvenile idiopathic arthritis - enthesis related | RA rheumatoid arthritis | | | | |

| Pylum | Class | Order | Family | Genus | Species | AS | PsA | Ps | ReA | JIA-E | JIA-P | IBD | CD-I | T1D | RA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Overall diversity | | | | | | △ | ▽ | ▽ | ▽ | ◁▷ | | | | ▽ | |
| Actinobacteria | | | | | | | | ▼ | | | | ▲ | | | |
| | Actinobacteria | | | | | | | ▼ | | | | | | | |
| | | Actinomycetales | Actinomycetaceae | *Actinomyces* | | ▼ | | | | | | | | | |
| | | Bifidobacteriales | Bifidobacteriaceae | *Bifidobacterium* | *bifidum* | | | | | | | | | | ▼ |
| | | Coriobacteriales | Coriobacteriaceae | *Eggerthella* | *lenta* | | | | | | | | | | ▲ |
| | | | | *Gordonibacter* | *pamelaeae* | | | | | | | | | | ▲ |
| | | Bifidobacteriales | Bifidobacteriaceae | | | | | | ▲ | | | | | | |
| | | | | *Bifidobacterium* | | | | | | ▲ | | | | | |
| | | | | | *adolescentis* | | | | | ▲ | | | | | |
| | | | | | *dentium* | | | | | | | | | | ▲ |
| Firmicutes | | | | | | ▲ | | | | | ▽ | | | ▼ | |
| | Unclassified | | | | | | | ▼ | | | | | | | |
| | Clostridia | | | | | | | | | | | | | ▼ | |
| | | Unclassified | | | | | ▼ | ▼ | | | | | | | |
| | | Clostridiales | | | | | | | | | | | | ▼ | |
| | | | Unclassified | | | | ▼ | | | | | | | | |
| | | | Lachnospiraceae | | | ▲ | | | | | | ▼ | | ▼ | ▲ |
| | | | | *Unclassified* | | | ▼ | ▼ | | ▼ | | | | | |
| | | | | *Fusicatenibacter* | *saccharivorans* | | | | | ▼ | | | | | |
| | | | | *Clostridium* | *clostridioforme* | | | | | ▼ | | | | | |
| | | | | *Roseburia* | | ▲ | | | | | | | | | |
| | | | | *Coprococcus* | | ▲ | ▼ | ▼ | | | | | | | |
| | | | | *Pseudobutyrivibrio* | | | ▼ | | | | | | | | |
| | | | Ruminococcaceae | | | ▲ | | | | | | | | ▼ | |
| | | | | *unclassified* | | | | ▲ | | | | | | | |
| | | | | *unclassified* | | | ▼ | ▼ | | | | | | | |
| | | | | *Ruminococcus* | | ▲ | ▼ | | | | | | | | |
| | | | | | *lactaris* | | | | | | | | | | ▲ |
| | | | Clostridiaceae | *Faecalibacterium* | | | | | | ▼ | | | | ▼ | |
| | | | | | *prausnitzii* | | | | | ▼ | | | ▼ | | |
| | | | | *Clostridium* | *perfringens* | | | | | | | | | | ▲ |
| | | | | | *asparagiforme* | | | | | | | | | | ▲ |
| | | | Peptostreptococcaceae | | | | | | ▼ | | | | | | |
| | | | | *Sporacetigenium* | | | ▼ | | | | | | | | |
| | | | Eubacteriaceae | | | | | | | | | | | ▼ | |
| | | | | *Eubacterium* | | | | | | | | | | ▼ | |
| | Bacilli | Lactobacillales | Lactobacillaceae | *Lactobacillus* | *salivarius* | | | | | | | | | | ▲ |
| | | | Streptococcaceae | *Streptococcus* | | ▼ | | | | | | | | | |
| | | Gemellales | Gemellaceae | | | ▼ | | | | | | | | | |
| | Negativicutes | Selenomonadales | Veillonellaceae | | | ▼ | | | | | | | | | |
| | | | | *Megamonas* | *hypermegale* | | | | | | | | | | ▼ |
| | | | | *Veillonella* | | | | | | | | | | | ▼ |
| | Erysipelotrichia | | | | | | | ▼ | | | | | | | |
| | | Erysipelotrichales | | | | | | ▼ | | | | | | | |
| | | | Erysipelotrichaceae | | | | | ▼ | | | | | | | |
| | | | | *Coprobacillus* | | | | ▼ | | | | | | | |
| Verrucomicrobia | | | | | | | ▼ | | | | | | | | |
| | Verrucomicrobiae | | | | | | ▼ | | | | | | | | |
| | | Verrucomicrobiales | | | | | ▼ | | | | | | | | |
| | | | Verrucomicrobiaceae | *Akkermansia* | | | ▼ | | | | | | | | |
| Bacteroidetes | | | | | | ▲ | | | | | ◁▷ | ▼ | | ▲ | |
| | Unclassified | | | | | | | ▼ | | | | | | | |
| | Bacteroidetes | | | | | | | | | | | | | ▲ | |
| | | Bacteroidales | | | | | | | | | | | | ▲ | |
| | | | Unclassified | | | | ▲ | | | | | | | | |
| | | | Rikenellaceae | | | ▲ | | | | | | | | ▲ | |
| | | | | *Alistipes* | | | ▼ | | | | | | | | |
| | | | Porphyromonadaceae | | | ▲ | | ▼ | | | | | | ▼ | |
| | | | | *Parabacteroides* | | ▲ | ▼ | ▼ | | | | | | | |
| | | | Bacteroidaceae | | | ▲ | | | | | | | | ▲ | |
| | | | | *Bacteroides* | | | | | | | | | | ▲ | |
| | | | | | *ovatus* | | | | | | | | | ▲ | |
| | | | | | *vulgatis* | | | | | | | | | ▼ | |
| | | | | | *fragilis* | | | | | | | | | ▼ | |
| | | | Prevotellaceae | | | ▼ | | | ▼ | | | | | | |
| | | | | *Prevotella* | | | | ▼ | | | | | | | |
| Proteobacteria | | | | | | ▼ | | | | | ◁▷ | ▲ | | | |
| | Epsilonproteobacteria | Campylobacterales | Campylobacteraceae | *Campylobacter* | | ◀▶ | | | | | | | | | |
| | | Enterobacteriales | Enterobacteriaceae | *Escherichia* | *coli* | | | | | | | | ▲ | | |
| | | | | *Shigella* | | ◀▶ | | | | | | | | | |
| | | | | *Salmonella* | | ◀▶ | | | | | | | | | |
| | | | | *Yersinia* | | ◀▶ | | | | | | | | | |
| | | | | *Klebsiella* | *pneumoniae* | ◀▶ | | | | | | | | | ▼ |
| | | Pasteurellales | Pasteurellaceae | *Haemophilus* | *sp.* | | | | | | | | | | ▼ |
| | Beta Proteobacteria | Burkholderiales | Unclassified | | | | | ▼ | | | | | | | |
| | | | Sutterellaceae | *Sutterella* | *wadsworthensis* | | | | | | | | | | ▼ |

[1]References: [47][48][49][50][51][52][53][54]

There is good cause to suspect the gut microbial environment as a driver of pathogenesis in SpA. Recent 16S metagenomic sequencing has revealed differences in the microbial composition of AS [47], ERA [48] and PsA[49] patients compared to healthy controls. Reactive arthritis (ReA) has definitely been associated with gut infections by *Campylobacter, Salmonella, Shigella, and Yersinia*. Interestingly, there exists an established association between human immunodeficiency virus (HIV) and the development of SpA symptoms resembling ReA. Although HIV patients are immune compromised, few musculoskeletal infections occur, and those that do, tend to occur later in disease when CD4 T cell counts are particularly low[55]. This would argue in favour of infections at remote sites such as the gut or skin initiating an immune response in the joints indirectly, perhaps through the trafficking of immune cells. Anti-retroviral drugs have proven very effective in treating HIV, and have greatly reduced the incidence of ReA in HIV patients. Although there may be a direct link between HIV infection and the onset of ReA, another plausible theory is that the chance of infection from other pathogens, particularly at barrier sites such as the gut or skin, is reduced in uncompromised individuals, thereby removing the increased likelihood of developing ReA.

Ileocolonoscopic evidence of gut inflammation exists in 70% of AS patients (4% to 6% developing inflammatory bowel disease (IBD)), 80% of enterogenic ReA patients, and 70% of undifferentiated SpA patients [56]. In addition, bone marrow oedema of the SI joint is increased when gut inflammation is present in axial SpA [57], and fecal calprotectin, a measure of gut inflammation, is elevated [58] in SpA and correlates with disease activity in AS [59]. IgA, the predominant antibody involved in mucosal defence, is also increased in AS patient serum [60] correlating with disease activity

[61][62][63][64]. The interaction between microbiota, disease status and genetic background is also evidenced in animal models. In a germ-free environment, HLA-B27 transgenic rat models of AS fail to induce disease[65] and the SKG mouse model results in reduced disease severity[66]. Furthermore the gut microbiota of Lewis rats transgenic for HLA-B27 and human b2-microglobulin differs significantly from that of wild type.[67]

IL-17 is thought to play a protective role in mucosal defence against the gut pathogens *Salmonella typhimurium, Listeria monocytogenes, Citrobacter rodentium* (strongly resembling human enteropathogenic Escherichia coli (EPEC)[68]), *Klebsiella pneumoniae* (also found on the skin and can infect the lungs), and *Candida albican*s (also inhabiting skin) [69]. However, the full list of gut microbes capable of inducing an IL-17 response is far from known. It is also worth noting that a human genetic condition exists resulting in autosomal recessive deficiency of IL-17RA, consequently abolishing responses to IL17A and IL17F. The main recurring infection associated with this condition is candidiasis.

## 1.6 Advances in single-cell technologies

Recent advances in technology have enabled the measurement of many cell attributes simultaneously and with high throughput, resulting in granular acquisition of high-dimensional data. When assessing protein expression in cells, multi-colour fluorescence activated flow cytometry can now routinely capture 14-30 parameters per cell, cytometry by time of flight (CyTOF) over 100, and feature barcoding used in single cell RNA sequencing (scRNA-seq) is currently only limited by the number of oligo-tagged antibodies available for this technology. The more widespread adoption of

scRNA-seq is being driven by decreasing sequencing costs, however cost still remains a prohibitive barrier for many experimental designs. This technology can be used to simultaneously quantify  in a hypothesis-free way the RNA expression of individual genes within single cells. Although the end stage of this technology always involves sequencing, multiple processing techniques exist to isolate cells and prepare RNA libraries for sequencing, and each carries with it advantages and disadvantages. The two scRNA-seq techniques used in this thesis were the 10x Genomics droplet based (10x) approach, capable of cost effectively analysing a large number of cells at a lower read depth; and the Smart-seq2(SS2) plate based approach, only able to analyse a lower number of cells for the same price but obtaining full length transcript data, typically at higher read depth[70].

## 1.7 Project aims

1.7.1 To determine whether aberrant or enhanced immune responses exist in SpA patients towards known commensal or pathogenic microbes typically found in the human gut.

1.7.2 To use single-cell RNA sequencing to interrogate the cell phenotype and clonality elicited by a type 17.1 immune response towards *S. typhimurium* in SpA synovial fluid, and contrast this to *ex vivo* unstimulated CD4 and CD8 memory T cells in blood and synovial fluid.

1.7.3 To validate single cell sequencing findings in a larger patient cohort, characterising the degree of phenotypic and clonal overlap between individual patients.

# I. References


1. McGonagle, D., Aydin, S.Z., Gül, A., Mahr, A., and Direskeneli, H. (2015). 'MHC-I-opathy'—unified concept for spondyloarthritis and Behçet disease. Nature Reviews Rheumatology *11*, 731–740.

2. Brown, M.A., Kennedy, L.G., MacGregor, A.J., Darke, C., Duncan, E., Shatford, J.L., Taylor, A., Calin, A., and Wordsworth, P. (1997). Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. Arthritis Rheum. *40*, 1823–1828.

3. Castillo-Gallego, C., Aydin, S.Z., Emery, P., McGonagle, D.G., and Marzo-Ortega, H. (2013). Brief Report: Magnetic Resonance Imaging Assessment of Axial Psoriatic Arthritis: Extent of Disease Relates to HLA–B27. Arthritis & Rheumatism *65*, 2274–2278.

4. Colmegna, I., Cuchacovich, R., and Espinoza, L.R. (2004). HLA-B27-Associated Reactive Arthritis: Pathogenetic and Clinical Considerations. Clin. Microbiol. Rev. *17*, 348–369.

5. Blum, J.S., Wearsch, P.A., and Cresswell, P. (2013). Pathways of Antigen Processing. Annual Review of Immunology *31*, 443–473.

6. Ridge, J.P., Rosa, F.D., and Matzinger, P. (1998). A conditioned dendritic cell can be a temporal bridge between a CD4 + T-helper and a T-killer cell. Nature *393*, 474–478.

7. Laidlaw, B.J., Craft, J.E., and Kaech, S.M. (2016). The multifaceted role of $CD4^+$ T cells in $CD8^+$ T cell memory. Nature Reviews Immunology *16*, 102–111.

8. Eickhoff, S., Göbel, A., Gerner, M.Y., Klauschen, F., Komander, K., Hemmi, H., Garbi, N., Kaisho, T., Germain, R.N., and Kastenmüller, W. (2015). Robust Anti-viral Immunity Requires Multiple Distinct T Cell-Dendritic Cell Interactions. Cell *162*, 1322–1337.

9. Castellino, F., Huang, A.Y., Altan-Bonnet, G., Stoll, S., Scheinecker, C., and Germain, R.N. (2006). Chemokines enhance immunity by guiding naive CD8 + T cells to sites of CD4 + T cell–dendritic cell interaction. Nature *440*, 890–895.

10. Call, M.E., Pyrdol, J., Wiedmann, M., and Wucherpfennig, K.W. (2002). The Organizing Principle in the Formation of the T Cell Receptor-CD3 Complex. Cell *111*, 967–979.

11. Dong, D., Zheng, L., Lin, J., Zhang, B., Zhu, Y., Li, N., Xie, S., Wang, Y., Gao, N., and Huang, Z. (2019). Structural basis of assembly of the human TCR–CD3 complex. Nature, 1–9.

12. Willcox, B.E., and Willcox, C.R. (2019). γδ TCR ligands: the quest to solve a 500-million-year-old mystery. Nat Immunol *20*, 121–128.

13. Roth, D.B. (2014). V(D)J Recombination: Mechanism, Errors, and Fidelity. Microbiol Spectr *2*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5089068/ [Accessed September 8, 2019].

14. Jenkins, M.K., Chu, H.H., McLachlan, J.B., and Moon, J.J. (2010). On the Composition of the Preimmune Repertoire of T Cells Specific for Peptide–Major Histocompatibility Complex Ligands. Annual Review of Immunology *28*, 275–294.

15. Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., *et al.* (2017). Identifying specificity groups in the T cell receptor repertoire. Nature *547*, 94–98.

16. Bowness, P. (2015). Hla-B27. Annual Review of Immunology *33*, 29–48.

17. Cortes, A., Pulit, S.L., Leo, P.J., Pointon, J.J., Robinson, P.C., Weisman, M.H., Ward, M., Gensler, L.S., Zhou, X., Garchon, H.-J., *et al.* (2015). Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nat Commun *6*, 7146.

18. Thomson, W., Barrett, J.H., Donn, R., Pepper, L., Kennedy, L.J., Ollier, W.E.R., Silman, A.J.S., Woo, P., and Southwood, T. (2002). Juvenile idiopathic arthritis classified by the ILAR criteria: HLA associations in UK patients. Rheumatology (Oxford) *41*, 1183–1189.

19. Winchester, R., Minevich, G., Steshenko, V., Kirby, B., Kane, D., Greenberg, D.A., and FitzGerald, O. (2012). HLA associations reveal genetic heterogeneity in psoriatic arthritis and in the psoriasis phenotype. Arthritis & Rheumatism *64*, 1134–1144.

20. Bowes, J., Budu-Aggrey, A., Huffmeier, U., Uebe, S., Steel, K., Hebert, H.L., Wallace, C., Massey, J., Bruce, I.N., Bluett, J., *et al.* (2015). Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis. Nature Communications *6*, ncomms7046.

21. Londono, J., Santos, A.M., Peña, P., Calvo, E., Espinosa, L.R., Reveille, J.D., Vargas-Alarcon, G., Jaramillo, C.A., Valle-Oñate, R., Avila, M., *et al.* (2015). Analysis of HLA-B15 and HLA-B27 in spondyloarthritis with peripheral and axial clinical patterns. BMJ Open *5*, e009092.

22. Brewerton, D.A., Hart, F.D., Nicholls, A., Caffrey, M., James, D.C.O., and Sturrock, R.D. (1973). ANKYLOSING SPONDYLITIS AND HL-A 27. The Lancet *301*, 904–907.

23. Schlosstein, L., Terasaki, P.I., Bluestone, R., and Pearson, C.M. (1973). High Association of an HL-A Antigen, W27, with Ankylosing Spondylitis. New England Journal of Medicine *288*, 704–706.

24. Brown, M.A., Kenna, T., and Wordsworth, B.P. (2016). Genetics of ankylosing spondylitis—insights into pathogenesis. Nat Rev Rheumatol *12*, 81–91.

25. Robinson, P.C., Claushuis, T.A.M., Cortes, A., Martin, T.M., Evans, D.M., Leo, P., Mukhopadhyay, P., Bradbury, L.A., Cremin, K., Harris, J., *et al.* (2015). Genetic Dissection of Acute Anterior Uveitis Reveals Similarities and Differences in Associations Observed With Ankylosing Spondylitis. Arthritis & Rheumatology *67*, 140–151.

26. Neumann-Haefelin, C. (2013). HLA-B27-mediated protection in HIV and hepatitis C virus infection and pathogenesis in spondyloarthritis: two sides of the same coin? Current Opinion in Rheumatology *25*, 426.

27. McMichael, A.J. (2007). Triple bypass: complicated paths to HIV escape. J Exp Med *204*, 2785–2787.

28. Elahi, S., Dinges, W.L., Lejarcegui, N., Laing, K.J., Collier, A.C., Koelle, D.M., McElrath, M.J., and Horton, H. (2011). Protective HIV-specific CD8+ T cells evade Treg cell suppression. Nat Med *17*, 989–995.

29. Benjamin, R., and Parham, P. (1990). Guilt by association: HLA-B27 and ankylosing spondylitis. Immunology Today *11*, 137–142.

30. Allen, R.L., O'Callaghan, C.A., McMichael, A.J., and Bowness, P. (1999). Cutting Edge: HLA-B27 Can Form a Novel β2-Microglobulin-Free Heavy Chain Homodimer Structure. J Immunol *162*, 5045–5048.

31. Wong-Baeza, I., Ridley, A., Shaw, J., Hatano, H., Rysnik, O., McHugh, K., Piper, C., Brackenbridge, S., Fernandes, R., Chan, A., *et al.* (2013). KIR3DL2 Binds to HLA-B27 Dimers and Free H Chains More Strongly than Other HLA Class I and Promotes the Expansion of T Cells in Ankylosing Spondylitis. Journal of Immunology *190*, 3216–3224.

32. Ridley, A., Hatano, H., Wong-Baeza, I., Shaw, J., Matthews, K.K., Al-Mossawi, H., Ladell, K., Price, D.A., Bowness, P., and Kollnberger, S. (2016). Activation-induced killer cell immunoglobulin-like receptor 3DL2 binding to HLA-B27 licenses pathogenic T cell differentiation in spondyloarthritis. Arthritis and Rheumatology *68*, 901–914.

33. Bowness, P., Ridley, A., Shaw, J., Chan, A.T., Wong-Baeza, I., Fleming, M., Cummings, F., McMichael, A., and Kollnberger, S. (2011). Th17 Cells Expressing KIR3DL2+ and Responsive to HLA-B27 Homodimers Are Increased in Ankylosing Spondylitis. J Immunol *186*, 2672–2680.

34. Thjodleifsson, B., Geirsson, árni J., Björnsson, S., and Bjarnason, I. (2007). A common genetic background for inflammatory bowel disease and ankylosing spondylitis: A genealogic study in Iceland. Arthritis & Rheumatism *56*, 2633–2639.

35. Lee, Y., Awasthi, A., Yosef, N., Quintana, F.J., Xiao, S., Peters, A., Wu, C., Kleinewietfeld, M., Kunder, S., Hafler, D.A., *et al.* (2012). Induction and molecular signature of pathogenic TH17 cells. Nat Immunol *13*, 991–999.

36. Wendling, D., Cedoz, J.-P., Racadot, E., and Dumoulin, G. (2007). Serum IL-17, BMP-7, and bone turnover markers in patients with ankylosing spondylitis. Joint Bone Spine *74*, 304–305.

37. Shen, H., Goodall, J.C., and Hill Gaston, J.S. (2009). Frequency and phenotype of peripheral blood Th17 cells in ankylosing spondylitis and rheumatoid arthritis. Arthritis & Rheumatism *60*, 1647–1656.

38. Jandus, C., Bioley, G., Rivals, J.-P., Dudler, J., Speiser, D., and Romero, P. (2008). Increased numbers of circulating polyfunctional Th17 memory cells in patients with seronegative spondylarthritides. Arthritis & Rheumatism *58*, 2307–2317.

39. Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., van der Heijde, D., McInnes, I., van Laar, J.M., Landewé, R., Wordsworth, P., *et al.* (2013). Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. The Lancet *382*, 1705–1713.

40. Poddubnyy, D., Hermann, K.-G.A., Callhoff, J., Listing, J., and Sieper, J. (2014). Ustekinumab for the treatment of patients with active ankylosing spondylitis: results of a 28-week, prospective, open-label, proof-of-concept study (TOPAS). Ann Rheum Dis *73*, 817–823.

41. Helander, H.F., and Fändriks, L. (2014). Surface area of the digestive tract – revisited. Scandinavian Journal of Gastroenterology *49*, 681–689.

42. Consortium, T.H.M.P. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

43. Kuleshov, V., Jiang, C., Zhou, W., Jahanbani, F., Batzoglou, S., and Snyder, M. (2016). Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome. Nat Biotech *34*, 64–69.

44. Sekirov, I., Russell, S.L., Antunes, L.C.M., and Finlay, B.B. (2010). Gut Microbiota in Health and Disease. Physiological Reviews *90*, 859–904.

45. Kato, L.M., Kawamoto, S., Maruya, M., and Fagarasan, S. (2014). Gut TFH and IgA: key players for regulation of bacterial communities and immune homeostasis. Immunol Cell Biol *92*, 49–56.

46. Peterson, C.T., Sharma, V., Elmén, L., and Peterson, S.N. (2015). Immune homeostasis, dysbiosis and therapeutic modulation of the gut microbiota. Clin Exp Immunol *179*, 363–377.

47. Costello, M.-E., Ciccia, F., Willner, D., Warrington, N., Robinson, P.C., Gardiner, B., Marshall, M., Kenna, T.J., Triolo, G., and Brown, M.A. (2015). Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis: Gut Microbiome and AS-Related Genes. Arthritis & Rheumatology *67*, 686–691.

48. Stoll, M.L., Kumar, R., Morrow, C.D., Lefkowitz, E.J., Cui, X., Genin, A., Cron, R.Q., and Elson, C.O. (2014). Altered microbiota associated with abnormal humoral immune responses to commensal organisms in enthesitis-related arthritis. Arthritis research & therapy *16*, 486.

49. Scher, J.U., Ubeda, C., Artacho, A., Attur, M., Isaac, S., Reddy, S.M., Marmon, S., Neimann, A., Brusca, S., Patel, T., *et al.* (2015). Decreased Bacterial Diversity

Characterizes the Altered Gut Microbiota in Patients With Psoriatic Arthritis, Resembling Dysbiosis in Inflammatory Bowel Disease. Arthritis & Rheumatology *67*, 128–139.

50. Ogdie-Beatty (2015). The Microbiome of Reactive Arthritis in a Guatemalan Cohort. ACR Meeting Abstracts. Available at: http://acrabstracts.org/abstract/the-microbiome-of-reactive-arthritis-in-a-guatemalan-cohort/ [Accessed January 31, 2016].

51. Hissink Muller, P., Westedt, P., Budding, A., Allaart, C., Brinkman, D., Kuijpers, T., Van den Berg, J., Van Suijlekom-Smit, L., Van Rossum, M., De Meij, T., *et al.* (2013). PReS-FINAL-2160: Intestinal microbiome in polyarticular juvenile idiopathic arthritis: a pilot study. Pediatr Rheumatol Online J *11*, P172.

52. Willing Willing-Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease .pdf.

53. Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Novelo, L.L., Casella, G., Drew, J.C., Ilonen, J., Knip, M., Hyöty, H., *et al.* (2011). Toward defining the autoimmune microbiome for type 1 diabetes. ISME J *5*, 82–91.

54. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., *et al.* (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med *21*, 895–905.

55. Louthrenoo, W. (2008). Rheumatic manifestations of human immunodeficiency virus infection. Current Opinion in Rheumatology *20*, 92–99.

56. Veys, E.M., Mielants, H., De Vos, M., and Cuvelier, C. (1996). Spondylarthropathies: from gut to target organs. Baillière's Clinical Rheumatology *10*, 123–146.

57. Praet, L.V., Jans, L., Carron, P., Jacques, P., Glorieus, E., Colman, R., Cypers, H., Mielants, H., Vos, M.D., Cuvelier, C., *et al.* (2014). Degree of bone marrow oedema in sacroiliac joints of patients with axial spondyloarthritis is linked to gut inflammation and male sex: results from the GIANT cohort. Ann Rheum Dis *73*, 1186–1189.

58. Cypers, H., Varkas, G., Beeckman, S., Debusschere, K., Vogl, T., Roth, J., Drennan, M.B., Lavric, M., Foell, D., Cuvelier, C.A., *et al.* (2016). Elevated calprotectin levels reveal bowel inflammation in spondyloarthritis. Ann Rheum Dis *75*, 1357–1362.

59. Duran, A., Kobak, S., Sen, N., Aktakka, S., Atabay, T., and Orman, M. (2016). Fecal calprotectin is associated with disease activity in patients with ankylosing spondylitis. Bosn J Basic Med Sci *16*, 71–74.

60. Veys, E.M., and van Leare, M. (1973). Serum IgG, IgM, and IgA levels in ankylosing spondylitis. Annals of the rheumatic diseases *32*, 493.

61. Cowling et al. (1980). Association of inflammation with raised serum IgA in ankylosing spondylitis.

62. Franssen, M.J., Van de Putte, L.B., and Gribnau, F.W. (1985). IgA serum levels and disease activity in ankylosing spondylitis: a prospective study. Annals of the rheumatic diseases *44*, 766–771.

63. Mackiewicz, A., Khan, M.A., Reynolds, T.L., van der Linden, S., and Kushner, I. (1989). Serum IgA, acute phase proteins, and glycosylation of alpha 1-acid glycoprotein in ankylosing spondylitis. Annals of the rheumatic diseases *48*, 99–103.

64. Collado, A., Sanmarti, R., Serra, C., Gallart, T., Cañeté, J.D., Gratacos, J., Vives, J., and Muñoz-Gomé, J. (1991). Serum Levels of Secretory IgA in Ankylosing Spondylitis. Scandinavian Journal of Rheumatology *20*, 153–158.

65. Taurog, J.D., Richardson, J.A., Croft, J.T., Simmons, W.A., Zhou, M., Fernández-Sueiro, J.L., Balish, E., and Hammer, R.E. (1994). The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. The Journal of experimental medicine *180*, 2359–2364.

66. Rehaume, L.M., Mondot, S., Aguirre de Cárcer, D., Velasco, J., Benham, H., Hasnain, S.Z., Bowman, J., Ruutu, M., Hansbro, P.M., McGuckin, M.A., *et al.* (2014). ZAP-70 Genotype Disrupts the Relationship Between Microbiota and Host, Leading to Spondyloarthritis and Ileitis in SKG Mice. Arthritis & Rheumatology *66*, 2780–2792.

67. Lin, P., Bach, M., Asquith, M., Lee, A.Y., Akileswaran, L., Stauffer, P., Davin, S., Pan, Y., Cambronne, E.D., Dorris, M., *et al.* (2014). HLA-B27 and Human β2-Microglobulin Affect the Gut Microbiota of Transgenic Rats. PLoS One *9*. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4139385/ [Accessed January 30, 2015].

68. Schauer, D.B., Zabel, B.A., Pedraza, I.F., O'Hara, C.M., Steigerwalt, A.G., and Brenner, D.J. (1995). Genetic and Biochemical Characterization of Citrobacter rodentium sp. nov. J. CLIN. MICROBIOL. *33*, 5.

69. Jin, W., and Dong, C. (2013). IL-17 cytokines in immunity and inflammation. Emerg Microbes Infect *2*, e60.

70. Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Medicine *9*, 75.

# Chapter 2: Materials and methods

## 2.1 Patient and control recruitment

Full informed patient consent was obtained from SpA and rheumatoid arthritis (RA) control donors for the collection of blood and synovial fluid samples when undergoing intra-articular aspiration at Oxford University Hospitals. The study and associated protocols were approved by the Oxford Research Ethics committee (Immune Function in Inflammatory Arthritis study; ethics reference number 06/Q1606/139). Patient and control metadata is shown in Table 2-1.

# Table 2-1 SpA, RA and healthy control metadata

| Subject ID | Diagnosis† | Sample(s) collected § | Age | Sex | CRP | BASDAI / DAS—28—CRP / PGA¶ | HLA-B27 / RF / ACCP status ‡ | Comorbidity §§ | DMARD | NSAID treatment | Biologic treatment | Steroid treatment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AS-1559 | AS | PB | 30 | M | 0.3 | - | B27+ | Uveitis, PA | Previous | Yes | Adalimumab | No |
| AS-1614 | AS | PB | 25 | F | - | BASDAI: 5.9 | B27+ | No | No | No | No | No |
| AS-1620 | AS | PB | 19 | M | - | BASDAI: 1.9 | B27+ | No | No | No | Adalimumab | No |
| AS-1626 | AS | PB | 80 | M | 16.5 | BASDAI: 2.7 | B27+ | No | No | No | No | No |
| AS-1627 | AS | SF | 54 | F | 0.9 | - | B27- | IBD | Yes | No | No | No |
| AS-1629 | AS | PB | 47 | M | 5 | BASDAI: 6 | B27+ | Uveitis | No | No | Adalimumab | No |
| AS-1630 | AS | PB | 39 | M | - | BASDAI: 1.5 | B27+ | No | No | Yes | No | No |
| AS-1631 | AS | PB | 34 | M | 6.1 | BASDAI: 3.2 | B27+ | No | No | Yes | No | No |
| AS-1632 | AS | PB | 45 | M | - | BASDAI: 2.4 | B27+ | No | No | Yes | No | No |
| AS-1635 | AS | PB | 70 | M | 0.6 | BASDAI: 0.76 | B27- | Uveitis | No | No | No | No |
| AS-1642 | AS | PB | 37 | M | 2.1 | BASDAI: 1.5 | B27+ | No | No | Yes | Adalimumab | No |
| AS-1644 | AS | PB | 20 | F | 3.7 | BASDAI: 7.5 | B27+ | No | No | Yes | No | Yes |
| AS-1647 | AS | PB | 44 | F | 0.3 | - | B27+ | Uveitis | Previous | No | Etanercept | No |
| AS-1652 | AS | PB | 23 | M | 9 | BASDAI: 5.1 | B27+ | PA, Enthesitis | Yes | Previous | No | Yes |
| AS-1653 | AS | PB | 56 | M | 0.8 | BASDAI: 3 | B27+ | IBD | No | Yes | Infliximab | Yes |
| AS-1654 | AS | PB | 47 | M | 51 | BASDAI: 2.9 | B27+ | No | No | Yes | No | No |
| AS-1655 | AS | PB | 52 | F | 23 | BASDAI: 5.77 | B27+ | IBD | Yes | No | Infliximab | Previous |
| AS-1657 | AS | PB | 50 | M | - | BASDAI: 8.32 | B27- | No | No | No | Adalimumab | No |
| AS-1660 | AS | PB | 53 | F | 2.9 | BASDAI: 6.1 | B27- | No | No | Yes | To start | Previous |
| AS-1661 | AS | PB | 46 | F | 5.8 | - | B27+ | PA, Psoriasis | No | No | No | Yes |
| AS-1665 | AS | PB, SF | 62 | F | 57.3 | - | B27+ | PA, Enthesitis | Yes | Yes | Ustekinumab | No |
| PsA-1505B2 | PsA | PB, SF | 35 | M | 1 | PGA: 2/5 | - | Psoriasis | Previous | No | No | No |
| PsA-1604 | PsA | PB, SF | 37 | M | 7.6 | - | - | Psoriasis | No | Previous | No | No |
| PsA-1607 | PsA | SF | 41 | M | - | - | - | No | No | No | No | No |
| PsA-1614 | PsA | PB, SF | 50 | F | 2.6 | - | - | Psoriasis | No | No | To start | No |
| PsA-1616 | PsA | PB | 48 | M | 0.5 | - | - | No | Yes | No | Adalimumab | No |
| PsA-1618 | PsA | PB | 50 | M | 1.8 | - | - | Psoriasis, IBD | Yes | No | Etanercept | No |
| PsA-1619 | PsA | PB | 49 | M | 1.4 | - | B27+ | Psoriasis | Yes | No | No | No |
| PsA-1621 | PsA | PB | 67 | F | 4.2 | - | - | No | Yes | Yes | Adalimumab | Previous |
| PsA-1622 | PsA | PB | 36 | M | 2 | - | - | Psoriasis, Nail | Yes | Previous | No | Previous |
| PsA-1640B2 | PsA | PB, SF | 56 | F | - | PGA: 3/5 | B27- | No | Previous | No | No | No |
| PsA-1718B3 | PsA | PB, SF | 31 | F | 12.5 | PGA: 4/5 | - | Psoriasis | Previous | No | No | No |
| PsA-1719 | PsA | PB, SF | 33 | M | 36.6 | PGA: 3/5 | B27- | Psoriasis | No | No | No | No |
| PsA1728 | PsA | PB, SF | 72 | F | 43.2 | PGA: 3/5 | - | No | Previous | No | No | Previous |
| PsA-1734 | PsA | PB, SF | 51 | M | 4.5 | PGA: 1.5/5 | - | Psoriasis, Nail, Enthesitis | Yes | Yes | No | No |
| PSA1801 | PsA | PB,SF | 54 | F | 9.9 | PGA: 3/5 | - | Psoriasis, Nail | No | Yes | No | No |
| ReA-1602 | ReA | PB, SF | 23 | M | - | - | B27+ | Chlamydia infection | No | No | No | No |
| RA-1503B2 | RA | PB, SF | 57 | F | 10.7 | - | RF+ ACCP+ | No | Yes | No | No | No |
| RA-1505 | RA | PB | 69 | M | 3.4 | - | - | No | Yes | No | No | No |
| RA-1506 | RA | PB | 54 | M | 13.3 | DAS-28-CRP: 1.7 | RF- ACCP+ | No | Yes | No | No | No |
| RA-1507 | RA | PB | 64 | M | 3.6 | - | RF- ACCP+ | No | Yes | No | No | No |
| RA-1508 | RA | PB | 63 | M | 17.7 | - | - | No | Yes | No | No | No |
| RA-1509 | RA | PB | 55 | F | 3.9 | - | RF- ACCP- | No | No | Yes | No | No |
| RA-1510 | RA | PB | 56 | F | 2 | DAS-28-CRP: 1.72 | RF+ ACCP+ | No | Yes | No | No | No |
| RA-1511 | RA | PB | 74 | M | 43.3 | - | RF+ | No | Yes | No | No | No |
| RA-1512 | RA | PB | 76 | M | 10.6 | DAS-28-CRP: 2.59 | - | No | Yes | No | No | No |
| RA-1513 | RA | PB | 69 | F | 4.2 | - | RF- ACCP- | No | Yes | Yes | No | No |
| RA-1514 | RA | PB | 65 | F | 0.7 | - | - | No | Yes | No | No | No |
| RA-1515 | RA | PB | 60 | F | 6.2 | - | - | No | Yes | No | No | No |
| RA-1516 | RA | PB, SF | 82 | F | 3.1 | - | - | No | Yes | No | No | No |
| RA-1518 | RA | PB | 70 | F | 16.7 | DAS-28-CRP: 4.41 | RF+ ACCP+ | No | Yes | No | No | No |
| RA-1519 | RA | PB | 47 | F | 41.9 | - | RF+ | No | Yes | No | No | No |
| RA-1617 | RA | PB, SF | 52 | F | 1.9 | - | RF- ACCP- | No | Yes | No | No | No |
| RA-1801 | RA | PB, SF | 56 | M | 36.6 | - | RF+ ACCP+ | No | Yes | Yes | No | No |
| HC-1603 | HC | PB | 53 | F | - | - | - | No | No | No | No | No |
| HC-1604 | HC | PB | 47 | M | - | - | - | No | No | No | No | No |
| HC-1704 | HC | PB | 27 | M | - | - | - | No | No | No | No | No |
| HC-1706 | HC | PB | 32 | M | - | - | - | No | No | No | No | No |
| HC-1707 | HC | PB | 27 | M | - | - | - | No | No | No | No | No |
| HC-B1 | HC | PB | 35 | F | - | - | - | No | No | No | No | No |
| HC-B2 | HC | PB | 32 | M | - | - | - | No | No | No | No | No |
| HC-B3 | HC | PB | 64 | M | - | - | - | No | No | No | No | No |
| HC-B4 | HC | PB | 34 | M | - | - | - | No | No | No | No | No |
| HC-B5 | HC | PB | 26 | M | - | - | - | No | No | No | No | No |
| HC-B6 | HC | PB | 54 | M | - | - | - | No | No | No | No | No |
| HC-B7 | HC | PB | 21 | F | - | - | - | No | No | No | No | No |
| HC-B8 | HC | PB | - | - | - | - | - | No | No | No | No | No |
| HC-B9 | HC | PB | - | - | - | - | - | No | No | No | No | No |

† AS: ankylosing spondylitis; PsA: psoriatic arthritis; ReA: reactive arthritis; RA: rheumatoid arthritis; HC: healthy control
§ PB: peripheral blood; SF: synovial fluid
¶ PGA: physician global assessment; a dash (-) in this table indicates data unavailable
‡ RF: rheumatoid factor; ACCP: anti cyclic citrullinated peptide test
§§ PA: peripheral arthritis

## 2.2 Laboratory methods and materials

### 2.2.1 Mononuclear cell separation

Peripheral blood and synovial fluid was collected in 10 mL tubes containing sodium heparin (17 IU of Heparin/ml of blood volume, BD vacutainer). Samples were first centrifuged at 850 x*g* for 10 minutes to collect blood plasma and synovial fluid supernatant. The majority of excess synovial fluid supernatant was then removed from the remaining synovial fluid sample, before being resuspended, diluted with 1 mL R0 (RPMI 1640 (Sigma) supplemented with 100 Units/mL Penicillin, 100 µg/mL Streptomycin (Gibco)) and passed through a 70 µm cell strainer (Corning). Remaining blood sample was diluted 1:1 in R0 and filtered synovial fluid sample was also diluted to a matching volume with R0, then both were separated using Histopaque (Sigma) by density centrifugation at 544 x*g* for 20 minutes and no brake. The cellular interface was collected, washed twice in R0 at 850 x*g* for 10 minutes and 544 x*g* for 5 minutes respectively all at room temperature (RT). Cells were then resuspended in R10 (R0 supplemented with 10% Fetal Bovine Serum (Sigma) and L-glutamine (2mM) (Gibco)) at 37°C and counted after staining with Trypan blue (Sigma) using a hemocytometer under a light microscope.

### 2.2.2 Cryopreservation of mononuclear cells

PBMC or SFMC were centrifuged at 450 x *g* for 5 minutes and resuspended in freezing medium (10% dimethyl sulphoxide in 90% fetal bovine serum (all Sigma-Aldrich)) at 1x10^8 cells / mL in 2mL cryovials (Greiner Bio One) before being placed in a pre-chilled Mr Frosty container and transferred into a -80 freezer overnight. For long term storage vials were transferred on dry ice to liquid nitrogen storage.

### 2.2.3 CD154 stimulation assay

#### *2.2.3.1 Microbial lysate preparation*

Bacterial lysates were provided as a kind gift from Ahmed Hegazy (Oxford University), prepared as previously described[1]. Briefly, bacteria were cultured overnight at 37°C, washed in sterile PBS, heat-inactivated and subjected to 3 freeze-thaw cycles. Bacterial lysates were titrated using a CFSE dilution assay and an optimal concentration used for stimulation assays (5-15 μg/ml).

#### *2.2.3.2 Microbial lysate and LPS stimulation for ICS staining*

Separated peripheral blood and synovial fluid mononuclear cells were centrifuged at 544 x*g*, resuspended in R5 (RPMI 1640 (Sigma) supplemented with 5% human serum (John Radcliffe Blood Bank), 100 Units/mL Penicillin, 100 μg/mL Streptomycin and L-glutamine (2mM) (Gibco)) and plated in 96-well plates (Greiner bio-one) at a concentration of 1.5 x $10^6$ cells per well. Wells were individually stimulated with bacterial lysates from *Bacteroides vulgatus* (Bv 1447), *Bifidobacterium animalis subsp. lactis* (Bi-07), *Clostridium difficile* (OXF1003, Toxin AB-), *E. coli* (Nissle 1917), *Faecalibacterium prausnitzii* (A2-165), *Lactobacillus acidophilus* (NCFM), *Roseburia intestinalis* (M50/1), *Ruminococcus obeum* (A2-162), *Salmonella enterica serovar typhimurium* (NCTC 12023), *Staphylococcus aureus* (NCTC 6571); heat treated *Mycobacterium tuberculosis* (H37Ra) and *Candida albicans* (both InvivoGen); 1ug/mL SEB (Sigma) as a positive control; 100ng/mL LPS from *Salmonella enterica* serotype *typhimurium* (Sigma) and no stimulation as negative controls, and incubated for 16 hours at 37°C and 5% CO2, adding 5 μg/ml of Brefeldin-A (Biolegend) after the first hour of stimulation. After incubation, cells were centrifuged for 5 minutes at 544 x*g*, washed, then stained with CD3-BV510 (OKT3), CD4-PE/Dazzle 594 (RPA-T4),

CD45RA-BV711 (HI100) and Fixable Viability Dye eFluor 780 (eBioscience) for 20 minutes at 4°C. Cells were then washed twice, permeabalized (BD Cytofix/Cytoperm) at RT for 30 minutes, washed again and stained for 3 hours at 4°C with: CD154-FITC (24-31), TNFα-Pacific Blue (Mab11), IFNγ-AF700 (B27), IL2-BV650 (MQ1-17H12), GMCSF-PerCp/Cy5.5 (BVD2-21C11), IL-17A-PE (eBio64CAP17, eBioscience) and IL-22-PE-Cyanine7 (22URTI, eBioscience). After intracellular staining, cells were washed twice before being fixed in Dulbecco's Phosphate Buffered Saline (Sigma) supplemented with 1% Fetal Bovine Serum (Sigma) and 2% formaldehyde.

Washing steps prior to permeabilization used Dulbecco's Phosphate Buffered Saline (Sigma) supplemented with 1% Fetal Bovine Serum (Sigma) and Perm/wash buffer (BD) thereafter.

For titrations, cells were stimulated with either SEB as described above, or 100ng/mL PMA and 1ug/mL Ionomycin for 4 hours together with GolgiStop (1:1500) and GolgiPlug (1:1000) (both BD Cytofix/Cytoperm Plus kit) in R10. FMO with FITC Mouse IgG1k Isotype control was used to detect expected background noise for CD154-FITC.

#### *2.2.3.3 S. typhimurium lysate stimulation for sorting live cells by FACS*

Separated synovial fluid mononuclear cells were centrifuged at 544 x*g*, resuspended in R5 (RPMI 1640 (Sigma) supplemented with 5% human serum (John Radcliffe Blood Bank), 100 Units/mL Penicillin, 100 µg/mL Streptomycin and L-glutamine (2mM) (Gibco)) and plated in 96-well plates (Greiner bio-one) at a concentration of 1.5 x $10^6$ cells per well. Wells were individually stimulated with bacterial lysate from *Salmonella enterica serovar typhimurium* (NCTC 12023) and incubated for 16 hours at 37°C and

5% CO2 in the presence of CD40 blocking antibody (Biolegend). After incubation, cells were centrifuged for 5 minutes at 544 x*g*, washed, then stained with CD3-FITC (SK7), CD4-APC (RPA-T4), CD45RA-AF700(HI100) and CD154-PE (5C8) for 20 minutes at 4°C. Cells were then washed twice, resuspended in chilled R10 and passed through the 35 um cell strainer snap cap of a Falcon test tube (Corning).

Washing steps used Dulbecco's Phosphate Buffered Saline (Sigma) supplemented with 1% Fetal Bovine Serum (Sigma).

#### *2.2.3.4 Fluorescence activated flow cytometry of fixed samples*

Fixed cells were transferred into 1.2mL cluster tubes (Corning) and analysed using a BD LSR Fortessa flow cytometer calibrated daily with BD FACSDiva CS&T Research Beads (BD Biosciences). Compensation was performed using One Comp eBeads (eBioscience) stained with the same antibodies used to stain cells. Results were analysed using FlowJo (Treestar).

#### *2.2.3.5 Fluorescent activated cell sorting of S. Typhimurium stimulated SFMC.*

Cells in chilled R10 were brought to the Flow Cytometry Facility of the Experimental Medicine Division at the John Radcliffe Hospital on ice for sorting with a FACSAria III cell sorter[1]. 4',6-diamidino-2-phenylindole (DAPI) was added 15 minutes prior to sorting for viability staining. Results were analysed using FlowJo (Treestar).

[1] Helen Ferry performed FACS sorting

#### *2.2.3.6 Statistical calculations for CD154 assay*

When assessing cytokine specific CD154 responses, mean frequencies of Th-memory cells expressing CD154, or CD154 in addition to any combination of the cytokines TNFα, IFNγ, IL-17A, GM-CSF, IL-22 or IL-2 among total Th-memory cells are measured. Where SpA is specified, this refers to pooled samples consisting of all AS, PsA and ReA samples available from either blood or synovial fluid. No synovial fluid samples were available from healthy controls. Due to differences in sample sizes between groups, unweighted means analysis was used during statistical calculations. All calculations were performed using GraphPad Prism 7.0b for Macintosh.

Significance was calculated using 2-way analysis of variance (ANOVA) with Sidak's correction for multiple comparisons when comparing immune responses to different lysates by cells sampled from different locations (blood or synovial fluid) but the same volunteer group (SpA, HC or RA); or different volunteer groups but the same sample type. Any further statistical calculations are detailed in figure legends.

### 2.2.4 Single-cell 10x droplet and plate based RNA processing

#### *2.2.4.1 Cell separation of unstimulated ex vivo CD4 and CD8 memory T cells for scRNA-seq2*

Separated SFMCs and PBMCs (section 2.2.1) were stained in staining buffer (RNAse-free PBS, 2mM EDTA) using fluorescently conjugated antibodies: CD4-APC (RPA-T4), CD8a-PE (RPA-T8),CD3-FITC (SK7) and CD45RA-BV421 (HI100) (all BioLegend) for surface markers; and for viability with Fluor780(eBioscience) viability dye. PBMCs and SFMCs were then separately sorted into CD3+CD45RA-CD4+CD8- and CD3+CD45RA-

[2] Nicole Yager performed cell separation and FACS sorting

CD4-CD8+ cells and these enriched fractions re-combined in a 1:1 ratio separately for each sample type.

*2.2.4.2 10x droplet based scRNA-seq3*

After counting viable cells, pooled CD4/CD8 memory T cells (section 2.2.4.1 ) from blood were loaded into one channel of a 10x-Genomics Chromium controller, and those from synovial fluid were loaded into a second channel. For 2 of the patients, approximately 25 000 pooled CD4/CD8 cells were loaded, and for a third patient (PSA1607), approximately 50000 pooled CD4/CD8 cells were loaded. The standard 10x Genomics protocol was then followed to process cells using the 10x 5' Kit, which allowed for concurrent TCR and gene expression profiling of the same cell. The total time taken between acquiring a patient sample and loading the sample onto the 10x chip was approximately 4 hours for all samples. The Illumina HiSeq 4000 was then used to sequence pooled samples to a read depth of approximately 30000 reads per cell for gene expression libraries and 8500 reads per cell for V(D)J enriched T cell libraries. [2]

*2.2.4.3 Smart-seq2 plate based scRNA-seq4*

Mononuclear cells from paired synovial fluid and blood samples from 4 patients were separated according to section 2.2.1 before sorting CD3+CD45RA-CD8+ and CD3+CD45RA-CD4+ cells into separate wells of 96-well full-skirted plates (Eppendorf) containing 10μL of a 2% Dithiothreitol (DTT, 2M Sigma-Aldrich), RTL

[3] Cell loading ,10x protocol and sequencing carried out at Wellcome Centre for Human Genetics (WCHG)
[4] Cell separation was performed by Nicole Yager and SS2 protocol carried out at the Wellcome Sanger Institute.

lysis buffer (Qiagen) solution. After sealing cell lysates, they were mixed and spun down for storage at −80 °C. The Illumina Nextera XT DNA library prep kit was then used to prepare paired-end multiplexed sequencing libraries following Smart-seq2 (SS2) library preparation protocol 21. Barcoded libraries from each of the different plates were then sequenced using the Illumina HiSeq 2500. One of the 4 paired samples came from PsA patient (PSA1607) and cells from this paired sample were also processed using the 10x scRNA-seq pipeline (2.2.4.2). [2]

#### *2.2.4.4 10x droplet based scRNA-seq data mapping and pre-processing*

The 10x Genomics software package Cell Ranger (v2.1) was used in combination with the GRCh38 human reference genome to map and quantify sequencing reads, producing a filtered gene expression matrix containing only cell-associated barcodes and a count of the number of mRNA transcript molecules detected per gene for each barcode. The aggregated gene expression matrix was generated by combining the unstimulated CD4 and CD8 memory T cell matrix with the *Salmonella*-reactive CD154+Th memory cell matrix from patient PSA1607 using the Cell Ranger "aggr" pipeline.

#### *2.2.4.5 Quality control of individual 10x droplet based scRNA-seq gene expression matrices for ex vivo CD4 and CD8 memory T cells from both blood and synovial fluid.*

Gene expression matrices were first filtered to remove cells with > 10% mitochondrial genes, < 500 or > 3500 genes, and > 25000 UMI. Cells were further filtered to include only cells with corresponding CDR3 TCR data and to exclude potential multiplets, defined as cells with greater than 1 beta chain or 2 alpha chains or cells having both CD4 and CD8 gene expression (given the sorting strategy used).[2]

#### *2.2.4.6 Quality control of 10x droplet based scRNA-seq gene expression matrix for synovial Salmonella-reactive Th memory cells from patient PSA1607*

The gene expression matrix representing only *Salmonella*-reactive Th memory cells was first filtered to remove cells with > 10% mitochondrial genes, < 500 or > 5000 genes. Cells were further filtered where corresponding CDR3 TCR data was available to exclude as doublets cells with greater than 1 beta chain or 2 alpha chains.

#### *2.2.4.7 Quality control of 10x droplet based scRNA-seq aggregated gene expression matrix of synovial ex vivo CD4 and CD8 memory T cells and synovial Salmonella-reactive Th memory cells from patient PSA1607*

Gene expression matrices were first filtered to remove cells with > 10% mitochondrial genes, < 500 or > 2500 genes. Cells were further filtered where corresponding CDR3 TCR data was available to exclude as doublets cells with greater than 1 beta chain or 2 alpha chains.

#### *2.2.4.8 Quality control of SS2 plate based single cell gene expression matrices*

Cells with more than 5 median absolute deviations (MAD) 8.35% of their mRNA originating from mitochondrial genes; a total number of reads <500,000 or > 5000000; a total number of counts >3 median absolute deviations (MAD); or with number of genes <1000 or > 6000, were filtered out prior to downstream analysis. Where matching TCR data for a cell was available, any cells with greater than 1 beta chain or 2 alpha chains were additionally filtered.[2]

#### *2.2.4.9 Droplet based integrated gene expression analysis of peripheral blood and synovial fluid from 3 patients*

Cell Ranger output included 6 expression matrices (3 patients, each with paired blood and synovial fluid samples) and downstream analyses of these matrices was carried out

using R (3.6.0) and the Seurat package (v 3.0.2, satijalab.org/seurat). After quality control filtering and filtering to only include cells with available V(D)J sequencing information, data were subsampled to include an equal number of cells from blood and synovial fluid from all patients (6542 cells per sample, totaling 39252 cells). All 6 data sets were then individually normalised and variable genes discovered using the SCTransform function with default parameters, providing total number of UMIs and mitochondrial fraction as factors to regress out. The FindIntegrationAnchors function command was subsequently run with default parameters (dims = 1:30) to discover integration anchors across all samples. Any TCR genes were excluded from anchor features discovered prior to downstream analysis, and the IntegrateData function was run on this reduced anchorset with default additional arguments. ScaleData and RunPCA were then performed on the integrated assay to compute 30 principal components (PC). Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction was carried out and Shared Nearest Neighbor (SNN) graph constructed using dimensions 1:30 as input features and default PCA reduction. Clustering was performed on the Integrated assay at a resolution of 0.7 with otherwise default parameters which yielded a total of 16 clusters, each composed of cells originating from both blood and synovial fluid from all 3 patients and classified by differentially expressed genes.

To compare synovial (1750) and blood (828) cells from synovial and blood enriched clones respectively, all significantly enriched clones from both tissues were isolated from the previously filtered integrated data set using the SubsetData function. The integrated assay of this subset was scaled using the ScaleData function before running principal component analysis. Dimensionality reduction was calculated with UMAP

using the first 20 principal components and clustering set at a resolution of 0.4 to reflect the smaller cell numbers relative to the parent data set. This yielded 7 clusters. [2]

#### *2.2.4.10 Droplet and SS2 based integrated gene expression analysis of ex vivo CD4 and CD8 memory T cells from blood and synovial fluid of patient PSA1607*

For validation of sequencing data across platforms, quality control filtered 10x expression matrices from 1 patient (PSA1607) were subsampled to include an equal number of cells from blood and synovial fluid (11192 cells from each, totaling 22384 cells). Quality control filtered SS2 data matrices from this same patient were similarly subsampled to include 433 cells from both blood and synovial fluid. The same steps outlined above for integration of droplet based only data sets from 3 patients (2.2.4.9) were used to integrate these 4 data sets (blood and synovial from both 10x and SS2 platforms) for patient PSA1607. Apart from additionally regressing out platform when scaling data, the same optional arguments specifying 30 principal components and a cluster resolution of 0.7 were used. This yielded 13 clusters which were again classified by differentially expressed genes, bearing gene expression signatures that overlapped with clusters identified in the 10x only integration of 3 patients.[2]

#### *2.2.4.11 Analysis of 10x droplet based scRNA-seq gene expression matrix for synovial Salmonella-reactive Th memory cells from patient PSA1607*

The quality control filtered 10x expression matrix representing Salmonella-reactive CD154+Th memory cells was used to construct an RNA assay containing log normalised count data (using the default scale factor of 10000 ) which was also scaled using the ScaleData function, providing total number of UMIs and mitochondrial fraction as factors to regress out. A separate "SCT" assay was constructed for the same dataset, using the SCTransform function with argument to regress out mitochondrial

fraction and identify 3000 variable features. 75 principal components were then calculated using the SCT assay as argument to the Seurat RunPCA function and the SCT assay was set as default; a shared nearest neighbor (SNN) graph was constructed using the FindNeighbors function and RunUMAP used to create a 2-dimensional representation of the dataset (both Seurat functions). Finally clusters were computed using the Seurat FindClusters function, passing 0.8 as the resolution argument, which identified 7 clusters.

An IL17A-subset of this object was constructed using the Seurat "subset" function, passing it the arguments "subset = IL17A > 0, slot=data" , then recalculating 75 principal components, and re-running the FindNeighbors, FindClusters and RunUMAP Seurat functions, this time providing a resolution argument of 1.0 to the FindClusters function. This identified 3 clusters within the IL17A-subset.

#### *2.2.4.12 TCR mapping and analysis*

Chromium 10x V(D)J single-cell sequencing data was mapped and quantified using the software package Cell Ranger (v2.1) against the GRCh38 reference provided by 10x Genomics with that release. The generated consensus annotations files for each patient and sample type (blood or synovial fluid) were then used to construct clonality tables and input files for further downstream analysis. Full-length, paired T cell receptor (TCR) nucleotide sequences from SS2 data were constructed using the TraCeR program[3], and further mapped to V, D and J genes as well as CDR3 nucleotide and amino acid sequences using the online IMGT-HighVQuest tool[4]. The software package GLIPH v 1.0 was used to construct and assess specificity groups[5]. As the GLIPH algorithm only makes use of single CDR3 beta chain amino acid sequences to associate clones of common specificity, multiple beta chain sequences within the same

partition were treated as multiplets and provided as separate individual sequences to GLIPH (annotated with a "v" suffix). Partitions containing only alpha chain sequences were excluded from GLIPH input. The VDJtools 1.1.824 software package was used to construct circle plots illustrating V(D)J gene usage.

To assess clonal enrichment, the proportion of cells having the same clone was compared between sample types for each clone using a Fisher's exact test with Benjamini & Hochberg (1995) correction for multiple comparisons (R Stats Package). A cell's clonotype was defined as the combined alpha and beta chain CDR3 nucleotide sequences for that cell. As it was not possible to deduce beta and alpha chain pairing for partitions with multiple beta chains, these partitions were treated as a single clone.[2]

#### *2.2.4.13 Plate based ssRNA-seq expression quantification*

To assess the expression from the SS2 data, raw reads were pseudo-mapped and counted using kallisto v0.43.123 based on the annotation made by ENSEMBL(v90) of the human reference genome (GRCh38). To obtain per gene counts, all of the transcript counts were summarised using scater v1.6.326. [2]

# I. References


1. Hegazy, A.N., West, N.R., Stubbington, M.J.T., Wendt, E., Suijker, K.I.M., Datsi, A., This, S., Danne, C., Campion, S., Duncan, S.H., *et al.* (2017). Circulating and Tissue-Resident CD4+ T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. Gastroenterology *153*, 1320-1337.e16.

2. Penkava, F., Velasco-Herrera, M.D.C., Young, M.D., Yager, N., Lara, A.L., Guzzo, C., Maroof, A., Mamanova, L., Cole, S., Efremova, M., *et al.* (2019). Single-cell sequencing reveals a clonal expansion of pro-inflammatory synovial CD8 T cells expressing tissue homing receptors in psoriatic arthritis. bioRxiv, 704494.

3. Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat Meth *13*, 329–332.

4. Aouinti, S., Malouche, D., Giudicelli, V., Kossida, S., and Lefranc, M.-P. (2015). IMGT/HighV-QUEST Statistical Significance of IMGT Clonotype (AA) Diversity per Gene for Standardized Comparisons of Next Generation Sequencing Immunoprofiles of Immunoglobulins and T Cell Receptors. PLOS ONE *10*, e0142353.

5. Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., *et al.* (2017). Identifying specificity groups in the T cell receptor repertoire. Nature *547*, 94–98.

# Chapter 3: Fluorescence based single cell analysis of microbe reactive T cell responses in SpA

## 3.1 Introduction

Th17 cells are well known for their function in regulating microbial populations at barrier sites such as the gut and skin, but are also capable of inducing potentially undesired inflammation promoting autoimmunity[1]. A role for Th17 cells in the pathogenesis of SpA is suggested by genetic associations with the type 17 immune pathway; the efficacy of monoclonal antibody intervention targeting IL17 (Secukinumab[2]); and the ability of SpA animal models to transfer disease phenotype through CD4 T cells[3]. Evidence supporting an altered gut microbiome in SpA, and chronic inflammation after infection with pathogenic gut bacteria in ReA (discussed in chapter 1.5 ), further points to the interaction between CD4 T cells and the gut microbiota as a potential driver of disease in SpA.

Previous work in mice provides evidence that tolerance to commensal microbiota may be lost after exposure to pathogenic bacteria capable of damaging the intestinal epithelial barrier[4–6] . The aim of this chapter is to test whether aberrant or enhanced Th immune responses exist in SpA patients towards known commensal and pathogenic microbes typically found in the gut and at other barrier sites of humans. Flow cytometric analysis is used to measure the qualitative cytokine response of individual cells reactive to microbial antigen, while simultaneously quantifying the frequency of those cells within the CD4+ Th (T helper) memory compartment.

To assess immune responses, careful consideration was given to the type of assay chosen, from which tissue location cells would be sourced, the type of stimulation used

and appropriate experimental controls. A previously described[6–8] functional assay was adapted to measure both the frequency and phenotype of blood and synovial fluid Th cells reactive to a panel of microbial lysates (Table 3-1). This assay made use of CD154 expression on CD4 T cells, which is transiently upregulated upon TCR-mediated activation, to identify CD4 T cells reactive to particular microbial lysates. The function of cells reactive to microbes was simultaneously assessed by adding Brefeldin A during overnight stimulation and subsequently staining for the intracellular cytokines: TNFa, IFNg, IL-2, IL-17, GM-CSF and IL-22 (see Methods). Utilizing CD154 rather than cytokine expression as a marker of activation allowed for unbiased cytokine profiling of each individual cell in response to antigen stimulation, and for the possibility of FACS sorting live antigen-reactive cells required for downstream single-cell RNA sequencing (see Chapter 4).

Where possible, in addition to studying frequencies of circulating microbe reactive T cells, T cells from synovial fluid were also assayed. PsA, RA and less frequently AS patients can require therapeutic aspiration of synovial exudate from large joints such as the knees, providing an ideal opportunity to study cells from this inflammatory environment. As these cells are already in suspension, any unintended effects associated with tissue dissociation procedures are avoided. It was known from preliminary experiments that the vast majority of T cells present in SpA and RA synovial fluid express a CD45RA-negative memory phenotype, whereas the fraction of $CD45RA^-$ T cells from peripheral blood is typically lower and can vary greatly between individuals. To target the $CD45RA^-$ memory compartment and avoid biasing comparisons between synovial fluid and peripheral blood, a CD45RA antibody was included during staining and used to gate out $CD45RA^+$ cells before calculating the frequency of activated (CD154+) and cytokine expressing cells.

The panel of microbial lysates (Table 3-1) used for stimulations included the pathogenic bacteria *Salmonella typhimurium* (ST), a known causative agent of reactive arthritis; *Clostridium difficile*, the main cause of antibiotic related colitis; *Candida albicans,* a fungal microbe normally present in the intestine and on skin, but particularly susceptible to dysregulation in the absence of proper IL17 signalling[1,9]; *Staphylococcus aureus* and *Mycobacterium tuberculosis,* both well-characterised pathogenic bacteria associated with infection of skin and lung barrier sites respectively and capable of inducing strong IL17 and IFNg responses respectively; and a representative cross-section of common aerobic and anaerobic bacterial species from 4 different phyla found in the healthy gut (*Bifidobacterium lactis*, *Bacteroides vulgatus*, *Escherichia coli*, *Faecalibacterium prausnitzii*, *Lactobacillus acidophilus*, *Roseburia intestinalis*, and *Ruminococcus obeum*). In addition *Chlamydia trachomatis* lysate was used when stimulating cells from a patient newly diagnosed with *Chlamydia* induced ReA. The superantigen SEB was used as a positive control representative of strong TCR mediated activation, while medium only and LPS derived from *Salmonella enterica* were used as negative controls for unstimulated and potential bystander induced activation respectively.

**Table 3-1 Lysates used for stimulation assay**

| Lysate abbreviation | Source microbe | Phyla:family |
|---|---|---|
| BL | *Bifidobacterium animalis subsp. lactis* (Bi-07) | Actinobacteria:Bifidobacteriaceae; |
| BV | *Bacteroides vulgatus* (Bv1447) | Bacteroidetes:Bacteroidaceae; |
| CA | *Candida albicans* | Fungi:Saccharomycetaceae; |
| CD | *Clostridium difficile* (OXF1003, Toxin AB-) | Firmicutes:Clostridiaceae; |
| CT | *Chlamydia trachomatis* | Chlamydiae:Chlamydiaceae; |
| EC | *Escherichia coli* (Nissle 1917) | Proteobacteria:Enterobacteriaceae |
| FP | *Faecalibacterium prausnitzii* (A2-165) | Firmicutes:Ruminococcaceae; |
| LA | *Lactobacillus acidophilus* (NCFM) | Firmicutes:Lactobacillaceae |
| MT | *Mycobacterium tuberculosis* (H37Ra) | Actinobacteria:Mycobacteriaceae |
| RI | *Roseburia intestinalis* (M50/1) | Firmicutes:Lachnospiraceae |
| RO | *Ruminococcus obeum* (A2-162) | Firmicutes:Lachnospiraceae |
| SA | *Staphylococcus aureus* (NCTC 6571) | Firmicutes:Staphylococcaceae |
| ST | *Salmonella enterica serovar typhimurium* (NCTC12023) | Proteobacteria:Enterobacteriaceae |

## 3.2 Aim

To test whether aberrant or enhanced immune responses exist in SpA patients towards known commensal or pathogenic microbes typically found in the human gut.

## 3.3 Results

### 3.3.1 CD154 assay development and considerations

#### *3.3.1.1 Quantity of cells required and sample availability*

The purpose of the CD154 assay used for this study was to detect both the frequency and phenotype of microbe reactive Th memory cells in PBMC and SFMC samples. The frequency of CD4 T cells in circulation reactive to a particular antigen can be very rare, and functional subsets of reactive cells rarer still. Based on previous findings [6,7], 1.5x10^6 PBMC were sufficient to detect microbe reactive responses to all microbial lysates listed in Table 3-1 from healthy individuals. A total of approximately 23 million PBMC or SFMC per volunteer were therefore required to perform stimulations using each lysate from Table 3-1, in addition to positive and negative controls. This was considered a reasonable number of mononuclear cells to obtain from 30 mL of whole blood, but also limited the number of lysates that could be tested. Occasionally the number of cells obtained was too low to test all lysates. For example where less than 30mL of blood was obtained from a patient, or where synovial fluid was largely acellular. In such situations a positive and negative control was always used and an appropriate number of lysates chosen to maintain 1.5x10^6 cells per stimulation.

This study assessed PBMC from 31 SpA patients to provide enough power to detect equivalent or enhanced immune responses to those previously observed in 30 healthy individuals[6]. In addition 14 healthy volunteers took part in this study, which provided sufficient power to detect an increased mean frequency of CD154+ Th memory cells reactive to any lysates also found capable of inducing a CD154 response in SpA PBMC (discussed in more detail in section 3.3.2). Less data was available regarding expected frequencies of microbe reactive CD4 T cells from synovial fluid and synovial fluid samples were more difficult to obtain. However, statistical power when comparing

responses between blood and synovial fluid was aided by paired samples from the same patient usually being available. Unfortunately the small number of synovial fluid samples from RA patients obtained during this study would have resulted in underpowered statistical calculations and the decision was made not to include RA synovial fluid samples as part of the analysis for this thesis. Ongoing work will attempt to increase the number of RA synovial fluid samples assayed to provide statistically meaningful results.

**Table 3-2 Number of patient and control samples for each type of stimulation**

| Stimulation | SpA blood | HC blood | RA blood | SpA synovial fluid | RA synovial fluid |
|---|---|---|---|---|---|
| No antigen | 30 | 12 | 6 | 9 | 4 |
| LPS-SE | 31 | 14 | 16 | 8 | 4 |
| BV | 30 | 14 | 15 | 8 | 4 |
| RI | 30 | 14 | 15 | 8 | 4 |
| RO | 30 | 14 | 15 | 8 | 4 |
| FP | 30 | 14 | 15 | 8 | 4 |
| LA | 30 | 14 | 15 | 8 | 4 |
| BL | 30 | 14 | 17 | 8 | 4 |
| CA | 31 | 14 | 16 | 8 | 4 |
| EC | 31 | 14 | 17 | 8 | 4 |
| CD | 30 | 14 | 15 | 8 | 4 |
| ST | 31 | 14 | 17 | 9 | 4 |
| CT | 18 | 7 | 0 | 5 | 0 |
| SA | 31 | 14 | 17 | 9 | 4 |
| MT | 31 | 14 | 17 | 7 | 4 |
| SEB | 31 | 14 | 17 | 9 | 4 |

#### *3.3.1.2 Minimising the effect of non-specific antibody binding and potential artefacts*

Effectively analysing greater than 1x10^6 cells using flow cytometry presented several issues not commonly encountered in lower throughput assays. Plate based acquisition, although desirable for speed and automation, proved unreliable. A great deal of forward and side scatter variation when plotted against time was observed if wells were loaded with more than 2x10^5 cells. In addition, each well required a "dead" volume of approximately 50ul to avoid the uptake of air, reducing the maximum number of cells which could be analysed per test. For these reasons, all acquisitions using the BD LSRFortessa Cell Analyzer were performed in "tube mode". To further minimise any dead volume, samples were first placed into a 1.2mL polypropylene cluster tube before being placed into a standard 5mL FACS tube (see methods). This allowed samples to be loaded at high concentration in a small volume, increasing the speed at which each sample could be acquired.

Antibodies (see methods) were first titrated against unstimulated and SEB stimulated healthy cone blood to ensure an adequate degree of separation could be achieved between positive and negative populations when staining 1.5x10^6 PBMC in 50ul of wash buffer.

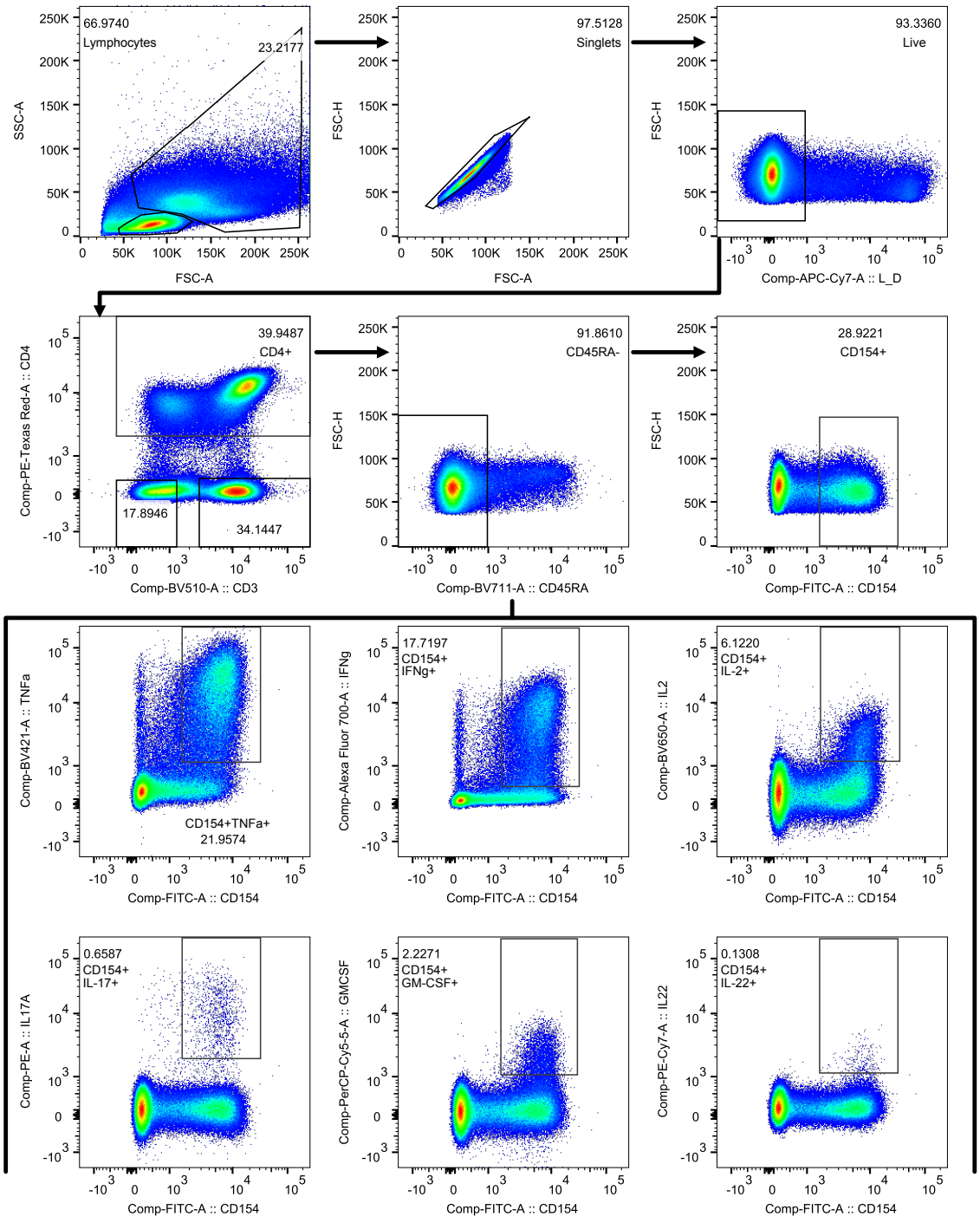


**Figure 3-1 SEB stimulation of mononuclear cells upregulates CD154, TNFa, IFNg, IL-2, IL-17, GM-CSF and IL-22 expression on Th memory cells.**

Flow cytometric gating of 1.5x10^6 SFMC from an AS patient stimulated overnight with SEB and subsequently stained for surface markers: CD3, CD4, CD45RA, and viability dye; and intracellularly for: CD154, TNFα, IFNγ, IL-2, IL-17A, GM-CSF and IL-22. After gating on lymphocytes (SSC-A vs. FSC-A) and single cells (FSC-H vs FSC-A) with reduced or no uptake of viability dye, CD4+CD45RA- cells were further gated on expression of CD154 alone and in combination with cytokines. Both CD4+CD3+ and CD4+CD3- cells were included when gating for CD4+ T cells due to the down regulation of CD3 after stimulation with SEB. Plots are representative of positive control gating strategy used for both PBMC and SFMC stimulated overnight with SEB.

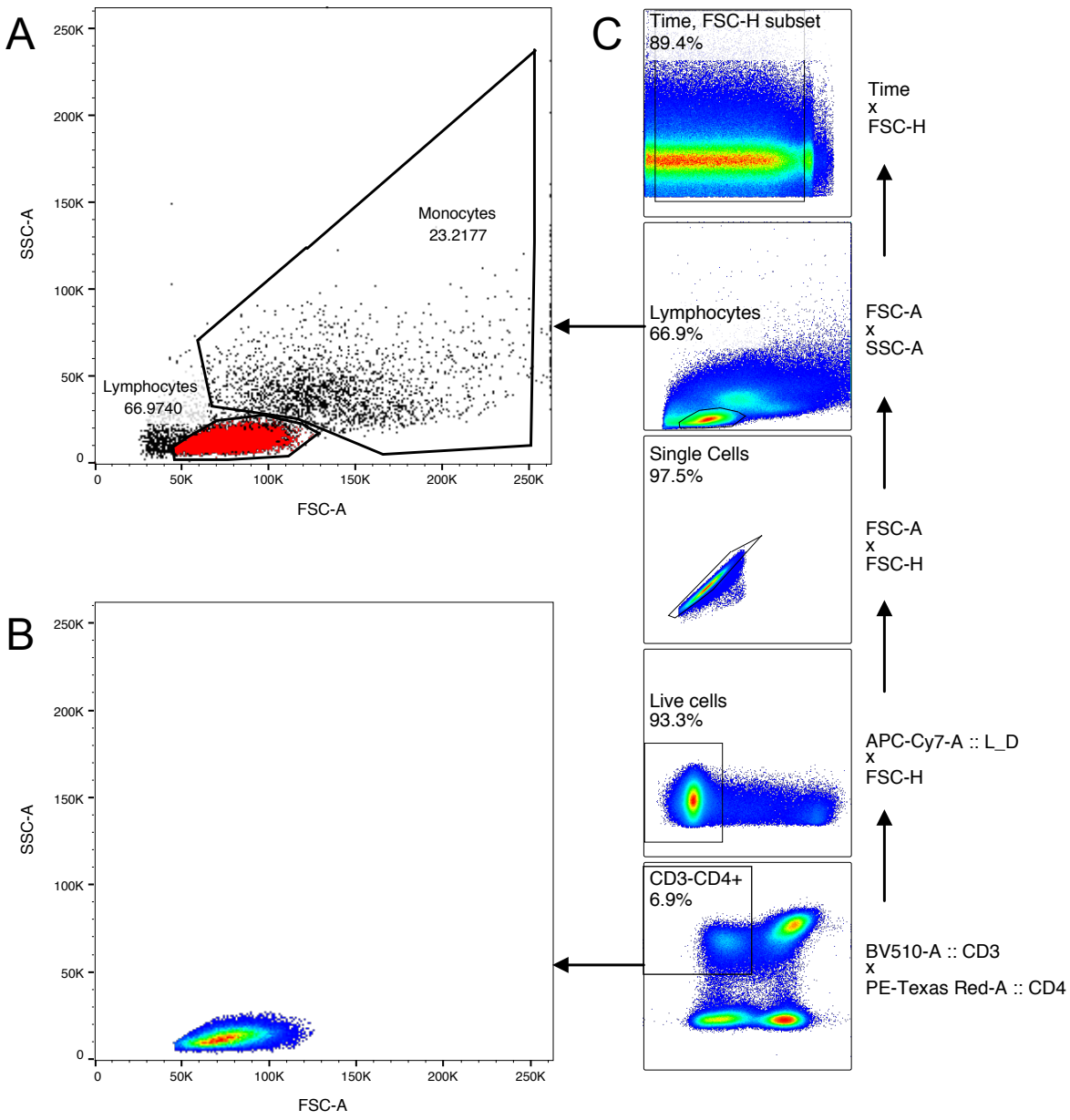


**Figure 3-2 CD3 is downregulated after stimulation with SEB**

FACS plots showing the distribution of CD3-CD4+ mononuclear cells within the upstream FSC-A x SSC-A lymphocyte gate after a 16 hour stimulation with SEB. (A) Overlay of cells from B (red) onto time gated cells showing location within a FSC-A x SSC-A plot. (B) Pseudocolour dot plot of the FSC-A x SSC-A distribution of CD3-CD4+ gated cells from C, with highest to lowest concentration indicated in descending order by red, yellow, green and blue. (C) Hierarchical gating of the CD3-CD4+ population depicted in B. Cells from the same SEB stimulated and stained sample from Figure 3-1 are shown.

SEB stimulation of PBMC and SFMC was used to induce CD154 and cytokine expression for positive control staining (Figure 3-1). Due to the downregulation of CD3 after stimulation with SEB[10], unstimulated cells were used for the titration of surface markers, and the CD3+CD4+ gate of Th cells was extended to include CD3- cells when gating SEB stimulated cells. Figure 3-2 confirms that these CD3- cells have a very similar distribution compared to other cells within the lymphocyte gate, and are therefore highly unlikely to be due to unexcluded CD4+ monocytes. If these CD3- cells were in fact CD4+ monocytes, they would be expected to cluster towards the FSC-A x SSC-A monocyte gate, which they do not. A FITC isotype control was used in lieu of the FITC conjugated CD154 antibody to estimate expected non-specific binding when staining for CD154 or CD154 and cytokines. This was shown to be minimal (Table 3-3)

**Table 3-3 Expected background noise from non-specific CD154-FITC antibody binding within gated populations***

| Gate | Mean (%) | Standard Deviation (SD) | Max (Mean + 3SD) | Cells per 100 000 Th-memory cells | |
|---|---|---|---|---|---|
| | | | | Mean | Max |
| $CD154^+TNF\alpha^+$ | 0.00139 | 0.001704 | 0.00651 | 1-2 | 6-7 |
| $CD154^+IFN\gamma^+$ | 0.00047 | 0.000953 | 0.00333 | 0-1 | 3-4 |
| $CD154^+IL\text{-}2^+$ | 0.00028 | 0.000907 | 0.00300 | 0-1 | 3 |
| $CD154^+IL\text{-}17A^+$ | 0.00051 | 0.000777 | 0.00284 | 0-1 | 2-3 |
| $CD154^+GM\text{-}CSF^+$ | 0.00035 | 0.000671 | 0.00237 | 0-1 | 2-3 |
| $CD154^+IL\text{-}22^+$ | 0.00019 | 0.000555 | 0.00186 | 0-1 | 1-2 |
| Total $CD154^+$ | 0.00690 | 0.006320 | 0.02586 | 6-7 | 25-26 |

*Calculations based on n=37 samples.

As can be seen in Figure 3-1 (and Figure 3-3), there is a level of CD154 upregulation which associates poorly with cytokine production. To maintain the sensitivity of results when characterising cytokine production of CD154+ cells, and to ensure that only cells confidently classified as upregulating CD154 in response to antigen stimulation were assessed, a conservative gating approach excluding low levels of CD154 upregulation was used. This approach captured the vast majority of events double positive for both CD154 and cytokine staining, while helping to exclude potential false positive events likely attributable to non-specific binding.

In addition, to mitigate the effects of clogging or back-pressure which can occasionally occur during flow cytometric acquisition, flow stability gating was manually applied as an initial gate on forward scatter versus time. This typically excluded the first 10 seconds of acquisition before a stable flow rate could be achieved. An example of this time gating is seen in Figure 3-2C.

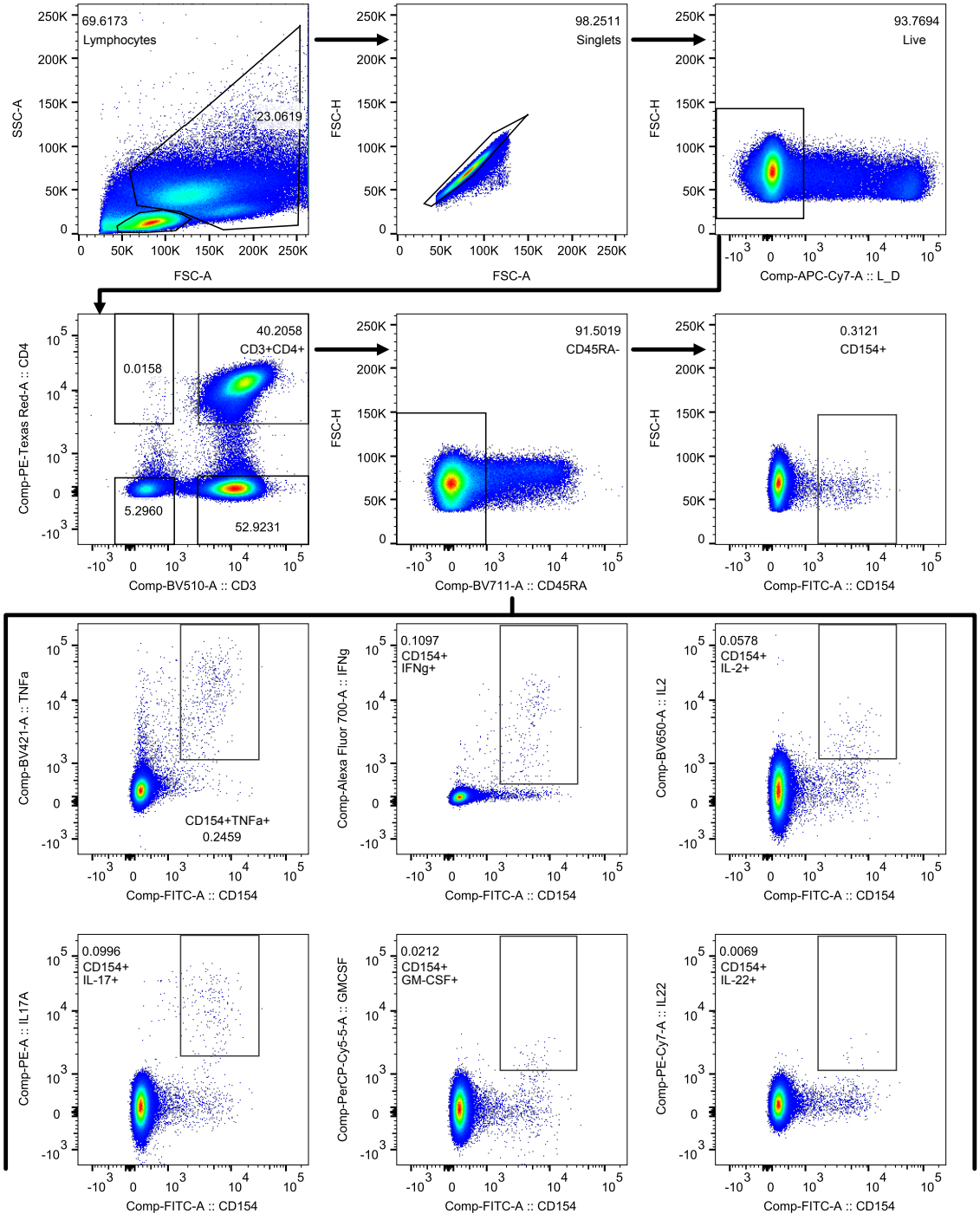


**Figure 3-3 Representative staining of SFMC from AS patient stimulated with *S. aureus* for 16 hours.**

Flow cytometric gating of 1.5x10^6 SFMC from an AS patient stimulated overnight with *S. aureus* lysate and subsequently stained for surface markers: CD3, CD4, CD45RA, and viability dye; and intracellularly for: CD154, TNFα, IFNγ, IL-2, IL-17A, GM-CSF and IL-22. After gating on lymphocytes (SSC-A vs. FSC-A) and single cells (FSC-H vs FSC-A) with reduced or no uptake of viability dye, CD3+CD4+CD45RA- cells were further gated on expression of CD154 alone, and in combination with cytokines. Plots are representative of the gating strategy used for both PBMC and SFMC stimulated overnight with medium only, microbial lysates or LPS.

#### *3.3.1.3 Titration and quantification of lysates used.*

The concentration of each lysate used was titrated as previously described[6] by a collaborator, Ahmed Hegazy, to obtain an optimal level of stimulation for each lysate. All samples described in this chapter were analysed together in one batch, however 2 healthy control and 14 RA blood samples were processed, stimulated and acquired using flow cytometry by A. Hegazy using the same protocols outlined in methods. To ensure consistent comparisons could be made between these stimulations, and with previous work by the same collaborator, the same lysates were used at the same concentrations previously described for all stimulations discussed in this thesis[6].

#### *3.3.1.4 Use of fresh PBMC and SFMC*

Upregulation of CD154 expression after stimulation with the lysates used as part of this study are greatly reduced when stimulation of PBMC occurs in the presence of MHC-II blocking antibodies [6], supportive of TCR mediated activation being reliant on the presence of antigen presenting cells during stimulations. This study similarly found that lysate stimulation of CD4 T cells alone, or stimulation of PBMC thawed from liquid nitrogen containing a reduced number of monocytes, was not capable of inducing a measurable CD154 response. Therefore, only freshly obtained PBMC or SFMC were used for this study.

### 3.3.2 Peripheral blood responses to microbial lysates

There was no significant difference in the baseline frequency of CD154+ Th memory cells present in unstimulated PBMC when compared with PBMC stimulated with LPS for any volunteer group (Figure 3-4). This suggested that LPS stimulation alone could not induce a level of bystander activation sufficient to effect CD154+ expression.

When healthy control PBMC were stimulated with microbial lysates, the frequency of CD154+ Th memory cells was higher than LPS stimulation for all lysates derived from pathogenic microbes as well as the commensal microbe *B. lactis* and the non-pathogenic nissle 1917 strain of *E. coli*. (Figure 3-5). No significant response was detected against *B. vulgatus*, *R. intestinalis*, *F. prausnitzii* or *L. acidophilus* lysates. Increased frequencies of CD154+ Th memory cells reactive to these lysates in healthy peripheral blood has previously been reported[6] when 30 volunteers were assayed. It is likely that more healthy volunteers would be required to detect responses against these lysates in this study. However, as 31 SpA patient PBMC samples were assayed, and as the same concentration of the same lysates were used for these experiments as previously described[6], it was expected that any equivalent or enhanced response to these lysates in SpA blood samples would be detected if present.

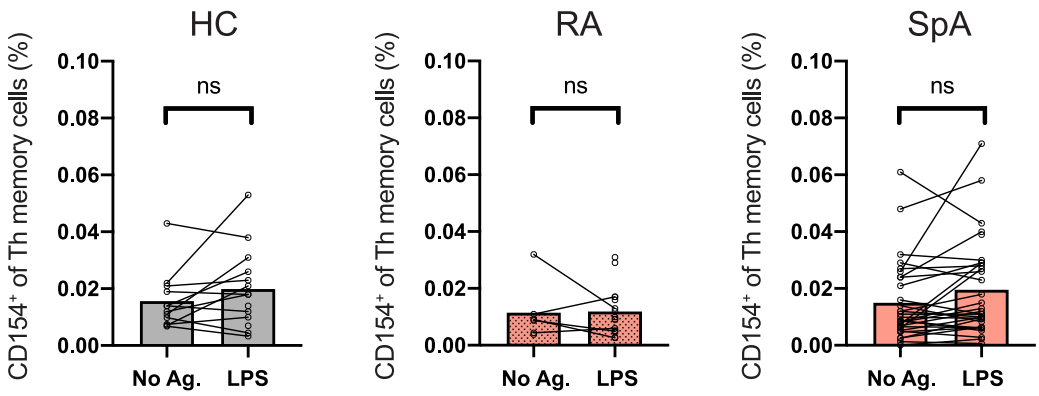


**Figure 3-4 LPS stimulation of PBMC does not alter the frequency of CD154+ Th memory cells in SpA, RA or healthy controls.**

Mean frequency of healthy control, RA and SpA Th memory cells staining positive for CD154 expression after 16 hours of stimulation with LPS or no stimulation. Paired samples are joined with lines. Statistics: Wilcoxon matched-pairs signed rank test calculated on paired samples only. Number of pairs for HC, RA and SpA is 12, 5 and 30 respectively. No significant differences were found. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; HC: healthy control; RA: rheumatoid arthritis; SpA: spondyloarthropathy.

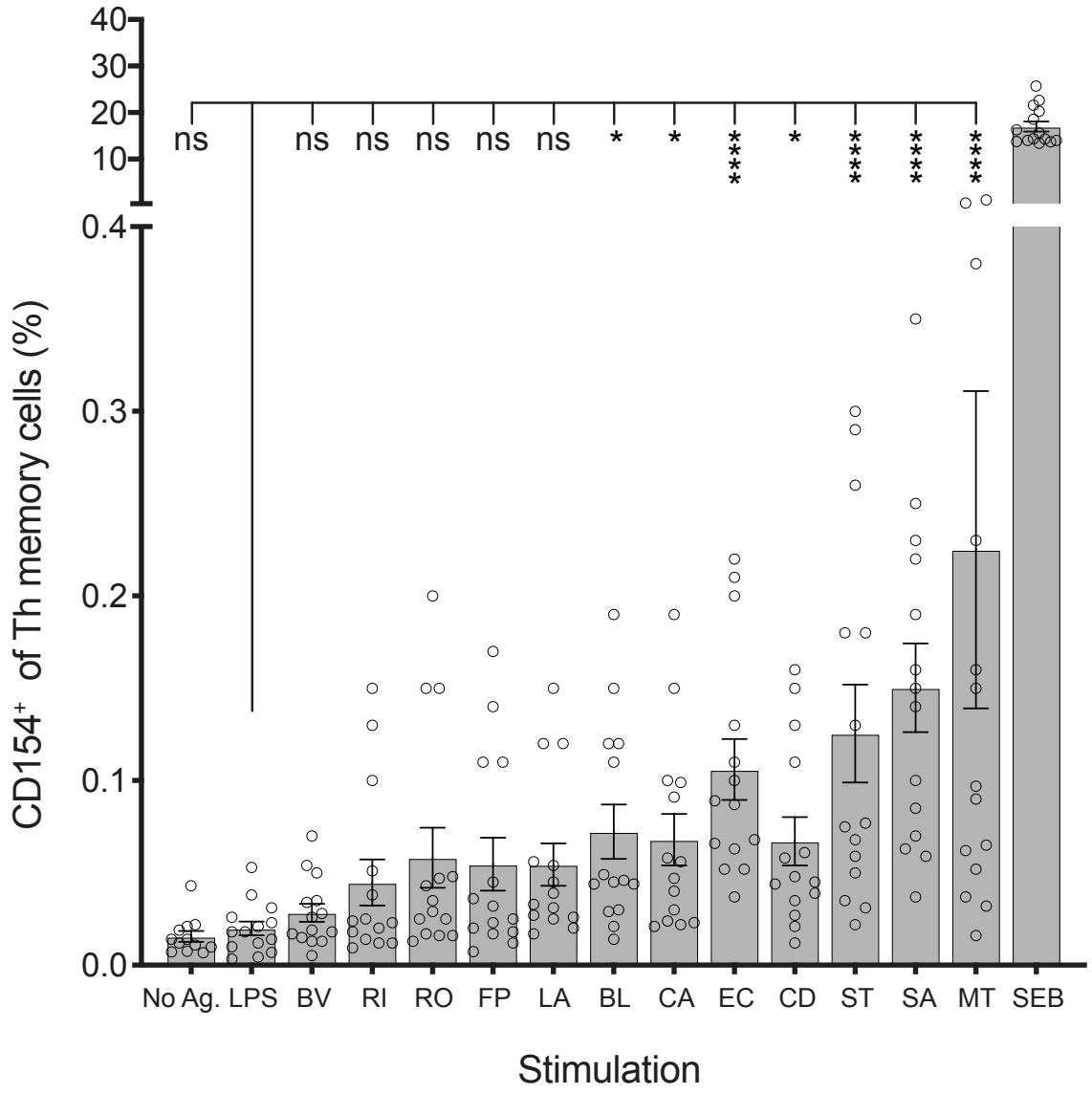


**Figure 3-5 Healthy blood contains circulating Th memory cells reactive to *B. lactis*, *C. albicans*, *E. coli*, *C. difficile*, *S. typhimurium*, *S. aureus* and *M. tuberculosis*.**

Mean frequency (± SEM) of healthy control Th memory cells staining positive for CD154 expression after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation is shown. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=14 for LPS, n=12-14 for lysates and no stimulation. *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

When comparing CD154+ Th memory frequencies between SpA PBMC stimulated with either microbial lysates or LPS, a response higher than LPS stimulation could only be detected towards *Candida, E. coli*, *S. typhimurium, S. aureus* and *M. tuberculosis.* This suggests that circulating Th memory cells capable of mounting a TCR mediated immune response against the commensal microbes assayed as part of this study (other than E. coli) are largely absent in SpA peripheral blood (Figure 3-6). This finding was mirrored in RA blood results (Figure 3-6). When comparing blood immune responses side by side between SpA, RA and healthy controls (Figure 3-7), a reduced frequency of *E. coli*, *S. typhimurium* and *M. tuberculosis* reactive cells is evident in both SpA and RA compared to healthy blood. SpA patients and healthy volunteers did however share a similar mean frequency of CD154+ cells reacting to the skin pathogen *S. aureus* in peripheral blood, whereas this frequency was lower in RA patient blood.

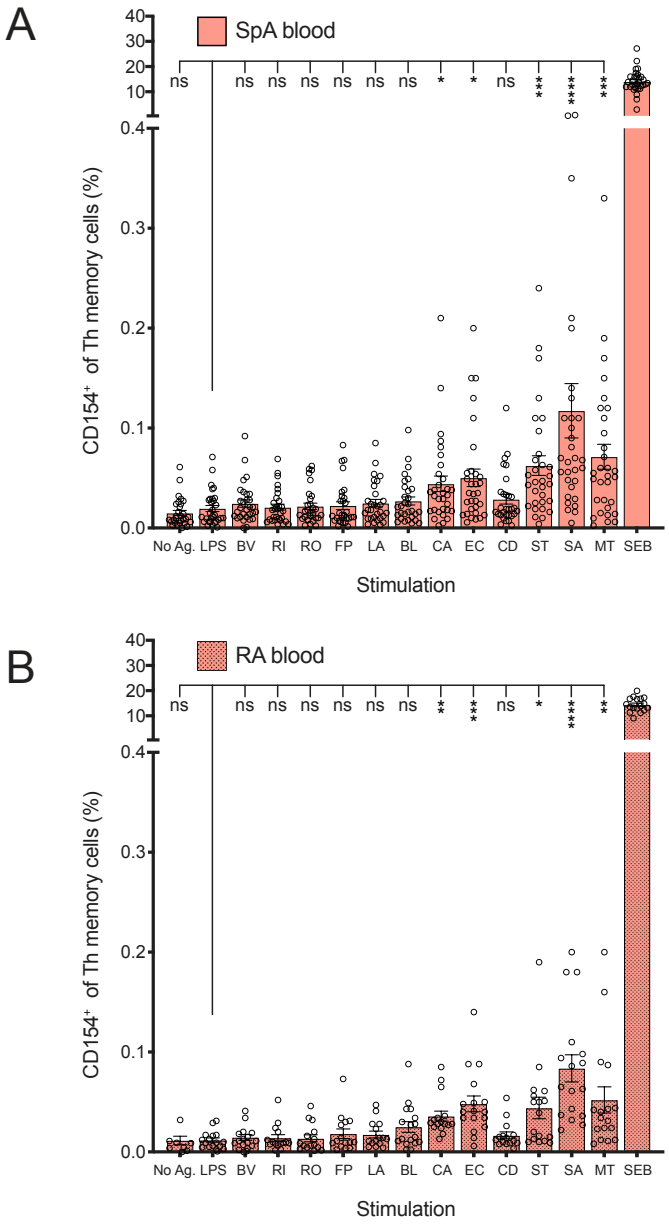


**Figure 3-6 Th memory cells were detected in both SpA and RA blood reactive to *C. albicans*, *E. coli*, *S. typhimurium*, *S. aureus* or *M. tuberculosis*; but not to *B. vulgatus*, *R. intestinalis*, *R. obeum*, *F. prausnitzii*, *L. acidophilus* or *C. difficile.***

Mean frequency (± SEM) of SpA (A) and RA (B) Th memory cells staining positive for CD154 expression after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium (No Ag.) in comparison to negative control LPS stimulation are shown. CD154 responses higher than LPS control stimulation were only detected against pathogenic microbes *C. albicans*, *S. typhimurium*, *S. aureus* and *M. tuberculosis*; and the commensal microbe *E. coli* (Nissle 1917). Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; A/B: n=31/16 for LPS, 30/6 for no antigen and 30-31/15-17 for lysates. SEB was included as a positive assay control but excluded from comparisons. Statistics: $*P \leq .05$; $**P \leq .01$; $***P \leq .001$; $****P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

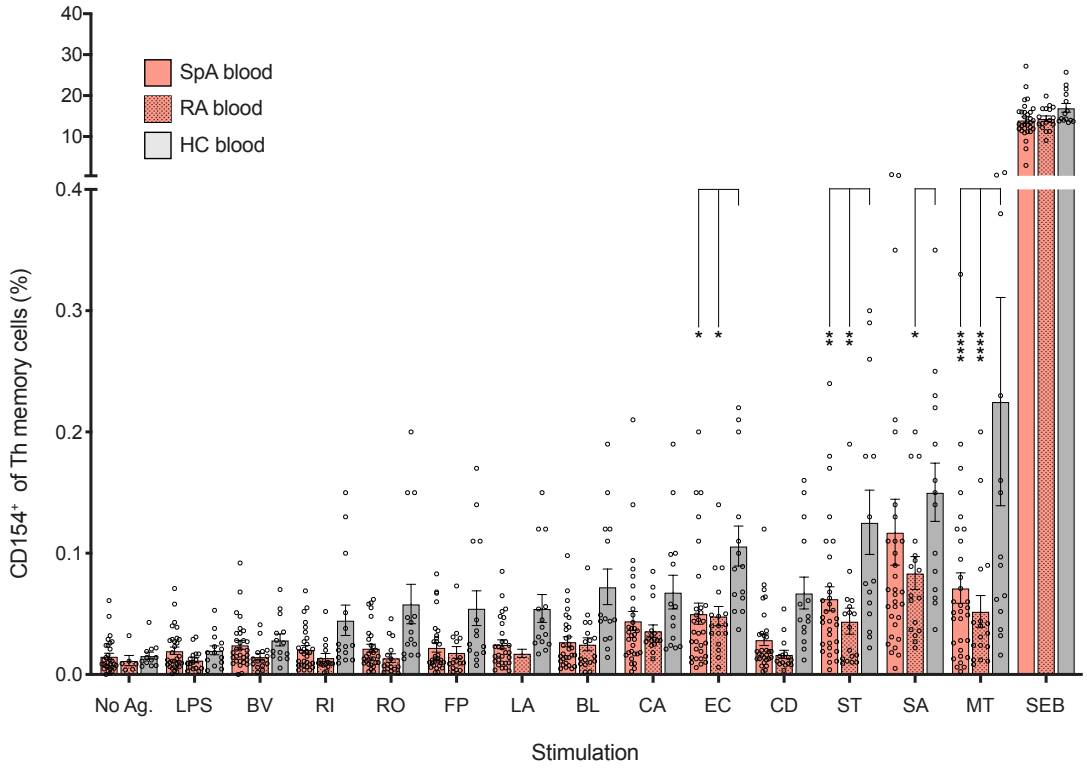


**Figure 3-7 There are fewer *E. coli*, *S. typhimurium* and *M. tuberculosis* reactive Th memory cells in SpA and RA blood compared to healthy controls.**

Mean frequency (± SEM) of SpA, RA and HC CD154+ peripheral blood Th memory cells after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. The mean *S. aureus* reactive frequency was higher in HC compared to RA but not SpA blood. No significant difference between SpA and RA frequencies were found for any stimulation. Statistics: 2-way ANOVA with Tukey's correction for multiple comparisons; Lysate and LPS stimulations n=30-31,15-17 and 14 for SpA, RA and HC respectively. SEB was included as a positive assay control but excluded from comparisons. Statistics: $*P \leq .05$; $**P \leq .01$; $***P \leq .001$; $****P \leq .0001$, no stars = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

#### *3.3.2.1 S. typhimurium and M. tuberculosis reactive Th memory cells in SpA blood have an impaired IL-2 response.*

Figure 3-8 compares the cytokine responses of Th memory cells found to be CD154+ in response to microbial lysate stimulation in blood. This answers a complimentary but different question compared to results in Figure 3-7. Whereas Figure 3-7 showed the frequency of cells reacting to microbial lysate, Figure 3-8 shows what percentage of those cells reactive to microbial lysate are also producing cytokines in response to antigen specific stimulation. Lysates were only included where the frequency of CD154+ Th memory cells in SpA and RA blood reactive to those lysates had previously been shown to be above LPS stimulation (Figure 3-6), and where a sufficient number of CD154+ cells were reactive to lysates for a meaningful comparison (> 0.03% of Th memory cells). This cut-off was set to be at least 4 times higher than the potential level of background non-specific staining (0.0069%, Table 3-3) and with consideration for the expected number of Th memory cells present within a sample (typically between 1x10^5 and 2x10^5). Therefore at least 30 cells of 1x10^5 would need to stain positive for CD154 to be included for analysis.

These results indicate that even though there is no significant difference in the TNFa and IFNg response to *S. typhimurium* and *M. tuberculosis* in SpA blood, a reduced frequency of Th memory cells reactive to these microbes are capable of eliciting an IL-2 response compared to both RA and healthy control samples. In addition, a lower frequency CD154+ cells producing IL-2 in response to *E. coli* and *S. aureus* stimulation was found in SpA compared with RA, however this difference was not evident when compared to healthy control samples. It is therefore more likely to be a feature of RA rather than SpA.

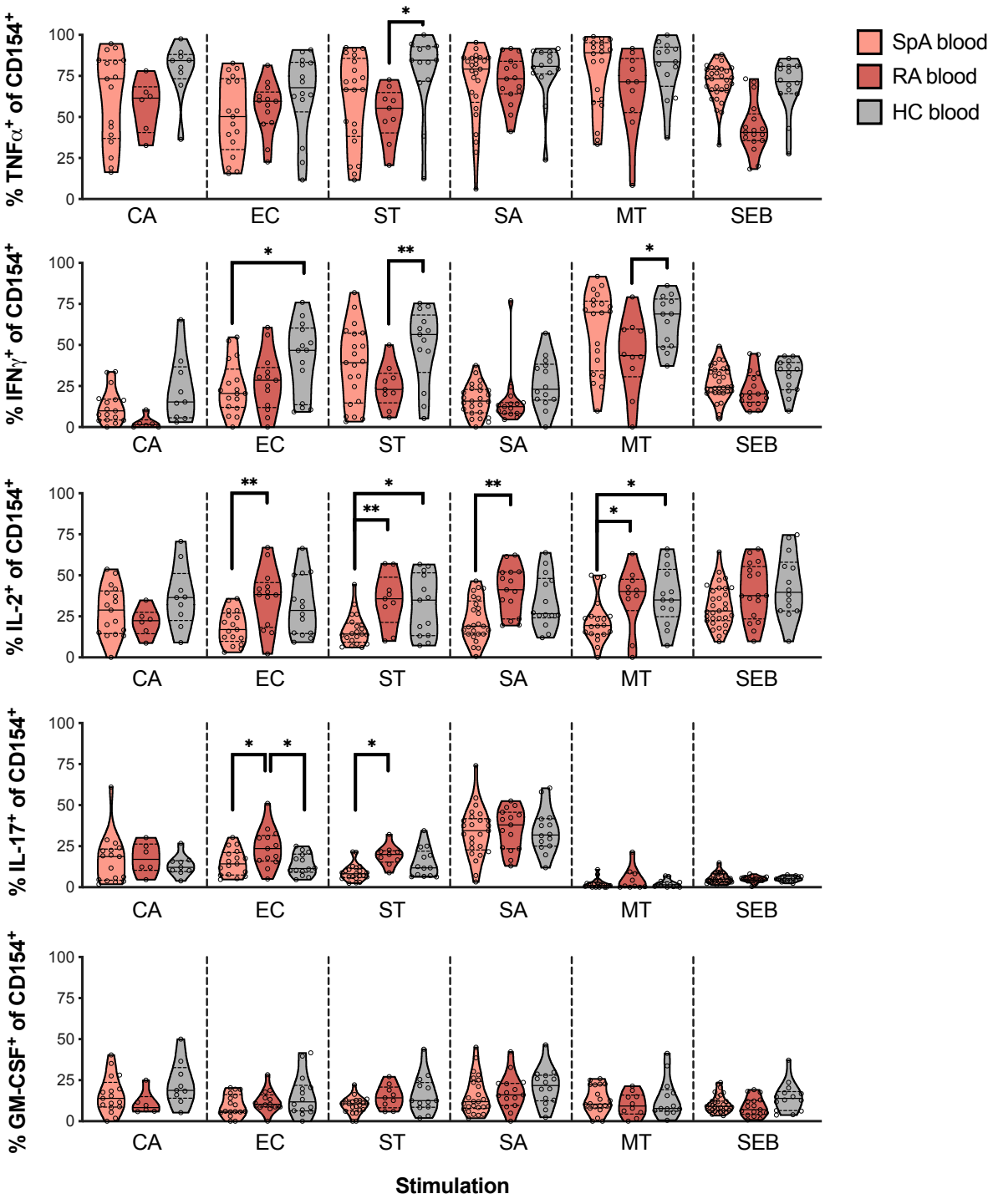


**Figure 3-8 *S. typhimurium* and *M. tuberculosis* reactive Th memory cells in SpA blood have an impaired IL-2 response.**

Proportion of microbe reactive CD154+ Th memory cells expressing TNFa, IFNg, IL-2, IL-17 or GM-CSF in SpA, RA and HC blood after 16 hour stimulation with microbial lysates or SEB. Samples were only included where the frequency of CD154+ cells as a percentage of Th memory cells was greater than 0.03%. Solid line: median, dashed lines: quartile. Statistics: 2-way ANOVA with Tukey's correction for multiple comparisons; Lysate stimulations n=17-25, 6-15 and 9-14 for SpA, RA and HC samples respectively. SEB was included as a positive assay control but excluded from comparisons. Statistics: *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$, no stars = not significant. Abbreviations: CA:*Candida albicans*; EC:*Escherichia coli*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

The mean frequency of Th memory cells reactive to *S. typhimurium* and *M. tuberculosis* was similar between SpA and RA (Figure 3-7). The lower IL-2 response in SpA therefore does not seem related to a difference in the overall number of cells reactive to these lysates. There is also no significant difference in the SpA IL-2 response to *C. albicans*, suggesting that the observed impaired IL-2 response in SpA blood is specific to particular microbes.

### 3.3.3 Synovial fluid responses to microbial lysates

The frequency of unstimulated CD154+ Th memory cells in SpA synovial fluid was significantly higher than in blood (Figure 3-9A). Stimulation with LPS seemed to increase this frequency further (Figure 3-9B). When stimulating with microbial lysates, no lysate was found to increase the frequency of CD154+ Th memory cells above the frequency observed when stimulating with LPS alone. This observation was also true for Th memory cells staining positive for both CD154 and TNFa after stimulation (Figure 3-10).

A response above LPS stimulation for co-expression of CD154 and IFNg was only seen when stimulating with *S. typhimurium* (Figure 3-11), and co-expression of CD154 with IL-17 only after stimulation with either *S. typhimurium* or *S. aureus* (Figure 3-12). CD154+IL-2+ responses above LPS were seen for pathogenic microbes *C. albicans*, *S. typhimurium*, *S. aureus* and *M. tuberculosis* as well as the commensal microbes *E. coli* (nissle 1917) and *B. lactis* (Figure 3-13). *C. albicans* stimulation also elicited a CD154+GM-CSF+ response above LPS, as did stimulation with *S. typhimurium*, *S. aureus* and *M. tuberculosis* (Figure 3-14). *S. aureus* was the only microbe capable of inducing a CD154+IL-22+ response above LPS stimulation (Figure 3-15Figure 3-14).

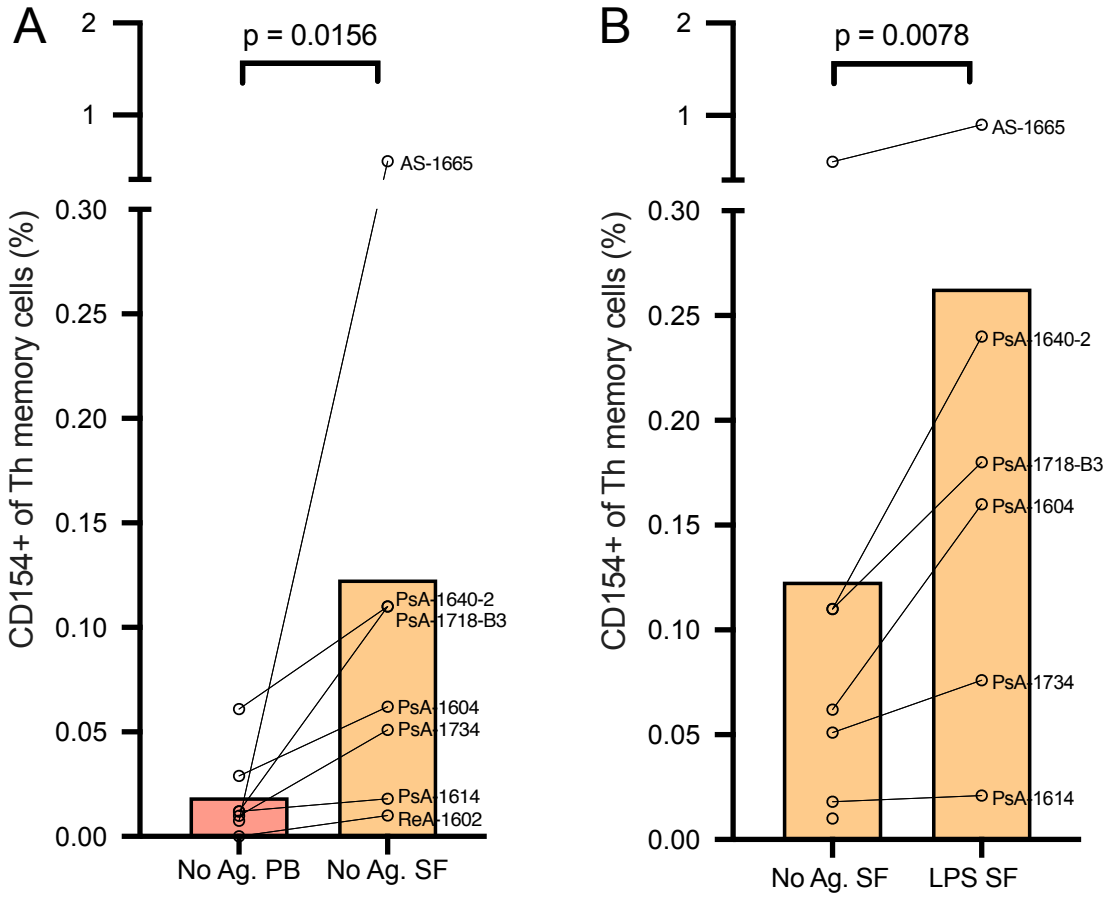


**Figure 3-9 A higher frequency of Th memory cells in SpA synovial fluid express CD154 without stimulation compared to blood.**

Mean frequency of SpA Th memory cells staining positive for CD154 expression. (A) Comparison of frequencies between peripheral blood and synovial fluid in the absence of any stimulation. (B) Effect of LPS stimulation on frequency of Th memory cells expressing CD154 in synovial fluid. Paired samples are joined with lines. Statistics: Wilcoxon matched-pairs signed rank test calculated on paired samples only. Only samples where matching unstimulated blood and synovial fluid from the same patient was available are shown. Number of pairs in A:7 ; B:6. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide.

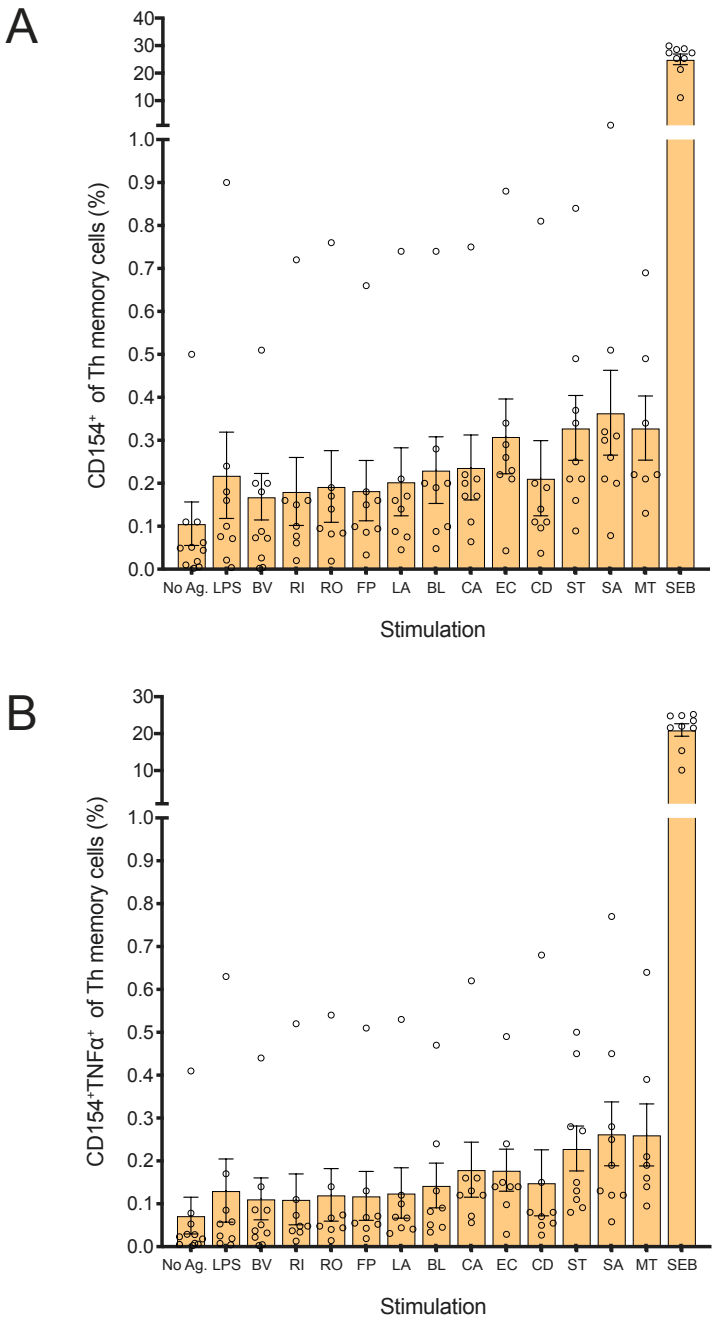


**Figure 3-10 Microbial lysate stimulation of SFMC did not increase the frequency of CD154+ or CD154+TNFa+ Th memory cells above LPS stimulation alone.**

Mean frequency (± SEM) of SpA Th memory cells from SFMC staining positive for CD154 expression alone (A), or co-expression of CD154 and TNFa (B) after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. No significant CD154+ or CD154+TNFa+ responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation were found. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=7-9 for lysates and no stimulation, n=8 for LPS; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

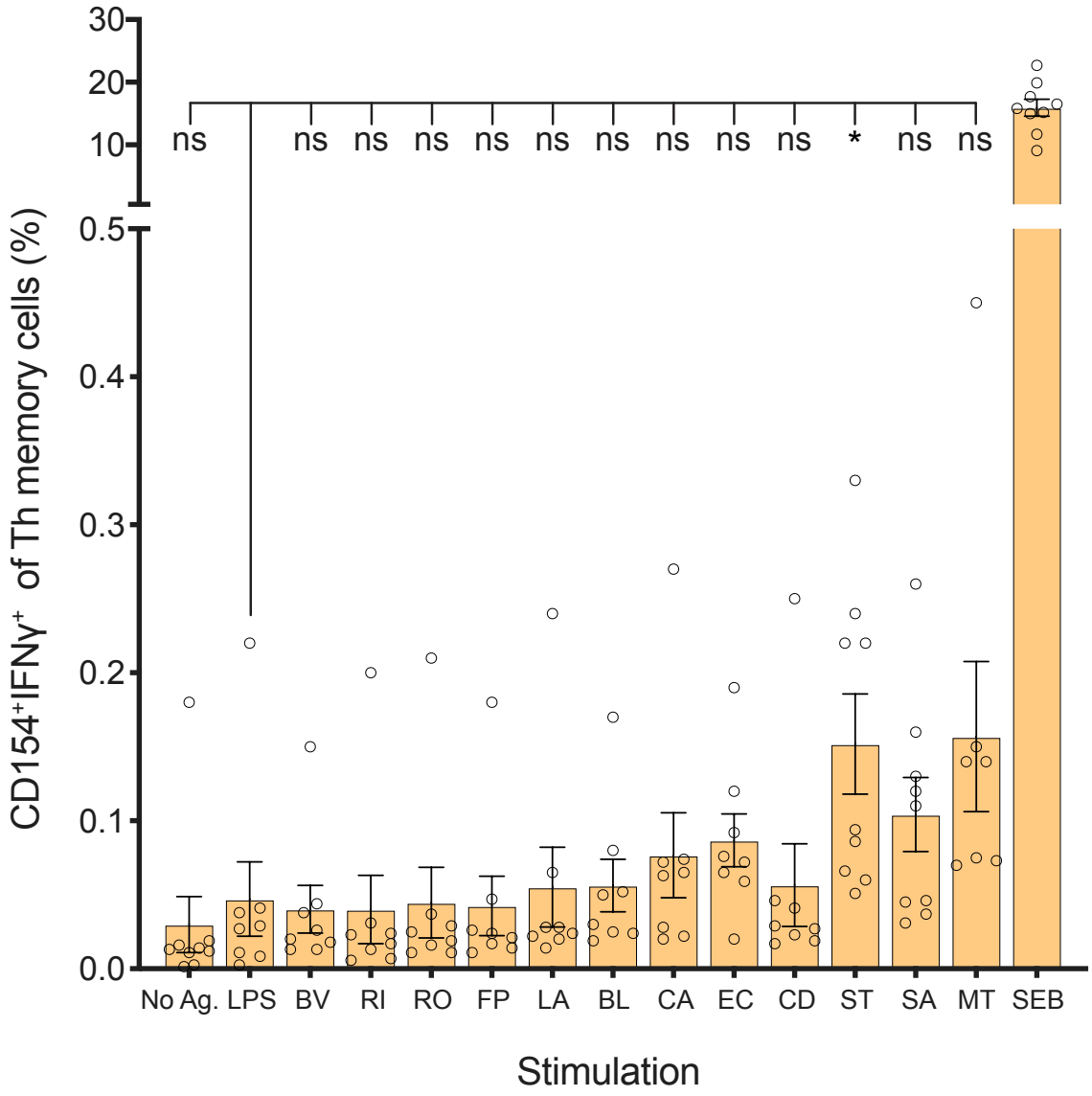


**Figure 3-11 *S. typhimurium* reactive Th memory cells capable of eliciting an IFNg response are present in SpA synovial fluid.**

Mean frequency (± SEM) of SpA Th memory cells from SFMC staining positive for co-expression of CD154 and IFNg after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation is shown. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=7-9 for lysates and no stimulation, n=8 for LPS. *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

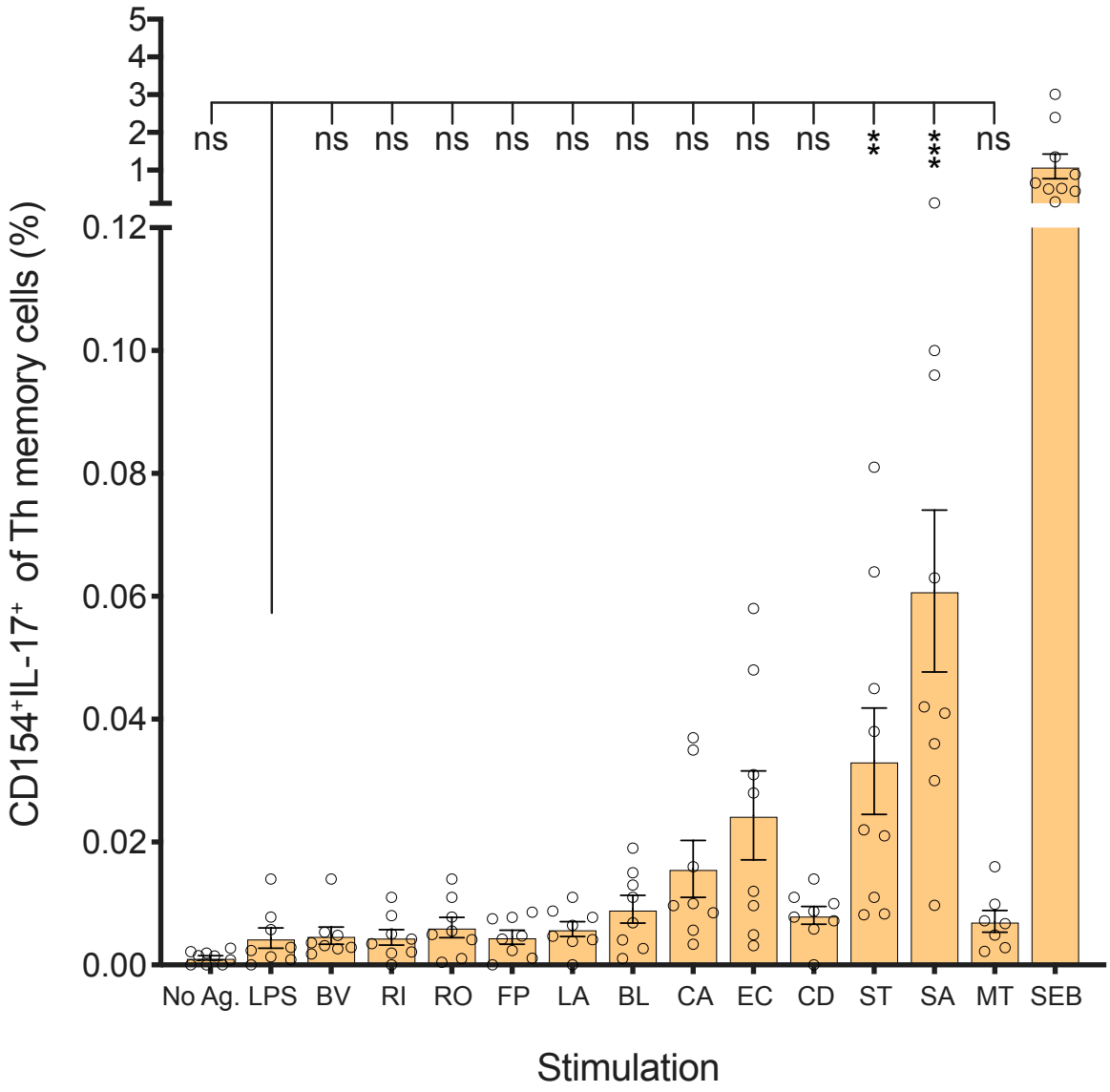


**Figure 3-12 *S. typhimurium* and *S. aureus* reactive Th memory cells capable of eliciting an IL-17 response are present in SpA synovial fluid.**

Mean frequency (± SEM) of SpA Th memory cells from SFMC staining positive for co-expression of CD154 and IL-17 after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation is shown. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=7-9 for lysates and no stimulation, n=8 for LPS. *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

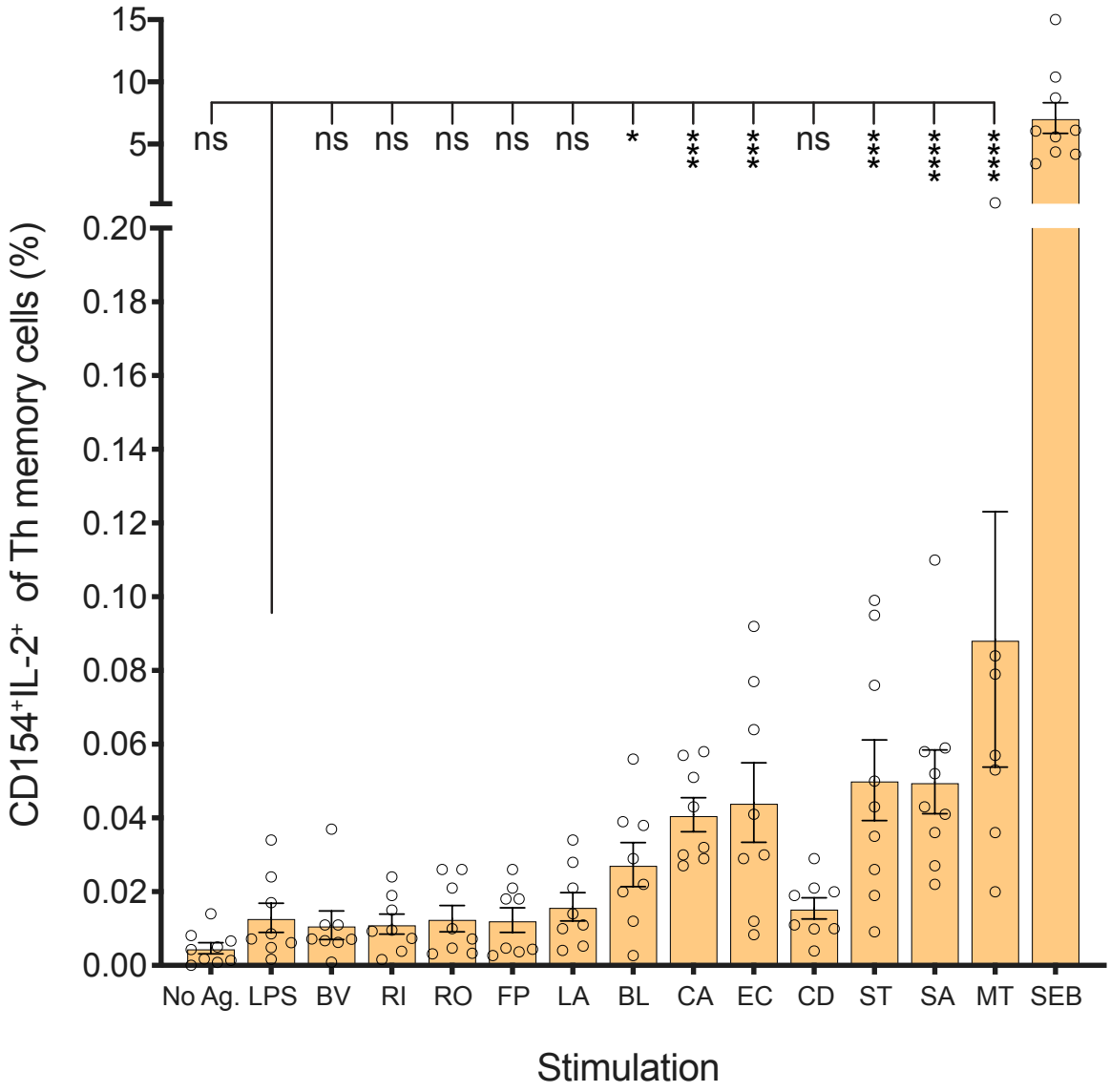


**Figure 3-13 *B. lactis*, *C. albicans*, *E. coli*, *S. typhimurium*, *S. aureus* and *M. tuberculosis* reactive Th memory cells capable of eliciting an IL-2 response are present in SpA synovial fluid.**

Mean frequency (± SEM) of SpA Th memory cells from SFMC staining positive for co-expression of CD154 and IL-2 after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation is shown. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=7-9 for lysates and no stimulation, n=8 for LPS. *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

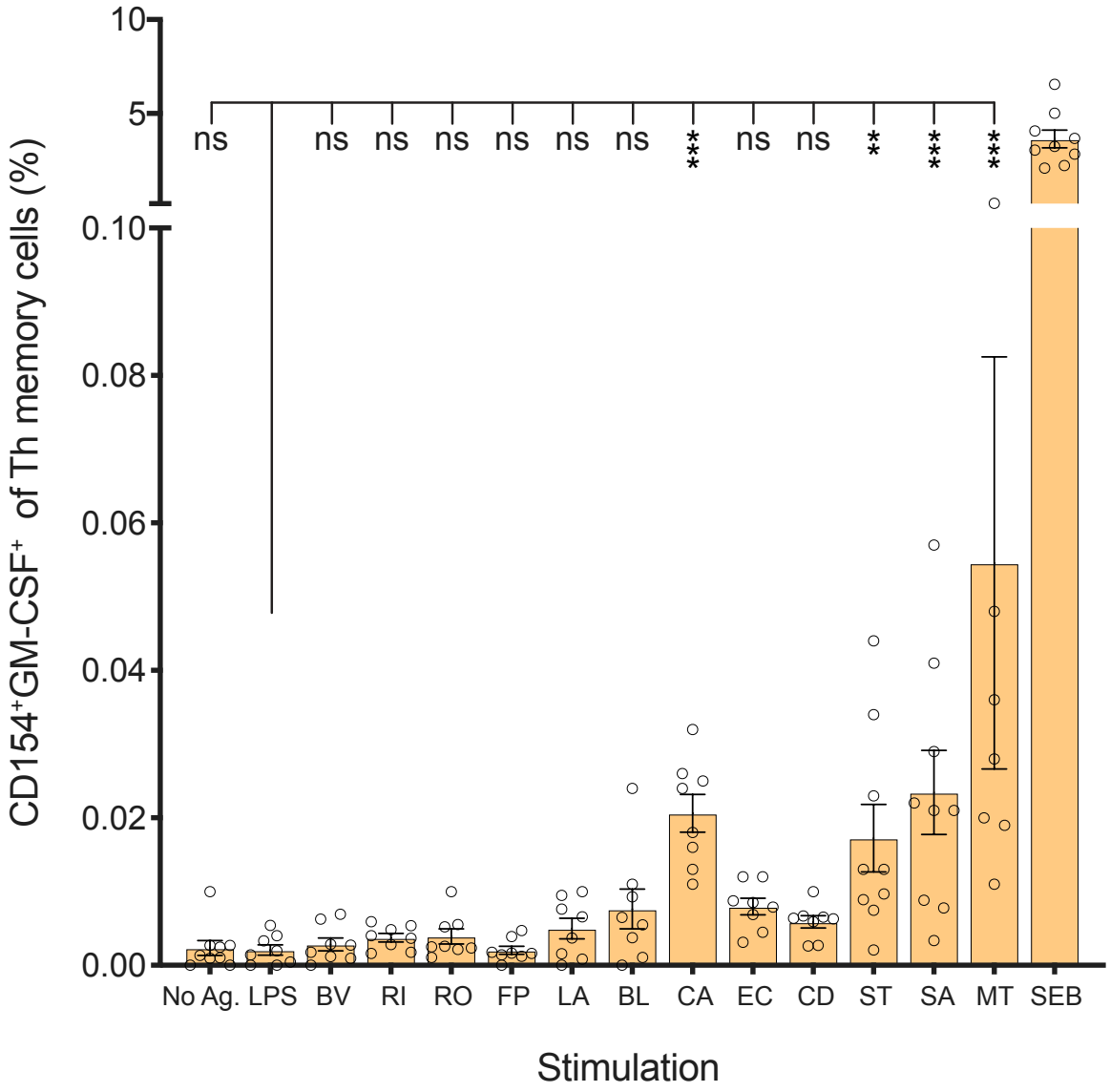


**Figure 3-14 *C. albicans*, *S. typhimurium*, *S. aureus* and *M. tuberculosis* reactive Th memory cells capable of eliciting a GM-CSF response are present in SpA synovial fluid.**

Mean frequency (± SEM) of SpA Th memory cells from SFMC staining positive for co-expression of CD154 and GM-CSF after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation is shown. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=7-9 for lysates and no stimulation, n=8 for LPS. *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

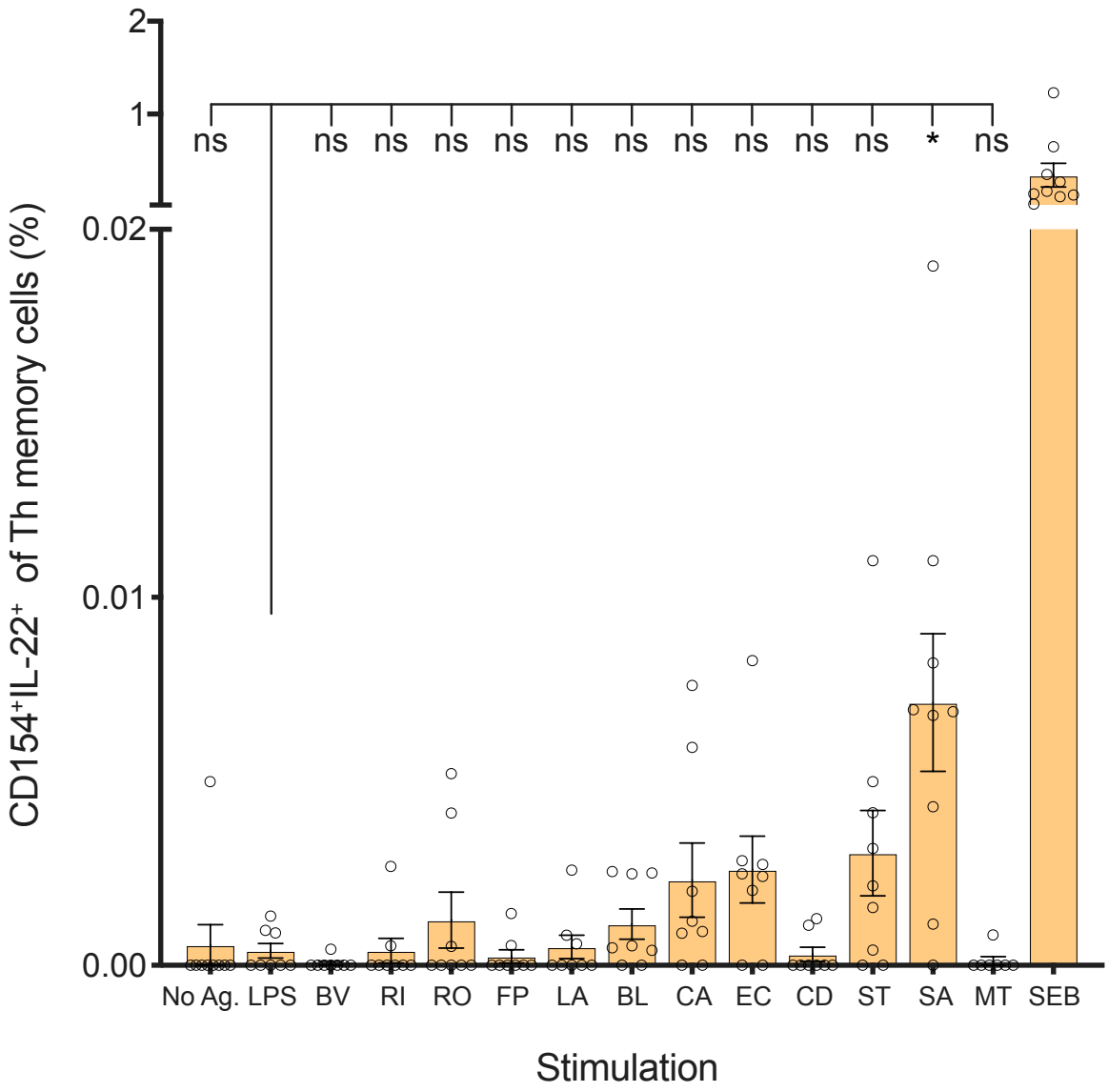


**Figure 3-15 *S. aureus* reactive Th memory cells capable of eliciting a IL-22 response are present in SpA synovial fluid.**

Mean frequency (± SEM) of SpA Th memory cells from SFMC staining positive for co-expression of CD154 and IL-22 after 16 hours of stimulation with microbial lysates, LPS, SEB or no stimulation. Significance of CD154 responses against microbial lysates or culture medium without stimulation relative to control LPS stimulation is shown. Statistics: Kruskal-Wallis test with Dunn's correction for multiple comparisons; n=7-9 for lysates and no stimulation, n=8 for LPS. *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$; ns = not significant. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B.

### 3.3.4 Comparing synovial fluid responses to blood

Cytokine responses in synovial fluid were further compared to responses in peripheral blood, to determine if there was an enrichment of Th memory cells reactive to particular microbes depending on sample type.

When only considering synovial fluid responses, IL-2 was the most common cytokine expressed after stimulation with microbial lysates at levels higher than stimulation with LPS. This included responses towards commensal as well pathogenic microbes (Figure 3-13). Comparing these results to stimulations of PBMC showed an enriched response in synovial fluid towards the same set of microbes (Figure 3-16), while the response to LPS was not significantly enriched in synovial fluid for IL-2. A Wilcoxon matched pairs signed rank test on paired blood and synovial fluid samples supported this finding for all lysates other than *B. lactis* (although the direction of change was consistent).

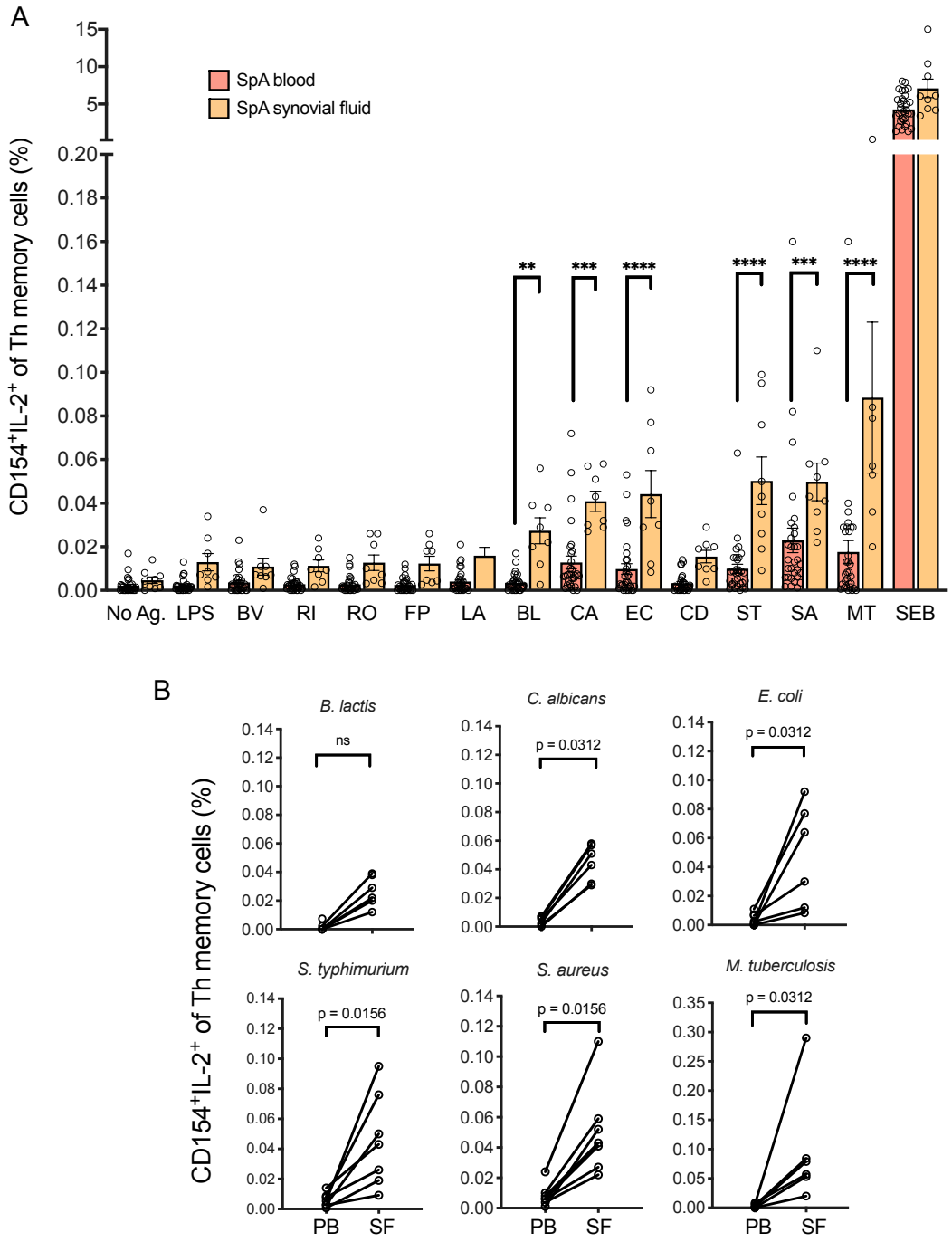


**Figure 3-16 BL, ST, EC, SA, MT and CA reactive IL2$^+$ Th memory cells are enriched in SpA synovial fluid.**

(A) Mean frequencies (± SEM) of SpA microbe-reactive peripheral (red) and synovial fluid (yellow) Th memory cells co-expressing CD154 and IL-2 after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics calculated using 2-way ANOVA and Sidak correction for multiple comparisons; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$. (B) SpA blood and synovial fluid Wilcoxon matched pairs signed rank tests for stimulations found to be significant from A. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B; PB: peripheral blood, SF: synovial fluid.

The IL-17 response in SpA synovial fluid was enriched only for Th memory cells reactive to *S. typhimurium* and *S. aureus*, with findings again supported in paired samples (Figure 3-17). *S. typhimurium* was also the only microbe capable of inducing a statistically significant IFNg response above LPS stimulation when SFMC was stimulated (Figure 3-11). The frequency of CD154+ Th memory cells co-expressing IL-17 and IFNg was therefore assessed (Figure 3-18). This showed that a large proportion of IL-17+ responses were also IFNg+ for *S. typhimurium* stimulations (mean frequency 0.02 and 0.03 for IL-17+IFNg+ and IL-17+ respectively), whereas this was less evident for *S. aureus* (mean frequency 0.02 and 0.06 for IL-17+IFNg+ and IL-17+ respectively). Interestingly, the IL-17+IFNg+ response to stimulation with the non-pathogenic *E. coli* nissle 1917 strain (which has a close genetic relationship to *S. typhimurium*) was also enriched in SpA synovial fluid. When comparing SpA, RA and healthy control CD154+IL-17+IFNg+ responses in blood (Figure 3-19), a significantly reduced response to the same 3 lysates (ST, EC and SA) was seen in both SpA and RA compared to healthy controls. In addition, SpA IL-17+IFNg+ responses to the commensal microbes *R. obeum* and *B. lactis* was also reduced compared to healthy control blood.

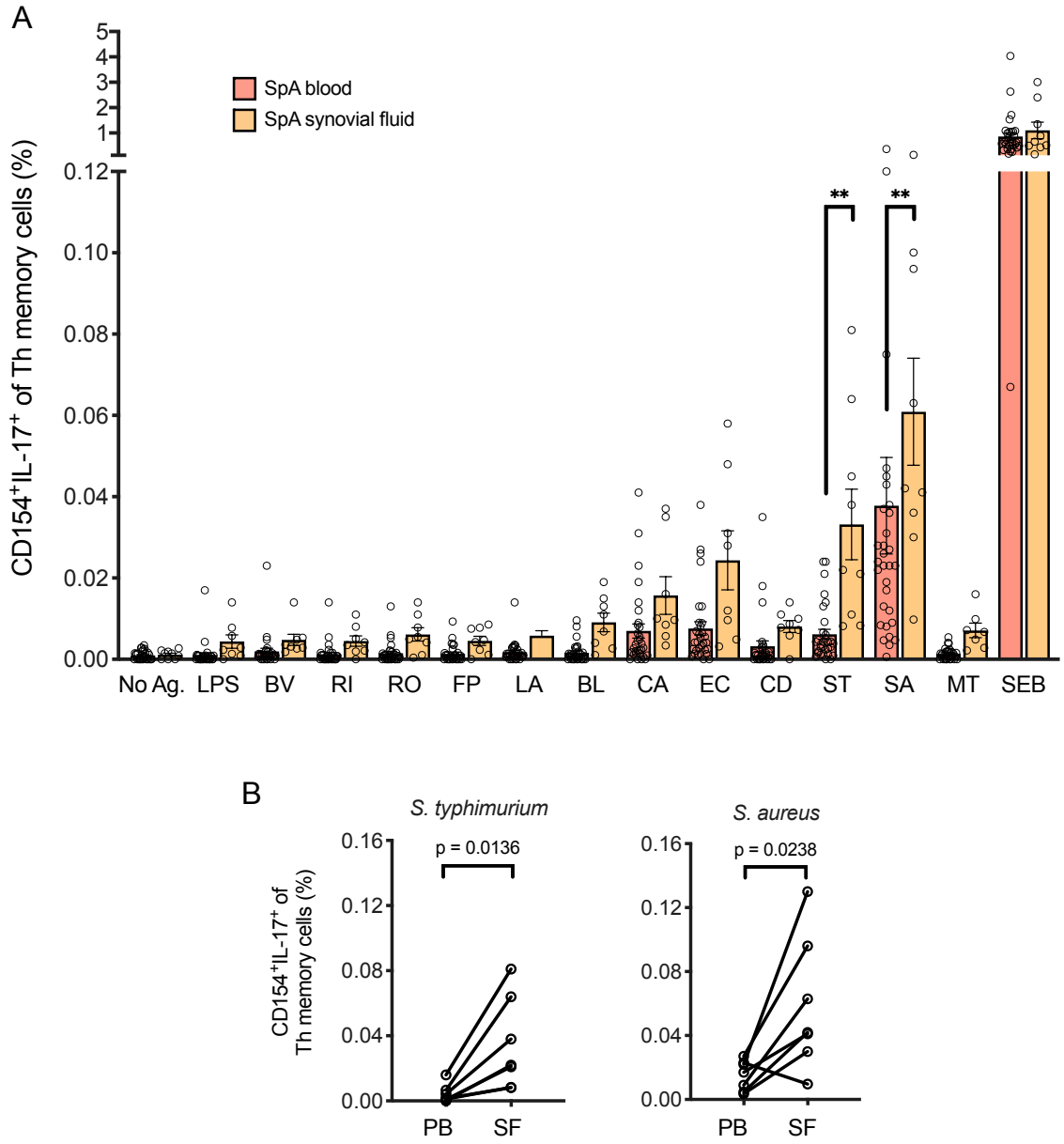


**Figure 3-17 *S. typhimurium* and *S. aureus* reactive IL-17$^+$ Th memory cells are enriched in SpA synovial fluid.**

(A) Mean frequencies (± SEM) of SpA microbe-reactive peripheral (red) and synovial fluid (yellow) Th memory cells co-expressing CD154 and IL-17 after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics calculated using 2-way ANOVA and Sidak correction for multiple comparisons; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$. (B) SpA blood and synovial fluid Wilcoxon matched pairs signed rank tests for stimulations found to be significant from A. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B; PB: peripheral blood, SF: synovial fluid.

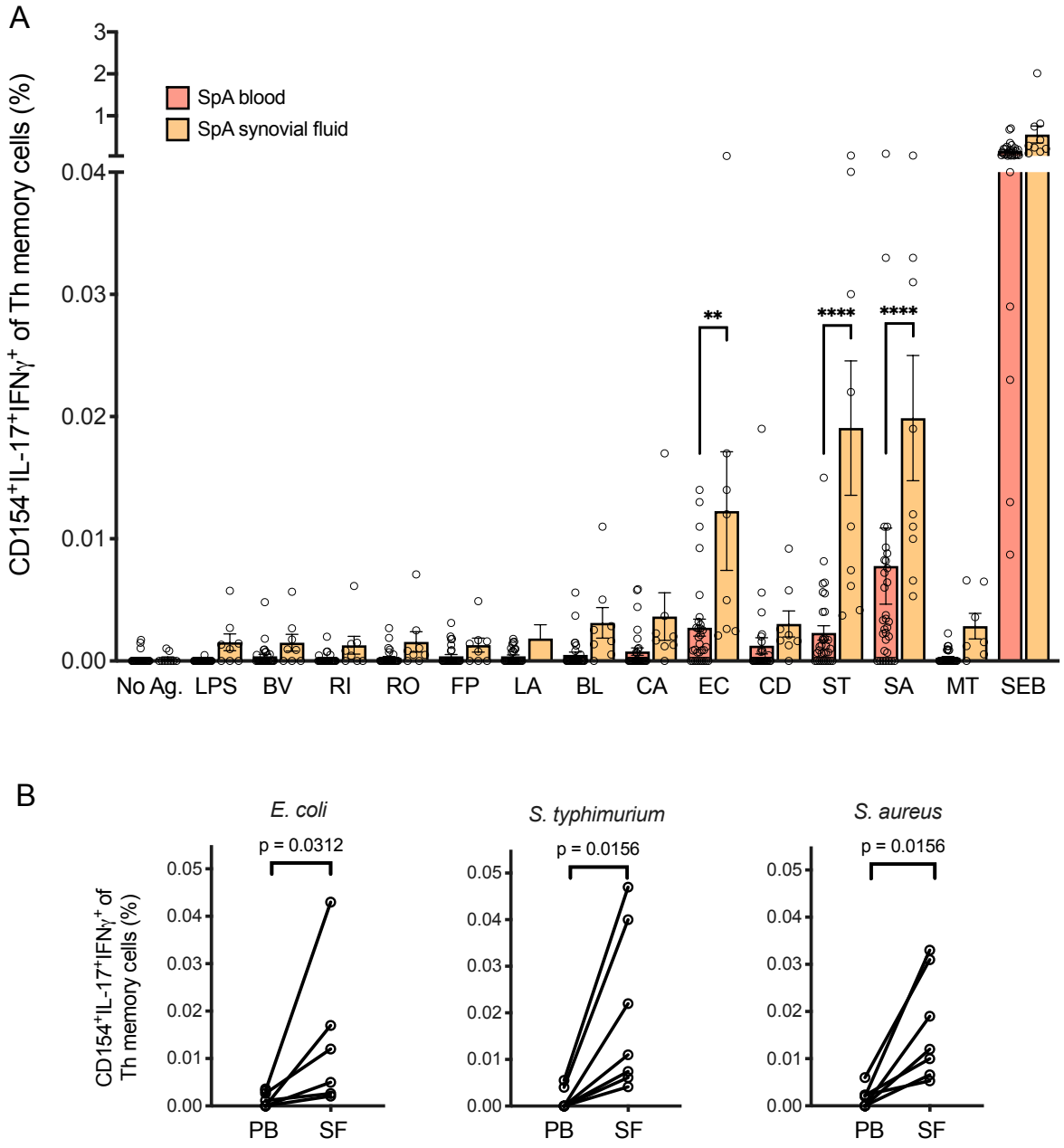


**Figure 3-18 *S. typhimurium*, *S. aureus* and *E. coli* reactive IL-17$^+$IFNγ$^+$ Th memory cells are enriched in SpA synovial fluid.**

(A) Mean frequencies (± SEM) of SpA microbe-reactive peripheral (red) and synovial fluid (yellow) Th memory cells co-expressing CD154, IL17A and IFNγ after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics calculated using 2-way ANOVA and Sidak correction for multiple comparisons; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$. (B) SpA blood and synovial fluid Wilcoxon matched pairs signed rank tests for EC, ST and SA stimulations from A. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B; PB: peripheral blood, SF: synovial fluid.

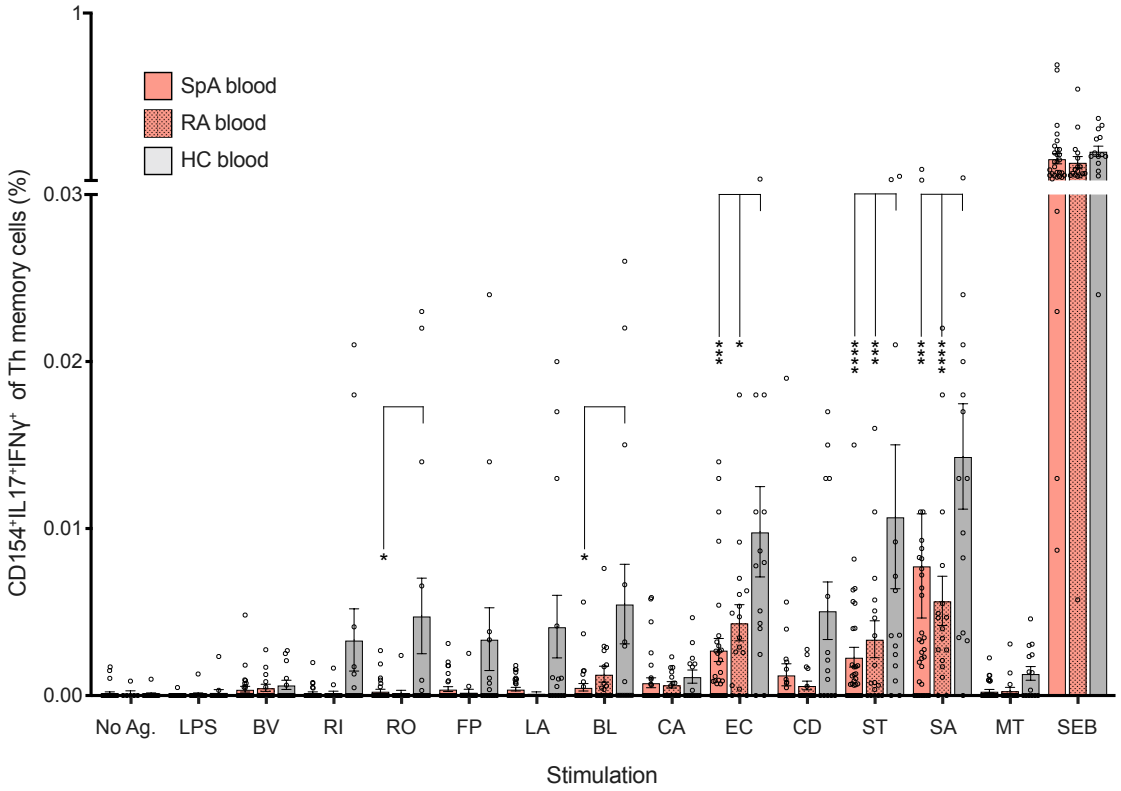


**Figure 3-19 SpA and RA patients have a lower frequency of IL17A+IFNγ+ microbe reactive Th memory cells in peripheral blood compared with healthy controls**

Mean frequencies (± SEM) of SpA (red), RA (red chequered) and healthy control (grey) microbe-reactive peripheral Th memory cells co-expressing CD154, IL17A and IFNγ after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics: calculated using 2-way ANOVA and Sidak correction for multiple comparisons; no significant difference between SpA and RA peripheral blood frequencies was found for any stimulation; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$.

Given that the IL-2 response to stimulation with these microbes was also enriched in synovial fluid (Figure 3-16), the frequency of CD154+ Th memory cells co-expressing IL-17, IFNg and IL-2 was explored. Lysates from all 3 microbes: *S. typhimurium*, *S. aureus* and *E. coli*, were able to elicit an enriched response in synovial fluid (Figure 3-20). The contrast between blood and synovial *S. typhimurium* CD154+ IL-17+IFNg+IL-2+ responses was particularly striking.

A GM-CSF response was only enriched in synovial fluid for CD154+ Th memory cells reactive to *M. tuberculosis*, with one individual showing a particularly high response in synovial fluid. This individual had virtually no GM-CSF response to M. tuberculosis in blood or when SFMC was stimulated with LPS, demonstrating how strongly microbe reactive Th memory cells of a particular phenotype can localise to a particular site of the body (Figure 3-21).

IL-22 expression by Th memory cells stimulated with lysates was generally very low, however a noticeable difference in the enrichment of IL-22 responses towards *S. typhimurium* and *S. aureus* was still observed (Figure 3-22). Although paired responses were not found to be significantly enriched in synovial fluid for *S. typhimurium*, this was likely due to a lack of statistical power in determining a smaller degree of enrichment.

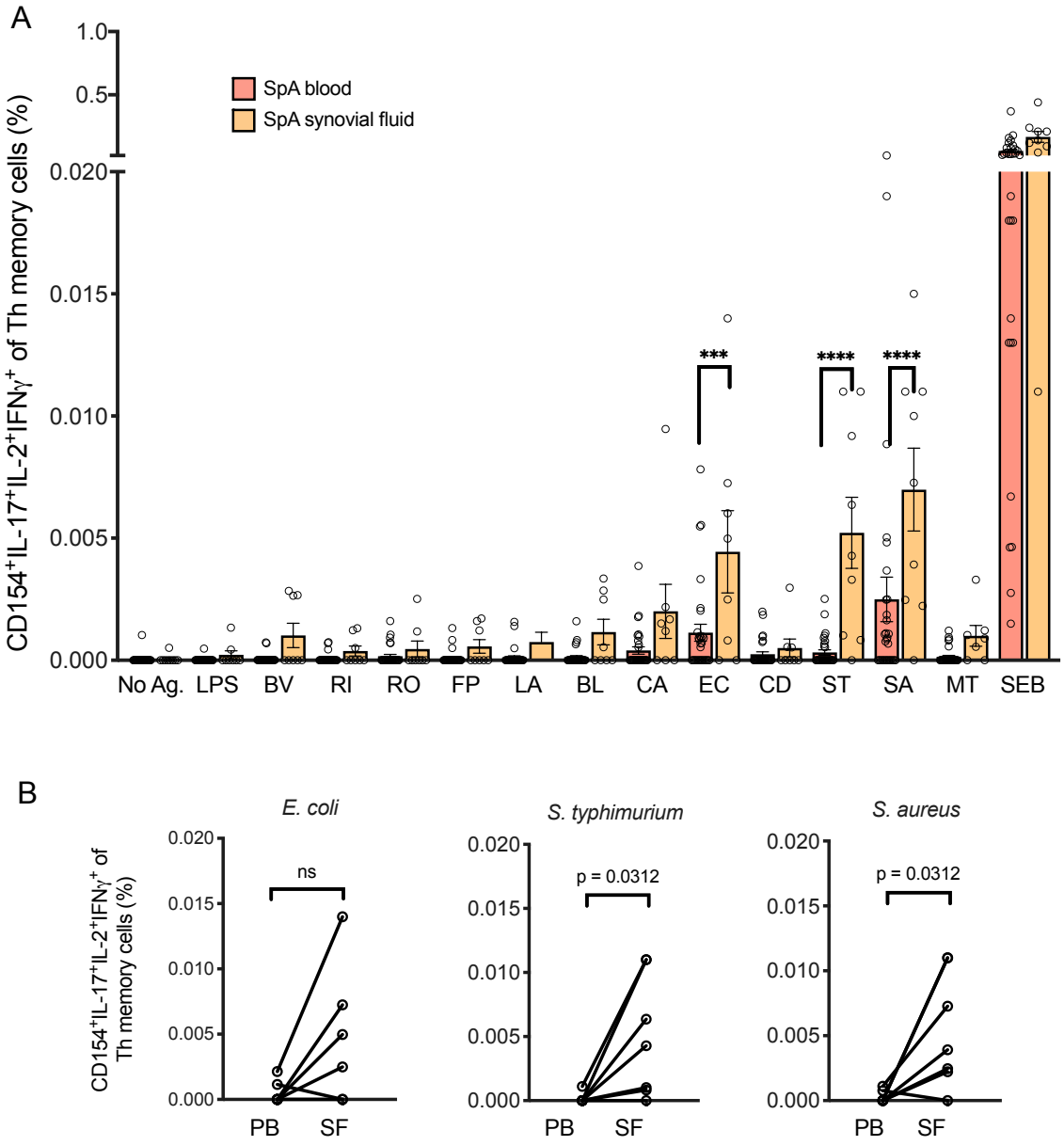


**Figure 3-20 *S. typhimurium* and *S. aureus* reactive IL-17$^+$IL-2$^+$IFNγ$^+$ Th memory cells are enriched in SpA synovial fluid.**

(A) Mean frequencies (± SEM) of SpA microbe-reactive peripheral (red) and synovial fluid (yellow) Th memory cells co-expressing CD154, IL17A, IFNγ and IL-2 after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics calculated using 2-way ANOVA and Sidak correction for multiple comparisons; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$. (B) SpA blood and synovial fluid Wilcoxon matched pairs signed rank tests for EC, ST and SA stimulations from A. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B; PB: peripheral blood, SF: synovial fluid.

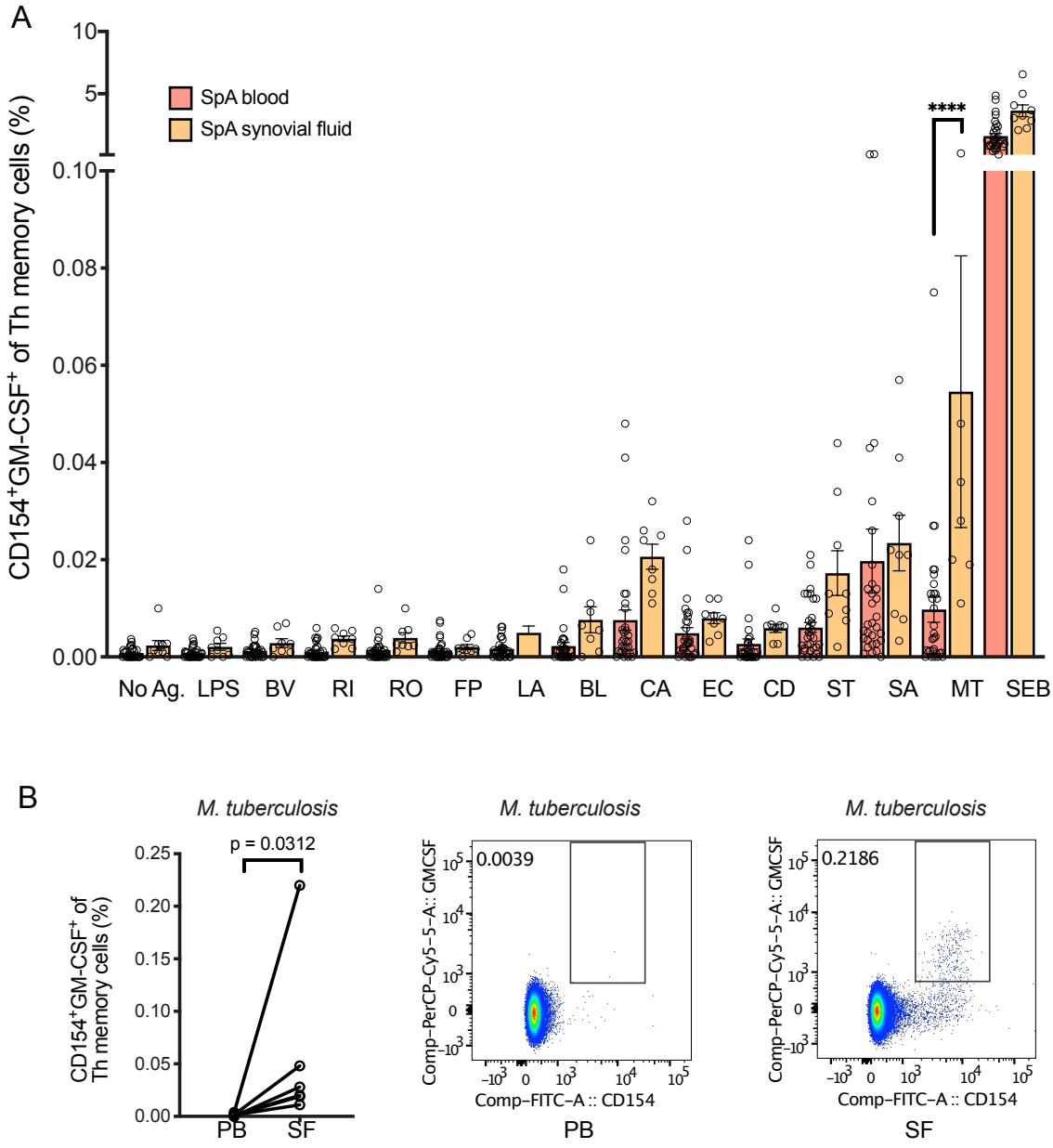


**Figure 3-21 *M. tuberculosis* reactive GM-CSF$^+$ Th memory cells are enriched in SpA synovial fluid.**

(A) Mean frequencies (± SEM) of SpA microbe-reactive peripheral (red) and synovial fluid (yellow) Th memory cells co-expressing CD154 and GM-CSF after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics calculated using 2-way ANOVA and Sidak's correction for multiple comparisons; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$. (B) SpA blood and synovial fluid Wilcoxon matched pairs signed rank tests for MT stimulation from A, and FACS plots for the patient sample with highest enrichment of CD154+CM-CSF+ Th memory cells in synovial fluid. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B; PB: peripheral blood, SF: synovial fluid.

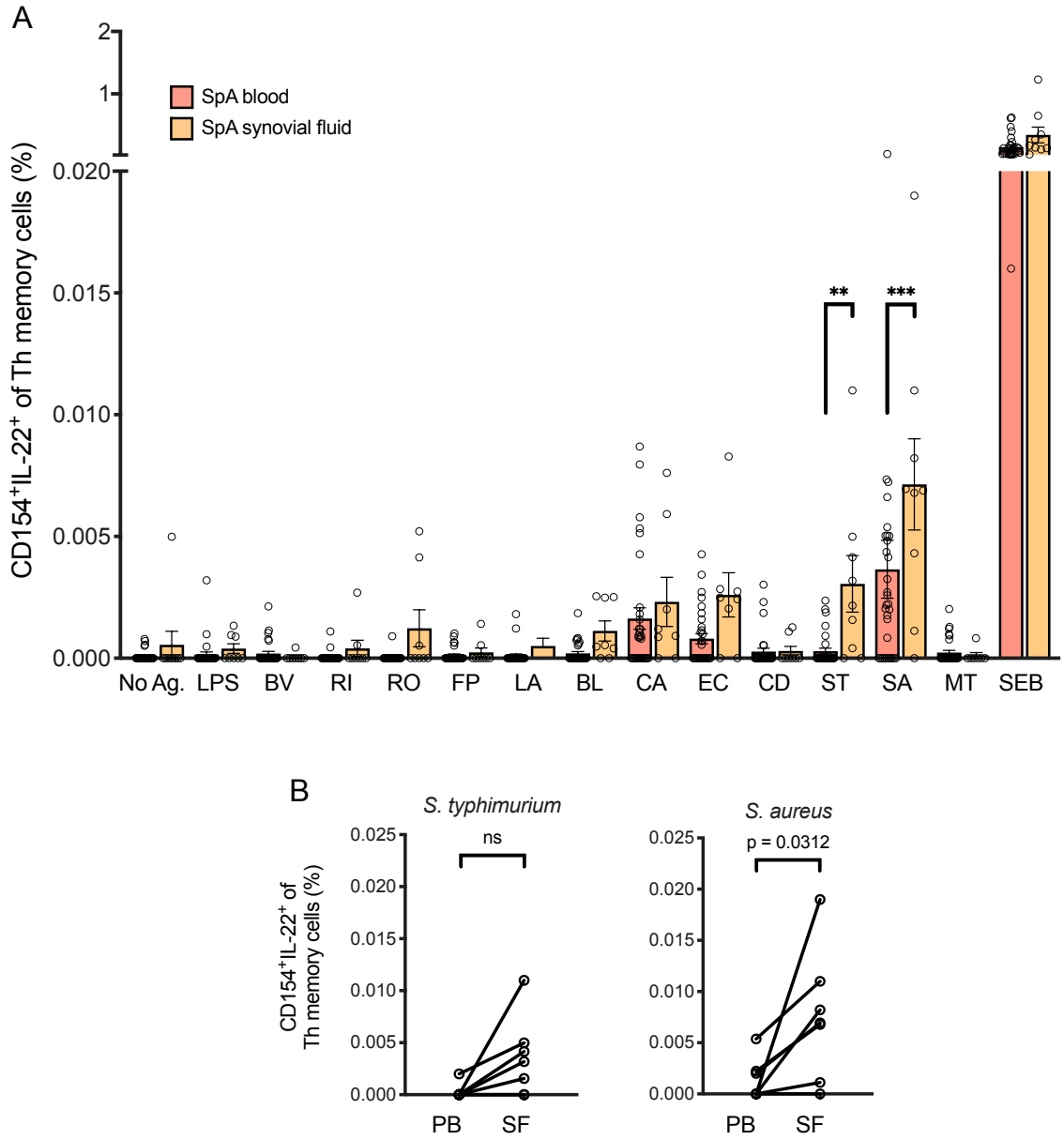


**Figure 3-22 *S. aureus and S. typhimurium* reactive IL-22$^+$ Th memory cells are enriched in SpA synovial fluid**

(A) Mean frequencies (± SEM) of SpA microbe-reactive peripheral (red) and synovial fluid (yellow) Th memory cells co-expressing CD154 and IL-22 after 16 hours of stimulation with microbial lysates, LPS, SEB or no antigen. Each circle represents an individual patient. Statistics calculated using 2-way ANOVA and Sidak correction for multiple comparisons; *$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; ****$P \leq .0001$. (B) SpA blood and synovial fluid Wilcoxon matched pairs signed rank tests for ST and SA stimulations from A. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB: Staphylococcal enterotoxin B; PB: peripheral blood, SF: synovial fluid

### 3.3.5 Strongly enriched synovial fluid cytokine responses towards *C. trachomatis* in a new onset reactive arthritis patient.

A rare opportunity presented itself to analyse immune responses in both the peripheral blood and synovial fluid of a newly diagnosed reactive arthritis patient testing positive for *C. trachomatis* infection (Figure 3-23). A greater than 20 fold increase in the frequency of Th memory cells co-expressing CD154 in combination with either IL-2, IL-17 or GM-CSF was found in the synovial fluid of this patient. Interestingly this patient also had a higher frequency of *S. typhimurium* and *S. aureus* reactive Th memory cells producing the same cytokines in synovial fluid. A very low frequency of Th memory cells were found to be CD154+ in synovial fluid without stimulation for this patient, the lowest observed for any synovial fluid sample (Figure 3-9). Unfortunately there was insufficient SFMC available from this patient for more stimulations than those seen in Figure 3-23A however the full panel of lysates was used for stimulation of PBMC (Figure 3-23B).

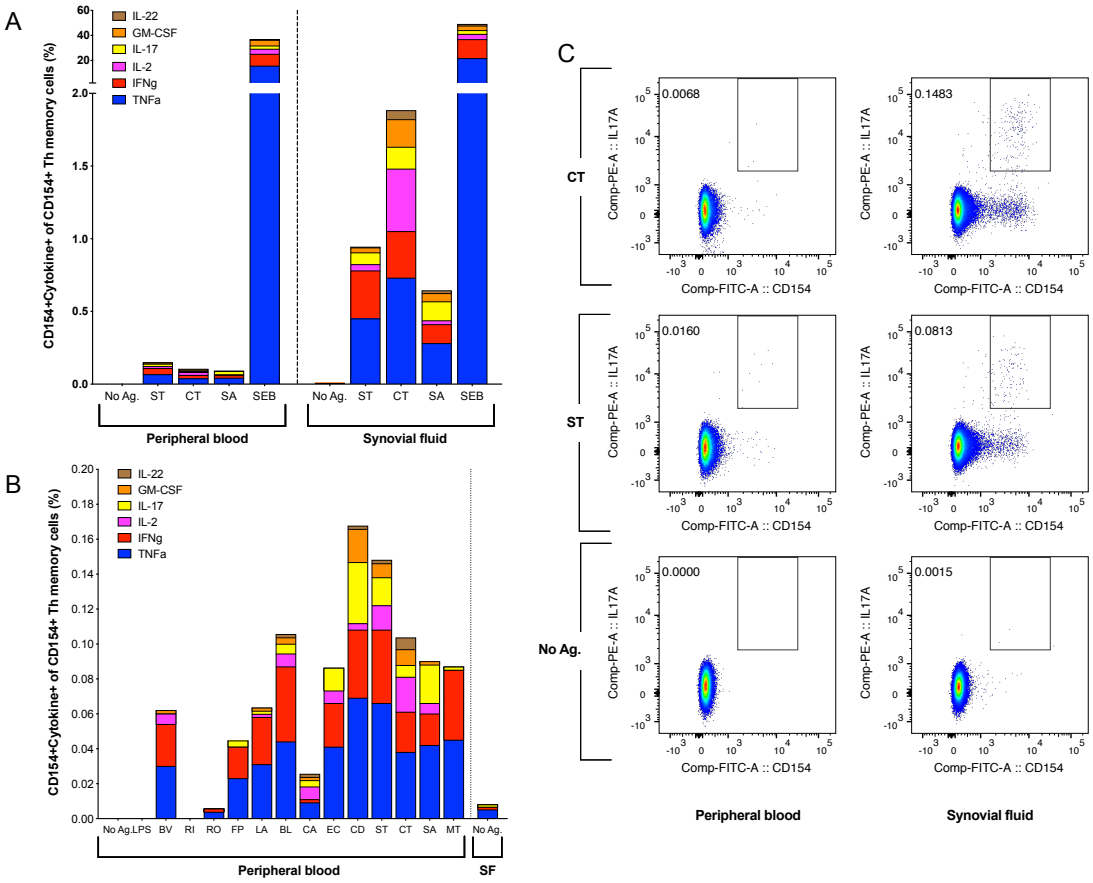


**Figure 3-23 A new onset reactive arthritis patient shows strongly enriched synovial fluid responses towards *C. trachomatis* in synovial fluid.**

Frequency of peripheral blood (left) and synovial fluid (right) Th memory cells co-expressing CD154 with either TNFa, IFNg, IL-2, IL-17, GM-CSF or IL-22 after 16 hours of stimulation with microbial lysates, SEB or no antigen. Responses to each cytokine are stacked for each stimulation. Synovial fluid fold change increases in the frequency of Th memory cells reactive to C. trachomatis for each cytokine: TNFa:19.2; IFNg:13.9; IL-2:21.5; IL-17:22.2; GM-CSF:21; IL-22:9.2. (A) Focus on *C. trachomatis*, *S. aureus* and *S. typhimurium* responses (there was insufficient SFMC for stimulation with complete panel of lysates). (B) Zoomed in view of microbe-reactive peripheral blood responses from A, comparing full panel of microbial lysates assayed as well as unstimulated peripheral blood and synovial fluid responses. (C) Flow cytometry dot plots showing percentage of CD154+IL17+ cells from CD3+CD4+CD45RA- gated populations after stimulation with CT, ST and No Antigen from A. Abbreviations: No Ag.: No antigen; LPS: Lipopolysaccharide; BV: *Bacteroides vulgatus*; RI: *Roseburia intestinalis*; RO:*Ruminococcus obeum*; FP:*Faecalibacterium prausnitzii*; LA: *Lactobacillus acidophilus*; BL:*Bifidobacterium lactis*; CA:*Candida albicans*; EC:*Escherichia coli*; CD:*Clostridium difficile*; ST:*Salmonella typhimurium*; CT:*Chlamydia trachomatis*; SA:*Staphylococcus aureus*; MT:*Mycobacterium tuberculosis*; SEB:Staphylococcal enterotoxin B; SF:synovial fluid.

## 3.4 Discussion

The first part of this chapter described the utility of employing CD154 expression in combination with intracellular cytokine staining to identify functional immune responses of microbe-reactive Th memory cells. CD154 was co-expressed by the majority of TNFa, IFNg, IL-2, IL-17, GM-CSF and IL-22 expressing Th memory cells in blood and synovial fluid after stimulation with SEB or microbial lysates. This provided a way to quantify cytokine responses as a proportion of total Th memory cells expressing CD154 (microbe-reactive cells), or to directly compare the frequency of cells co-expressing CD154 and cytokine(s) between patients and controls or sample types within SpA. LPS stimulation was used to control for potential indirect bystander activation of CD4 memory T cells. However, increased expression of CD154 by Th memory cells was observed when stimulating SFMC with LPS compared to no stimulation and must be taken into account when interpreting results of the CD154 assay. No such difference was found for blood responses to LPS, suggesting there may be an altered cellular or chemical environment in synovial fluid leading to the observed effects of LPS stimulation (Figure 3-9). There was also a significantly higher frequency of CD154+ Th memory cells present in synovial fluid in comparison to blood in the absence of any stimulation (Figure 3-9) which likely explained part of the increased response to LPS. Although varying between patients, a large proportion of these unstimulated CD154+ cells in synovial fluid were also expressing TNFa, and to a lesser extent IFNg (mean frequency of Th memory cells CD154+: 0.106%; CD154+TNFa+: 0.072%; CD154+IFNg+: 0.030%. From Figure 3-10A and B, and Figure 3-11 respectively). This raises the interesting possibility of an existing population of Th memory cells reacting to some antigen already present in patient synovial fluid, and would be worthy of further investigation.

As LPS was being used as a negative control stimulation, it was difficult to distinguish Th memory cell responses towards microbial lysates from existing or LPS induced responses in synovial fluid by the expression of CD154 alone. No CD154+ response above LPS stimulation could indicate that no reactivity to a particular microbial lysate was present in synovial fluid. On the other hand it may indicate that Th memory cells reactive to peptides within the lysate being assayed were already present in the unstimulated/LPS stimulated populations. The frequency of Th memory cells co-expressing CD154 and IL-2, IL-17, GM-CSF or IL-22 was however much lower in unstimulated or LPS stimulated samples relative to microbial stimulations ( Figure 3-12 to Figure 3-15 ). Thus robust comparisons utilising the CD154 assay for synovial fluid samples were possible when restricting the analysis to cells co-expressing these cytokines with CD154.

Perhaps the most striking observation when comparing immune responses in blood was the reduced frequency of circulating microbe reactive Th cells in both SpA and RA compared to healthy control blood (Figure 3-7). This was especially apparent in the CD154+ response towards *M. tuberculosis* and *S. typhimurium*, and may be indicative of a corresponding impaired immune response against those microbes. However, Th memory CD154+ responses to these microbes were still detected in SpA and RA blood (Figure 3-6), and when considering the percentage of these cells also expressing TNFa, IFNg, IL-17 and GM-CSF, the type of cytokine response in SpA towards *M tuberculosis* was not different to healthy control samples (Figure 3-8). The only significantly different cytokine response in SpA blood relative to both RA and healthy control blood was a lower frequency of IL-2+ Th memory cells reactive to *S. typhimurium* and *M. tuberculosis*.

An alternative explanation for the lower frequency of microbe reactive Th memory cells in blood would be the preferential trafficking of these cells to sites of inflammation, thereby depleting numbers in peripheral blood. Indeed, when comparing SpA patient peripheral blood and synovial fluid samples, there was clear enrichment in the latter of Th memory cells reactive to specific lysates and producing specific cytokine responses. Two distinct types of enrichment in synovial fluid emerged: an enrichment of IL-17+IFNg+ (type-17.1) responses against *E. coli*, *S. typhimurium* and *S. aureus*, and an enriched GM-CSF associated response towards *M. tuberculosis*. These microbes were also associated with an enriched IL-2 response in synovial fluid. In contrast, *C. albicans* and *B. lactis* reactive Th memory cells producing IL-2 were also enriched in synovial fluid, but responses to these microbes did not associate with a differentiated type-17 or type-1 cell fate.

The IFNg response to stimulation with *S. typhimurium* was substantial, similar to that seen in response to *M. tuberculosis* (Figure 3-11) which is well characterised for its ability to induce strong type-1 IFNg responses in CD4 T cells[11]. Almost two thirds of CD154+IL-17+ Th memory cells reactive to *S. typhimurium* also produced IFNg, and such type "17.1" cells have previously been associated with a pathogenic phenotype in autoimmunity[12]. The strongly enriched IL17+IFNg+ responses discovered in synovial fluid against *S. typhimurium* are particularly interesting given *Salmonella's* association with reactive arthritis.

Findings from the isolated case of a patient diagnosed with *Chlamydia* induced reactive arthritis supported the notion that Th memory cells reacting to infection at remote barrier sites can localise to the joint in SpA. Of particular interest was the overlap in both frequency and cytokine profiles of *S. typhimurium* and *C. trachomatis* reactive Th memory cells in synovial fluid, and the comparatively low frequency of these cells

found in peripheral blood (Figure 3-23). It is therefore possible that cross-reactivity towards peptides in both *C. trachomatis* and *S. typhimurium* lysates exists. As both microbes are associated with the onset of reactive arthritis, which itself can lead to chronic inflammation resembling other forms of SpA, it is tempting to speculate that the induction of a type-17.1 immune response is a unifying feature of ReA causitive microbes. Patients diagnosed with other forms of SpA do not present with clinically diagnosed infections. However, synovial responses to the probiotic *E. coli* nissle 1917 in this study demonstrated the capacity for lysate from non-pathogenic microbes to induce a type-17.1 response in SpA patients. It is therefore conceivable that a Th17 response to gut microbes intended to maintain homeostasis in the gut environment is exacerbated or otherwise perturbed given the genetic background of SpA patients, resulting in increased local gut inflammation and the migration of this response via Th cells to other sites of inflammation (such as the joints) in SpA patients. The next chapter aims to further characterise the type-17.1 immune response towards *S. typhimurium* using single cell RNA sequencing.

# I. References


1. Gaffen, S.L., Jain, R., Garg, A.V., and Cua, D.J. (2014). The IL-23-IL-17 immune axis: from mechanisms to therapeutic testing. Nat Rev Immunol *14*, 585–600.

2. Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., van der Heijde, D., McInnes, I., van Laar, J.M., Landewé, R., Wordsworth, P., *et al.* (2013). Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. The Lancet *382*, 1705–1713.

3. Breban, M., Fernández-Sueiro, J.L., Richardson, J.A., Hadavand, R.R., Maika, S.D., Hammer, R.E., and Taurog, J.D. (1996). T cells, but not thymic exposure to HLA-B27, are required for the inflammatory disease of HLA-B27 transgenic rats. Journal of Immunology *156*, 794–803.

4. Littman, D.R., and Pamer, E.G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host Microbe *10*, 311–323.

5. Belkaid, Y., Bouladoux, N., and Hand, T.W. (2013). Effector and Memory T cell Responses to Commensal Bacteria. Trends Immunol *34*, 299–306.

6. Hegazy, A.N., West, N.R., Stubbington, M.J.T., Wendt, E., Suijker, K.I.M., Datsi, A., This, S., Danne, C., Campion, S., Duncan, S.H., *et al.* (2017). Circulating and Tissue-Resident CD4+ T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. Gastroenterology *153*, 1320-1337.e16.

7. Frentsch, M., Arbach, O., Kirchhoff, D., Moewes, B., Worm, M., Rothe, M., Scheffold, A., and Thiel, A. (2005). Direct access to $CD4^+$ T cells specific for defined antigens according to CD154 expression. Nature Medicine *11*, 1118.

8. Bacher, P., Schink, C., Teutschbein, J., Kniemeyer, O., Assenmacher, M., Brakhage, A.A., and Scheffold, A. (2013). Antigen-Reactive T Cell Enrichment for Direct, High-Resolution Analysis of the Human Naive and Memory Th Cell Repertoire. The Journal of Immunology *190*, 3967–3976.

9. Puel, A., Cypowyj, S., Bustamante, J., Wright, J.F., Liu, L., Lim, H.K., Migaud, M., Israel, L., Chrabieh, M., Audry, M., *et al.* (2011). Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity*. Science *332*, 65–68.

10. Chattopadhyay, P.K., Yu, J., and Roederer, M. (2005). A live-cell assay to detect antigen-specific CD4+ T cells with diverse cytokine profiles. Nature Medicine *11*, 1113–1117.

11. Nikitina, I.Y., Panteleev, A.V., Kosmiadi, G.A., Serdyuk, Y.V., Nenasheva, T.A., Nikolaev, A.A., Gorelova, L.A., Radaeva, T.V., Kiseleva, Y.Y., Bozhenko, V.K., *et al.* (2018). Th1, Th17, and Th1Th17 Lymphocytes during Tuberculosis: Th1 Lymphocytes Predominate and Appear as Low-Differentiated CXCR3+CCR6+

Cells in the Blood and Highly Differentiated CXCR3+/−CCR6− Cells in the Lungs. The Journal of Immunology *200*, 2090–2103.

12. Boniface, K., Blumenschein, W.M., Brovont-Porth, K., McGeachy, M.J., Basham, B., Desai, B., Pierce, R., McClanahan, T.K., Sadekova, S., and Malefyt, R. de W. (2010). Human Th17 Cells Comprise Heterogeneous Subsets Including IFN-γ–Producing Cells with Distinct Properties from the Th1 Lineage. The Journal of Immunology *185*, 679–687.

# Chapter 4: Characterising the CD4 T cell immune response towards Salmonella in SpA synovial fluid using single-cell RNA sequencing

## 4.1 Introduction

In the previous chapter, S*almonella* lysate was identified as capable of inducing a type 17 response in SpA Th memory cells, with an enriched frequency of *Salmonella*-reactive memory Th17 cells found in the synovial fluid of SpA patients. This chapter expands on those findings by using single-cell RNA sequencing (scRNA-seq) to determine the gene expression profile and clonality of CD154+ *Salmonella*-reactive Th memory cells from the synovial fluid of a psoriatic arthritis patient. Results are further contrasted with the gene expression and clonality of unstimulated CD4 and CD8 memory T cells from the same patient sample. "CD40LG" is more commonly used when referring to the gene encoding CD154, and is therefore also used in this thesis when describing gene expression of mRNA transcripts.

The technique used to isolate *Salmonella*-reactive Th memory cells for single-cell sequencing is first discussed, as it relies on surface staining of live cells rather than intracellular staining of fixed and permeabalized cells used in the previous chapter. Although recent developments allow for single-cell quantification of RNA transcripts in fixed cells[1] and for quantification of intracellular proteins alongside RNA transcripts in permeabalized cells[2], such techniques are still largely untested. The vast majority of published scRNA-seq data relates to sequencing of live cells[1].

The percentage of *Salmonella*-reactive Th memory cells containing RNA transcripts for the cytokines assayed in chapter 3 is then compared with chapter 3 results, and with the

gene expression of these cytokines in unstimulated memory T cell populations originating from the same sample. The gene expression profile of unstimulated *ex vivo* memory T cell populations is addressed in more depth in chapter 5.

Finally, differential gene expression (DGE) is used to explore populations identified by the unsupervised clustering of *Salmonella*-reactive Th memory cells sequenced to a depth of approximately 90 thousand reads per cell. Results from the gene expression profile of these populations is presented, and pseudo-time analysis is used to identify lineage trajectories between populations. To further interrogate the IL17A response specifically, cells expressing IL17A gene transcripts are subsetted and unsupervised clustering of these cells is used to reveal 3 distinct Type 17 sub-populations. One of the populations identified is distinguished by the expression of both IL17A and IFNG gene transcripts. IL17A+IFNG+ *Salmonella*-reactive Th memory cells were found to be highly enriched in the synovial fluid of SpA patients in chapter 3. Results from this chapter therefore provide a detailed gene expression profile characterising these cells, and their relationship to the total Th memory pool of cells upregulating CD154 in response to stimulation with *Salmonella* lysate.

## 4.2 Aim

To use single cell RNA sequencing as a hypothesis-free approach to characterise a PsA patient's synovial CD4 T cell response to *Salmonella* stimulation, and to contrast this with the ex vivo CD4 and CD8 T cell phenotypes from the same patient.

## 4.3 Results

### 4.3.1 The frequency of viable *Salmonella*-reactive Th memory cells which can be detected by both surface and intracellular staining of CD154 is similar.

In the previous chapter, intracellular staining was used to identify upregulation of CD154 and cytokines in Th memory cells reactive towards microbial lysates. This required permeabilization of cells and the addition of Brefeldin A to prevent transport of CD154 to the cell membrane. However, a different technique to stain CD154 was required to preserve the viability of cells, and therefore be compatible with sorting of live CD154+ cells by FACS. CD154 expressed on the cell surface of Th cells is internalised and degraded after coming into contact with its binding partner CD40, an interaction which can be prevented by incorporation of a CD40 blocking antibody[3]. This technique was therefore employed to separate live *Salmonella*-reactive Th memory cells from SFMC for downstream scRNA-seq. SFMC were first stimulated overnight with *S. typhimurium* lysate in the presence of anti-CD40 antibody but without Brefeldin A, then surface stained with antibodies against CD154, CD3, CD45RA and CD4 adding DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride) for viability staining 15 minutes before sorting with FACS. The gating strategy used to separate CD154+ Th memory cells from the SFMC of patient PSA1607 is shown in Figure 4-1, where approximately 0.2% of CD4 T cells were identified as CD154+. This closely matched previous ICS staining of *Salmonella* stimulated SFMC from the same patient, where CD154+ cells accounted for 0.16% of CD4 T cells (data not shown).

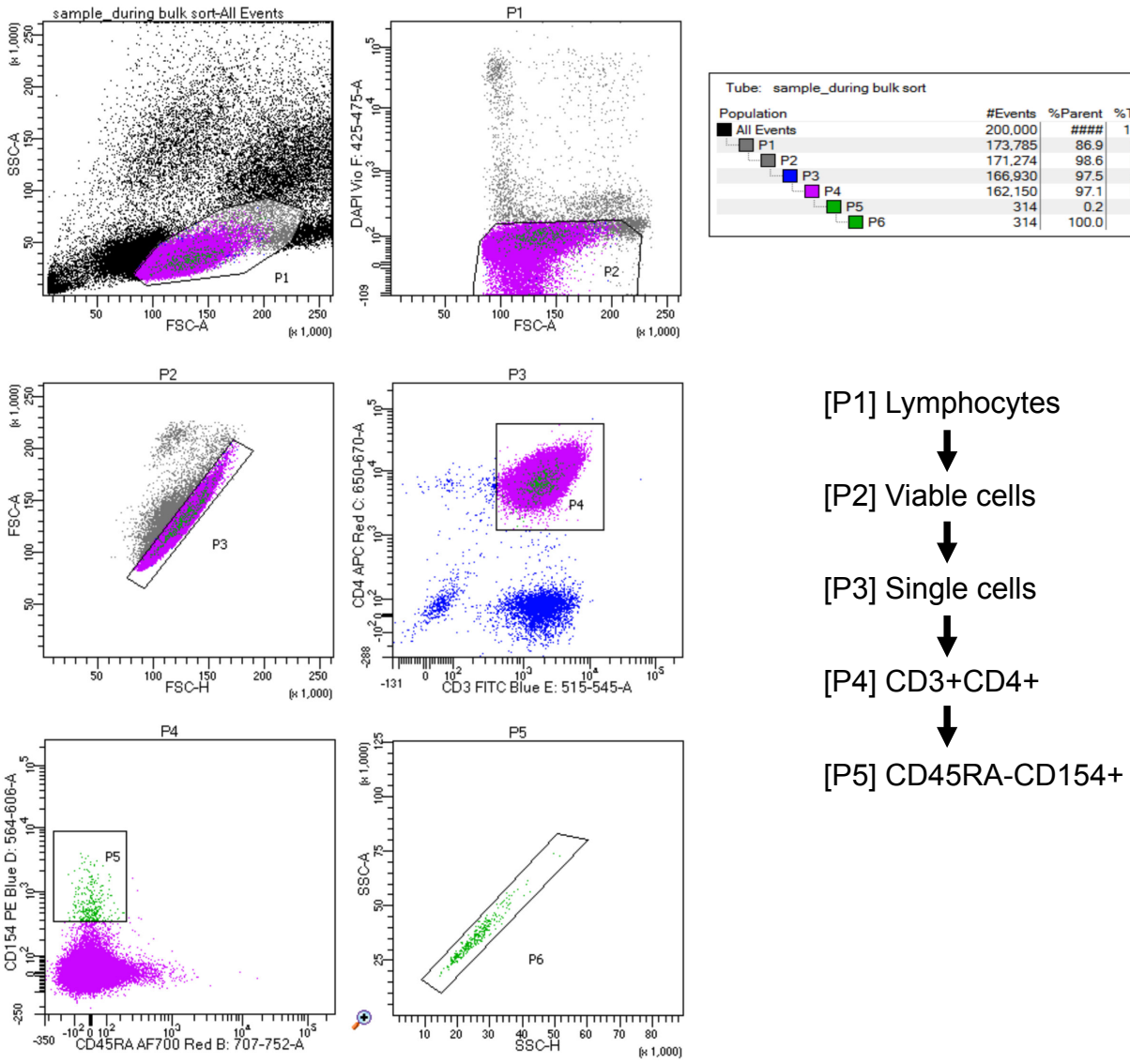


**Figure 4-1 CD154 surface expression was identified in 0.2% of live CD4+ T cells sorted by FACS from *Salmonella* stimulated SFMC.**

Dot plots depicting the FACS gating strategy used to sort live CD154+ Th memory cells from SFMC for downstream 10x scRNA-seq. SFMC were first separated from the synovial fluid of patient PSA1607 and stimulated overnight with *S. Typhimurium* lysate in the presence of a CD40 blocking antibody. SFMC were then surface stained for 20 minutes with CD3-FITC, CD4-APC, CD45RA-AF700 and CD154-PE. DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride) was additionally added 15 minutes prior to sorting to stain for viability.

### 4.3.2 Gene expression matrices captured by scRNA-seq are representative of the cell type, cell frequency and culturing conditions of cells entering the 10x scRNA-seq pipeline

Approximately 11000 live *Salmonella* stimulated CD154+ Th memory cells sorted from the SFMC of patient PSA1607 (Figure 4-1) were loaded[1] onto one well (channel) of a Chromium chip to enter the 10x 5 prime single-cell gene expression and V(D)J sequencing pipeline. This yielded gene expression data for 4098 cells sequenced to a read depth of 90070 reads per cell (Figure 4-2), and productive TCR clonality for 3640 cells. For comparison, and to characterise *ex vivo* memory T cell gene expression and clonality for the same patient, scRNA-seq was also carried out on unstimulated CD4+ and CD8+ memory T cells sorted from both PBMC and SFMC (Figure 4-3). To reduce costs, sorted CD4 and CD8 T cells from unstimulated cells were combined in a 1:1 ratio for loading onto a single 10x Chromium channel, with T cells originating from PBMC or SFMC loaded onto separate channels. For the synovial CD4/CD8 libraries this yielded gene expression for 14,669 cells, sequenced to a read depth of 37521 reads per cell, and productive TCR clonality from 15220 cells. Similar output was obtained from blood libraries, with gene expression data for 15828 cells at 36062 mean reads per cell and productive TCR clonality from 15287 cells.

[1] Cells were separated, stimulated and sorted by myself and brought to the WCHG 10x facility on ice, where they were further processed and loaded onto the 10x Chromium device according to standard 10x guidelines with the assistance of Hubert Slawinski and Moustafa Attar, who also provided this service for all other samples discussed in this thesis entering the 10x Genomics pipeline. The WCHG then provided fastq sequencing result files and summary reports of sequencing batches including output from the Cellranger pipeline. I then re-ran the Cellranger pipeline on all fastq files relating to specific samples from all sequencing batches as appropriate, and performed any additional processing (such as aggregation of multiple samples) to generate the final gene expression matrices used for downstream analysis.

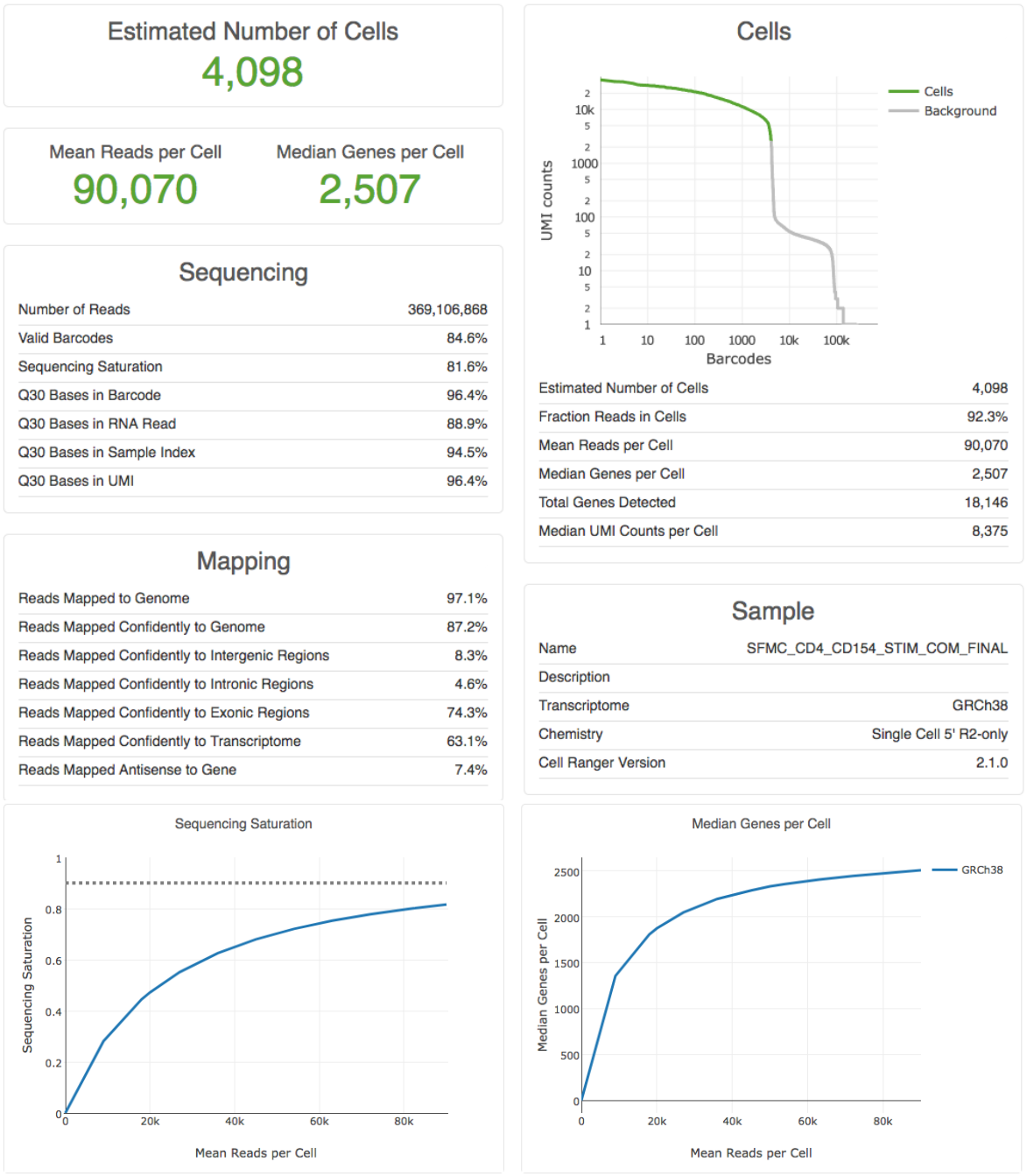


**Figure 4-2 Quality control summary data from 10x scRNA-seq of *Salmonella*-reactive synovial CD154+Th memory cells.**

Sequencing: data relating to read pairs sequenced including total read pairs and the percentage mapping to valid barcodes. Q30 Bases refers to the fraction of bases with a Q-score >=30. Mapping: fraction of reads mapping to different regions of the human genome and to genes of the transcriptome. Sequencing saturation graph: dashed line represents the estimated saturation beyond which additional sequencing depth would not be beneficial. Cells: statistics related only to reads mapping to barcodes considered to be associated with cells (green line in graph). Fraction reads in cells refers to the fraction of transcriptome reads with valid barcodes, where the barcode is also associated with a cell. Sample: additional metadata relating to this sequencing run, including the reference transcriptome to which reads were mapped and whether the chemistry used was 3 prime or 5 prime based. This figure is composed of summary data generated as part of the 10x Cellranger pipeline. Cells are from patient PSA1607.

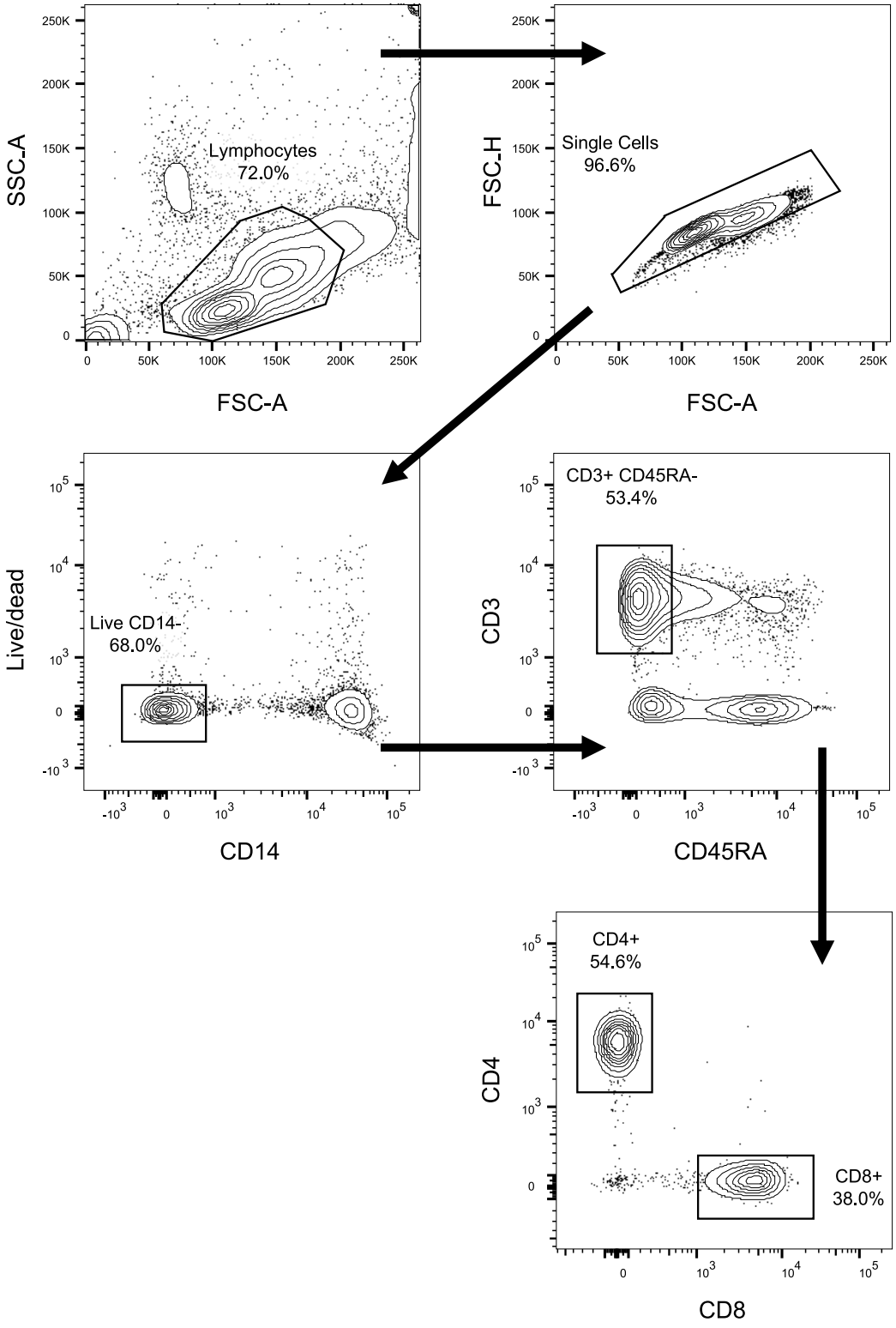


**Figure 4-3 Gating strategy for sorting unstimulated CD4 and CD8 memory T cells**

Dot plots depicting the FACS gating strategy used to sort live CD4+CD8- and CD8+CD4- memory T cells from unstimulated PBMC and SFMC for downstream 10x scRNA-seq. PBMC and SFMC were separated from peripheral blood or synovial fluid respectively then surface stained and washed prior to sorting (see methds). Unstimulated cells used in downstream ssRNA-seq were separated, sorted and transported to WCHG by Nicole Yager.

To make meaningful comparisons between synovial stimulated and unstimulated datasets, the Cell Ranger "aggr" function was used to normalise both libraries to a common read depth of 36616 reads/cell which identified 1421 median genes per cell. The gene expression matrix generated by the aggr function was then analysed using the Seurat R package. Cells were first filtered to exclude cells where transcripts for more than 2500 genes had been identified as these were rare and likely to be multiplets (partitions erroneously containing more than one cell during single-cell sequencing). The expectation that allelic exclusion would prevent more than 1 beta chain being transcribed and that at most 2 alpha chains could be transcribed per cell was further used to minimise multiplets. Cells where mitochondrial gene transcripts contributed to over 10% of transcripts (indicating dying or stressed cells) were also filtered. The SCTransform algorithm, which has demonstrated robustness in analysing gene expression datasets from different platforms sequenced at different depths[4], was then used to normalise and scale gene expression counts of both stimulated and unstimulated cells before calculating principal components (PC) across the unified dataset.

It is very useful when interpreting scRNA-seq data to reduce the high-dimensional gene expression matrix that is eventually generated into a 2-dimentional visual representation. To achieve this, the UMAP (Uniform Manifold Approximation and Projection) algorithm[5] was used to create a plot where each point represents a cell and the positioning of cells is representative of both local (within a cluster) and inter-cluster relationships[6]. The t-SNE algorithm[7] was also used for comparison and generated similar results (data not shown). Figure 4-4 shows the UMAP plot of CD4 and CD8 T cells originating from unstimulated SFMC as well as CD154+ CD4 T cells originating from *Salmonella*-stimulated SFMC. Unstimulated (left) and stimulated (right) cell populations clearly clustered very distinctly from one another. However, it is worth

noting that 12 ex vivo cells that had not undergone *in vitro* stimulation with *Salmonella* lysate still clustered with stimulated cells. This indicates that certain *ex vivo* cells had a phenotype resembling *Salmonella* stimulated CD154+ Th memory cells.

In addition 2 stimulated cells clustered with a small island of cells which most resembled cycling *ex vivo* cells by their increased expression of MKI67 and STMN1 when compared to all other cells (MKI67 logFC= 1.9, adj. p = 1.50E-74 ; STMN1 logFC= 3.1, adj. p= 3.32E-210; statistics: Seurat FindAllMarkers function with bimod[8][9] test).

As mentioned, sorted *ex vivo* CD4 and CD8 T cells were re-combined before entering the 10x pipeline at a known ratio of 1:1. These cells were also sorted as being either exclusively CD4+ or exclusively CD8+ before sequencing, therefore no double positive *ex vivo* CD4+CD8+ T cells were expected in the sequencing dataset. *Salmonella* stimulated SFMC were sorted only on positive expression of CD3, CD4 and CD154 within the memory (CD45RA-) lymphocyte gated population and were not negatively gated on CD8. Figure 4-5A and Figure 4-5B depict cells expressing at least one transcript of CD8A or CD4 respectively in the combined dataset of stimulated and unstimulated cells. Only 14 cells from the *Salmonella* stimulated population on the right had evidence of any CD8A gene expression. CD4 gene expression in the unstimulated fraction was inversely related to the vast majority of cells expressing CD8A, and cells expressing transcripts of either CD4 or CD8 also clustered separately within their local clusters. Of the 14,669 unstimulated synovial cells for which gene expression data was available, 2696 cells were quality control filtered leaving 11973 cells for downstream analysis. 51.6% of these cells expressed at least one CD8A or CD8B transcript. This was in agreement with the expected frequency of unstimulated CD8 T cells given the prior combination of CD4 and CD8 T cells in a 1:1 ratio, and formed the basis of

delineating CD8 T cells from CD4 T cells in downstream analysis. Although gene expression of CD4 associated with the stimulated fraction of cells as expected, CD4 gene expression was only detected in a limited number of these cells, reducing its utility as a cell classifier. There was however strong agreement between the percentage of stimulated cells expressing CD4 transcripts (27%) and the percentage of cells expressing CD4 transcripts among unstimulated cells lacking CD8 gene expression (28%).

Taken together, this initial analysis shows that the gene expression matrices captured by scRNA-seq were representative of the cell type, cell frequency and culturing conditions of cells entering the 10x scRNA-seq pipeline.

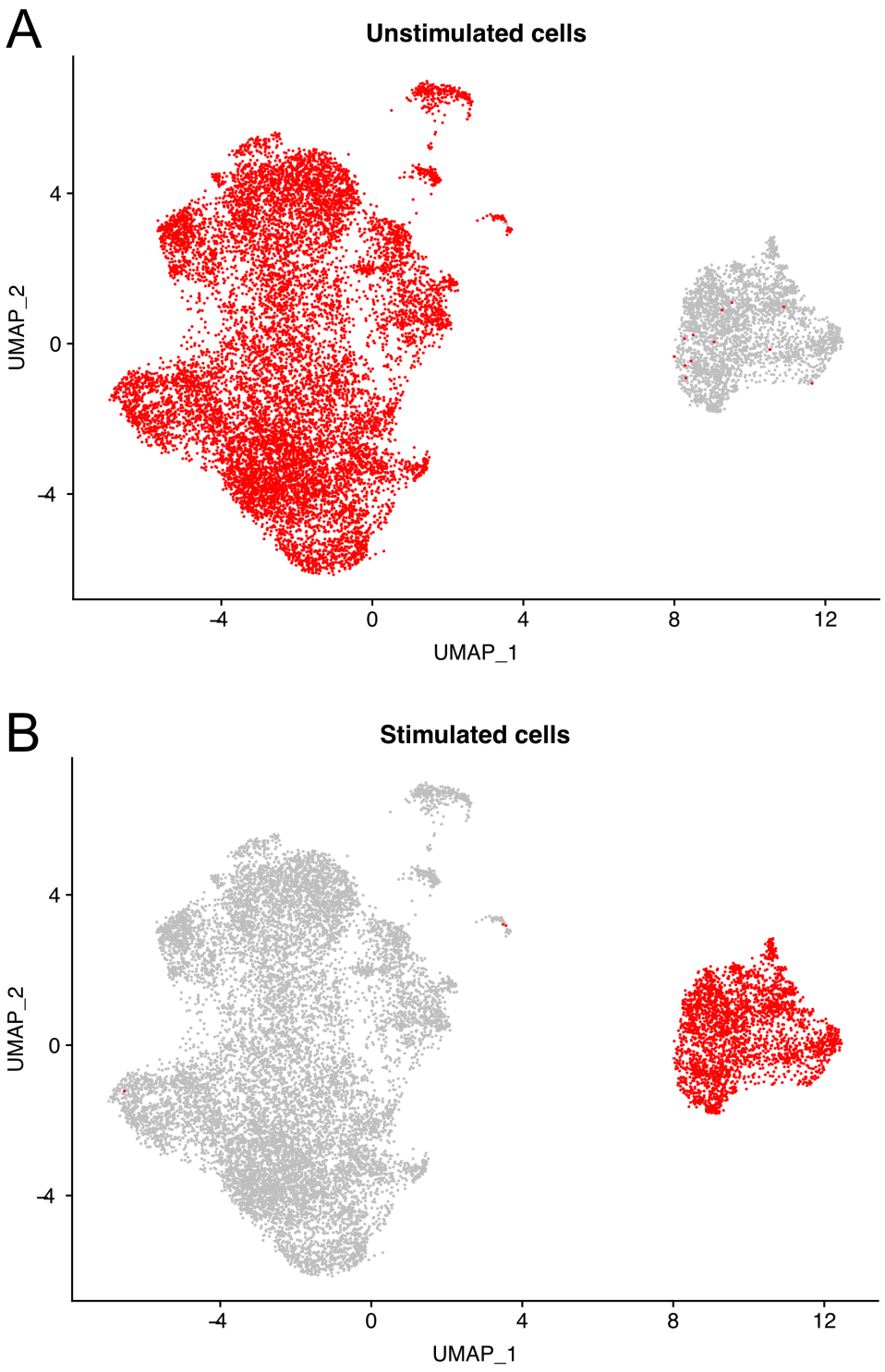


**Figure 4-4 *Salmonella* stimulated and unstimulated synovial T cells cluster separately with minimal overlap.**

A single UMAP showing (A): *ex vivo* CD4+ and CD8+ memory T cells sorted by FACS from SFMC and (B): CD154+ Th memory cells sorted by FACS from SFMC stimulated overnight with *S. Typhimurium* lysate. SFMC were separated from the same synovial fluid sample of patient PSA1607. Each sorted fraction was processed in a separate 10x channel producing 2 separate 5 prime RNA gene expression libraries. The Cellranger AGGR function was used to normalise sequencing depth between the 2 libraries. Cells were filtered for doublets and further normalised using the SCTransform function of the Seurat R package before principal component analysis and generation of the UMAP. Each dot represents a cell, and red is used to colour unstimulated cells in (A), and stimulated cells in (B).

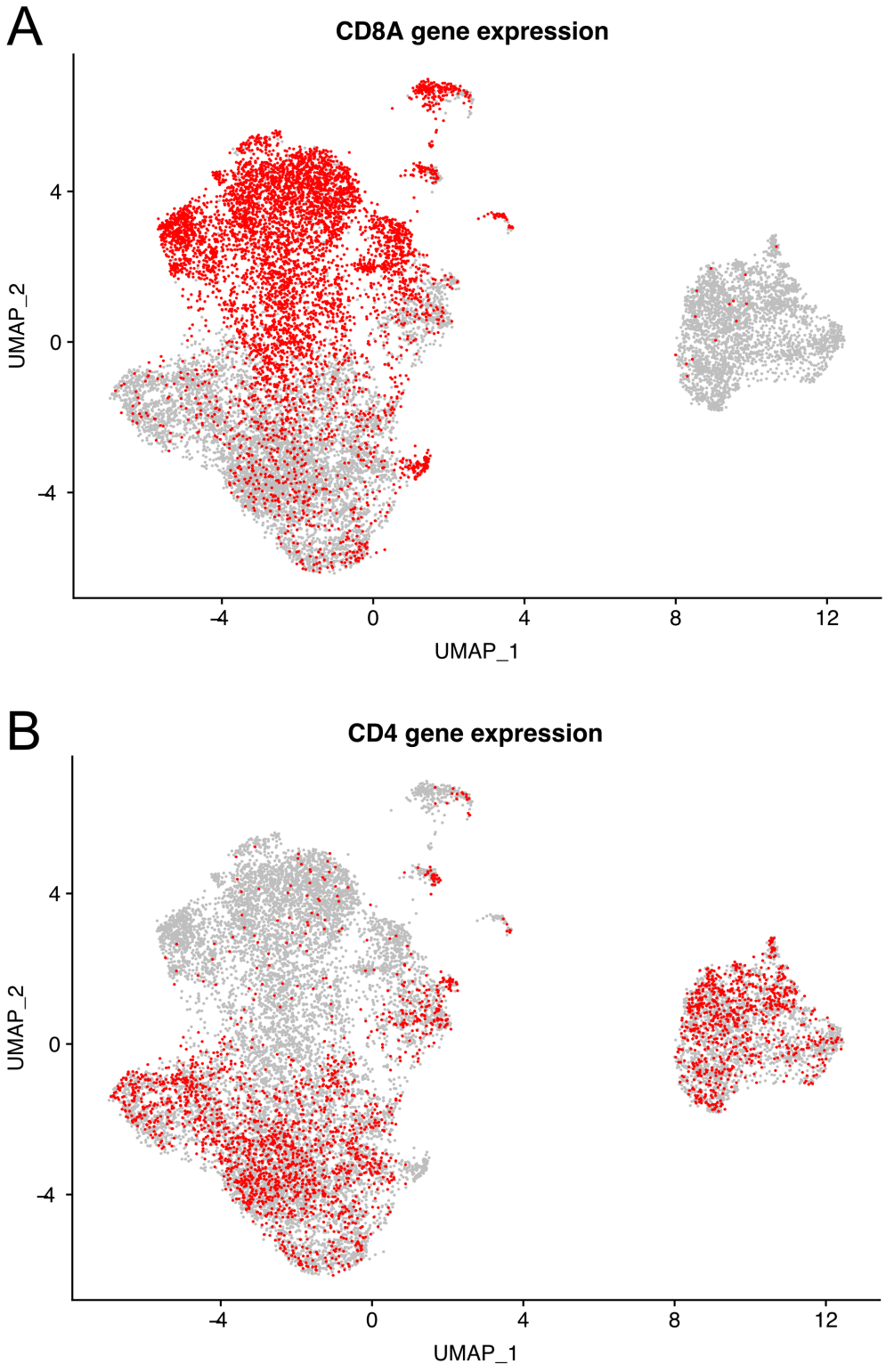


**Figure 4-5 The frequency of cells expressing CD4 and CD8 gene transcripts is consistent with the proportion of CD4+ or CD8+ T cells entering the 10x single-cell sequencing pipeline.**

UMAP plot (from figure Figure 4-4) highlighting the gene expression of (A): CD8A and (B): CD4. Any cell identified as having at least 1 transcript of the respective genes are marked red. Transcript frequency is based on counts after Cellranger normalisation of read depth between stimulated and unstimulated libraries. Each dot represents a cell.

### 4.3.3 CD40LG gene expression is present in 16% of *ex vivo* unstimulated synovial CD4 T-cells, a subset of which is clonally expanded in synovial fluid.

In the previous chapter, when stimulating SFMC overnight with microbial lysates, a negative control well of unstimulated cells was included as part of the assay. The mean percentage of SpA unstimulated CD4+ memory T cells from synovial fluid found to express CD154 by ICS was 0.12% where paired blood and synovial fluid was available (Figure 3-9) and 0.11% across all SpA synovial fluid samples. A previous ICS staining of unstimulated SFMC from patient PSA1607 was also available, where 0.04% of Th memory cells were found to express CD154. If the *ex vivo* surface expression of CD154 is similar to the overnight rested expression of CD154, it would be reasonable to predict that between 0.04% and 0.12% of the 5793 unstimulated Th memory cells (2 to 7 cells) would express the CD40LG gene. Surprisingly, CD40LG transcripts were detected in 932 (16%) of the 5793 unstimulated synovial Th memory cells from this patient, and 235 (3.8%) of 6180 unstimulated CD8 memory T cells (Figure 4-6A). The fraction of *Salmonella* stimulated Th memory cells expressing CD40LG was much higher (67%), as would be expected, but not universal. When looking at the relative expression of CD40L in both stimulated and unstimulated populations (Figure 4-6B), certain areas within the stimulated cluster had a higher relative gene expression of CD40LG compared to remaining stimulated and unstimulated cells.

To explore the relationship between clonality and CD40LG gene expression in synovial unstimulated cells, these cells were subsetted into CD40LG+ cells which also had CD4 or CD8 associated clonotype information available (1071 cells). The majority of clonotypes were represented by less than 3 cells, making it difficult to draw meaningful conclusions from clonal abundance alone regarding these cells. However, as clonal information was also available for a similar number of peripheral blood CD4 and CD8

memory T cells from the same patient acquired at the same time, this information was used to determine the relative frequency of clones in peripheral blood for comparison. 876 of the 1071 unstimulated CD40LG+ synovial cells belonged to clonotypes found exclusively in synovial fluid and not blood. All clonotypes represented by 3 or more CD40LG+ synovial memory T cells are listed in Table 4-1, in addition to information relating to each clone's enrichment in either synovial fluid or peripheral blood (analysed in more detail in chapter 5). Of the clonotypes listed in Table 4-1, all CD4 clones and all but 1 CD8 clone were over-represented in the CD40LG population relative to what would be expected by chance (16% and 3.8% of the synovial frequency of a CD4 or CD8 clone respectively). 29% to 75% of cells from CD4 clones listed in this table were CD40LG+, compared to only 2%-23% of the cells belonging to CD8 clones. One CD4 clone from Table 4-1 in particular, "CD4-1", was also found to be statistically enriched in this patient's synovial fluid compared with their peripheral blood. (Fisher's exact test, adj p-value = 0.032).

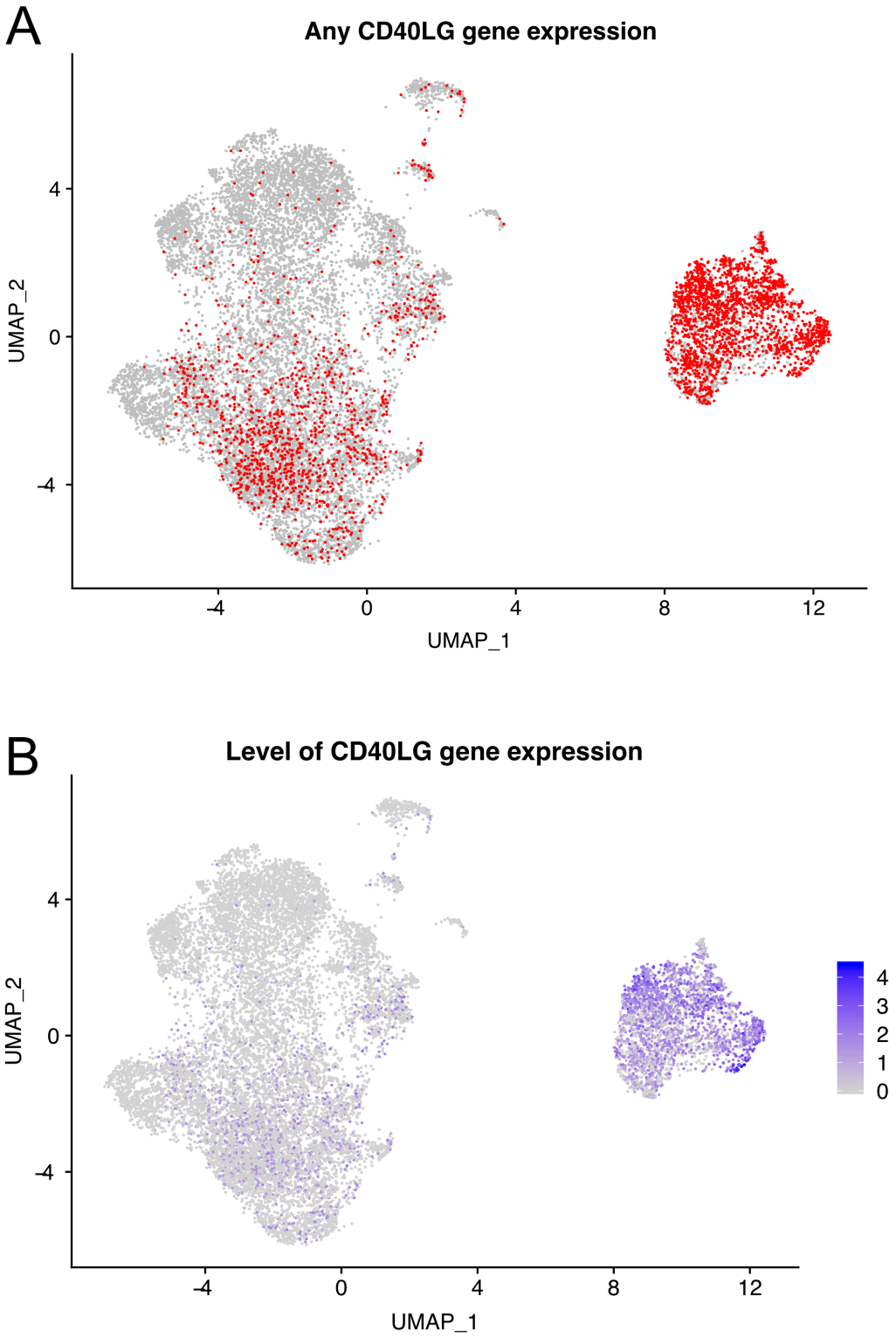


**Figure 4-6 Gene expression of CD40LG is primarily restricted to CD4 T cells and enhanced in synovial *Salmonella*-reactive Th memory cells.**

UMAP plot (from Figure 4-4) highlighting cells with any CD40LG (CD154) gene expression in red (A) and indicating the degree of gene expression in purple (B). A: Transcript frequency is based on counts after Cellranger normalisation of read depth between stimulated and unstimulated libraries. (B): Transcript frequency is based on (A) but further log normalised with a scale factor of 10000 using the Seurat "NormalizeData" function. Increased intensity of purple colour represents higher normalised transcript frequency.

## Table 4-1 CD40LG+ unstimulated cells include synovial enriched clones.

| Clone ID | CD4 / CD8 clone type | CD40LG+ cells with this clone | Expected CD40LG+ cells with this clone by chance | CD40LG+ difference from expected cell number | CD40LG fold-change difference from expected | Higher proportion of this clone in SF or PB | Proportion enrichment Fisher adj. p | Peripheral blood frequency | Peripheral blood proportion (%) | Synovial fluid frequency | Synovial fluid proportion (%) | beta cdr3 amino acid sequence | beta v gene | beta j gene | beta d gene | alpha cdr3 amino acid sequence | alpha v gene | alpha j gene |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CD4-1 | CD4+ | 4 | 2.24 | 1.76 | 1.7857 | SF | 0.032 | 0 | 0 | 14 | 0.1025 | CASSQRFSGAGELFF | TRBV14 | TRBJ2-2 | TRBD2 | CAVSALKAAGNKLTF | TRAV8-4 | TRAJ17 |
| CD4-2 | CD4+ | 3 | 1.60 | 1.40 | 1.8750 | SF | 1.000 | 1 | 0.0070 | 10 | 0.0732 | CASSSLSGTLGGYTF | TRBV27 | TRBJ1-2 | TRBD1 | CAVRQGLTGGGNKLTF | TRAV8-6 | TRAJ10 |
| CD4-3 | CD4+ | 3 | 1.28 | 1.72 | 2.3438 | SF | 1.000 | 1 | 0.0070 | 8 | 0.0585 | CASSYSNQRGQDTQYF | TRBV6-5 | TRBJ2-3 | TRBD2 | CAVSFKNQGGKLIF | TRAV12-2 | TRAJ23 |
| CD4-4 | CD4+ | 3 | 1.12 | 1.88 | 2.6786 | SF | 1.000 | 0 | 0 | 7 | 0.0512 | CASSTGSMGYTF | TRBV28 | TRBJ1-2 | TRBD1 | CAATRQTGANNLFF | TRAV29DV5 | TRAJ36 |
| CD4-5 | CD4+ | 3 | 1.12 | 1.88 | 2.6786 | SF | 1.000 | 0 | 0 | 7 | 0.0512 | CASSARVGNSPLHF | TRBV6-1 | TRBJ1-6 | TRBD1 | CVVNVNNDMRF | TRAV12-1 | TRAJ43 |
| CD4-6 | CD4+ | 3 | 0.64 | 2.36 | 4.6875 | SF | 1.000 | 0 | 0 | 4 | 0.0293 | CSARDPRAGGFDEQYF | TRBV20-1 | TRBJ2-7 | TRBD1 | *NA* | *NA* | *NA* |
| CD4-7 | CD4+ | 3 | 0.64 | 2.36 | 4.6875 | SF | 1.000 | 1 | 0.0070 | 4 | 0.0293 | CASSPSPGPEAIQYF | TRBV7-2 | TRBJ2-4 | TRBD2 | CAMRGTSGSARQLTF | TRAV14DV4 | TRAJ22 |
| CD4-8 | CD4+ | 3 | 0.64 | 2.36 | 4.6875 | SF | 1.000 | 0 | 0 | 4 | 0.0293 | CASGRPPTIGSPLHF | TRBV2 | TRBJ1-6 | *NA* | CALMLTQGGSEKLVF | TRAV9-2 | TRAJ57 |
| CD8-1 | CD8+ | 3 | 6.99 | -3.99 | 0.4291 | SF | 0.574 | 134 | 0.9352 | 184 | 1.3465 | CASSPPWVQETQYF | TRBV7-9 | TRBJ2-5 | *NA* | CAVSEDNDMRF | TRAV8-4 | TRAJ43 |
| CD8-2 | CD8+ | 11 | 4.71 | 6.29 | 2.3345 | SF | 0.012 | 67 | 0.4676 | 124 | 0.9074 | CASSTGTGELFF | TRBV27 | TRBJ2-2 | TRBD1 | CAVLPQSGGSYIPTF | TRAV22 | TRAJ6 |
| CD8-3 | CD8+ | 5 | 4.18 | 0.82 | 1.1962 | SF | 1.000 | 95 | 0.6630 | 110 | 0.8050 | CASRGDSKADTQYF | TRBV6-5 | TRBJ2-3 | TRBD2 | CGTGPYGGSQGNLIF | TRAV30 | TRAJ42 |
| CD8-4 | CD8+ | 4 | 1.79 | 2.21 | 2.2396 | PB | 1.000 | 58 | 0.4048 | 47 | 0.3439 | CASSAPLFTVARYEQYF | TRBV6-1 | TRBJ2-7 | *NA* | CAYSPTQGGSEKLVF | TRAV38-2DV8 | TRAJ57 |
| CD8-5 | CD8+ | 7 | 1.18 | 5.82 | 5.9423 | SF | 1.000 | 29 | 0.2024 | 31 | 0.2269 | CASRGDSKADTQYF | TRBV6-5 | TRBJ2-3 | TRBD2 | *NA* | *NA* | *NA* |

### 4.3.4 TNF and IFNG gene transcripts are preferentially expressed in either CD4 or CD8 *ex vivo* memory T cells respectively, and are strongly co-expressed in *S. Typhimurium*-reactive Th memory cells.

The UMAP plots depicted in Figure 4-7 to Figure 4-10 compare the frequency of cells expressing TNF, IFNG, IL2, IL17A, IL17F, IL22, GM-CSF and IL10 gene transcripts between *Salmonella*-reactive Th memory cells (referred to as "stimulated") and unstimulated CD4 and CD8 memory T cells from the synovial fluid of patient PSA1607.

Within the unstimulated memory T cell fraction, approximately 1/3 more CD4 T cells expressed TNF compared to CD8 T cells (6.63% and 4.5% respectively). 9.87% of the CD4 T cells expressing CD40LG also expressed TNF. IFNG was conversely expressed in a lower proportion of unstimulated memory CD4 T cells (5.13%) or CD4 T cells expressing CD40L (6.76%) compared to CD8 T cells (10.83%). Proportions did not differ greatly when only considering cells expressing CD40L and belonging to the CD4 T cell clones listed in Table 4-1, with 8% of these cells expressing TNF and 8% expressing IFNG.

38% and 44% of the *S. Typhimurium* stimulated Th memory cells expressed TNF or IFNG respectively, with 18% of these cells co-expressing both cytokines. In contrast, co-expression of both TNF and IFNG gene transcripts was detected in less than 1% of unstimulated CD4 or CD8 memory T cells.

Strikingly 16% (919 of 5728) of the unstimulated CD8 memory T cells belonged to clones statistically enriched in synovial fluid (Fisher test, adj. $p <= 0.05$) relative to peripheral blood for this patient, in contrast to 1.1% (61 of 5233) for unstimulated CD4 memory T cells (Table 4-2 All synovial enriched CD4 and CD8 clones for patient

PSA1607)[2]. Clonal enrichment of *ex vivo* unstimulated memory T cell populations are discussed further in the next chapter, but are referred to here for comparison with cytokine overlap.

Out of the 355 unstimulated CD4 memory T cells expressing TNF, 1.1% (4 cells) were of a clonotype enriched in synovial fluid relative to peripheral blood for this patient, which did not differ from what would be expected by chance. Similarly only 18.5% (49 of the 265) unstimulated CD8 T cells expressing TNF were of a synovial enriched CD8 clonotype, similar to what would be expected by chance (16%).

[2] Calculations based on cell counts of only cells with both gene expression and TCR data available after filtering, therefore cell frequency will differ slightly from table (1094 cells from 10x synovial enriched clones in table rather than 919).

## Table 4-2 All synovial enriched CD4 and CD8 clones for patient PSA1607

| CD4 / CD8 clone type | platform | Proportion enrichment in SF Fisher adj. p | Peripheral blood frequency | Peripheral blood proportion (%)* | Synovial fluid frequency | Synovial fluid proportion (%)* | beta cdr3 amino acid sequence | beta v gene | beta j gene | beta d gene | alpha cdr3 amino acid sequence | alpha v gene | alpha j gene |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CD8+ | 10x | 2.14E-112 | 40 | 0.2792 | 527 | 3.8566 | CASSWGLLTEAFF | TRBV28 | TRBJ1-1 | TRBD1 | *NA* | *NA* | *NA* |
| CD8+ | ss2 † | 0.000539821 | 1 | 0.2342 | 32 | 5.1948 | CASSWGLLTEAFF | TRBV28 | TRBJ1-1 | TRBD1 | CAGVGSDGQKLLF | TRAV27 | TRAJ16 |
| CD8+ | 10x | 6.39E-33 | 9 | 0.0628 | 153 | 1.1196 | CASSPTGGYHNEQFF | TRBV27 | TRBJ2-1 | TRBD1 | CAVRDPLNRDDKIIF | TRAV3 | TRAJ30 |
| CD8+ | 10x | 5.47E-13 | 13 | 0.0907 | 92 | 0.6733 | CASSLGLATPNTEAFF | TRBV5-1 | TRBJ1-1 | *NA* | CAVGDTGFQKLVF | TRAV21 | TRAJ8 |
| CD8+ | 10x | 0.0001943 | 0 | 0 | 23 | 0.1683 | CAISDPTGGTDTQYF | TRBV10-3 | TRBJ2-3 | TRBD1 | CAMREGYNFNKFYF | TRAV14DV4 | TRAJ21 |
| CD8+ | 10x | 0.000283237 | 1 | 0.0070 | 26 | 0.1903 | CASRGGGYGTDTQYF | TRBV5-1 | TRBJ2-3 | TRBD2 | CAVEDQGYQKVTF | TRAV2 | TRAJ13 |
| CD8+ | 10x | 0.005020934 | 15 | 0.1047 | 51 | 0.3732 | CASSGQETKETQYF | TRBV5-1 | TRBJ2-5 | TRBD1 | CAVRDQGSNYQLIW | TRAV1-1 | TRAJ33 |
| CD8+ | 10x | 0.007866394 | 0 | 0 | 17 | 0.1244 | CASSPTGGYHNEQFF | TRBV27 | TRBJ2-1 | TRBD1 | *NA* | *NA* | *NA* |
| CD8+ | 10x | 0.011564951 | 67 | 0.4676 | 124 | 0.9074 | CASSTGTGELFF | TRBV27 | TRBJ2-2 | TRBD1 | CAVLPQSGGSYIPTF | TRAV22 | TRAJ6 |
| CD8+ | 10x | 0.011717371 | 25 | 0.1745 | 65 | 0.4757 | CASSLDTHETSGGSSSYNEQFF | TRBV11-2 | TRBJ2-1 | *NA* | *NA* | *NA* | *NA* |
| CD8+ | 10x | 0.011717371 | 0 | 0 | 16 | 0.1171 | CATSDPNDRDTSGNTIYF;CASSQDPTKPQHF | TRBV24-1;TRBV3-1 | TRBJ1-3;TRBJ1-5 | TRBD1; | CAENPLSNAGGTSYGKLTF | TRAV13-2 | TRAJ52 |
| CD4+ | 10x | 0.012995977 | 1 | 0.0070 | 19 | 0.1390 | CAISESLEADSPLHF | TRBV10-3 | TRBJ1-6 | TRBD2 | CAMSHTGGFKTIF | TRAV12-3 | TRAJ9 |
| CD4+ | 10x | 0.012802485 | 2 | 0.0140 | 22 | 0.1610 | CASSLGTAGNYGYTF | TRBV7-9 | TRBJ1-2 | *NA* | CAMSINQAGTALIF | TRAV12-3 | TRAJ15 |
| CD4+ | 10x | 0.011717371 | 0 | 0 | 16 | 0.1171 | CASSPRADTQYF | TRBV5-8 | TRBJ2-3 | *NA* | CAAALSGNTPLVF | TRAV23DV6 | TRAJ29 |
| CD4+ | 10x | 0.032396089 | 0 | 0 | 14 | 0.1025 | CASSQRFSGAGELFF | TRBV14 | TRBJ2-2 | TRBD2 | CAVSALKAAGNKLTF | TRAV8-4 | TRAJ17 |
| * | Proportions are calculated as a fraction of all CD4 and CD8 clones. Proportions within clones of the same type (CD4/CD8) will be higher. | | | | | | | | | | | | |
| † | Paired alpha and beta chain discovered by SS2 sequencing of 10x clone of row above. Inferred to be the same clonotype as the 10x clone above, based on identical beta chain CDR3 and similar enrichment in synovial fluid. | | | | | | | | | | | | |
| *NA* = | Not available | | | | | | | | | | | | |

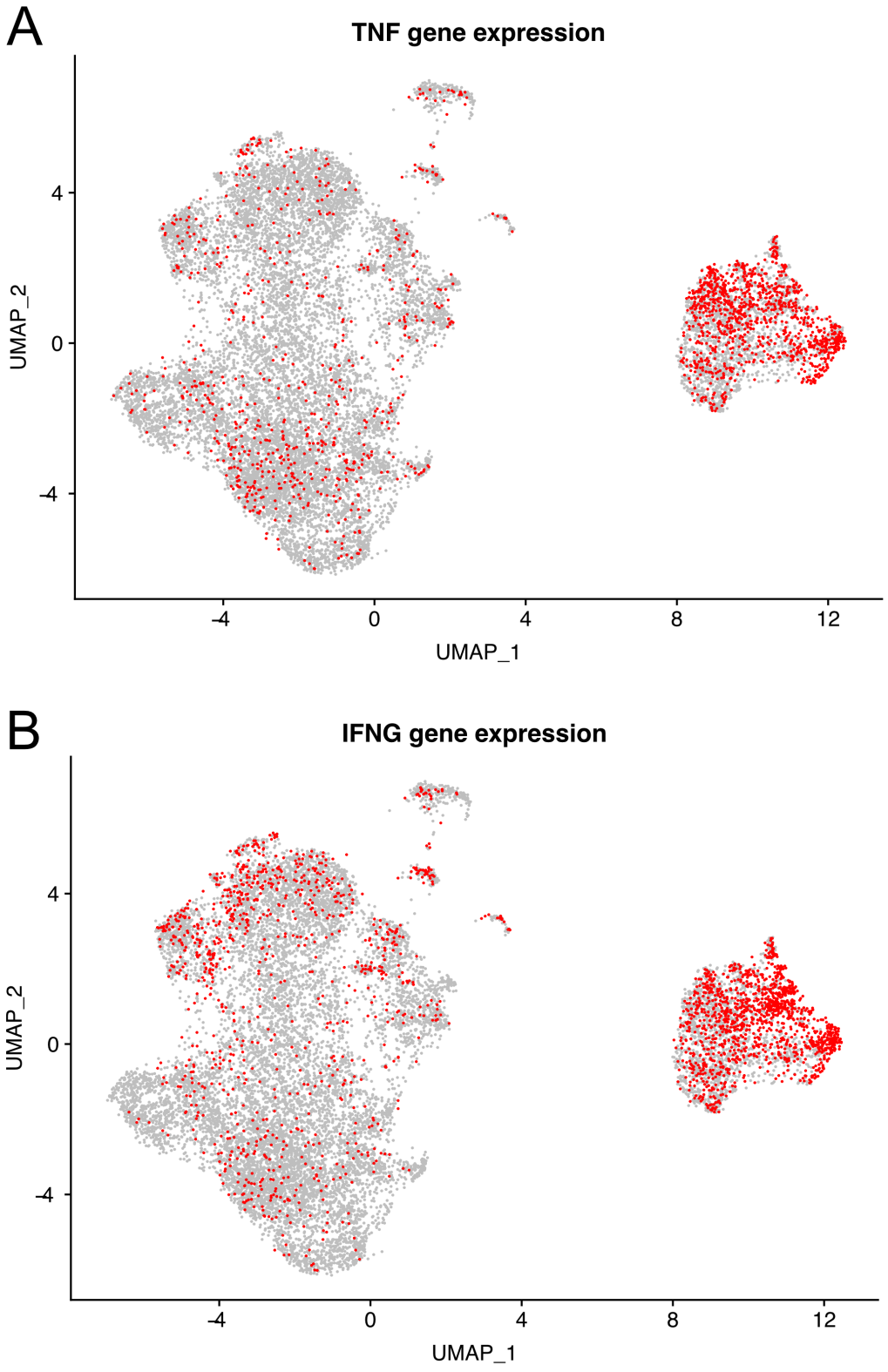


**Figure 4-7 TNF and IFNG gene transcripts are preferentially expressed in either CD4 or CD8 *ex vivo* synovial memory T cells respectively, and strongly co-expressed in *Salmonella*-reactive synovial Th memory cells.**

UMAP plot (from figure Figure 4-4) highlighting the gene expression of (A): TNF and (B): IFNG. Any cell identified as having at least 1 transcript of the respective genes are marked red. Transcript frequency is based on counts after Cellranger normalisation of read depth between stimulated and unstimulated libraries.

### 4.3.5 A distinct cluster of *Salmonella*-reactive Th memory cells express IL2, a portion of which co-express CSF2.

IL2 gene expression was present in only 0.242% of CD4 and 0.194% of CD8 unstimulated memory T cells but 7% of *Salmonella*-reactive Th memory cells (Figure 4-8). Each unstimulated CD4 memory T cell expressing IL2 was of a different clonotype, none of which were enriched in synovial fluid. Similarly only 2 of the 12 cells expressing IL2 in the unstimulated CD8 population were of the same clonotype.

Transcripts for CSF2, the gene encoding GM-CSF, were similarly found in only a handful of unstimulated cells (0.1% CD4 and CD8 memory T cells, each cell of a different unenriched clonotype), but in 2.95% of stimulated Th memory cells. Almost 1/3 of the stimulated cells expressing CSF2 also expressed IL2. *Salmonella*-stimulated Th memory cells expressing IL2 and to a lesser extent CSF2 clustered very distinctly within the UMAP plot, suggesting a distinct phenotype might be associated with these cells.

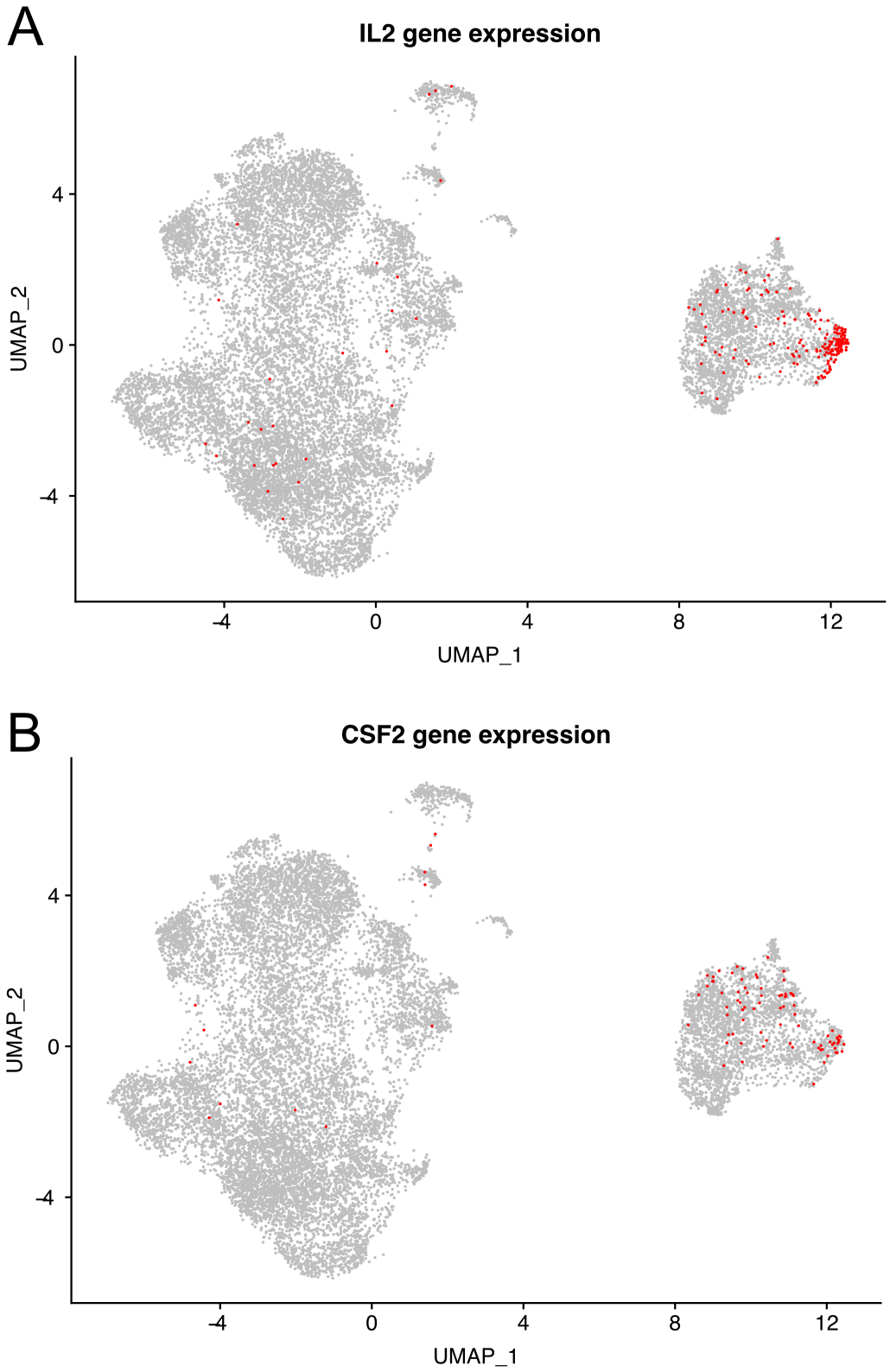


**Figure 4-8 A distinct cluster of *Salmonella*-reactive synovial Th memory cells express IL2, a portion of which co-express CSF2.**

UMAP plot (from figure Figure 4-4) highlighting the gene expression of (A): IL2 and (B): CSF2. Any cell identified as having at least 1 transcript of the respective genes are marked red. Transcript frequency is based on counts after Cellranger normalisation of read depth between stimulated and unstimulated libraries.

### 4.3.6 Th memory cells expressing IL17A and IL17F form a distinct cluster within *Salmonella*-reactive Th memory cells in close proximity and overlapping with cells expressing IL22.

As seen in Figure 4-9 cells expressing IL17A and IL17F transcripts clearly form distinct clusters within *Salmonella*-reactive Th memory cells, representing 2.8% and 6.1% of these cells respectively. There is also considerable overlap between these clusters, with 1.5% of *Salmonella*-reactive Th memory cells expressing both cytokines. In addition, a small but distinct cluster of cells expressing IL17A can be seen clustering in the same region as IL2 expressing cells. Cells expressing IL22 represented 11% of stimulated cells (Figure 4-10) and also formed a distinct cluster which is in close proximity to and partly overlapping with IL17A and IL17F clusters. The subset of *Salmonella*-reactive Th memory cells expressing IL17A and the related gene expression of this cluster is explored in detail later in this chapter.

Somewhat surprisingly, the 3 cells expressing IL17A and 1 cell expressing IL17F in the unstimulated fraction of cells were classified as CD8 rather than CD4 memory T cells. Each of these cells were of a different clonotype, none of which were enriched in synovial fluid. Only 1 CD8 and 2 CD4 unstimulated memory T cells expressed IL-22, also of different unenriched clonotypes and not co-expressing either IL17A or IL17F. 0.76% and 0.42% of unstimulated CD4 and CD8 memory T cells respectively expressed IL10 compared with 2.1% of *Salmonella* stimulated cells.

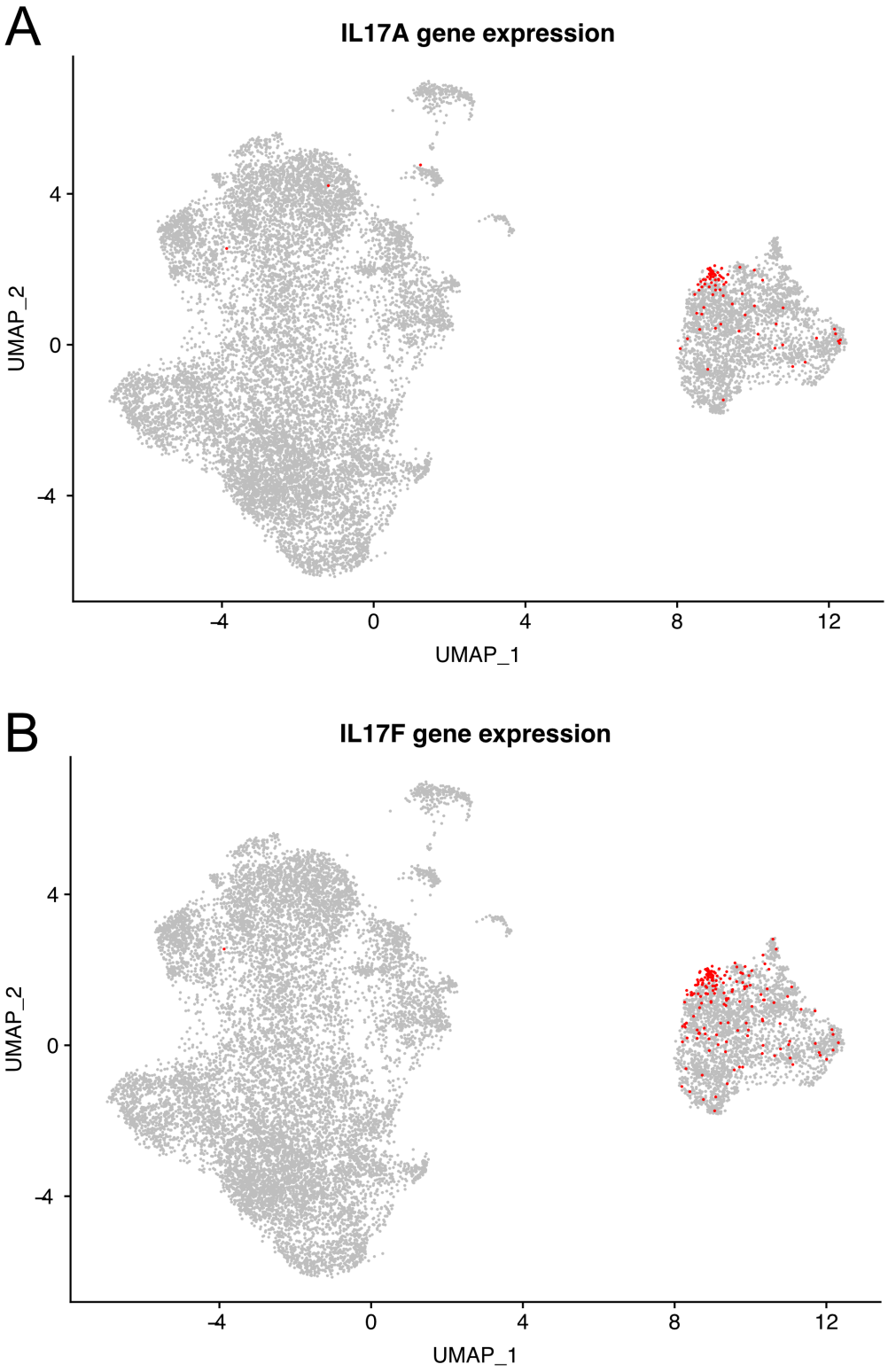


**Figure 4-9 Synovial Th memory cells expressing IL17A and IL17F form a distinct cluster among *Salmonella* stimulated cells but are very rare in unstimulated memory CD4 and CD8 T cells.**

UMAP plot (from figure Figure 4-4) highlighting the gene expression of (A): IL17A and (B): IL17F. Any cell identified as having at least 1 transcript of the respective genes are marked red. Transcript frequency is based on counts after Cellranger normalisation of read depth between stimulated and unstimulated libraries. The single cell expressing IL17F in the unstimulated fraction also expresses IL17A.

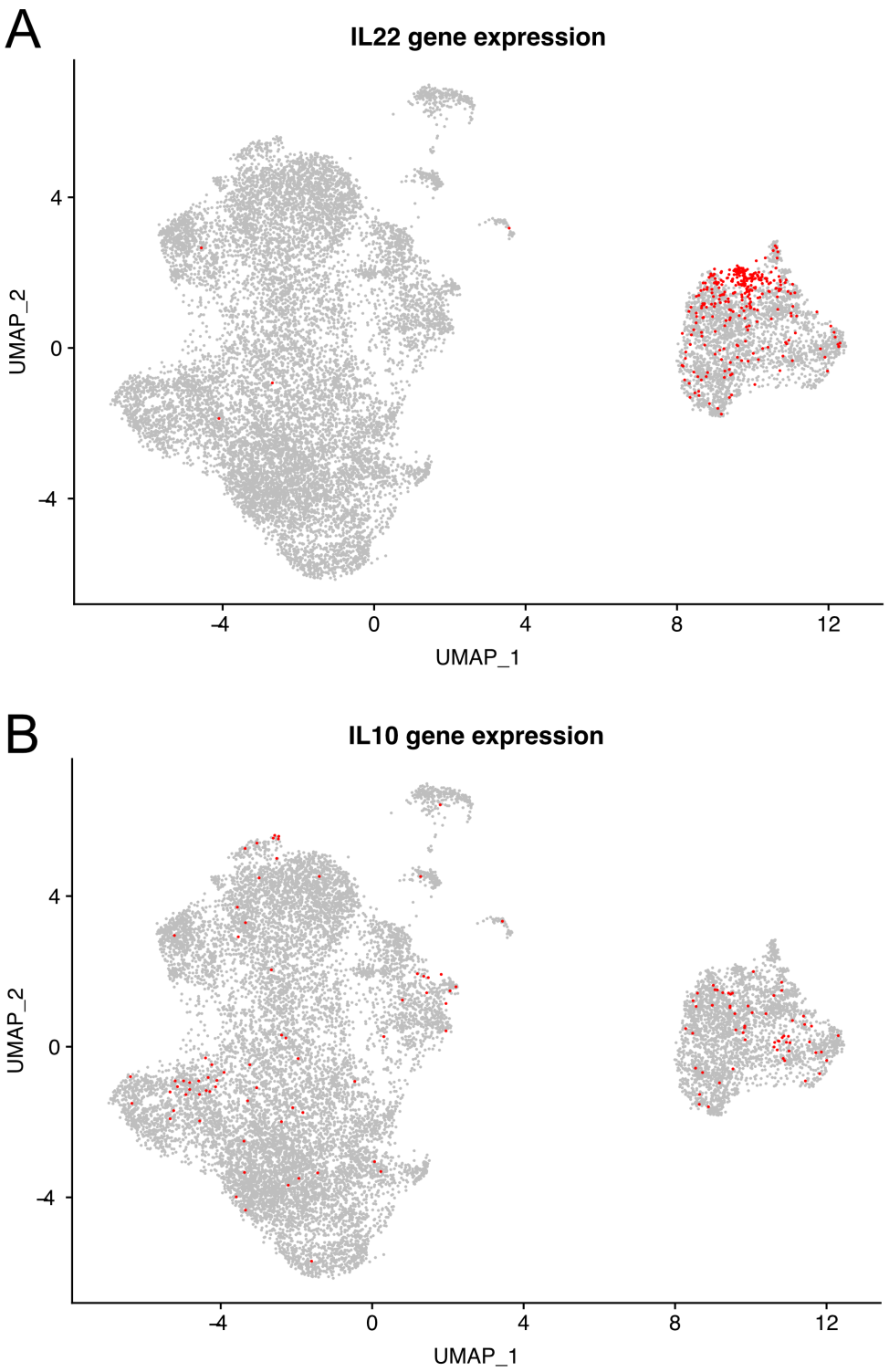


**Figure 4-10 Synovial Th memory cells expressing IL22 cluster in close proximity to and overlap with cells expressing IL17F and IL17A among *Salmonella* stimulated cells.**

UMAP plot (from figure Figure 4-4) highlighting the gene expression of (A): IL22 and (B): IL10. Any cell identified as having at least 1 transcript of the respective genes are marked red. Transcript frequency is based on counts after Cellranger normalisation of read depth between stimulated and unstimulated libraries.

### 4.3.7 Exploring the *Salmonella*-reactive Th memory dataset at a read depth of 90070 reads per cell.

In previous sections of this chapter, the read depth of stimulated and unstimulated datasets was normalised to 36062 reads per cell in order to make valid comparisons between them. However, at this read depth it is possible that some genes represented by a lower abundance of mRNA are not detected, or that the frequency of transcripts detected is not sufficient to properly capture the magnitude of gene expression. The following sections of this chapter will focus on the 10x scRNA-seq dataset of *Salmonella*-reactive synovial CD154+Th memory cells sequenced to a depth of 90070 reads per cell (Figure 4-2). This dataset reaches a sequencing saturation of 81.6%, very close to the estimated limit of saturation beyond which additional sequencing is not expected to be beneficial. The median genes per cell discovered also increases to 2507 from 1421 in the aggregated dataset. Most of the increase in median genes per cell is likely attributable to a higher number of genes being transcribed by an activated cell state, rather than undiscovered genes in the unstimulated dataset. This is because, despite a lower number of median genes found in the unstimulated dataset prior to aggregation (1333), sequencing saturation was still relatively high at 78%.

Filtering was applied to take account of the higher median genes per cell, this time excluding droplet partitions where the number of genes was > 5000 and as before, partitions where mitochondrial genes accounted for >10% of transcripts. Where clonal information was available for a cell, those cells associated with more than 1 TCR beta chain or more than 2 alpha chains were also excluded (Figure 4-11 and Figure 4-12).

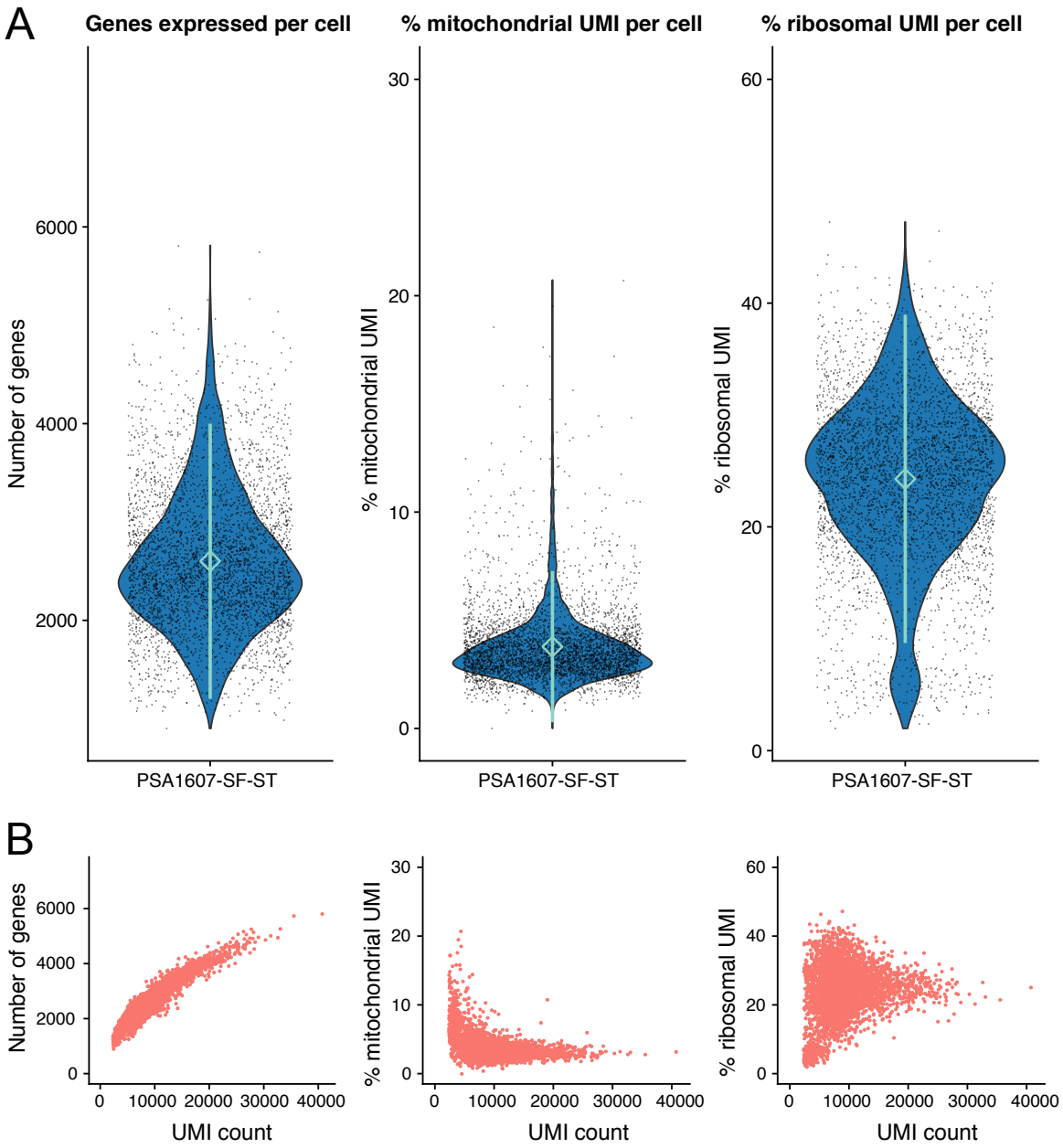


**Figure 4-11 Gene, mitochondrial and ribosomal content of *Salmonella*-reactive synovial CD154+Th memory cells**

A: Violin plots showing the number of genes expressed per sequenced cell and the percentage of mRNA transcripts attributed to mitochondrial or ribosomal genes. Diamond represents mean value, line ends are 2 standard deviations from the mean. B: Scatter graphs plotting the number of mRNA transcripts within a cell (UMI count) against: the number of genes detected in that cell; the percentage of mRNA transcripts attributed to mitochondrial genes; and the percentage of mRNA transcripts attributed to ribosomal genes. Each dot represents a cell. 4098 cells in total. UMI: Unique molecular identifier representing an mRNA transcript.

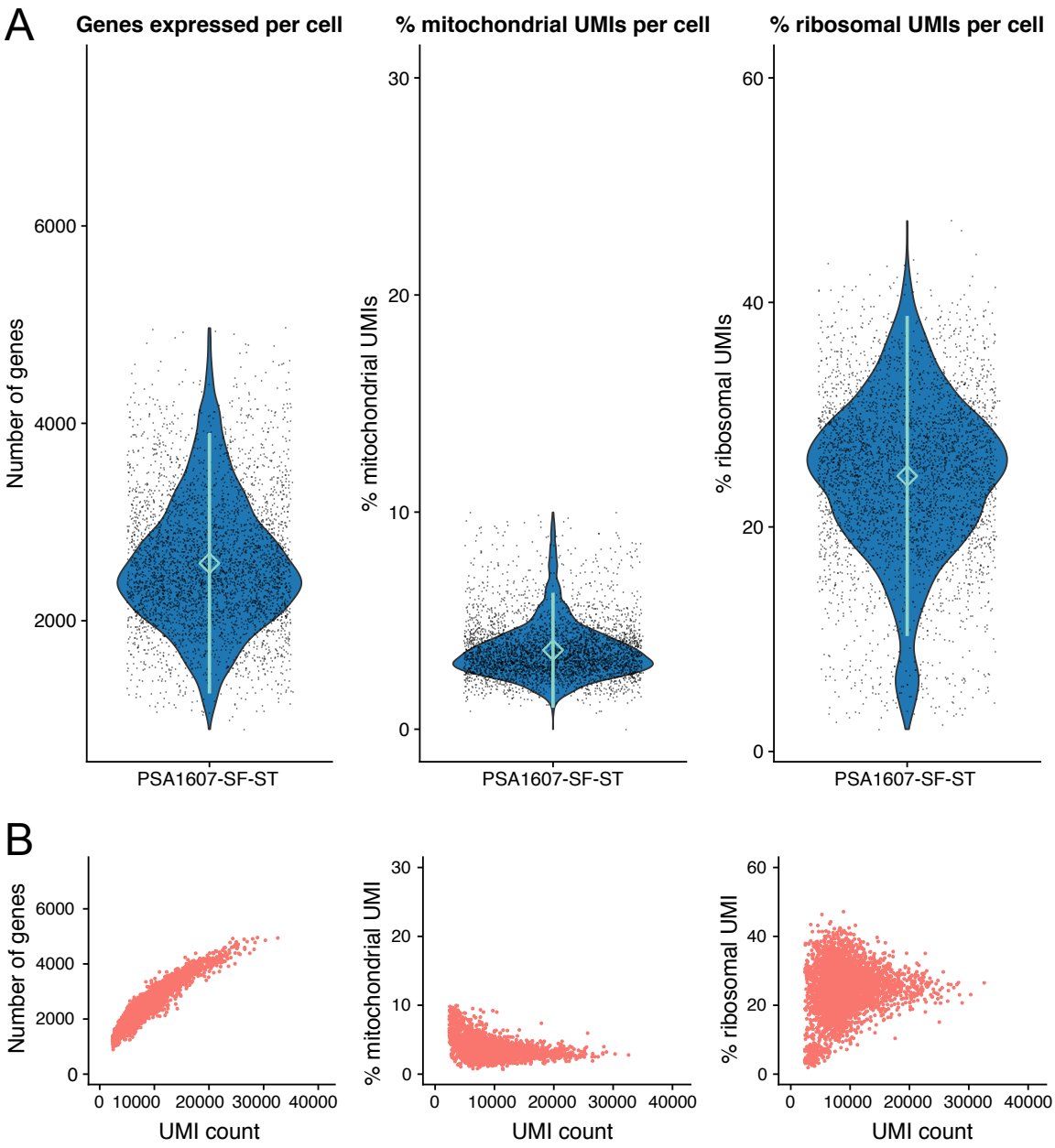


**Figure 4-12 *Salmonella*-reactive synovial CD154+Th memory cells remaining after filtering on number of genes and % of mitochondrial mRNA transcripts per cell.**

Same plots as those shown in Figure 4-11, excluding cells with < 500 or > 5000 genes per cell, cells containing > 10% mitochondrial transcripts and cells with more than 1 TCR beta chain or 2 TCR alpha chains where clonal information was available. A: Violin plots showing the number of genes expressed per sequenced cell and the percentage of mRNA transcripts attributed to mitochondrial or ribosomal genes. Diamond represents mean value, line ends are 2 standard deviations from the mean. B: Scatter graphs plotting the number of mRNA transcripts within a cell (UMI count) against: the number of genes detected in that cell; the percentage of mRNA transcripts attributed to mitochondrial genes; and the percentage of mRNA transcripts attributed to ribosomal genes. Each dot represents a cell. 3894 cells in total. UMI: Unique molecular identifier representing an mRNA transcript.

### 4.3.8 The frequency of cytokine producing cells measured by scRNA-seq was within the same range measured by ICS for all cytokines other than TNFa and IL22.

The previous chapter used a functional assay quantifying the frequency of CD154+ cells expressing TNFa, IFNg, IL-2, IL-17, GM-CSF and IL-22 at the protein level to assess Th memory immune responses towards microbial lysates. Intra-cellular staining of cytokines formed the basis of this assay which was analysed by FACS. This chapter describes the use of scRNA-seq to quantify both the frequency of cells expressing a particular gene, and the level of expression quantified by the number of transcripts detected per gene. The presence or frequency of a particular gene's RNA transcripts within a cell does not always correspond to the presence or level of protein encoded by that gene for the same cell. To interpret the RNA sequencing data from this chapter in the context of the previous chapter's results, it is therefore useful to first consider the relationship between protein expression and RNA expression between the two datasets. Figure 4-13 compares cytokine detection within cells measured by gene expression or intra-cellular staining of protein. Unfortunately, limited sample availability prevented the acquisition of ICS data for the exact same *Salmonella* stimulated SFMC sample from which CD154+ Th memory cells were sorted for scRNA-seq. Instead synovial ICS data from the same patient acquired on a different day (Figure 4-13A), and the complete range of synovial ICS data from all SpA patients (Figure 4-13B) is presented. The percentage of CD154+ Th memory cells where at least one transcript of IFNg, IL-2, IL-17 or GM-CSF was detected by scRNA-seq was within the range observed by ICS when considering all SpA synovial fluid samples. The percentage of cells expressing TNFa transcripts (50.1%) was lower than the range observed by ICS in all SpA patients (51.4%-92.3%). However, the percentage of CD154+ Th memory cells expressing TNFa as previously measured by ICS from the same patient was 55.4%, therefore towards the lower end of this spectrum and much closer to 50.1%. When

considering the percentage of CD154+ Th memory cells expressing IL-22, scRNA-seq detected IL-22 in far more cells (16.4%) than ICS was able to for any SpA sample assayed (12 fold higher than the highest frequency detected by ICS for any SpA patient (1.33%)).

The scRNA-seq results in Figure 4-13 are based on sequencing at a read depth of 90070 reads/cell, which was able to detect a higher frequency of cells expressing cytokines compared to the aggregated dataset, where reads were normalised to 36062 reads/cell (Table 4-3). A higher read depth of the aggregated dataset would have been unlikely to change conclusions drawn when comparing unstimulated and stimulated populations. However, the higher read depth was particularly useful when exploring the stimulated dataset in more depth, as it allowed for a higher number of cells expressing IL17A to be subsetted and separately analysed (discussed later in this chapter). Gene expression of CD40LG was also found in a higher frequency of *Salmonella*-stimulated cells sorted on CD154 expression by FACS (78% versus 67% at lower read depth) providing more confidence that this sorted fraction was indeed enriched for CD154+ cells.

**Table 4-3 An increased frequency of synovial *Salmonella*-reactive Th memory cells expressing cytokines was detected at a read depth of 90070 reads/cell**

| Cytokine | Frequency of cells expressing cytokine at 36062 reads/cell | Frequency of cells expressing cytokine at 90070 reads/cell | % Gain in frequency of cells detected |
|---|---|---|---|
| TNF | 37.5% | 50.1% | 33.5% |
| IFNG | 43.5% | 64.4% | 48.0% |
| IL2 | 7.0% | 10.0% | 44.0% |
| IL17A | 2.8% | 4.4% | 57.8% |
| IL17F | 6.1% | 9.7% | 59.1% |
| CSF2 | 3.0% | 5.1% | 72.0% |
| IL22 | 11.0% | 16.4% | 49.8% |
| IL10 | 2.1% | 4.3% | 104.2% |

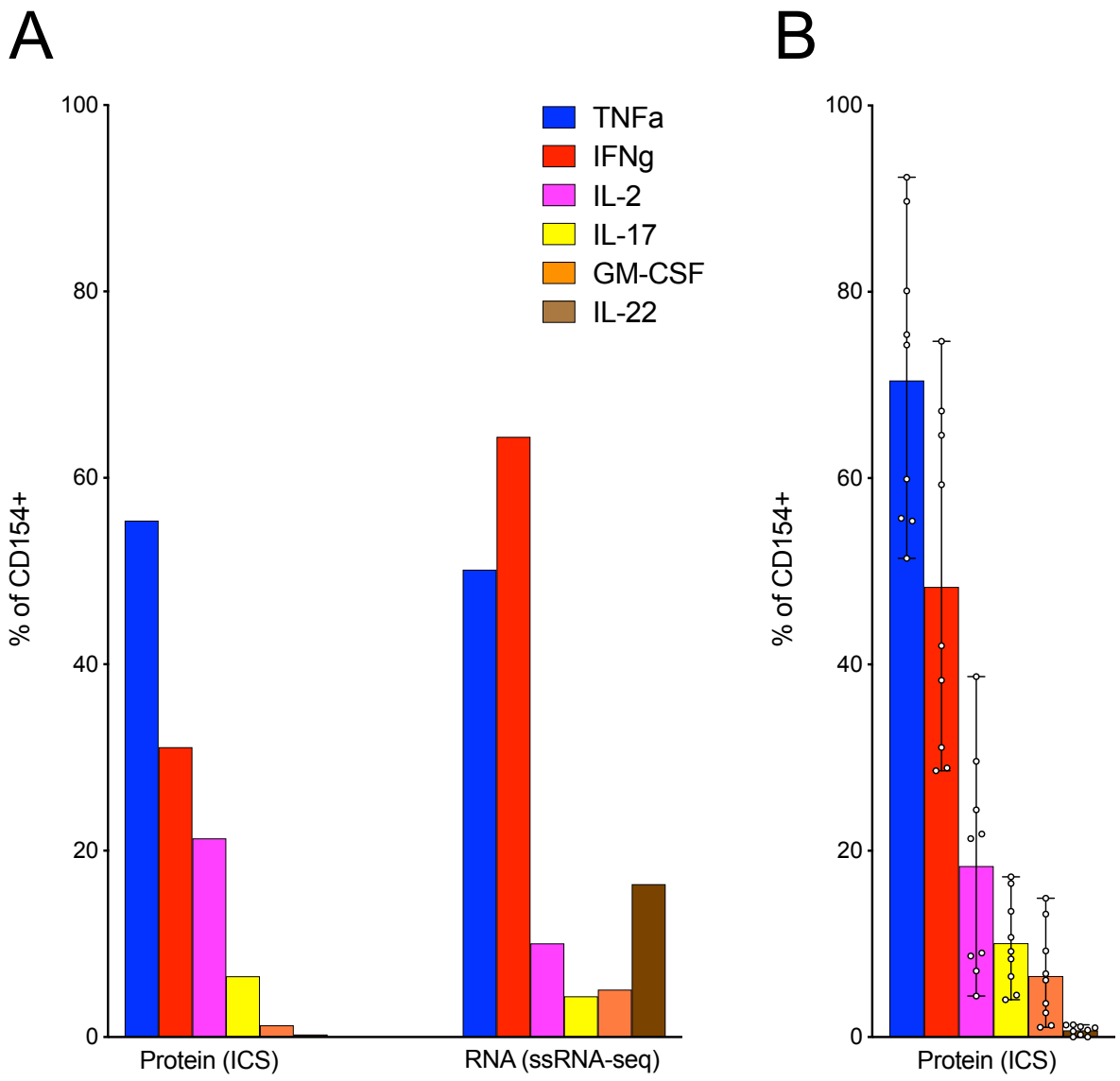


**Figure 4-13 Comparison of cytokine detection by intra-cellular staining and single cell RNA sequencing.**

(A) The percentage CD154+ Th memory cells producing TNFa (blue), IFNg (red), IL-2 (purple), IL-17A (yellow), GM-CSF (orange) or IL-22 (brown) after intracellular antibody staining (ICS) and flow cytometric analysis, compared with the percentage of CD154+ Th memory cells having detectable RNA transcripts for the same cytokines after single-cell RNA sequencing analysis (SS-RNA-seq). For both assays SFMC were stimulated overnight with *S. typhimurium* (ST). For ICS analysis of protein expression, stimulation occurred in the presence of brefeldin-A and ICS was used to detect both CD154 and cytokine expression. For SS-RNA-seq, SFMC were stimulated in the presence of a CD40 blocking antibody to allow for surface staining and sorting of CD154+ Th memory cells by FACS. Each assay was performed on separate SFMC isolated from the same patient (PSA1607) at different time points. (B) The mean percentage of CD154+ Th memory cells producing each cytokine from all SpA patients as measured by protein (ICS) assay. Each point represents a patient, bars indicate minimum and maximum range.

## 4.4 Seurat SCTransform-based analysis reveals 7 distinct clusters within synovial Salmonella-reactive Th memory cells from patient PSA1607

The filtered gene expression matrix constructed from scRNA-seq of CD154+ Th memory cells reactive to *Salmonella* (stimulated dataset) was first normalised using both the SCTransform[4] (SCT assay) function of the Seurat package and the more traditional workflow using log normalisation and scaling which aides visualisation of gene expression (RNA assay). These 2 normalised representations of the underlying gene expression matrix are stored in separate "assays" in the Seurat workflow, and can therefore be accessed independently during analysis. 75 principal components (PC) were calculated based on the SCT assay, and the degree of variation within the dataset explained by these PCs is illustrated in Figure 4-14. A shared nearest neighbour (SNN) graph was then constructed using the FindNeighbors function before using the RunUMAP function to create a 2-dimentional UMAP representation of the dataset. The Seurat FindClusters function was then run at a resolution of 0.8 to identify clusters by utilising a SNN modularity optimization based clustering algorithm[10]. The resolution parameter supplied to this algorithm can alter the number of clusters detected, with higher values indicating a higher degree of granularity is desired which can generate more clusters. The FindClusters function was run with resolutions ranging from 0.4 to 1.6 identifying between 5 and 14 clusters. A resolution of 0.8 was found to provide the most biologically meaningful distinction between clusters and was therefore chosen for downstream analysis. At this resolution, 7 clusters were identified which are illustrated by the UMAP plot in Figure 4-15.

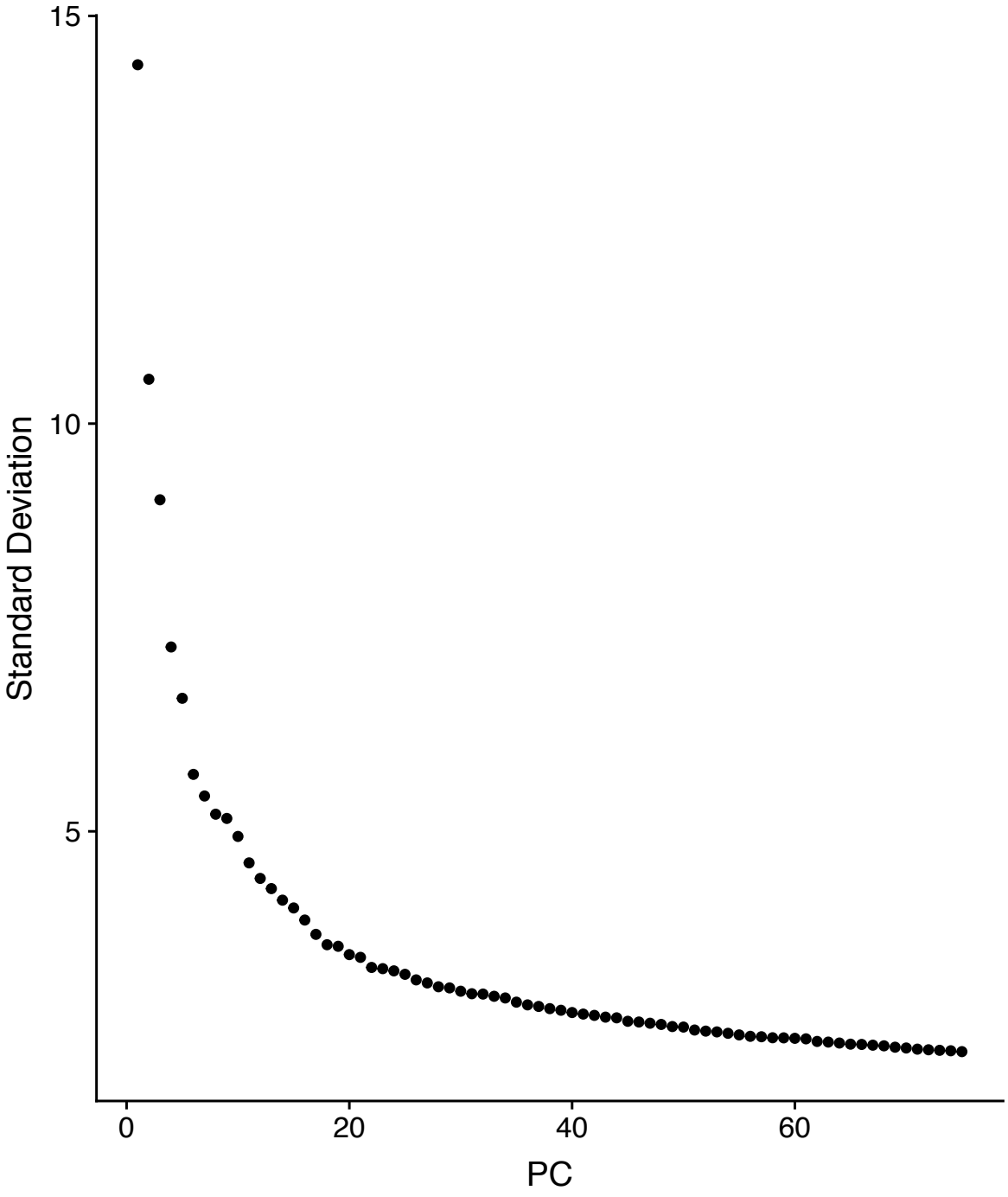


**Figure 4-14 75 principal components explain the majority of variance seen in the synovial *Salmonella*-reactive CD4 T cell dataset from patient PSA1607**

Elbow plot ranking of principal components (PC) by variance explained. Principal component analysis (PCA) was performed on the normalised transcript counts of genes in CD154+CD4+ Th memory cells stimulated with *S. typhimurium* lysate overnight. Normalisation performed using the SCTransform function in the R Seurat package on quality control filtered cells. 75 principal components were calculated.

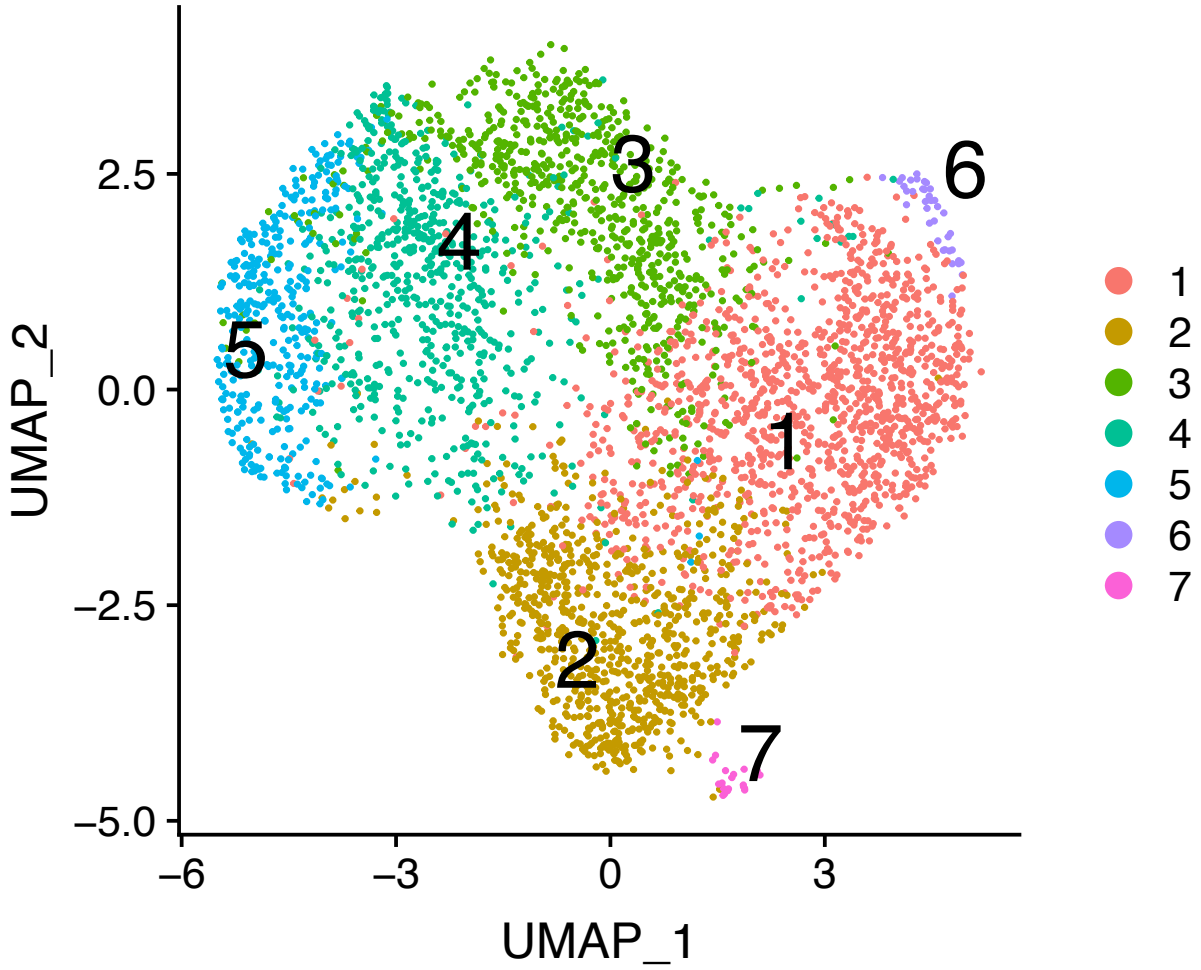


**Figure 4-15 Seurat SCTransform based UMAP reveals 7 distinct clusters within synovial *Salmonella*-reactive Th memory cells from patient PSA1607**

UMAP plot of the *Salmonella*-reactive Th memory cells from PSA patient 1607, highlighting 7 distinct clusters. Clusters were identified by the Seurat FindClusters function run at a resolution of 0.8 after performing the Seurat SCTransform workflow of normalisation, principal component calculation, SNN calculation and UMAP generation. Clusters are numbered according to the number of cells they represent from 1 (largest) to 7 (smallest) and uniquely coloured according to the legend shown.

### 4.4.1 Differential gene expression identifies distinct subsets of Th cells characterised primarily by cytokine expression

Key genes distinguishing each of the 7 clusters in Figure 4-15 were computed using the Seurat FindAllMarkers function based on the avgLogFC (average log fold change) of normalised RNA gene expression. Differentially expressed genes were identified by the Wilcoxon Rank Sum test and those genes with significant differential expression (adj. $p < 0.05$) for each cluster were then ordered by average logFC. The top 20 of these genes per cluster are shown in Figure 4-16.

4 of the 7 clusters (clusters 1,3,5 and 6) could be characterised based on cytokine expression alone, which is made more obvious in Figure 4-17 showing gene expression averaged over each cluster for key markers. Cluster 1, the largest cluster identified, overexpressed IL17F relative to all other clusters, while sharing expression of IL22 but not IFNG with cluster 3. Cluster 1 also overexpressed CCR6, known to be associated with Th17 cells, and shared expression of IL17A with clusters 5 and 6 even though IL17A expression was not statistically over-represented in any one of these clusters specifically. Thus cluster 1 most represented a TH17 subset, potentially incorporating a more specialized Th22 subset.

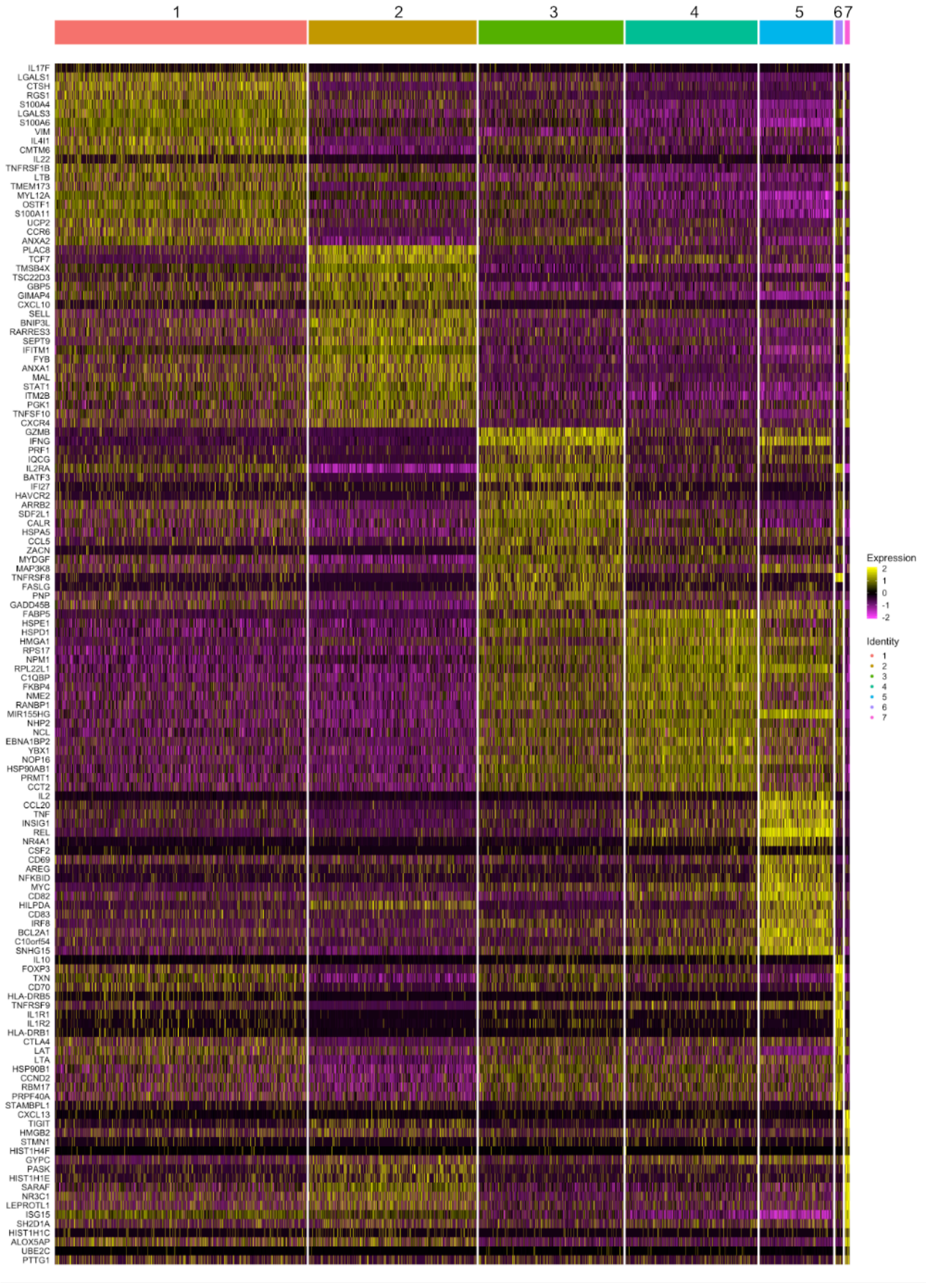


**Figure 4-16 Heatmap of Top 20 most differentially expressed genes per cluster for *Salmonella*-reactive Th memory cells from the synovial fluid of patient PSA1607.**

Top 20 most differentially expressed genes for each of the 7 clusters identified in Figure 4-15, calculated by Seurat FindAllMarkers function using Wilcoxon rank sum test. Only significantly differentially expressed genes are shown (adj. $p < 0.05$), ordered by avgLogFC in normalised gene expression from RNA assay. Gene expression is coloured by scaled, normalised RNA assay values per cell within clusters shown left to right from largest to smallest cluster.

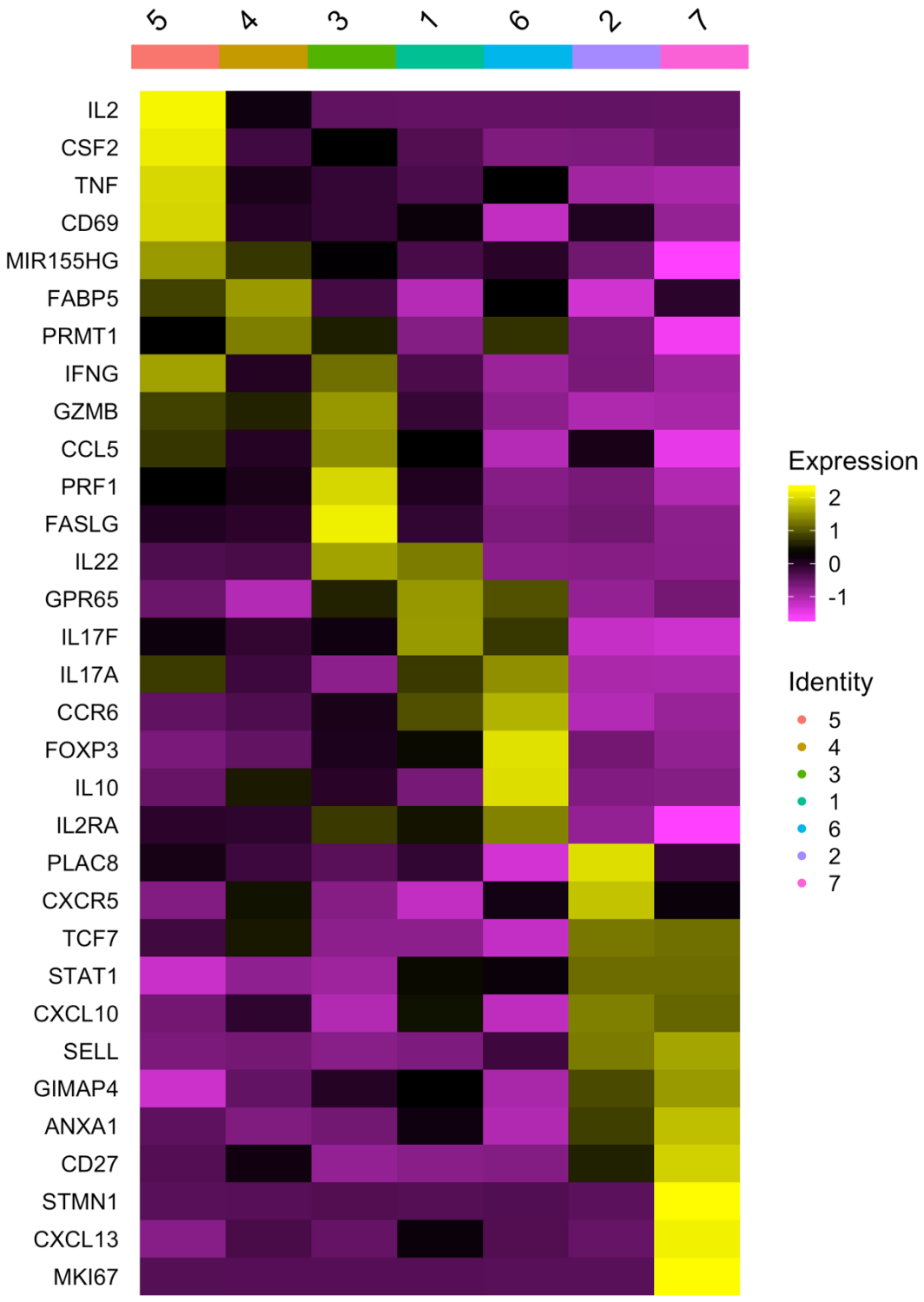


**Figure 4-17 Key cluster associated markers identify T-helper subsets in synovial *Salmonella*-reactive Th memory cells**

The average expression of key markers from the top 20 differentially expressed genes from Figure 4-16 in addition to GPR65 (cluster 1: avgLogFC = 0.51, adj.p = 2.29E-81); CD27 (cluster 2: avgLogFC=0.4 p= 9.78E-69 ; cluster 7: avgLogFC= 0.8, p=8.45E-03); CXCR5 (cluster 2: avgLogFC=0.25 p= 4.37E-32) and IL17A (not significant) are shown for each cluster. Clusters are numbered according to their size and arranged according to their spatial proximity to one another on the UMAP plot of Figure 4-15.

Additional genes overexpressed by cluster 1 included RGS1 (regulator of G-protein signalling 1) which plays a role in chemokine signalling, has been associated with multiple auto-immune diseases, and more recently with Th17 cells isolated from the central nervous system of diseased experimental autoimmune encephalomyelitis (EAE) mice[11]; CMTM6, which encodes a protein capable of stabilizing PD-L1 at the cell surface[12]; OSTF1 (osteoclast-stimulating factor 1) better known for its expression in osteoclasts where it indirectly induces osteoclast formation and bone resorption[13] ; TMEM173 (transmembrane protein 173), also known as STING (Stimulator of interferon genes) capable of inducing type I interferon responses; and CTSH (Cathepsin H) which encodes a proteinase involved in lysosomal degradation of proteins.

Cluster 3 could be distinguished by FASLG (Fas Ligand) gene expression and by the combined over-expression of both IFNG and IL22. In addition GZMB (granzyme B) and PRF1 (perforin) were most highly expressed by this cluster, associating it with a more CD4+ CTL (cytotoxic lymphocyte) phenotype[14].

Cluster 5 strongly overexpressed IL2 relative to all other clusters, and also had the highest overall expression of CSF2, TNF, and CD69. Cluster 5 also shared overlap with cluster 4 in the expression of MIR155HG ( a micro RNA (miRNA) known to be a highly upregulated upon TCR stimulation[15]) and FABP5 (Fatty Acid Binding Protein 5); and with cluster 3 in the expression of IFNG. Clearly cluster 5 represented a highly activated cluster producing a large number of pro-inflammatory cytokines. Clusters 5 and 4 also associated most strongly with clonally expanded cells, including the largest clone by cell frequency present in this dataset (Figure 4-18), which was also the most strongly enriched clone relative to unstimulated synovial Th memory cells from the same patient.

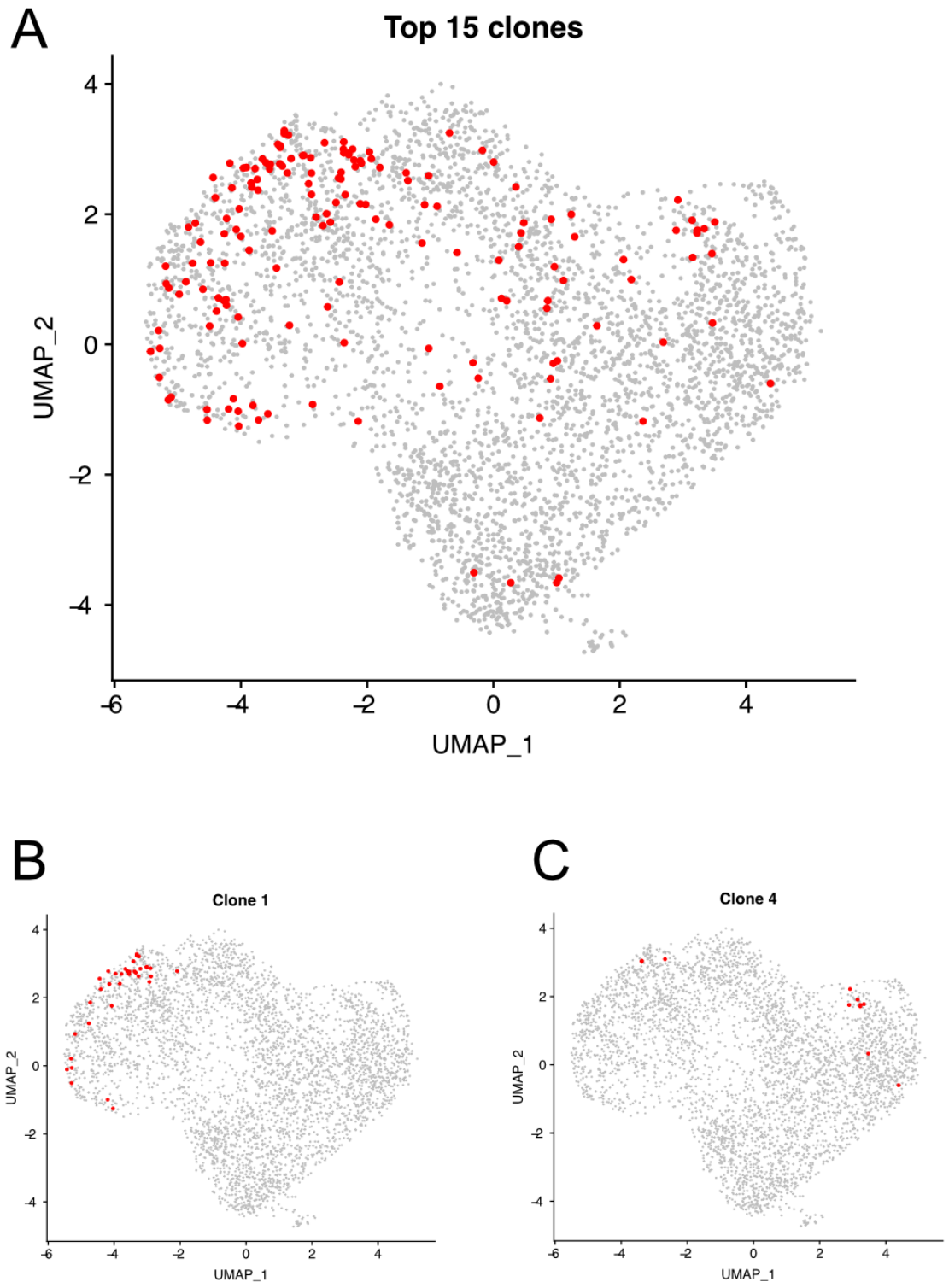


**Figure 4-18 UMAP plots highlighting the cell distribution of the top 15 most abundant clonotypes present in synovial *Salmonella*-reactive Th memory cells**

A: The cumulative representation of all cells belonging to the top 15 clones by cell frequency. B: Cells belonging to the most abundant clone, including 18, 15, 1 and 1 cells from clusters 5, 4, 3 and 2 respectively. C: 4th most abundant clone with 7, 3 and 1 cells from clusters 4, 1 and 3 respectively. Cells belonging to clones are highlighted in red with an increased dot size for better visualization and overlaid onto the UMAP plot from Figure 4-15. Each dot represents a cell.

Although consisting of very few cells, cluster 6 clearly overexpressed IL10 in addition to FoxP3 and CTLA4, identifying this cluster with a T regulatory (Treg) cell phenotype. Cluster 6 most closely associated with TH17 cluster 1, both in terms of its UMAP proximity to cluster 1 (Figure 4-15) and in its gene expression profile, including a tendency for cells within this cluster to express IL17A.

Clusters 2, 4 and 7 expressed very little of any cytokine. Cluster 7 could be distinguished by expression of STMN1, a gene upregulated in cycling cells, and although not found to be statistically significant in this small cluster of cells, also contained cells with high gene expression of MKI67, another marker of cycling cells.

Cluster 4 most strongly overexpressed FABP5 which is also known as Psoriasis-Associated Fatty Acid-Binding Protein Homolog and encodes a protein highly abundant in psoriatic skin[16]. FABP5 is also thought to play a role in CD8 T cell tissue residency[17]. In addition, increased gene expression in cluster 4 was seen for PRMT1 and MIR155HG (shared with cluster 5), both of which can play a role in Th differentiation in addition to T cell activation [18][19][20,21], and the heatshock protein genes HSPE1, HSPD1 and HSP90AB1. Many of the top 20 most differentially expressed genes in cluster 4 were also upregulated in clusters 3, 5 and 6 relative to clusters 1, 2 and 7 (Figure 4-16).

Cluster 2 was clearly and uniquely identifiable by increased gene expression of PLAC8. Among the top 20 genes distinguishing clusters in Figure 4-16, Cluster 2 bears little resemblance to any other cluster with the exception of cycling cell cluster 7, next to which it is situated on the UMAP plot of Figure 4-15. Clusters 7 and 2 share upregulation of TCF7 and CD27, key markers which have very recently been

associated with a "stem-like" phenotype of memory Th17 cells poised for differentiation into pathogenic Th17 cells[22]. CXCR5, a marker usually associated with T follicular helper cells, was also upregulated in these stem-like cells and is similarly upregulated in cluster 2. Both clusters 2 and 7 also overexpressed several genes associated with glucocorticoid signalling, including: NR3C1 encoding the glucocorticoid receptor; TSC22D3 (encoding the GILZ glucocorticoid (GC)-induced leucine zipper) which is upregulated by glucocorticoids and can help regulate TCR activation[23]; and ANXA1, encoding Annexin A1, also upregulated by glucocorticoids, and also having immunosuppressive effects[24]. The tissue homing related genes most upregulated in both clusters 2 and 7 are: CXCL10, a chemokine ligand for CXCR3[25]; SELL, which encodes CD62L/L-selectin, a homing receptor promoting entry into lymphoid tissues via high endothelial venules; and CXCR4 a chemokine receptor for CXCL12. Interestingly, GIMAP4, which has a genetic association with the related "MHC-I opathy" Behçet 's disease[26], is also upregulated in both clusters. The relative expression of all markers from Figure 4-17 can be seen in Figure 4-19.

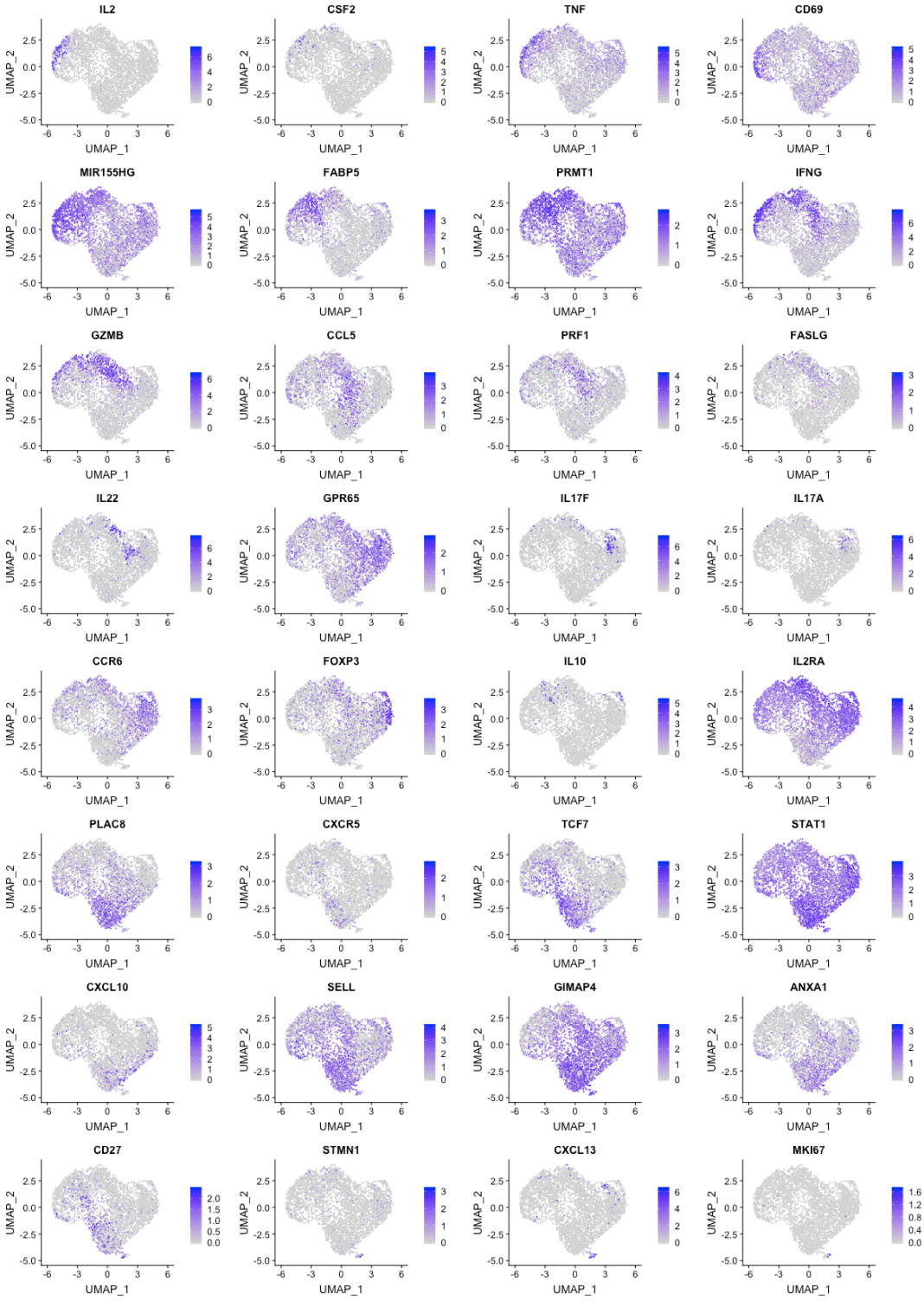


**Figure 4-19 Relative gene expression of key markers associated with *Salmonella*-reactive Th memory subsets identified from PSA synovial fluid.**

UMAP plots showing the relative gene expression of each of the genes from Figure 4-17 among *Salmonella*-reactive Th memory cells. Transcript frequency is based on log normalised counts from the RNA assay, normalised using a scale factor of 10000 by the Seurat "NormalizeData" function. Increased intensity of purple colour represents higher normalised transcript frequency. Each dot represents a cell.

### 4.4.2 Pseudotime analysis of *Salmonella*-reactive Th memory cells reveals 2 trajectories sharing a common progression from Th1 to Th17 memory subtypes before diverging into either regulatory or cycling cells

To better understand the relationship between clusters and to identify potential paths of differentiation between them, a pseudotime analysis was conducted using the Slingshot package[27]. This analysis was based on the 3000 most variable genes identified by the Seurat SCTransform function, further filtered to only include genes where at least 10 cells expressed greater than or equal to 3 gene transcripts. The slingshot function was then used to discover lineages in an unsupervised manner, before plotting these lineages back onto a UMAP dimensional reduction for visualisation (Figure 4-20). Two lineages were identified, both sharing a common path which traversed the Th1, FABP5+ and Th CTL clusters, before bifurcating at the Th17 cluster to terminate at the T regulatory IL10+ cluster, or continue through the PLAC8+ cluster towards cycling cells.

For systems where there is a smooth transition from one phenotypic state to another, pseudotime analysis can help to identify this transition. In the case of single cell analysis, this is achieved by using individual cells to represent points along a continuum or "lineage"[27]. Pseudotime is the name given to the underlying temporal variable which drives the progress between states and is calculated based on the gradual transcriptional changes observed between cells[27]. The 2 lineages identified in Figure 4-20 therefore represent the trajectories along which changes between states (clusters) is most likely to occur.

From Figure 4-17 it can be seen that cluster 5, although being most distinguished by the upregulation of IL2, TNF and CSF2, also contains cells expressing IFNG, GZMB and IL17A. One possible interpretation of this pseudotime analysis is that upon stimulation

with *Salmonella* lysate, Th memory cells reactive to this antigen first upregulate a wide spectrum of pro-inflammatory cytokines (cluster 5), before refining their effector programs to the more specialized end states seen in other clusters, perhaps based on a pre-programmed cell fate. It is also possible that the pseudotime analysis did not identify a temporal change which occurred during stimulation, but rather the "ghost" of such transitions which occurred previously. From a similar point of view, these trajectories may represent the path of least resistance in transitioning from one cell state to another, and reflect the potential for plasticity between cell phenotypes.

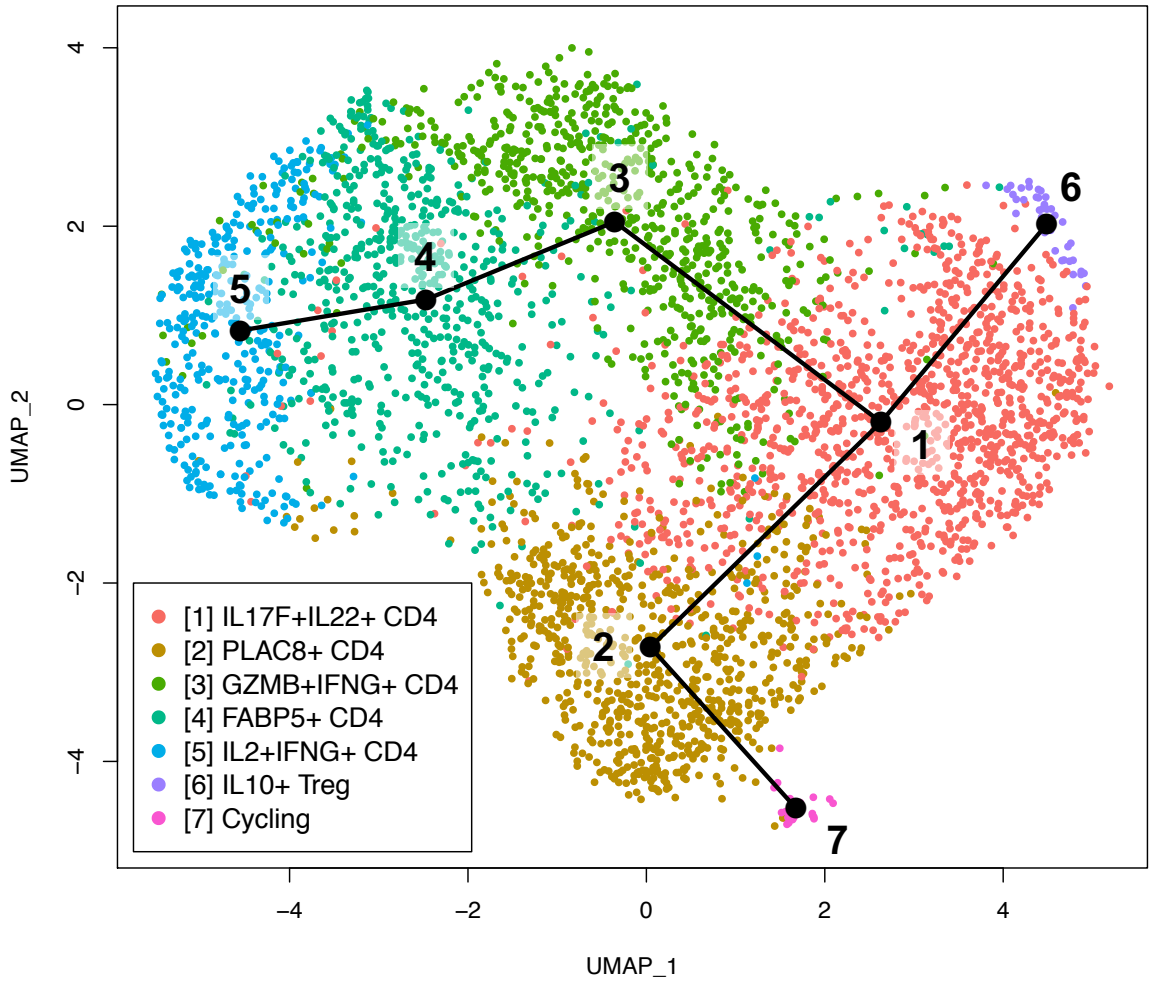


**Figure 4-20 Pseudotime analysis of *Salmonella*-reactive Th memory cells reveals 2 trajectories sharing a common progression from IL2+Th1 to IL17F+Th17 memory subtypes, eventually diverging into either regulatory or cycling cells.**

Lineage trajectories calculated by the Slingshot R package are depicted on the previously created UMAP 2 dimensional reduction of the dataset (Figure 4-15). To perform the analysis, a SingleCellExperiment object was first created from the SCTransformed assay of the dataset. This was subsetted to include only highly variable genes classified as being robustly expressed, which was then passed to the slingshot function to calculate lineages. Only the top 3000 genes identified by the Seurat SCTransform function as being able to explain the most variation between cells and filtered to only those genes represented by more than 3 transcripts in at least 10 cells within the dataset were included in the analysis.

### 4.4.3 IL17A+ *Salmonella* reactive Th memory cells can be further classified as IFNG+, IL17F+ or PPP1R3B+

In the previous chapter, IL-17A+, IL17A+IFNG+ and IL17A+IFNG+IL2+ subsets of *Salmonella*-reactive Th memory cells were all found to be enriched in the synovial fluid of SpA patients relative to their peripheral blood. Differential gene expression analysis of the scRNA-seq dataset did not attribute statistical significance to the overexpression of IL17A in any one of the 7 clusters analysed. However, 4.2% of the entire population of *Salmonella*-reactive Th memory cells did express IL17A (Table 4-3, Figure 4-13), predominantly in clusters 1, 6 and 5 (Figure 4-17)

To specifically characterise cells expressing IL17A, these cells were subsetted from the main dataset and named the "IL17A-subset". The workflow of calculating principal components, constructing an SNN graph, identifying clusters and creating a UMAP representation of the IL17A-subset was repeated. An increased resolution of 1 was provided as argument to the Seurat FindClusters function to obtain a higher degree of distinction between cells.

3 clusters were identified within the IL17A-subset and are depicted in the UMAP representation shown in Figure 4-21A. Differential gene expression performed on these clusters (Figure 4-22) was used to classify them as IL17F+, IFNG+ or PPP1R3B+ Th17 subsets. The distribution of these cells within the main UMAP of all synovial *Salmonella*-stimulated cells from patient PSA1607 (Figure 4-15) is shown in Figure 4-23.

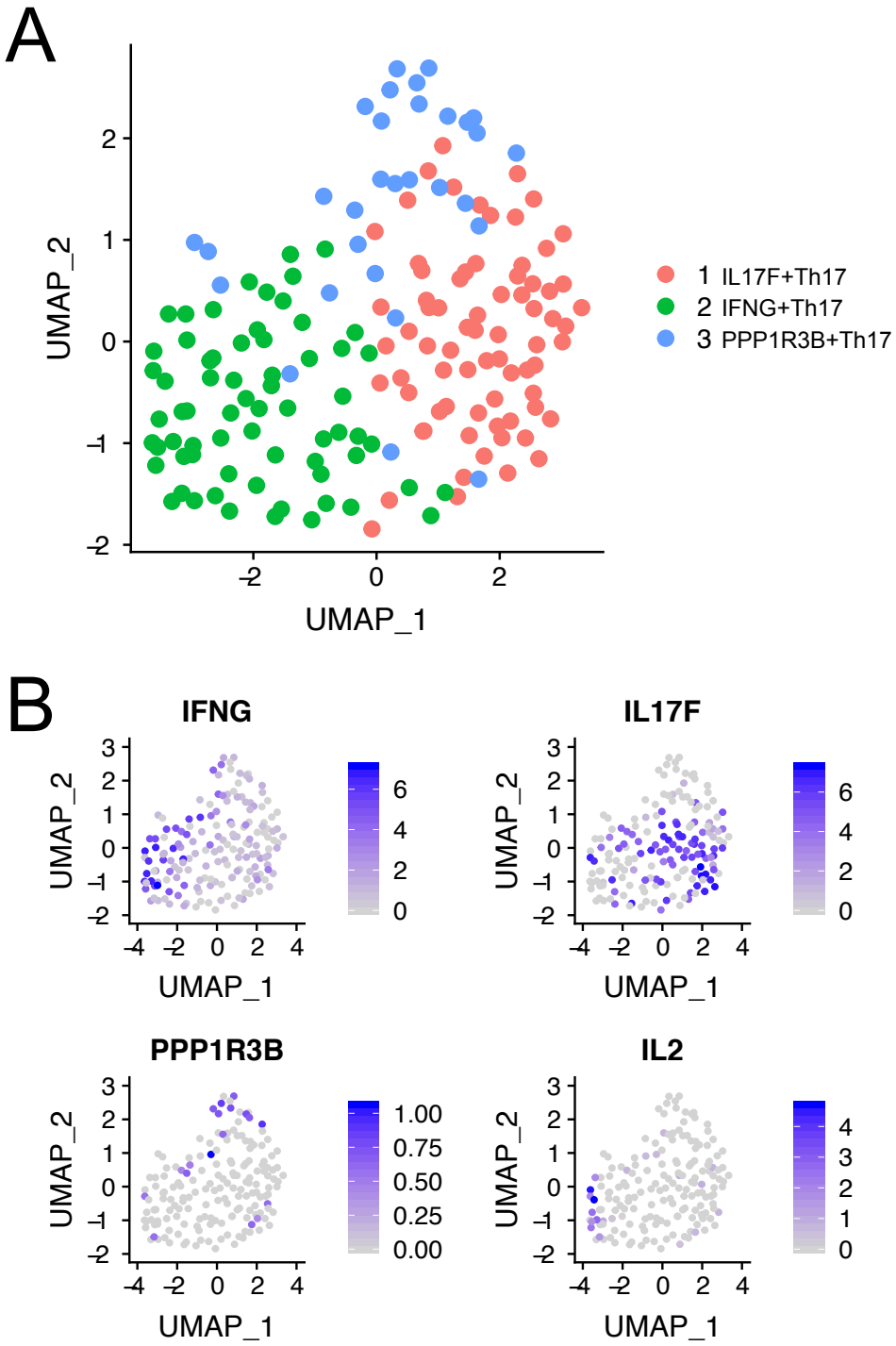


**Figure 4-21 IL17A+ *Salmonella*-reactive Th memory cells can be further classified as IFNG+, IL17F+ or PPP1R3B +**

A: UMAP plot of all synovia *Salmonella*-reactive Th memory cells from PSA patient 1607 where any gene expression of IL17A was detected. 3 clusters were identified by the Seurat FindClusters function run at a resolution of 1.0 after performing the Seurat SCTransform workflow of normalisation, principal component calculation, SNN calculation and UMAP generation. Clusters are numbered according to the number of cells they represent from 1 (largest) to 3 (smallest) and uniquely coloured according to the legend shown. 170 cells total: cluster 1 has 73 cells; cluster 2 has 67 cells; and cluster 3 has 30 cells. Each B: UMAP from (A) with relative gene expression for key cluster defining genes. Transcript frequency is based on log normalised counts from the RNA assay, normalised using a scale factor of 10000 by the Seurat "NormalizeData" function. Increased intensity of purple colour represents higher normalised transcript frequency. Each dot represents a cell.

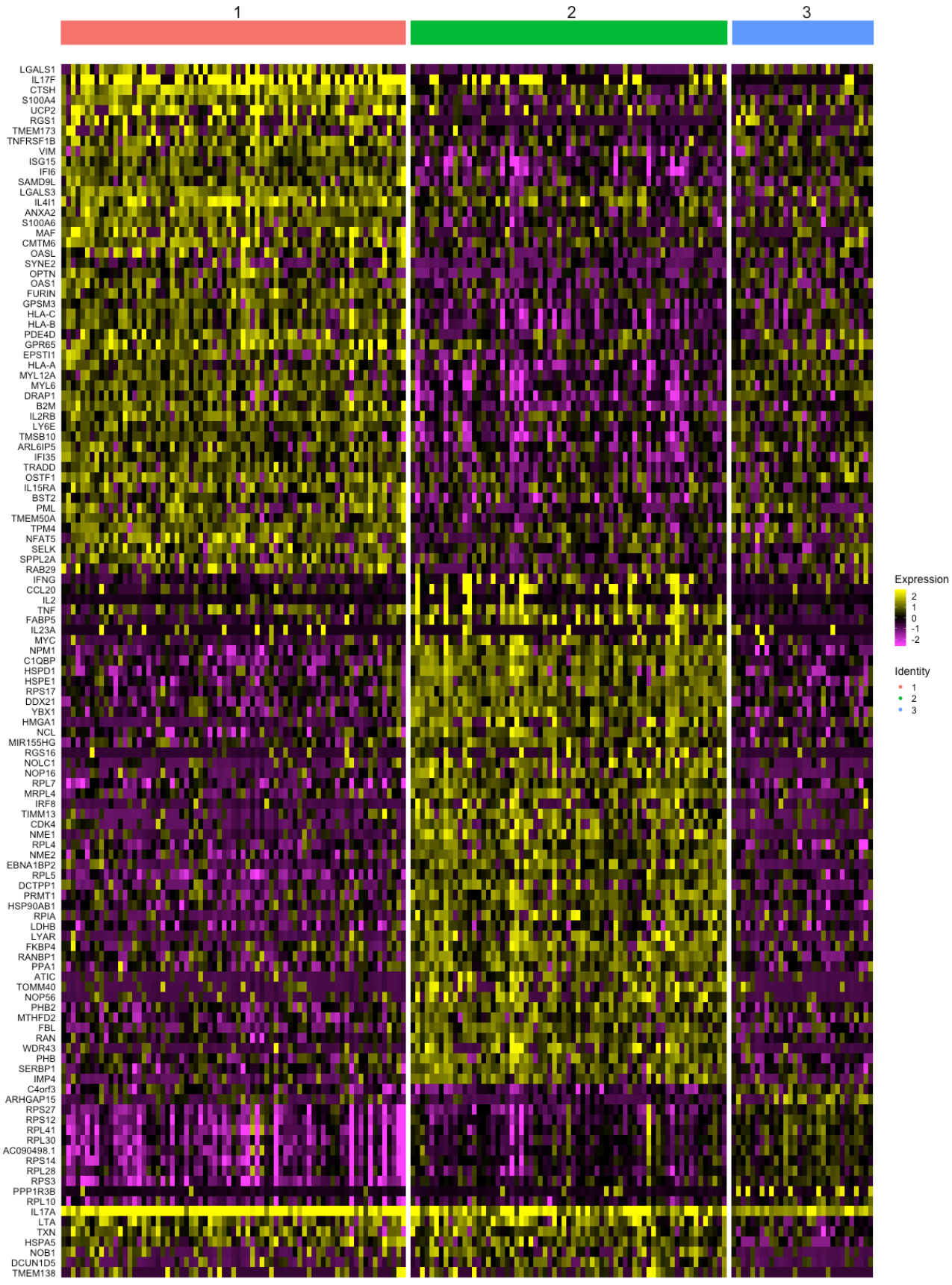


**Figure 4-22 Heatmap of Top 50 most differentially expressed genes per cluster for all *Salmonella*-reactive Th memory cells expressing IL17A from the synovial fluid of PSA patient 1607.**

Top 50 most differentially expressed genes for each of the 3 clusters identified in Figure 4-21, calculated by the Seurat FindAllMarkers function using bimod[9] test. Only significantly differentially expressed genes are shown (adj. $p < 0.05$), ordered by avgLogFC within clusters shown left to right from largest to smallest cluster.

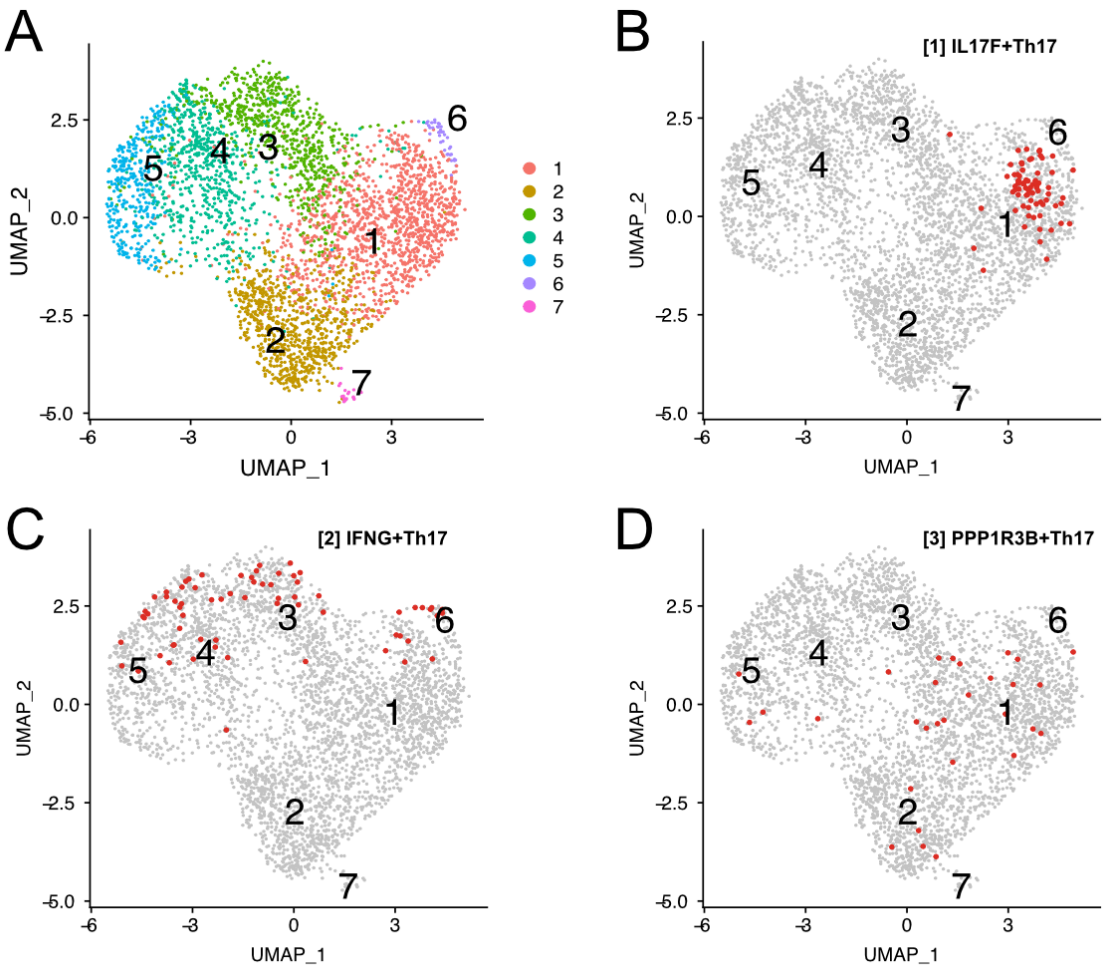


**Figure 4-23 Location of cells from the IL17A-subset on main UMAP plot**

A: UMAP plot of all synoviaL *Salmonella*-reactive Th memory cells from PSA patient 1607 provided for reference from Figure 4-15. B-D: Cells from clusters 1-3 respectively of the IL17A-subset highlighted red in a larger dot size for visualization. Each cell is represented by a dot. Relates to Figure 4-21 and Figure 4-22.

### 4.4.4 scRNA-seq of *Salmonella*-reactive synovial CD154+Th memory cells reveals distinct gene expression profiles for IFNG+ and IFNG- Th17 cells.

The Th17 sub-clusters identified within the IL17A-subset of cells each had distinct gene expression signatures. Gene expression of IFNG could be used to distinguish between the 2 largest clusters within this subset. IL2 expression was largely restricted to IFNG+Th17 cells which also expressed considerably more TNF and CCL20, while IL17F expression clustered more tightly with IFNG-Th17 cells. Even where cells expressed some IL17F within the IFNG+Th17 cluster, those cells typically expressed lower IFNG.

More cells from the IFNG+Th17 cluster also expressed: FABP5, MIR155HG and PRMT1, likening them to cluster 4 (see Figure 4-17) from the complete stimulation dataset (Figure 4-15). Other overexpressed genes of interest in the IFNG+Th17 sub-cluster included IRF8, MYC, NPM1, PHB and PHB2.

Thus the IFNG+Th17 cluster could be most likened to clusters 5 and 4 from the main analysis of all *Salmonella*-stimulated cells. However the distribution of cells from this cluster across the main UMAP plot (illustrated in Figure 4-23) suggests a continuum following one of the 2 trajectories identified by pseudotime analysis which terminated in Treg cluster 6 (Figure 4-20). Conversely, the PPP1R3B+Th17 cluster, which had largely downregulated expression of genes associated with either of the other 2 clusters from the IL17A-subset, appeared to follow a continuum representing the second pseudotime trajectory terminating in cycling cells. The IFNG-IL17F+Th17 subset on the other hand did not follow either trajectory and clustered very tightly with the largest Th17 cluster (cluster 1) from the main analysis of all *Salmonella*-stimulated cells. Unsurprisingly then, most of the genes overexpressed in the main Th17 cluster for the

total stimulated dataset (Figure 4-16: cluster 1) were also overexpressed in the IFNG-IL17F+Th17 cluster relative to the other 2 clusters within the IL17A-subset.

## 4.5 Discussion

The aim of this chapter was to use scRNA-seq to characterise *Salmonella*-reactive CD154+ Th memory cells from the synovial fluid of a PSA patient (PSA1607) in the context of unstimulated synovial *ex vivo* memory T cell populations from the same donor. The technique used to isolate Th memory cells by surface rather than intra-cellular expression of CD154 captured a similar frequency of CD154+ Th memory cells compared to previous ICS assays from chapter 3 (Figure 4-1). Initial scRNA-seq analysis also determined that a similar frequency of these CD154+ cells expressed genes encoding each of the cytokines assayed by ICS in chapter 3 (Figure 4-13, Table 4-3). A notable exception to this was IL22. There may be several reasons why IL22 gene expression was detected by RNA sequencing in a higher frequency of CD154+ Th memory cells compared with the detection of IL-22 protein by ICS. Firstly, a direct comparison of gene and protein expression of IL-22 for the same sample was not possible, and it may be that an ICS assay of the same sample would have yielded a more comparable level of protein expression. This seems unlikely however given that all 9 synovial patient samples measured by ICS after stimulation with *Salmonella* lysate shared a low frequency of IL22+ cells within the CD154+ Th memory compartment. It is also possible that antibody staining with an IL-22 antibody conjugated to a brighter fluorophore would be more sensitive in detecting IL-22, or that it takes longer for IL-22 protein to be produced following activation. However detection of IL-22 in the previous chapter varied depending on the microbial lysate used to stimulate cells, arguing against failure of the antibody to detect IL-22 and making a time delay less likely. For example, when assaying the ReA *Chlamydia* stimulated SFMC sample by ICS (Figure 3-24), 8.25% of CD154+ Th memory cells also expressed IL-22, a figure much closer (but still considerably lower) compared to the 16.4% detected by scRNA-

seq. Alternative explanations include IL22 RNA expression not being directly representative of IL-22 expression at the protein level due to suppression of translation by an unknown factor(s), or to a binding protein[28] blocking epitope exposure to the anti-IL-22 antibody used in the ICS assay. Despite the difference between protein and scRNA-seq quantification of IL22, the remaining cytokines assayed by scRNA-seq did fall within the range predicted by previous ICS protein staining, indicating that the cells sequenced by scRNA-seq were representative of the population of *Salmonella*-reactive Th memory cells assayed in chapter 3.

The high frequency of *ex vivo* unstimulated synovial CD4 T cells expressing CD40LG by scRNA-seq (16%) was striking. If these cells represent a population of activated CD4 T cells in the synovial fluid of PsA patients then their isolation and further characterisation could be very insightful. Further characterisation of these cells from the dataset in this thesis is ongoing. One possible explanation is that a degree of CD40LG expression may normally be present in CD4 T cells, which is only translated or expressed at the protein level upon proper TCR engagement. Assessing the clonality of unstimulated cells expressing CD40LG (Table 4-1) was encouraging but not conclusive, partly owing to the relatively small number of cells assayed by scRNA-seq and the diverse CD4 T cell TCR repertoire present in both synovial fluid and blood.

The finding that a small percentage of *ex vivo* memory T cells also expressed TNF and IFNG by scRNA-seq was also encouraging, however there was not a striking overlap between clonality or CD40LG expression and cytokine expression among unstimulated T cells. It is therefore possible that the small percentage of unstimulated T cells expressing these cytokines are doing so as by-standers of the pro-inflammatory environment present in the inflamed joint rather than responding directly to TCR engagement.

The in-depth (read depth of 90070 reads/cell) 10x analysis provided meaningful characterisation of both the CD154+Th response to *Salmonella* stimulation as a whole and specifically the CD154+ IL17A+ / IL17A+IFNG+ / IL17A+IFNG+IL2+ responses towards *Salmonella* found to be enriched in the synovial fluid of SpA patients in chapter 3. The CD154 response as a whole could be divided into 7 distinct groups of cells, characterised primarily by cytokine production, but also by less well known markers, such as PLAC8 and markers suggestive of a CTL phenotype. PLAC8 is not commonly associated with Th cells, and was first characterised for its presence in the spongiotrophoblast placental layer of mice[29]. The protein encoded by PLAC8, also known as ONZIN, is highly conserved in vetebrates and upregulated in conserved metabolic pathways, but its function seems to vary depending on organism and cell type[30]. One recent scRNA-seq study did however find a strong association with PLAC8 and airway Th2 cells involved in the immune response towards house dust mites[31]. In a separate series of investigations by Johnston et al.[32,33], PLAC8 was associated with a subtype of CD4+ T cells capable of clearing *Chlamydia* infection in an inducible nitric oxide synthase (iNOS) independent manner. *Chlamydia trachomatis*, like *Salmonella typhimurium*, is an intracellular pathogen which can instigate reactive arthritis. Interestingly, chapter 3 showed a great deal of overlap between immune responses towards Salmonella and Chlamydia, suggesting that the PLAC8 cell subset of *Salmonella*-reactive Th cells identified in this thesis may play an important role in immune defence against both of these pathogens. The PLAC8 cluster also uniquely overexpressed CXCR5, and in common with the cycling cell cluster TCF7, SELL, and to a lesser extent CD27, whereas cytokine effector expression was switched off.

Another interesting finding is that the same clones could be identified in multiple phenotypic clusters (Figure 4-18). For example, cells belonging to the largest and 4th largest clones detected among *Salmonella*-stimulated cells were present in clusters 5, 4, 3 and 2; and 4, 1 and 3 respectively. It was also interesting that although these clones appeared to cluster quite tightly based on UMAP dimensional reduction, the Seurat FindClusters function (which utilizes the Louvain[34] algorithm by default to identify communities), placed many of these cells in clusters quite distal to the location of the majority of cells for the same cluster on the UMAP plot. I.e. cells from clone 4 appear to be primarily located in cluster 1 on the UMAP projection, but the majority of cells for this clone are actually from cluster 4. This implies a relationship between these clusters exists. It was difficult to phenotypically classify cluster 4 by its gene expression profile and this cluster may represent an intermediary state between switching from one phenotype to another, possibly regulated by PRMT1 via its effect on STAT3/STAT5[18][35], and MIR155HG (Figure 4-17). The concept of Th plasticity (including Th17 plasticity and even Th17 plasticity specifically in synovial fluid of juvenile arthritis patients[36]) is not new, and the ability of the same clone to take on multiple phenotypes was explored by Becattini et al., who also proposed a model whereby polarized T cell responses resulted from the selective expansion of existing clones and that this selection process was dependent on the microenvironment [37][38].

Taking all of these findings into consideration, it seems plausible that a degree of plasticity exists between the highly activated and pro-inflammatory Th1 subset (cluster 5) and the Th17 subset (cluster1). Furthermore, given the bifurcation identified in pseudotime analysis, it also seems plausible that a subset of Th17 cluster 1 cells can then progress to a more Treg-like phenotype, and conversely that another subset of cells

can switch off their cytokine program and instead upregulate a proliferative program, adding to a CXCR5+PLAC8+ stem-like pool. Interestingly CXCR5+ Th memory cells have previously been associated with decreased cytokine expression attributable to hypoacetylation[39]. The increased expression of SELL (CD62L) also suggests that these PLAC8+ cells have the capacity to traffic towards lymph nodes and take on a more central memory phenotype.

While assessing the clonality of unstimulated memory T cells, a striking level of (paired alpha and beta chain) enrichment for specific CD8 clones in synovial fluid was observed. The following chapter is largely dedicated to exploring and validating this finding with additional patient samples processed using different single-cell sequencing platforms.

# I. References


1. Chen, J., Cheung, F., Shi, R., Zhou, H., and Lu, W. (2018). PBMC fixation and processing for Chromium single-cell RNA sequencing. J Transl Med *16*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6050658/ [Accessed September 25, 2019].

2. Gerlach, J.P., Buggenum, J.A.G. van, Tanis, S.E.J., Hogeweg, M., Heuts, B.M.H., Muraro, M.J., Elze, L., Rivello, F., Rakszewska, A., Oudenaarden, A. van, *et al.* (2019). Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. Sci Rep *9*, 1–10.

3. Frentsch, M., Arbach, O., Kirchhoff, D., Moewes, B., Worm, M., Rothe, M., Scheffold, A., and Thiel, A. (2005). Direct access to $CD4^+$ T cells specific for defined antigens according to CD154 expression. Nature Medicine *11*, 1118.

4. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. bioRxiv, 576827.

5. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 [cs, stat]. Available at: http://arxiv.org/abs/1802.03426 [Accessed September 28, 2019].

6. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnology *37*, 38–44.

7. Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579–2605.

8. McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467.

9. McDavid, A., Dennis, L., Danaher, P., Finak, G., Krouse, M., Wang, A., Webster, P., Beechem, J., and Gottardo, R. (2014). Modeling Bi-modality Improves Characterization of Cell Cycle on Gene Expression in Single Cells. PLOS Computational Biology *10*, e1003696.

10. Waltman, L., and Eck, N.J. van (2013). A smart local moving algorithm for large-scale modularity-based community detection. The European Physical Journal B *86*, 1–14.

11. Hoppmann, N., Graetz, C., Paterka, M., Poisa-Beiro, L., Larochelle, C., Hasan, M., Lill, C.M., Zipp, F., and Siffrin, V. (2015). New candidates for CD4 T cell pathogenicity in experimental neuroinflammation and multiple sclerosis. Brain *138*, 902–917.

12. Mezzadra, R., Sun, C., Jae, L.T., Gomez-Eerland, R., de Vries, E., Wu, W., Logtenberg, M.E.W., Slagter, M., Rozeman, E.A., Hofland, I., *et al.* (2017). Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. Nature *549*, 106–110.

13. Reddy, S., Devlin, R., Menaa, C., Nishimura, R., Choi, S.J., Dallas, M., Yoneda, T., and Roodman, G.D. (1998). Isolation and characterization of a cDNA clone encoding a novel peptide (OSF) that enhances osteoclast formation and bone resorption. J. Cell. Physiol. *177*, 636–645.

14. Patil, V.S., Madrigal, A., Schmiedel, B.J., Clarke, J., O'Rourke, P., Silva, A.D. de, Harris, E., Peters, B., Seumois, G., Weiskopf, D., *et al.* (2018). Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. Science Immunology *3*, eaan8664.

15. Baumjohann, D., and Ansel, K.M. (2013). MicroRNA-mediated regulation of T helper cell differentiation and plasticity. Nature Reviews Immunology *13*, 666–678.

16. Siegenthaler, G., Hotz, R., Chatellard-Gruaz, D., Didierjean, L., Hellman, U., and Saurat, J.H. (1994). Purification and characterization of the human epidermal fatty acid-binding protein: localization during epidermal cell differentiation in vivo and in vitro. Biochem J *302*, 363–371.

17. Pan, Y., Tian, T., Park, C.O., Lofftus, S.Y., Mei, S., Liu, X., Luo, C., O'Malley, J.T., Gehad, A., Teague, J.E., *et al.* (2017). Survival of tissue-resident memory T cells requires exogenous lipid uptake and metabolism. Nature *543*, 252–256.

18. Sen, S., He, Z., Ghosh, S., Dery, K.J., Yang, L., Zhang, J., and Sun, Z. (2018). PRMT1 Plays a Critical Role in Th17 Differentiation by Regulating Reciprocal Recruitment of STAT3 and STAT5. The Journal of Immunology *201*, 440–450.

19. Mycko, M.P., Cichalewska, M., Cwiklinska, H., and Selmaj, K.W. (2015). miR-155-3p Drives the Development of Autoimmune Demyelination by Regulation of Heat Shock Protein 40. J. Neurosci. *35*, 16504–16515.

20. Parry, R.V., and Ward, S.G. (2010). Protein arginine methylation: a new handle on T lymphocytes? Trends in Immunology *31*, 164–169.

21. Geoghegan, V., Guo, A., Trudgian, D., Thomas, B., and Acuto, O. (2015). Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. Nat Commun *6*, 1–8.

22. Karmaus, P.W.F., Chen, X., Lim, S.A., Herrada, A.A., Nguyen, T.-L.M., Xu, B., Dhungana, Y., Rankin, S., Chen, W., Rosencrance, C., *et al.* (2019). Metabolic heterogeneity underlies reciprocal fates of T H 17 cell stemness and plasticity. Nature *565*, 101.

23. Ayroldi, E., Migliorati, G., Bruscoli, S., Marchetti, C., Zollo, O., Cannarile, L., D'Adamio, F., and Riccardi, C. (2001). Modulation of T-cell activation by the glucocorticoid-induced leucine zipper factor via inhibition of nuclear factor κB. Blood *98*, 743–753.

24. Perretti, M., and D'Acquisto, F. (2009). Annexin A1 and glucocorticoids as effectors of the resolution of inflammation. Nat Rev Immunol *9*, 62–70.

25. Booth, V., Keizer, D.W., Kamphuis, M.B., Clark-Lewis, I., and Sykes, B.D. (2002). The CXCR3 Binding Chemokine IP-10/CXCL10: Structure and Receptor Interactions. Biochemistry *41*, 10418–10425.

26. Lee, Y.J., Horie, Y., Wallace, G.R., Choi, Y.S., Park, J.A., Choi, J.Y., Song, R., Kang, Y.-M., Kang, S.W., Baek, H.J., *et al.* (2013). Genome-wide association study identifies GIMAP as a novel susceptibility locus for Behçet's disease. Annals of the Rheumatic Diseases *72*, 1510–1516.

27. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477.

28. Lim, C., Hong, M., and Savan, R. (2016). Human IL-22 binding protein isoforms act as a rheostat for IL-22 signaling. Sci. Signal. *9*, ra95–ra95.

29. Galaviz-Hernandez, C., Stagg, C., de Ridder, G., Tanaka, T.S., Ko, M.S.H., Schlessinger, D., and Nagaraja, R. (2003). Plac8 and Plac9, novel placental-enriched genes identified through microarray analysis. Gene *309*, 81–89.

30. Daghino, S., Vietro, L.D., Petiti, L., Martino, E., Dallabona, C., Lodi, T., and Perotto, S. (2019). Yeast expression of mammalian Onzin and fungal FCR1 suggests ancestral functions of PLAC8 proteins in mitochondrial metabolism and DNA repair. Sci Rep *9*, 1–13.

31. Tibbitt, C.A., Stark, J.M., Martens, L., Ma, J., Mold, J.E., Deswarte, K., Oliynyk, G., Feng, X., Lambrecht, B.N., De Bleser, P., *et al.* (2019). Single-Cell RNA Sequencing of the T Helper Cell Response to House Dust Mites Defines a Distinct Gene Expression Signature in Airway Th2 Cells. Immunity *51*, 169-184.e5.

32. Johnson, R.M., Kerr, M.S., and Slaven, J.E. (2012). Plac8-Dependent and Inducible NO Synthase-Dependent Mechanisms Clear Chlamydia muridarum Infections from the Genital Tract. The Journal of Immunology *188*, 1896–1904.

33. Johnson, R.M., and Brunham, R.C. (2016). Tissue-Resident T Cells as the Central Paradigm of Chlamydia Immunity. Infection and Immunity *84*, 868–873.

34. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. *2008*, P10008.

35. Baumjohann, D., and Ansel, K.M. (2013). Identification of T follicular helper (Tfh) cells by flow cytometry. Available at: http://dx.doi.org/10.1038/protex.2013.060.

36. Zielinski, C.E., Mele, F., Aschenbrenner, D., Jarrossay, D., Ronchi, F., Gattorno, M., Monticelli, S., Lanzavecchia, A., and Sallusto, F. (2012). Pathogen-induced human TH17 cells produce IFN-γ or IL-10 and are regulated by IL-1β. Nature *484*, 514–518.

37. Cosmi, L., Cimaz, R., Maggi, L., Santarlasci, V., Capone, M., Borriello, F., Frosali, F., Querci, V., Simonini, G., Barra, G., *et al.* (2011). Evidence of the transient nature of the Th17 phenotype of CD4+CD161+ T cells in the synovial fluid of patients with juvenile idiopathic arthritis. Arthritis & Rheumatism *63*, 2504–2515.

38. Becattini, S., Latorre, D., Mele, F., Foglierini, M., De Gregorio, C., Cassotta, A., Fernandez, B., Kelderman, S., Schumacher, T.N., Corti, D., *et al.* (2015). Functional heterogeneity of human memory CD4+ T cell clones primed by pathogens or vaccines. Science *347*, 400–406.

39. Messi, M., Giacchetto, I., Nagata, K., Lanzavecchia, A., Natoli, G., and Sallusto, F. (2003). Memory and flexibility of cytokine gene expression as separable properties of human T H 1 and T H 2 lymphocytes. Nat Immunol *4*, 78–86.

# Chapter 5: Validation and characterisation of clonally expanded *ex vivo* synovial fluid CD4 and CD8 memory T cells in PsA patients using single-cell RNA sequencing.

## 5.1 Introduction

Chapters 3 and 4 focussed primarily on microbe-reactive, and in particular *Salmonella*-reactive, CD4 Th memory cell responses in SpA. CD154 was used as a marker of activation to identify CD4 memory T cells reactive to particular microbial lysates, and ssRNA-seq was then used to characterise the synovial Th response to *Salmonella* stimulation at high resolution. CD8 T cell responses to microbial lysate were not explored. This was partly due to the difficulty in establishing an assay capable of assessing a more general CD8 T cell response to whole microbial lysate rather than pre-defined immunogenic peptides. CD154 is not known as a marker of activation in CD8 T cells and upregulation of CD154 protein expression in CD8 T cells was not observed upon stimulation with microbial lysates.

However, when characterising the gene expression and clonality of *ex vivo* unstimulated CD4 and CD8 memory T cells from both the peripheral blood and synovial fluid of patient PSA1607, it was clear that a strong expansion of specific CD8 T cell clones was present in the synovial fluid of this patient. In fact 16% of all synovial CD8 T cells were represented by a small set of only 9 clonotypes (Table 4-2)[1], all of which were strongly enriched in synovial fluid in comparison to peripheral blood. Paired single alpha and single beta chain TCR data was available for 6 of these 9 clones

[1] The least enriched CD8 clonotype containing 16 cells included 2 beta chains and these cells and corresponding clone were excluded from calculations as doublets, leaving 9 10x clones.

from 10x sequencing data, and the paired alpha chain could be inferred from SS2 TCR data for an additional clone. Therefore 7 of the 9 clonotypes were very precisely defined by their paired alpha and beta chain expression. Although previous studies have found clonal expansions of CD8 T cells in the synovial fluid of PsA patients[1], never before had these cells been characterised to the degree of detail made possible by ssRNA-seq. This chapter is therefore focussed on validating the finding of paired alpha and beta chain TCR clonal enrichment in synovial fluid CD8 memory T cells of PsA patients, and characterising the phenotype of these cells using ssRNA-seq. This chapter also makes use of a recently developed algorithm known as GLIPH (grouping of lymphocyte interactions by paratope hotspots) to identify clonotypes that have a high probability of sharing specificity towards a common antigen [2], independently of the heterogeneity in MHC background and immune responses that exists between individuals.

Analysis of the gene expression and clonality of *ex vivo* unstimulated CD4 and CD8 memory T cells from  2 additional PsA patients is performed on data derived from the 10x scRNA-seq platform. The analysis of 3 more datasets generated from SS2 scRNA-seq of 3 extra patients is additionally used to support findings. Cells from patient PSA1607 were sequenced using both 10x and SS2 platforms, also allowing validation of findings across sequencing technologies.

The findings from this chapter are mirrored in a paper which has been accepted for review in the Nature Communications journal and is available for download on the BioRxiv [3] website. This chapter focusses on areas of the paper I am most responsible for.

## 5.2 Aim

To characterise the phenotype and clonality of *ex vivo* CD4 and CD8 memory T cells from the peripheral blood and synovial fluid of PsA patients using ssRNA-seq, and to validate the synovial enrichment of paired-chain CD8 memory T cell clonotypes in PsA patients.

## 5.3 Results

### *5.3.1 Integration of gene expression matrices representing ex vivo memory CD4 and CD8 T cells from the blood and synovial fluid of 3 PsA patients reveals 16 phenotypic clusters.*

Aligning or integrating multiple scRNA-seq datasets to identify overlap and heterogeneity between them is an area of research which has received a great deal of attention and continues to evolve. This chapter makes use of the recently released SCTransform algorithm[4] in combination with an integration process introduced into the Seurat version 3 pipeline to harmonise datasets from multiple patients into a single integrated dataset[5].

To complement the PSA1607 10x scRNA-seq gene expression paired datasets of unstimulated *ex vivo* peripheral blood and synovial fluid CD4 and CD8 memory T cells from chapter 4, paired peripheral blood and synovial fluid datasets from patients (PSA1505 and PSA1801) were also obtained by 10x scRNA-seq consisting of the same cell types processed in the same manner discussed in chapter 4. Thus 6 datasets originating from 6 libraries (peripheral blood and synovial fluid libraries from each of the 3 patients), each containing an approximately equal number of *ex vivo* CD4 and CD8 memory T cells were available for analysis. To ensure that no particular patient or sample source dominated the analysis, after quality control filtering and filtering to only include cells with matching V(D)J sequencing data available, all 6 datasets to be integrated were subsampled to a common number of cells (6542). The cell numbers used for downstream 10x integrated gene expression analysis, 10x TCR analysis and SS2 analysis are set out in Table 5-1. Approximately twice as many cells were sequenced for gene expression and TCR clonality for patient PSA1607 compared with the other 2 patients, and this is the only patient for which cellular samples were

processed using both the 10x and SS2 pipelines. To take advantage of this, a separate integration of PSA1607 datasets obtained from both scRNA-seq platforms is addressed later in this chapter. A separate analysis of all TCR sequencing data available from all patients was also undertaken and will be discussed later in this chapter.

**Table 5-1 Cell numbers used in downstream analysis**

| Technology | PsA Patients (n) | Cell Type | Synovial Cell (n) | Blood Cell (n) |
|---|---|---|---|---|
| 5' 10X Chromium GEX* | 3 | Sorted $CD45RA^-$ CD4/CD8 | 19,626 | 19,626 |
| 5' 10X TCR* | 3 | Sorted $CD45RA^-$ CD4/CD8 | 31,361 | 28,902 |
| SS2 GEX + TCR* | 4 | Sorted $CD45RA^-$ CD4/CD8 | 1,736 | 1,217 |

*One patient run in parallel across different technologies

The 6 filtered and subsampled 10x gene expression datasets from patients PSA1505, PSA1607 and PSA1801 were normalised using the SCTransform function and integrated using the Seurat integration pipeline. 16 clusters within the integrated dataset were then determined (detailed methods in Chapter 2), and are depicted on the 2-dimensional UMAP reduction in Figure 5-1. This single UMAP has been split to illustrate cells originating from peripheral blood in Figure 5-1A, and those originating from synovial fluid in Figure 5-1B. As can be seen in these plots, all 16 clusters contain cells from both peripheral blood and synovial fluid, suggesting that cells in the synovial fluid of these PsA patients have similar phenotypic counterparts in their blood. However, as part of the integration process, only "anchor" genes present in all integrated datasets are used to align the datasets[5]. Later in this chapter, the differences between pro-inflammatory and tissue homing gene expression between cell subsets of the same clusters is analysed in more detail.

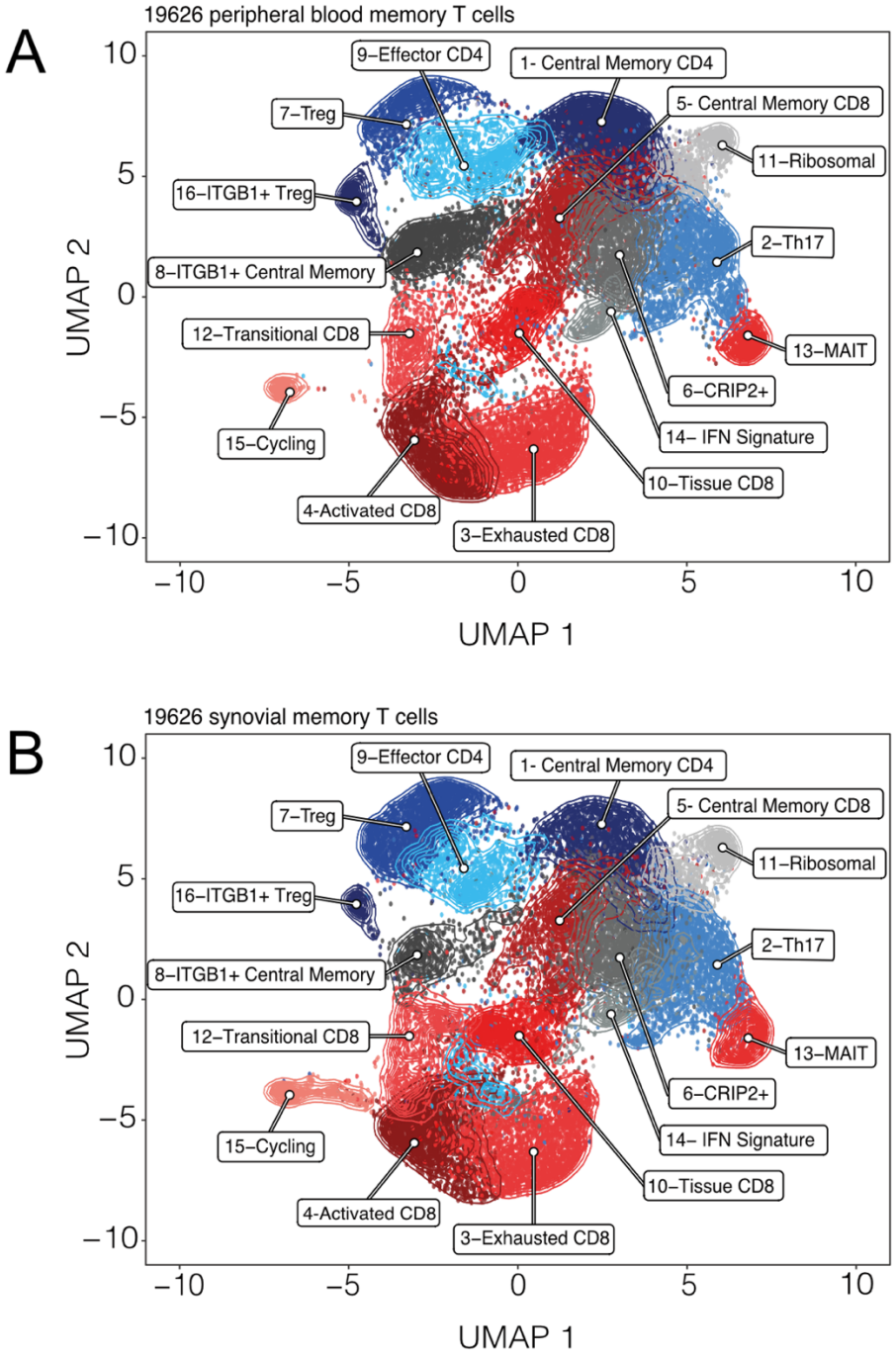


**Figure 5-1 Clustering of peripheral blood and synovial fluid *ex vivo* CD4 and CD8 memory T cells within the 10x integrated dataset.**

Split UMAP plot of integrated dataset, showing distribution of cells derived from peripheral blood (A) and synovial fluid (B) between the 16 clusters identified. Clusters were identified by the Seurat FindClusters function run at a resolution of 0.7, after performing Seurat SCTransform normalisation, principal component calculation, SNN calculation and UMAP generation on the integrated dataset. The integrated dataset contains an equal number of peripheral and synovial cells from each patient, selected by subsampling from quality control filtered datasets. Clusters are numbered according to the number of cells they represent from 1 (largest) to 16 (smallest) and coloured according to the dominant type of cell present in the cluster (Red=CD8, blue=CD4, and grey=mix of CD4 and CD8).[2]

[2]Hussein Al-Mossawi helped detail the split UMAP plot in this figure, especially contours.

Clusters in Figure 5-1 were annotated according to differential gene expression between them which is summarised by heatmaps from Figure 5-2 and Figure 5-3. When applying the same methodology described in chapter 4 to distinguish between CD4 and CD8 T cells, it was clear that certain clusters were strongly represented by CD8 T cells, other clusters by CD4 T cells and some clusters contained a considerable mix of both cell types. Figure 5-4 shows the distribution of CD4 and CD8 memory T cells among the 16 clusters identified. Clusters were determined in an unsupervised way (other than specifying resolution) and no effort was made to force cells into any particular cluster. As in chapter 4, multiple resolution parameters were tested, and the clusters discussed most in this thesis were maintained across a broad range of resolutions, with a resolution of 0.7 being finally adopted. The integration methodology documented in a previous version of Seurat (version 2) , named canonical correlation analysis (CCA), was also tested at different resolutions and identified similar clusters according to differential gene expression (data not shown).

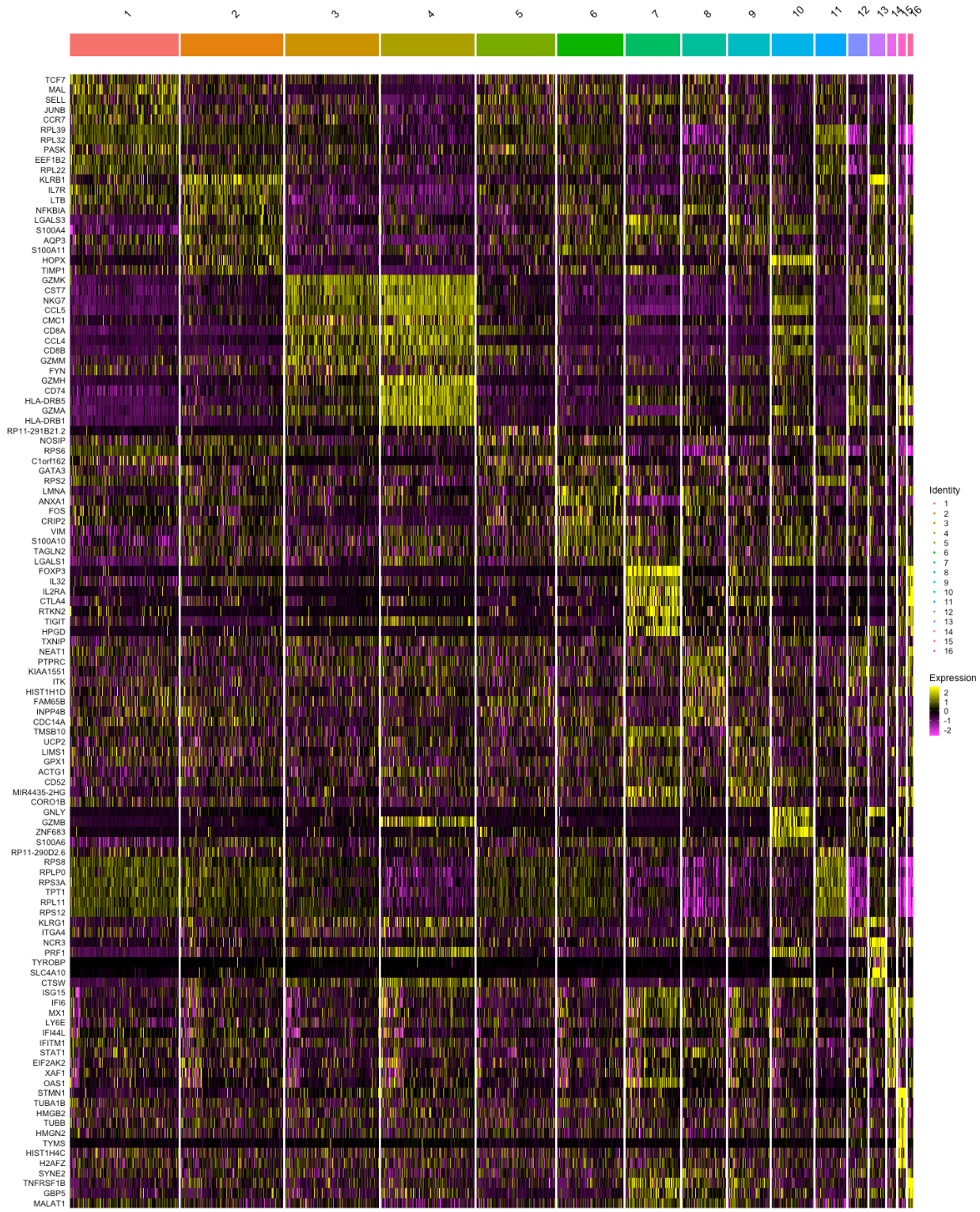


**Figure 5-2 Heatmap of Top 10 most differentially expressed genes per cluster of the 10x integrated dataset of peripheral and synovial *ex vivo* CD4 and CD8 memory T cells from 3 PsA patients.**

Top 10 most differentially expressed genes for each of the 16 clusters identified in Figure 5-1, calculated by Seurat FindAllMarkers function using Wilcoxon rank sum test. Only significantly differentially expressed genes are shown (adj. $p < 0.05$), ordered by avgLogFC in normalised gene expression from the integrated assay. Gene expression is coloured by scaled, normalised SCT assay values per cell within clusters shown left to right from largest to smallest cluster.

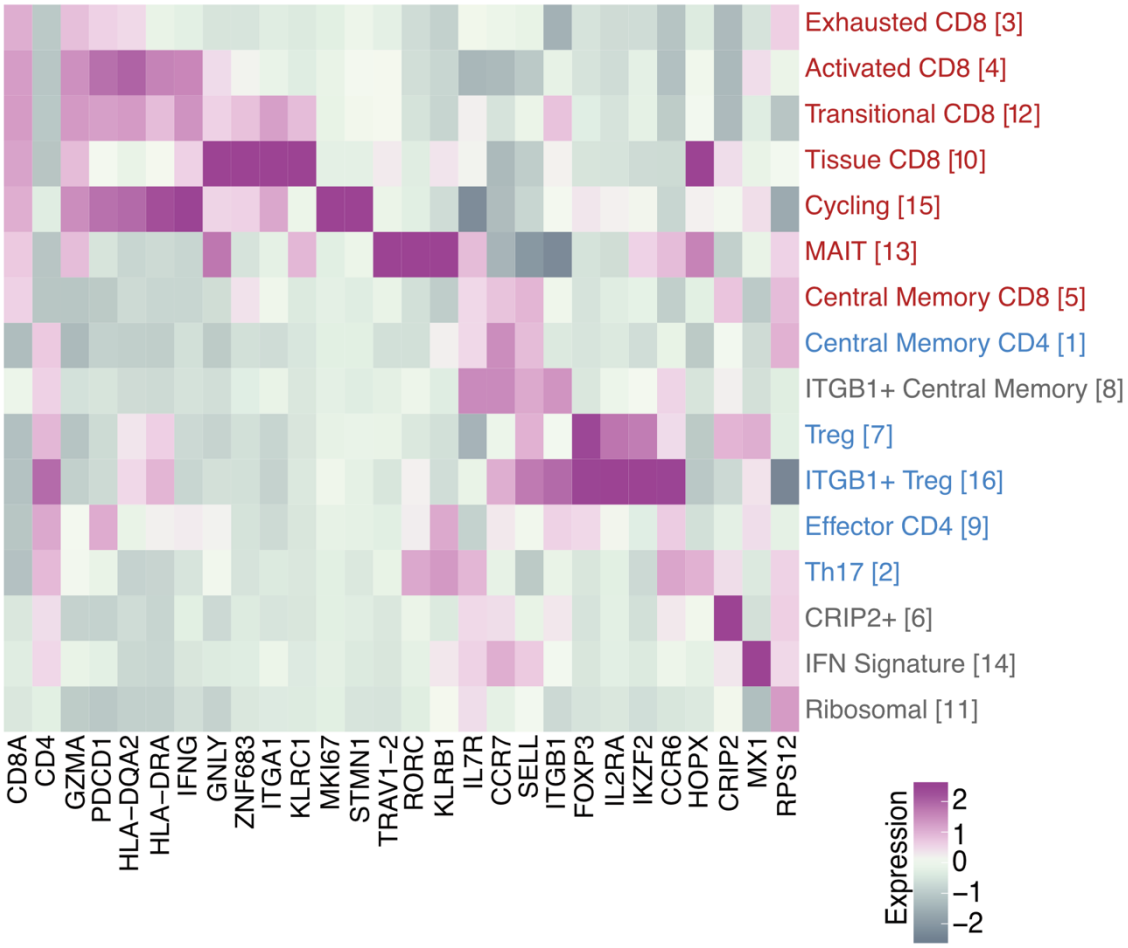


**Figure 5-3 Key markers defining clusters of the 10x integrated dataset**

The average expression of key markers chosen from genes differentially expressed between clusters to help define cluster phenotype. Clusters are arranged with regard to their spatial proximity to one another on the UMAP plot of Figure 5-1, and similarity between level of gene expression. Cluster names in red contain predominantly CD8 T cells, those coloured blue predominantly CD4 T cells and those with a grey colour contain a mix of both cell types. Average expression was computed using the Seurat AverageExpression function.

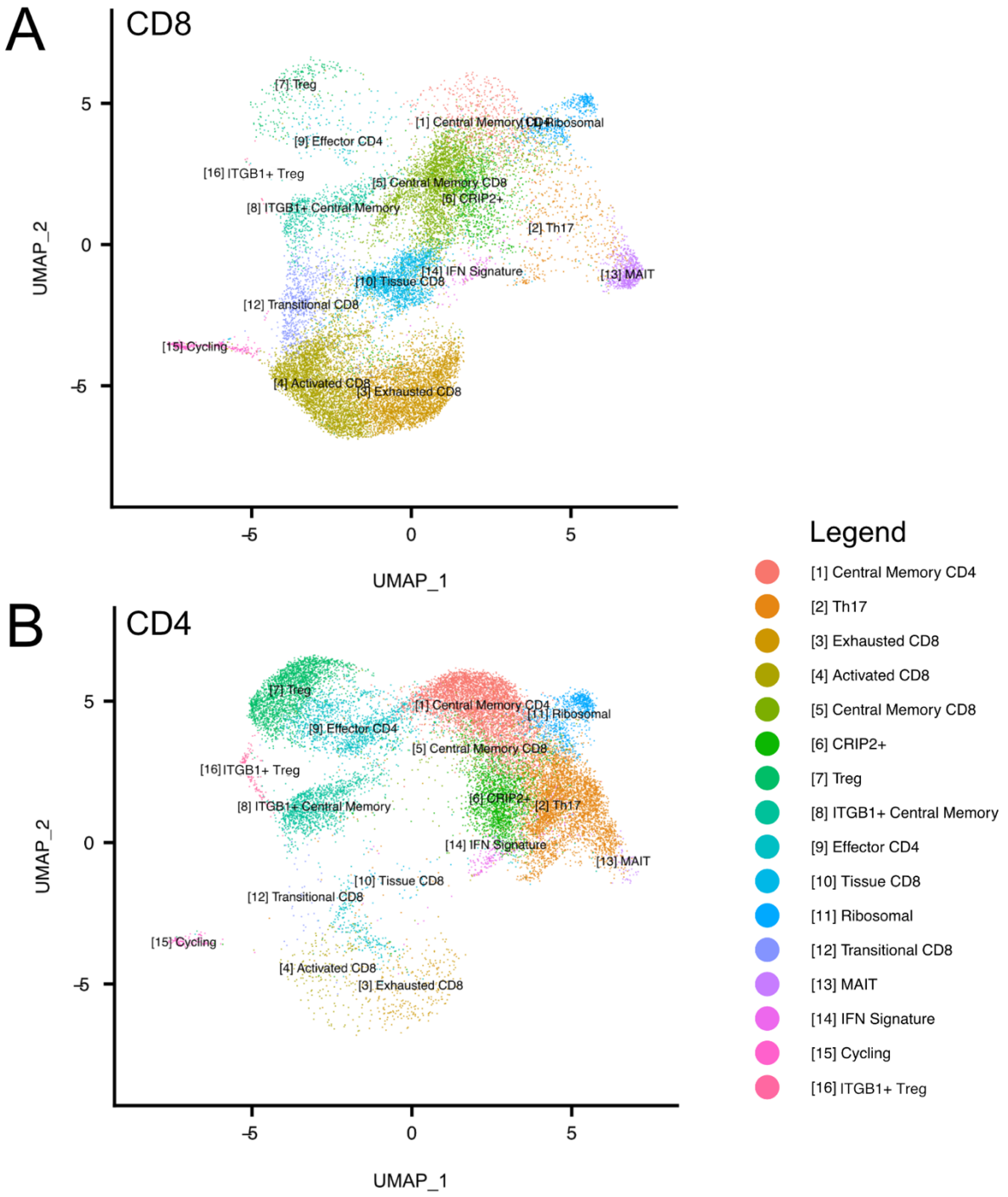


**Figure 5-4 Distribution of CD4 and CD8 *ex vivo* memory T cells within the integrated dataset.**

Split UMAP dot plots showing the distribution of (A) CD8 and (B) CD4 memory T cells between the 16 clusters identified within the 10x integrated dataset from Figure 5-1. Clusters are coloured according to the legend, which applies to both UMAP plots.

Cluster 1 was identified as containing primarily CD4 T cells and its increased expression of SELL and CCR7 associated this cluster with a central memory phenotype. TCF7 was also strongly upregulated in cluster 1, as it was in *Salmonella*-stimulated "stem-like" Th memory cells in chapter 4, however PLAC8 was not upregulated in any cluster, memory or otherwise. Clusters 5 and 8 similarly overexpressed SELL and CCR7 and were classified as central memory, with cluster 5 containing mainly CD8 T cells. Cluster 8 contained a mix of CD4 and CD8 T cells, but could be distinguished from other central memory clusters by an increased expression of ITGB1. This gene encodes the integrin beta chain of many adhesion molecules, but was also overexpressed by a number of other clusters. Mixed cluster 6 also appeared to have a similar gene expression profile to central memory clusters but lacked strong expression of SELL and its expression of CCR7 was not statistically different from other clusters. Instead cluster 6 overexpressed the CRIP2 gene. This gene was recently associated with a memory T cell precursor phenotype (Tcmp) by a separate study making use of scRNA-seq (see [6], supplemental table 2). The only remaining mixed CD4/CD8 clusters were cluster 14, associated with statistically significant overexpression of a large number of genes encoding interferon inducible proteins (MX1, MX2, IFI6, IFI16, IFI35, IFI44, IFI44L, IFIT1, IFIT3, IFITM1, IFITM2, ISG15, ISG20), and cluster 11 which overexpressed very little of anything other than ribosomal genes.

Remaining clusters consisting of primarily CD8 T cells could be defined as: MAIT cells, clearly distinguished by the overexpression of known associated genes TRAV1-2, KLRB1 (encoding CD161) and SLC4A10[7][8][9]; Cycling cells, overexpressing STMN1 and MKI67; Tissue and Activated clusters defined by their overexpression of ZNF683 and MHC-II genes respectively; a transitional cluster, expressing many of the

genes defining both the activated and tissue clusters; and what was termed an exhausted cluster, which shared overlapping gene expression with the activated cluster but expressed little IFNG and reduced GZMA. The Exhausted CD8 cluster did however continue to express high levels of GZMK (average logFC=1, adj.p=0) which has recently been linked by scRNA-seq to a "pre-exhausted" cell type[8], and the classification of exhaustion used in this thesis should be interpreted with a degree of flexibility.

Of the remaining CD4 clusters, 2 distinct FoxP3+ clusters (7 and 16) could be identified which were annotated as Treg clusters. Although sharing extensive overlap in gene expression, the smaller cluster additionally overexpressed ITGB1 relative to other clusters (average logFC = 0.55, adj. p= 7.69E-24) and was also positioned between Treg cluster 7 and the ITGB1+ Central Memory cluster. The 2 remaining CD4 clusters were classified as Th17 (cluster 2) based on increased CCR6 and RORC expression, and effector CD4 (cluster 9). However, these 2 clusters were difficult to classify due to generally low gene expression of any defining markers.

The annotated clusters were finally compared between patients to ascertain whether any of the clusters was peculiar to a particular patient, or whether the clustering was representative of cell types from all patients (Figure 5-5). Although some variation existed between cluster sizes between patients, all clusters contained cells from each patient derived from both synovial fluid and blood. It was also apparent that some clusters such as Treg cluster 7 and Cycling cluster 15, contained more cells in the synovial compartment for all 3 patients, whereas the ITGB1+ central memory cluster was larger in peripheral blood.

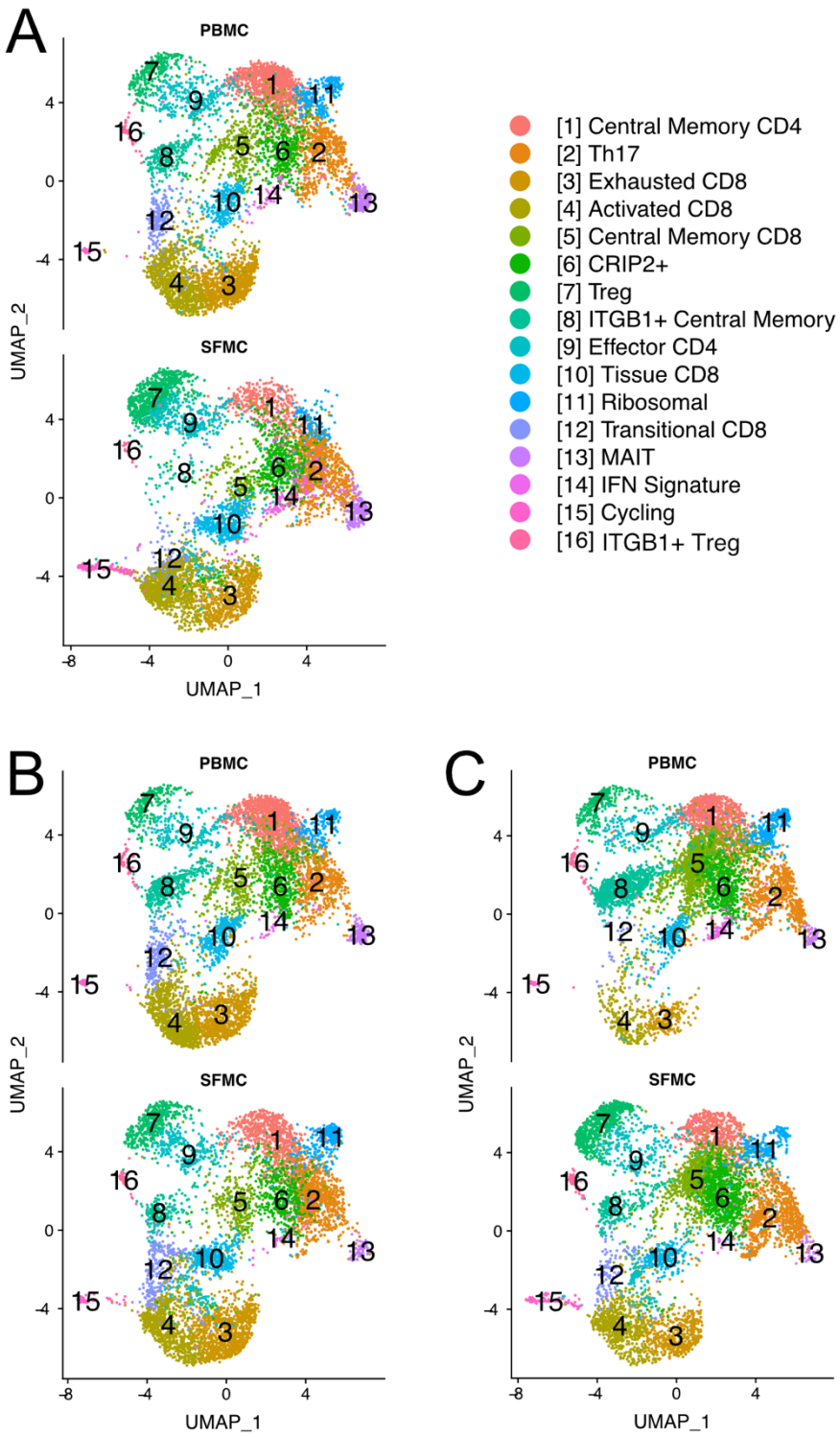


**Figure 5-5 Each cell cluster is represented by cells from both the peripheral blood and synovial fluid of all 3 patients within the 10x integrated dataset**

UMAP plots split between memory T cells derived from PBMC or SFMC for each of the 3 PSA patients represented by the integrated dataset (A: PSA1505, B:PSA1607, C:PSA1801). Clusters are coloured according to the legend, which applies to all UMAP plots. All clusters are represented by cells from each patient from both peripheral blood and synovial fluid.

### *5.3.2 Clonal analysis confirms that CD8 clonotypes are enriched in the synovial fluid of PsA patients.*

In chapter 4, a small number of CD8 memory T cell clones were discovered to be strongly enriched within the synovial fluid of patient PSA1607. However, additional evidence was required to show that this clonal enrichment was not peculiar to one patient. Paired alpha and beta chain CDR3 sequencing was therefore carried out on the *ex vivo* CD4 and CD8 memory T cells from both blood and synovial fluid of 2 additional PSA patients using the 10x 5 prime V(D)J pipeline. The output of this pipeline includes both a 10x "vloupe" file which can be browsed using the Loupe V(D)J Browser software supplied by 10x Genomics, in addition to several more detailed files describing the clonality detected for each cell which are generated by the Cell Ranger pipeline and available in the "outs" directory of this workflow. The contents of the consensus_annotations.json file from this directory were read into R using the jsonlite R package. The resulting object containing nested tables of information pertaining to clones was converted into a flat table where each row represented a cell with its associated barcode and contained all clonal information regarding that cell. This made it easy to work with the data and quickly ascertain all clonal information available regarding a particular cell. The enrichment of each clone was then determined by comparing the proportion of cells carrying a particular clone within synovial cells to the proportion of cells carrying that clone within peripheral blood cells. Comparing proportions rather than frequencies took into account any variation between the number of cells sequenced in synovial fluid compared to blood for each patient, and was preferred to subsampling to make maximal use of available clonal data. This calculation was also performed independently of whether cells had been assigned a CD4 or CD8 status, information which was attributed to clones after the calculation. Results are presented graphically in Figure 5-6.

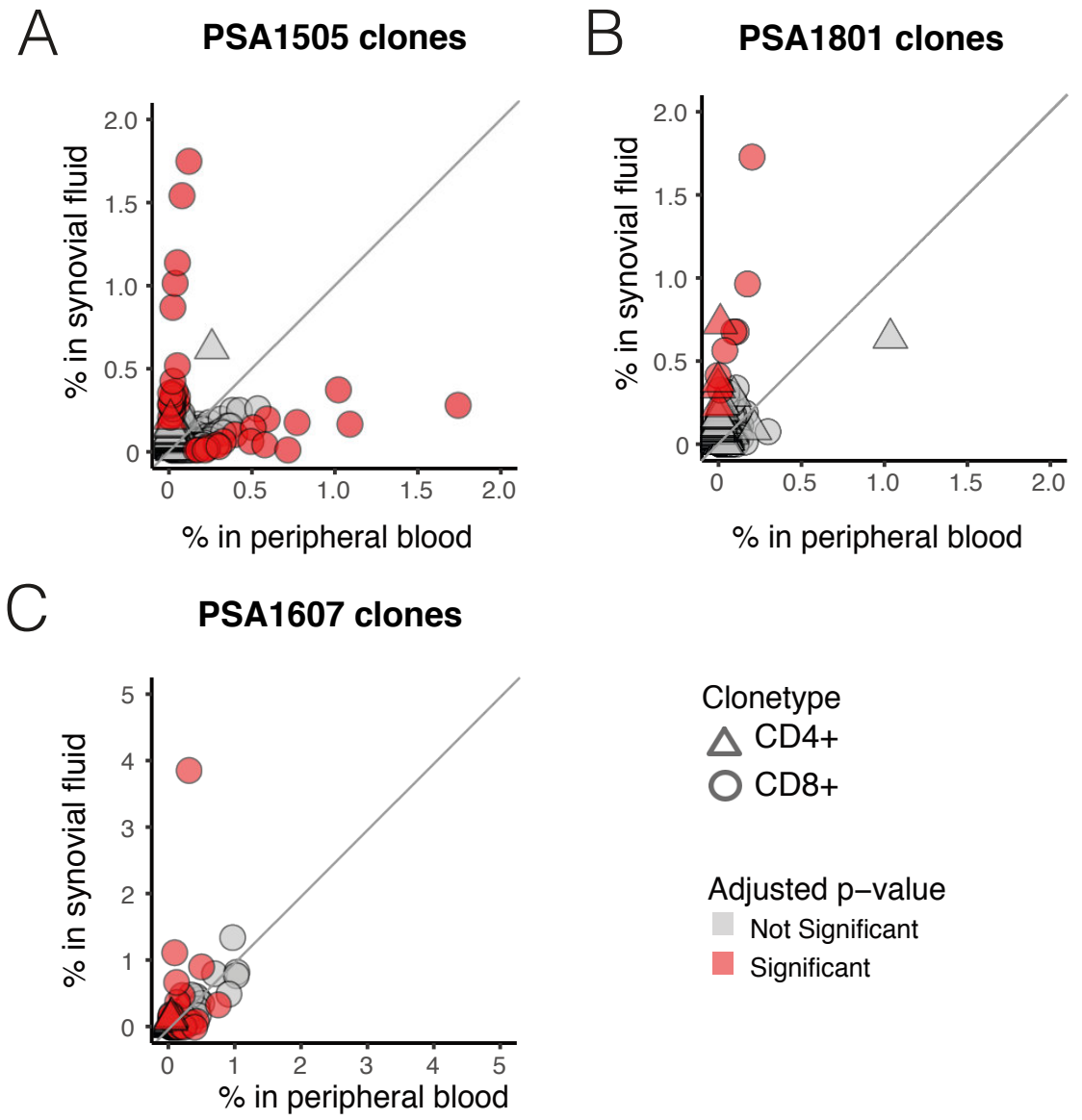


**Figure 5-6 Specific CD8 paired-chain clonotypes are strongly enriched in the synovial fluid of 3 PsA patients.**

Clonal enrichment in either blood or synovial fluid in three PsA patients (A) PSA1505, (B)PSA1801 and (C) PSA1607, based on 10x 5' V(D)J sequencing data. Circles represent CD8 clonotypes, triangles represent CD4 clonotypes. Clonotypes are defined as having the same CDR3 alpha and beta chain nucleotide sequences (paired-chain). Circles or triangles coloured red show significantly enriched clonotypes. Statistics: Fisher's exact test, Benjamini-Hochberg correction. Percentages are based on the fraction of all CD4 and CD8 clonotypes combined. Diagonal lines represent positions of no enrichment in either synovial fluid or peripheral blood. Clonotypes which could not be determined to be either CD4 or CD8 due to lack of available gene expression data are not shown.

In support of the findings from chapter 4, significantly enriched CD8 (and to a lesser extent CD4) clones were detected in the synovial fluid of all 3 PSA patients with 10x scRNA-seq data available (10x cohort). The p-values determined by using a Fisher's exact test for each clonotype were also compared with results calculated by the Loupe V(D)J Browser which uses the same statistical approach and were found to match, helping to confirm that no coding errors had taken place. In addition, peripheral blood enriched CD8 clones were also found for patients PSA1505 and PSA1607, while no CD4 clones were found to be enriched in blood for any patient. The percentages of cells present in either synovial fluid or peripheral blood for each individual clonotype depicted in Figure 5-6 are calculated without regard for CD4 or CD8 cell type. As CD4 and CD8 memory T-cells were present in both synovial fluid and peripheral blood in a 1:1 ratio, the percentage of cells from a particular CD4 or CD8 clone among only CD4 or CD8 memory T cells respectively will be approximately twice the percentage of what is shown in the graphs of Figure 5-6. This further emphasizes the degree of selective clonal enrichment present in synovial fluid, with a single clone from patient PSA1607 accounting for almost 8% of all synovial fluid CD8 memory T cells (Figure 5-6C).

Clonal information from the SS2 scRNA-seq analysis of an additional 3 patients (SS2 cohort) as well as patient PSA1607 was also obtained. The top enriched clone from the 10x dataset of PSA1607 was confirmed by its presence in the SS2 dataset from the same patient, where it was similarly enriched in synovial fluid. However, other clones with a lower degree of enrichment in the 10x dataset of PSA1607 were not assigned statistical significance in the SS2 dataset of this patient. This suggested that the smaller number of cells (427 from peripheral blood and 616 cells from synovial fluid) sequenced as part of the SS2 pipeline was underpowered to detect less enriched clones

for this patient.  Most likely for the same reason, no enriched clones were detected in the SS2 cohort. All of the SS2 datasets did however include clones representing between 1-2% of synovial cells which represented a much lower proportion of cells in peripheral blood (often 0%). This corresponds with levels seen in the 10x cohort for many of the smaller clonal enrichments, further supporting the idea that the SS2 dataset was underpowered to detect enrichment.

### *5.3.3 GLIPH identifies 5 convergence groups with shared specificity across all 6 patients from both 10x and SS2 cohorts.*

Although synovial enriched clonotypes were detected in all patients from the 10x cohort, there was no exact match for enriched clonotypes between patients. However, this does not automatically discount the possibility of the same peptide(s) being recognized by different clones from different patients. Heterogeneity between individuals, such as their MHC background and outcome of thymic selection can impact the specific TCR repertoire of any one individual. The GLIPH algorithm, recently published by Glanville et al. in 2017[2], attempts to overcome this heterogeneity and cluster TCRs into specificity groups (convergence groups) where all members of a particular convergence group (CRG) have a high probability of sharing specificity towards a common antigen. This algorithm was therefore employed to analyse synovial clonality from all 6 patients from both the 10x and SS2 cohorts to see whether there was any overlap between the TCRs of each patient, and whether that overlap included any of the clones already identified as enriched within synovial fluid from the 10x cohort.

The GLIPH algorithm currently only makes use of TCR beta chain sequences. Therefore, to maximise the utility of this algorithm, it was assumed that each unique beta chain represented a single cell, and unique beta chains were accordingly extracted from clonotypes which had previously been associated with multiple beta chains (likely multiplets) and included in analysis. All CDR3 beta chains present in the synovial clones of the 10x and SS2 cohorts were included in the GLIPH analysis[3].

[3] Only 10x clones, not SS2 clones, from patient PSA1607 were included to avoid any potential bias in GLIPH scoring.

# Table 5-2 Top 50* convergence groups (CRG) identified by GLIPH

| Convergence group identifier | CDR3s | Patients | Clones | Overall CRG Score | V gene p-value | CDR3 p-value | HLA p-value | Expansion p-value | Motif p-value | Size p-value | HLA | Motifs | Has motif(s) | Enriched clone count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CRG-1 | 543 | 6 | 570 | 4.45E-14 | 0.6946 | 0.0001 | 1 | 0.001 | 0.001 | 1.00E-05 | | NQNT | yes | 4 |
| CRG-2 | 564 | 6 | 588 | 5.68E-15 | 0.0888 | 0.0001 | 1 | 0.001 | 0.001 | 1.00E-05 | | LIRE | yes | 2 |
| CRG-3 | 344 | 6 | 357 | 1.94E-14 | 0.3024 | 0.0001 | 1 | 0.001 | 0.001 | 1.00E-05 | | SLKY | yes | |
| CRG-4 | 219 | 6 | 229 | 1.61E-14 | 0.2509 | 0.0001 | 1 | 0.001 | 0.001 | 1.00E-05 | | | | |
| CRG-5 | 85 | 6 | 88 | 1.33E-11 | 0.5543 | 0.0001 | 1 | 0.374 | 0.001 | 1.00E-05 | | | | |
| CRG-6 | 46 | 5 | 50 | 2.56E-14 | 0.7883 | 0.0001 | 0.33417 | 0.001 | 0.001 | 1.52E-05 | | | | |
| CRG-7 | 12 | 5 | 12 | 2.14E-13 | 0.6538 | 0.0001 | 0.33271 | 0.001 | 0.001 | 0.00015401 | | | | |
| CRG-8 | 16 | 5 | 16 | 3.49E-12 | 0.5104 | 0.0001 | 0.32873 | 0.038 | 0.001 | 8.56E-05 | | | | |
| CRG-9 | 22 | 5 | 24 | 2.31E-11 | 0.9706 | 0.0001 | 0.49908 | 0.099 | 0.001 | 7.51E-05 | | | | |
| CRG-10 | 13 | 5 | 13 | 1.01E-11 | 0.1401 | 0.0001 | 0.33032 | 0.444 | 0.001 | 7.70E-05 | | | | |
| CRG-11 | 7 | 5 | 7 | 1.09E-09 | 1 | 0.0001 | 0.33278 | 0.594 | 0.001 | 0.000862264 | | | | |
| CRG-12 | 16 | 5 | 17 | 5.57E-11 | 0.3051 | 0.0001 | 0.33335 | 1 | 0.001 | 8.56E-05 | | | | |
| CRG-13 | 36 | 4 | 36 | 1.80E-12 | 0.9603 | 0.0001 | 0.40011 | 0.022 | 0.001 | 3.33E-05 | | | | 2 |
| CRG-14 | 18 | 4 | 18 | 6.35E-13 | 0.9687 | 0.0001 | 0.39878 | 0.003 | 0.001 | 8.56E-05 | | | | 1 |
| CRG-15 | 5 | 4 | 5 | 3.27E-08 | 0.4959 | 0.8836 | 0.40087 | 0.002 | 0.001 | 0.001456437 | | SYQR | yes | |
| CRG-16 | 4 | 4 | 4 | 4.47E-06 | 1 | 0.0743 | 0.40019 | 1 | 0.001 | 0.002347221 | | NIRG | yes | |
| CRG-17 | 4 | 4 | 4 | 4.71E-06 | 1 | 0.0784 | 0.39952 | 1 | 0.001 | 0.002347221 | | PDSV | yes | |
| CRG-18 | 20 | 4 | 22 | 6.55E-15 | 0.1878 | 0.0001 | 0.06739 | 0.001 | 0.001 | 8.08E-05 | B*15(0.06739) | | | |
| CRG-19 | 13 | 4 | 14 | 5.57E-14 | 0.2834 | 0.0001 | 0.39895 | 0.001 | 0.001 | 7.70E-05 | | | | |
| CRG-20 | 6 | 4 | 6 | 8.06E-13 | 0.222 | 0.0007 | 0.06685 | 0.001 | 0.001 | 0.001212588 | B*15(0.06685) | | | |
| CRG-21 | 4 | 4 | 4 | 7.39E-10 | 0.3508 | 0.007 | 0.40077 | 0.005 | 0.001 | 0.002347221 | | | | |
| CRG-22 | 99 | 4 | 104 | 2.81E-13 | 0.848 | 0.0001 | 0.39833 | 0.013 | 0.001 | 1.00E-05 | | | | |
| CRG-23 | 6 | 4 | 7 | 8.84E-13 | 0.013 | 0.0001 | 0.39833 | 0.022 | 0.001 | 0.001212588 | | | | |
| CRG-24 | 5 | 4 | 6 | 6.97E-11 | 1 | 0.0001 | 0.068 | 0.11 | 0.001 | 0.001456437 | B*15(0.068) | | | |
| CRG-25 | 5 | 4 | 5 | 1.11E-10 | 0.1175 | 0.0013 | 0.06589 | 0.118 | 0.001 | 0.001456437 | B*15(0.06589) | | | |
| CRG-26 | 15 | 4 | 15 | 3.89E-12 | 0.062 | 0.0001 | 0.39692 | 0.26 | 0.001 | 9.51E-05 | | | | |
| CRG-27 | 16 | 4 | 16 | 2.18E-11 | 0.3542 | 0.0001 | 0.39816 | 0.282 | 0.001 | 8.56E-05 | | | | |
| CRG-28 | 9 | 4 | 11 | 4.34E-11 | 0.749 | 0.0001 | 0.06694 | 0.284 | 0.001 | 0.000476289 | B*15(0.06694) | | | |
| CRG-29 | 4 | 4 | 4 | 2.54E-08 | 1 | 0.0066 | 0.06531 | 0.392 | 0.001 | 0.002347221 | B*15(0.06531) | | | |
| CRG-30 | 4 | 4 | 4 | 6.39E-08 | 0.3436 | 0.0077 | 0.39974 | 0.402 | 0.001 | 0.002347221 | | | | |
| CRG-31 | 5 | 4 | 5 | 3.45E-08 | 1 | 0.0021 | 0.39867 | 0.442 | 0.001 | 0.001456437 | | | | |
| CRG-32 | 6 | 4 | 6 | 1.48E-09 | 0.6332 | 0.0003 | 0.20046 | 0.502 | 0.001 | 0.001212588 | | | | |
| CRG-33 | 6 | 4 | 6 | 5.65E-10 | 1 | 0.0002 | 0.06665 | 0.546 | 0.001 | 0.001212588 | B*15(0.06665) | | | |
| CRG-34 | 61 | 4 | 63 | 2.41E-12 | 0.9599 | 0.0001 | 0.06728 | 0.584 | 0.001 | 1.00E-05 | B*15(0.06728) | | | |
| CRG-35 | 18 | 4 | 19 | 5.84E-13 | 0.0258 | 0.0001 | 0.06574 | 0.629 | 0.001 | 8.56E-05 | B*15(0.06574) | | | |
| CRG-36 | 8 | 4 | 8 | 8.45E-11 | 0.3697 | 0.0001 | 0.06745 | 0.638 | 0.001 | 0.000829941 | B*15(0.06745) | | | |
| CRG-37 | 19 | 4 | 19 | 1.05E-11 | 0.3472 | 0.0001 | 0.06632 | 0.639 | 0.001 | 0.000111229 | B*15(0.06632) | | | |
| CRG-38 | 9 | 4 | 10 | 5.15E-11 | 0.3677 | 0.0001 | 0.06698 | 0.686 | 0.001 | 0.000476289 | B*15(0.06698) | | | |
| CRG-39 | 40 | 4 | 41 | 5.26E-11 | 0.9919 | 0.0001 | 0.3991 | 0.705 | 0.001 | 2.95E-05 | | | | |
| CRG-40 | 13 | 4 | 13 | 3.32E-11 | 0.9931 | 0.0001 | 0.06775 | 1 | 0.001 | 7.70E-05 | B*15(0.06775) | | | |
| CRG-41 | 11 | 4 | 11 | 3.89E-11 | 0.372 | 0.0001 | 0.06795 | 1 | 0.001 | 0.000240521 | B*15(0.06795) | | | |
| CRG-42 | 7 | 4 | 7 | 1.81E-10 | 0.2429 | 0.0002 | 0.06736 | 1 | 0.001 | 0.000862264 | B*15(0.06736) | | | |
| CRG-43 | 7 | 4 | 7 | 2.73E-10 | 0.751 | 0.0001 | 0.06582 | 1 | 0.001 | 0.000862264 | B*15(0.06582) | | | |
| CRG-44 | 7 | 4 | 8 | 3.08E-10 | 0.8498 | 0.0001 | 0.0657 | 1 | 0.001 | 0.000862264 | B*15(0.0657) | | | |
| CRG-45 | 5 | 4 | 5 | 1.25E-09 | 0.1113 | 0.0018 | 0.06711 | 1 | 0.001 | 0.001456437 | B*15(0.06711) | | | |
| CRG-46 | 4 | 4 | 4 | 6.24E-08 | 1 | 0.0063 | 0.06596 | 1 | 0.001 | 0.002347221 | B*15(0.06596) | | | |
| CRG-47 | 4 | 4 | 4 | 7.03E-08 | 1 | 0.0069 | 0.06781 | 1 | 0.001 | 0.002347221 | B*15(0.06781) | | | |
| CRG-48 | 4 | 4 | 4 | 7.88E-08 | 1 | 0.0078 | 0.06723 | 1 | 0.001 | 0.002347221 | B*15(0.06723) | | | |
| CRG-49 | 4 | 3 | 4 | 1.49E-08 | 1 | 0.0073 | 0.20011 | 0.068 | 0.001 | 0.002347221 | | | | 1 |
| CRG-50 | 2 | 3 | 3 | 5.76E-07 | 0.1896 | 0.0359 | 0.19909 | 0.228 | 0.001 | 0.029106402 | | | | 1 |
| | | | | | | | | | | | | | | |
| * Convergence groups are listed according to sorting by (in order): number of patients represented, enriched clone count, motif identified, expansion p-value and finally CRG score (lower is better). | | | | | | | | | | | | | | |

GLIPH assigns clones to CRGs by considering only unique beta chain CDR3 sequences present in the entire dataset. It is deliberately blinded to information about the individual person any particular sequence comes from or variable and joining genes surrounding the CDR3 region (nor does it try to deduce V/J genes). Instead, after the GLIPH algorithm has assigned CDR3 sequences to CRGs, those CRGs are evaluated for their likelihood of being valid specificity groups. This is ultimately represented by an overall "CRG score" which is a conflation of scores from individual tests such as the degree of V gene sharing and the size of a CRG. More information regarding these tests can be found on the GitHub website hosting the GLIPH program[4]. Summarised output listing the top 50 convergence groups from this analysis is provided in Table 5-2. CRGs were ranked as detailed in the table, with 5 convergence groups identified as containing clones from all 6 patients. These 5 convergence groups are listed at the top of the table and named CRG-1 to CRG-5 in this thesis. The proportion of cells with matching CDR3 beta chain sequences for CRG-1 to CRG-5 is shown for each patient in Figure 5-7. Three of these top 5 convergence groups were also assigned 4-mer motifs which were overrepresented relative to what would be expected in a naive TCR repertoire, and CRG-1 contained the largest number of enriched clones, 2 of which were the most enriched clones in the synovial fluid of patients PSA1505 and PSA1607. CRG-1 was therefore identified as a convergence group of particular interest, and cells belonging to clones associated with this CRG were targeted for further downstream analysis.

[4] https://github.com/immunoengineer/gliph

### *5.3.4 CRG-1 cells belong to clonotypes enriched for TRBJ1-1 and TRBV28 gene usage compared to all cells from either peripheral blood or synovial fluid*

The overall variable and joining gene usage of CRG-1 clones relative to all clones in peripheral blood and synovial fluid was first assessed to provide a more general perspective on clonal enrichment. Clear enriched usage of TRBV28 (coming primarily from the most enriched PSA1607 clone) and also TRBJ1-1 which had more evenly distributed gene usage among the 3 patients (data not shown),[5] was evident in cells from CRG-1 (Figure 5-8). However CRG-1 cells also made use of many other variable genes and it was clear that the GLIPH algorithm had indeed, by design, ignored this heterogeneity when determining members of CRG-1.

[5] After scaling gene usage to take into account variation in the number of cells sequenced for V(D)J clonality from each patient (in particular, patient PSA1607 which had almost 2-fold the number of cells sequenced compared to PSA1505 or PSA1801)

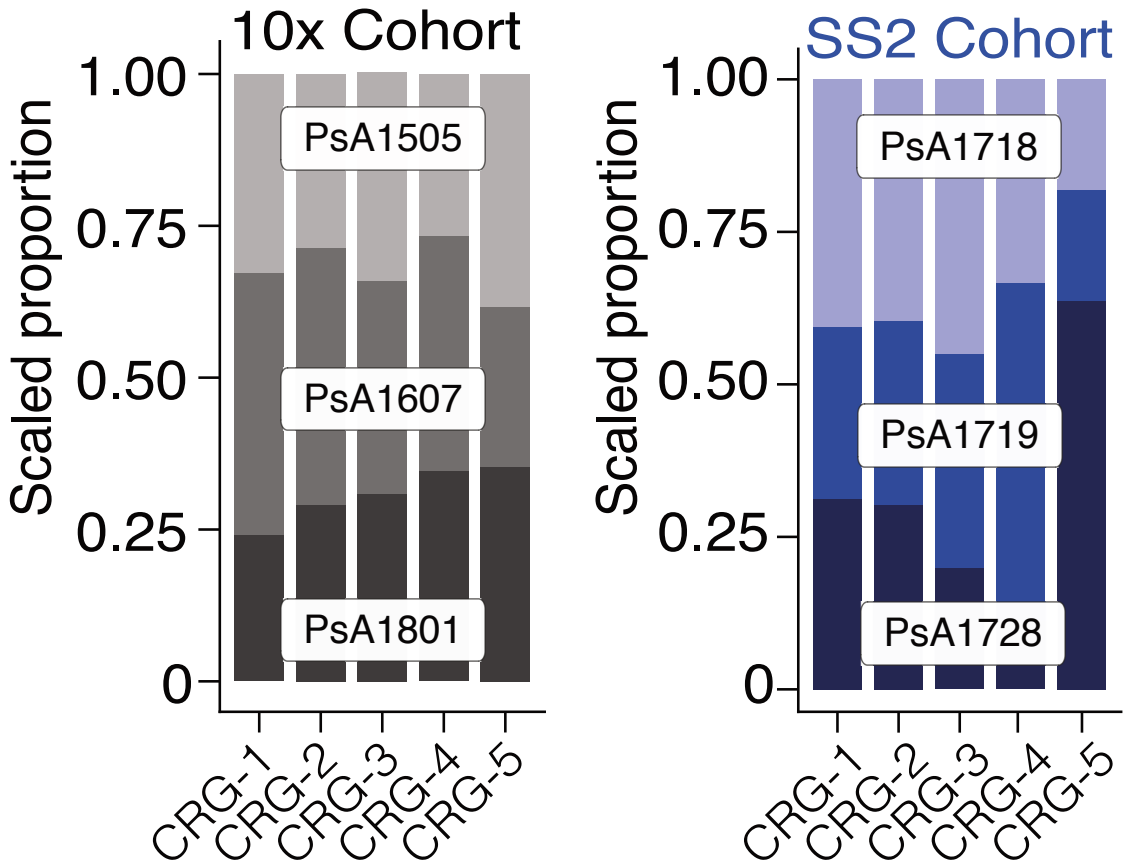


**Figure 5-7 CRG-1 contains a similar proportion of cells from all patients for both the 10x and SS2 patient cohorts.**

The proportion of cells from convergence groups (x-axis) belonging to each patient within either the 10x or SS2 cohorts. Proportions are scaled to take into account any difference in the frequency of a particular patient's cells relative to the whole dataset. The 5 convergence groups which contained CDR3 beta chain sequences from all 6 patients are shown. Colours represent patients. Different shades of grey represent different patients from the 10x cohort, and shades of blue represent different patients from the SS2 cohort.

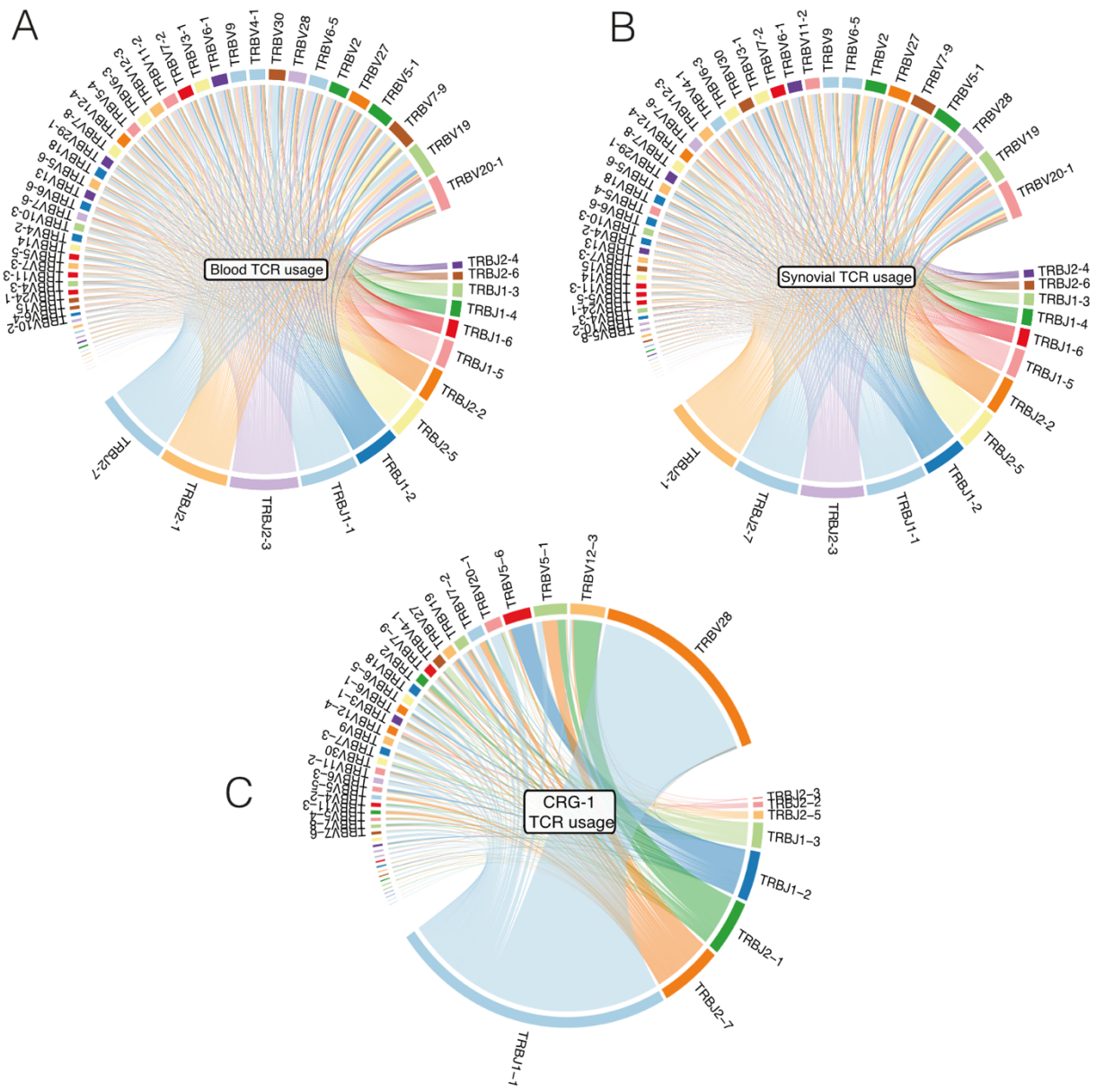


**Figure 5-8 CRG-1 cells belong to clonotypes enriched for TRBJ1-1 and TRBV28 gene usage compared to all cells from either peripheral blood or synovial fluid.**

Circle plots depicting the frequency of TCR beta chain variable and joining gene usage from 10x 5 prime data for memory T cells from (A) peripheral blood, (B) synovial fluid and (C) cells belonging to convergence group CRG-1. Arcs represent different variable and joining genes scaled to their frequency within the sample. Ribbons joining V and J genes represent pairings and the size of each ribbon is scaled to the particular pairing frequency within the sample.

### *5.3.5 Cells belonging to CRG-1 were most associated with Tissue and Activated CD8 T cell clusters*

Cells belonging to CRG-1 were next identified within the UMAP plot of all synovial cells forming part of the 10x integrated dataset to visualise their spatial distribution ( Figure 5-9A). The Tissue and Activated clusters, the 10th and 4th largest clusters by cell number in the integrated dataset, were also the 2 clusters containing the most cells from CRG-1, clearly disproportionate to what would be expected by chance (Figure 5-9B). Volcano plots of these 2 clusters (Figure 5-10) show that both clusters have an increased expression of effector and chemokine genes GZMB and CCL5 compared to all other clusters. However GNLY(encoding granulysin) is upregulated in the Tissue cluster but not the Activated cluster. Conversely, GZMK is only upregulated in the Activated cluster, together with many MHC-II genes. IFNG expression, although only detected in a small number of cells, was observed in both of these clusters (Figure 5-3), which together with the overexpression of other effector genes suggested that cells from both clusters may be actively participating in an immune response within the joints of these patients. By extension, the CRG-1 cells identified by GLIPH as having shared specificity could also be actively participating in the same immune response.

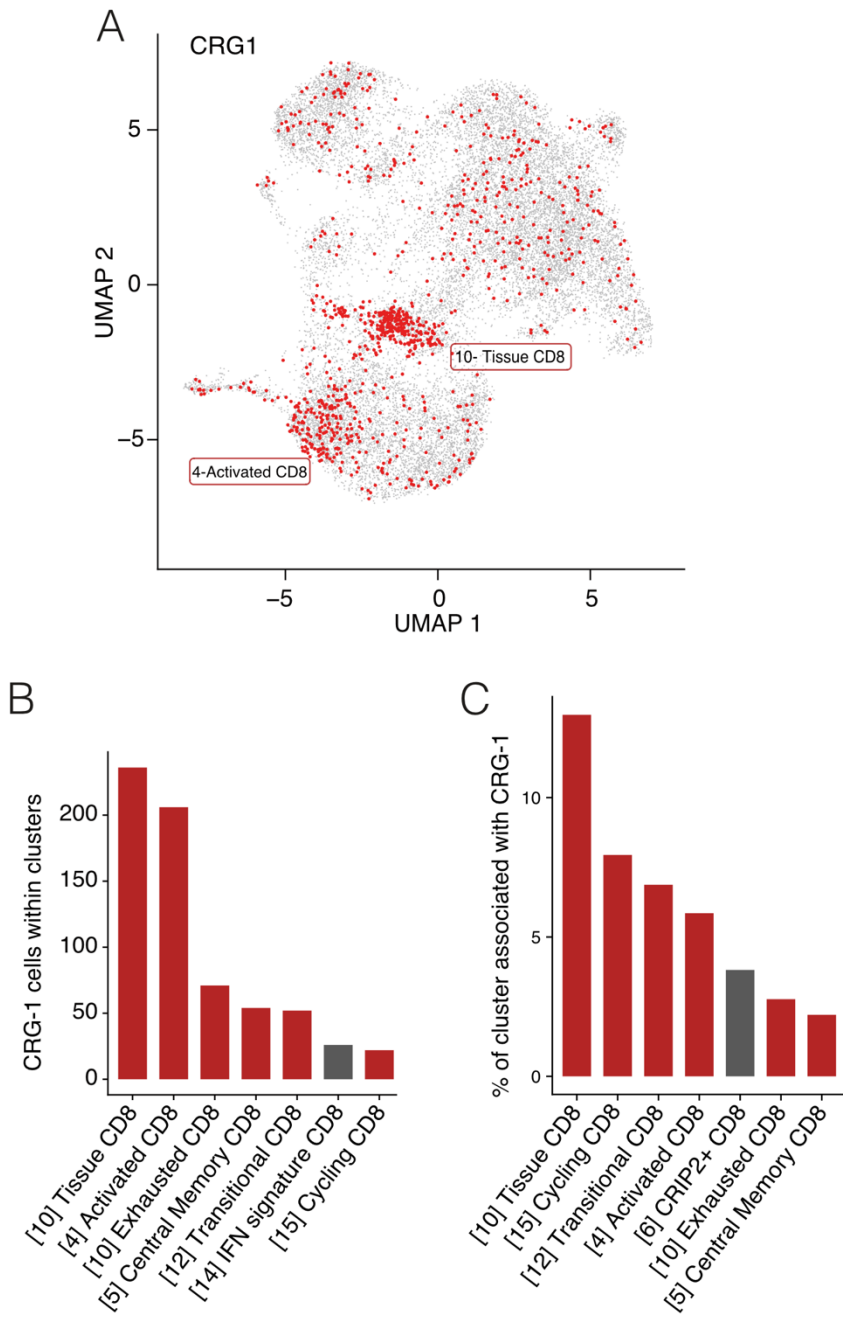


**Figure 5-9 Cells belonging to CRG-1 were most associated with Tissue and Activated CD8 T cell clusters among synovial cells from the 10x integrated dataset.**

(A): UMAP of synovial cells from the 10x integrated dataset (Figure 5-1B), with synovial cells from GLIPH convergence group CRG-1 highlighted in red. Each dot represents a cell. (B): The number of synovial CRG-1 cells within each of the synovial clusters. (C): The percentage of the CD8 cell component of each cluster which belonged to cells from CRG-1. Red bars indicate clusters designated as CD8. Grey bars indicate mixed clusters. Only top 7 clusters are shown for B and C.

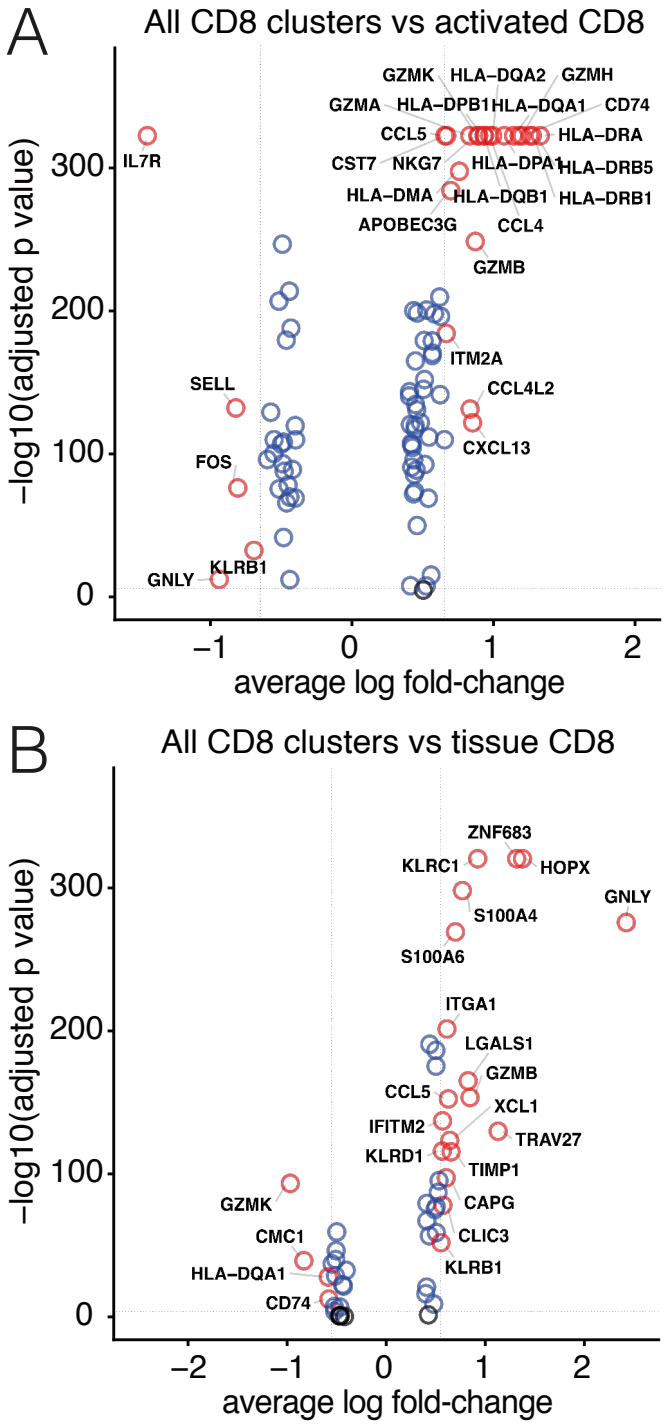


**Figure 5-10 The Activated and Tissue CD8 T cell clusters share upregulated GZMB and CCL5, but differ in the expression of GNLY and GZMK.**

Volcano plots showing differential gene expression of synovial fluid CD8 T cells in Activated CD8 cluster 4 (A) and Tissue CD8 cluster 10 (B) versus all other synovial fluid CD8 T cells in remaining clusters. Statistical significance calculated using Wilcoxon rank sum test. To help visualisation, only genes where the adjusted p-value < 0.0001 and average logFC is (A): < -0.55 or > 0.55 (B): < -0.65 or > 0.65 are labelled with the gene name and represented by red circles. Each circle represents a gene. Genes for which the average logFC was found to be between -0.4 and 0.4 were excluded from analysis.

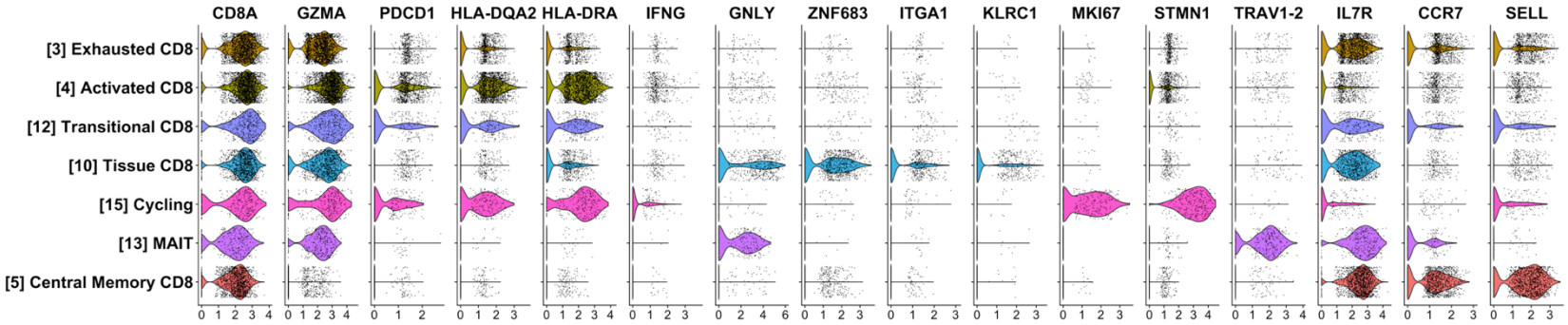


**Figure 5-11 Violin plots of the gene expression of key markers defining CD8 clusters among synovial cells of the integrated 10x dataset.**

Violin plots of the normalised count of gene transcripts for genes used to distinguish between clusters consisting primarily of CD8 *ex vivo* memory T cells among synovial cells of the 10x integrated dataset. Colours correspond to colours used to identify clusters from Figure 5-1. Each dot represents a cell. Numbered X-axis for each gene represents normalised number of transcripts detected for an individual cell within a cluster.

### *5.3.6 Pseudotime analysis of CD8 CRG-1 memory T cells reveals two trajectories in which the Tissue cluster represents an intermediate developmental state giving rise to activated, cycling and exhausted CD8 populations[3].*

When considering the percentage of each cluster associated with CRG-1 cells, it was clear that aside from Tissue and Activated clusters, CRG-1 cells also accounted for a considerable proportion of Cycling and Transitional cluster cells (Figure 5-9C). To better understand the relationship between CRG-1 cells across clusters, pseudotime analysis was carried out on only CRG-1 cells which formed a part of the CD8 component of any CD8 cluster. MAIT cells were excluded from the analysis as they were considered unlikely to form a meaningful lineage with other CD8 T cells, and their Th17-like phenotype could potentially interfere with the analysis. The pseudotime analysis revealed two trajectories sharing a common path from Central Memory through Tissue, Transitional and Activated CD8 clusters before bifurcating into either Cycling or Exhausted CD8 clusters (Figure 5-12). Details of how the analysis was conducted are included in the figure legend.

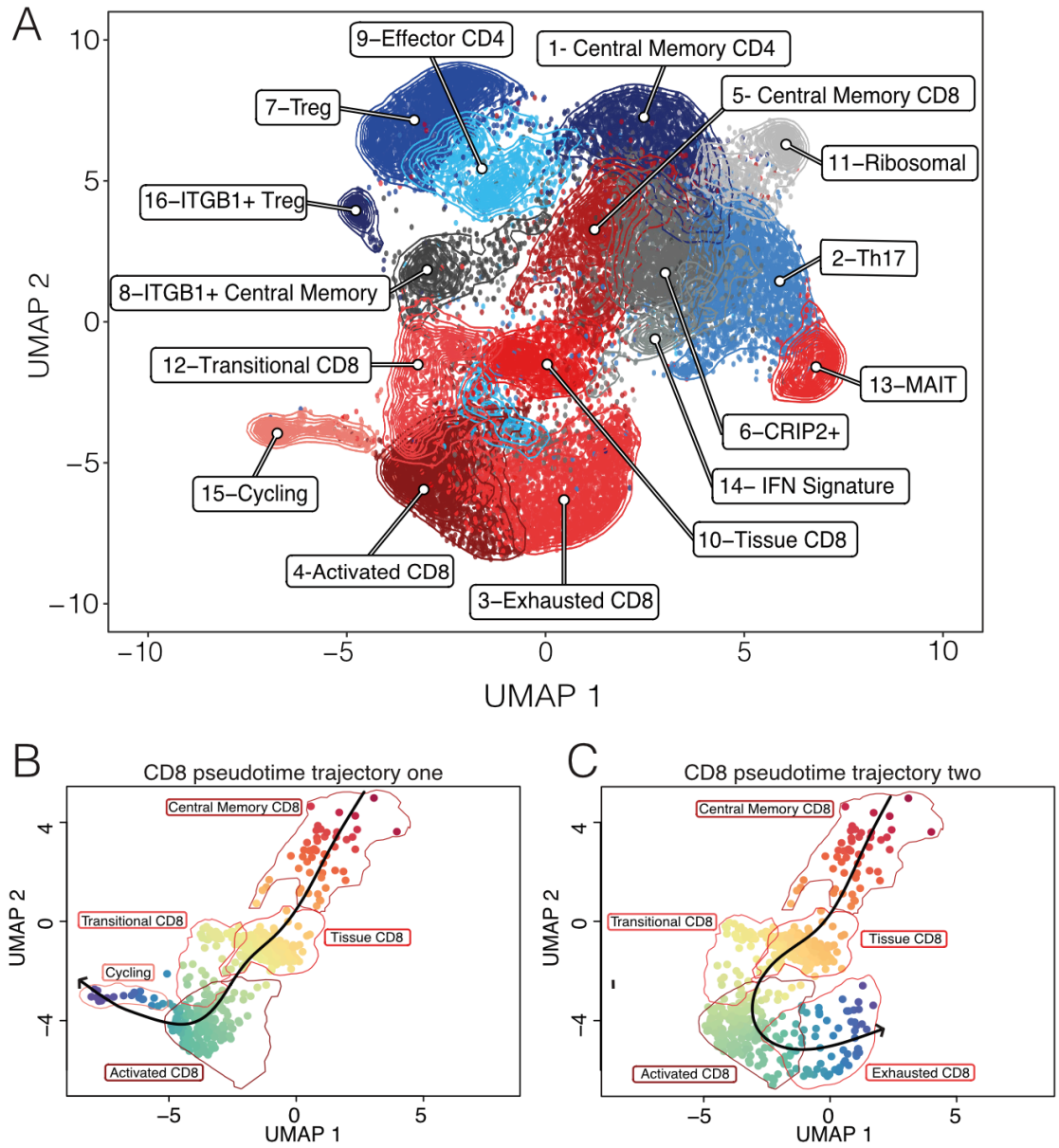


**Figure 5-12 Pseudotime analysis of CRG-1 CD8 memory T cells reveals two trajectories in which the Tissue cluster represents an intermediate developmental state giving rise to activated, cycling and exhausted CD8 populations.**

(A): The UMAP of synovial cells from Figure 5-1B shown for easy reference. (B and C): The location of synovial CD8 T cells from CRG-1 on the UMAP plot in (A) with 2 lineage trajectories sharing common paths from Central Memory to Activated CD8 clusters but then diverging into cycling cells (B) or Exhausted CD8 T cells (C). The 2 trajectories were calculated by the Slingshot R package and are overlaid on top of the CRG-1 cells in B and C using a black line with arrowhead. To perform the analysis, synovial cells belonging to both CRG-1 and any of the CD8 clusters 3, 4, 5, 10, 12 and 15 were first subsetted from the main integrated synovial dataset (A). The SCTransformed assay from this subset was then converted into a SingleCellExperiment object. This object was then further subsetted to include only highly variable genes classified as being robustly expressed, which was then passed to the slingshot function to calculate lineages. Only those genes represented by more than 3 transcripts in at least 10 cells within the dataset were considered robustly expressed, and only genes identified by Seurat as being highly variable and common to all patient and PB/SF data subsets were included for analysis, excluding any TCR genes. Cells from the MAIT cell cluster were excluded from analysis. Cells are coloured by pseudotime from red to blue.

### *5.3.7 Integrated analysis of 10x and SS2 datasets from patient PSA1607 identifies 13 clusters represented by cells sequenced across both platforms.*

Patient PSA1607 had the single most strongly synovial enriched clone of all 3 patients within the 10x cohort. The highest number of cells from both blood and synovial fluid were also sequenced for this patient using the 10x platform, and a relatively large (1043) number of cells from this patient were also sequenced using the SS2 platform. This provided an opportunity to explore the phenotype of the maximally expanded synovial clone for patient PSA1607 in more detail. The procedure for integrating the 10x and SS2 datasets from this patient were similar to those used when integrating all 10x datasets. Both 10x and SS2 datasets were first subjected to quality control, and the remaining cells were then subsampled for each platform to include an equal number of cells from both blood and synovial fluid (11192 for 10x and 433 cells for SS2). These datasets were then integrated, making use of the SCTransform algorithm for normalisation. The platform (10x/SS2) was additionally regressed out when scaling the integrated dataset (PSA1607-integrated). More details are available in methods chapter 2. At the same resolution (0.7) used when finding clusters for the main integrated dataset from the whole 10x cohort (PSA-integrated), 13 clusters rather than 16 were obtained (Figure 5-13). This is likely due to the smaller number of cells in the PSA1607-integrated dataset (23250 in comparison to 39252 cells). Differential gene expression analysis however identified clusters which could be confidently associated with key clusters identified in the PSA-integrated dataset (Figure 5-14).

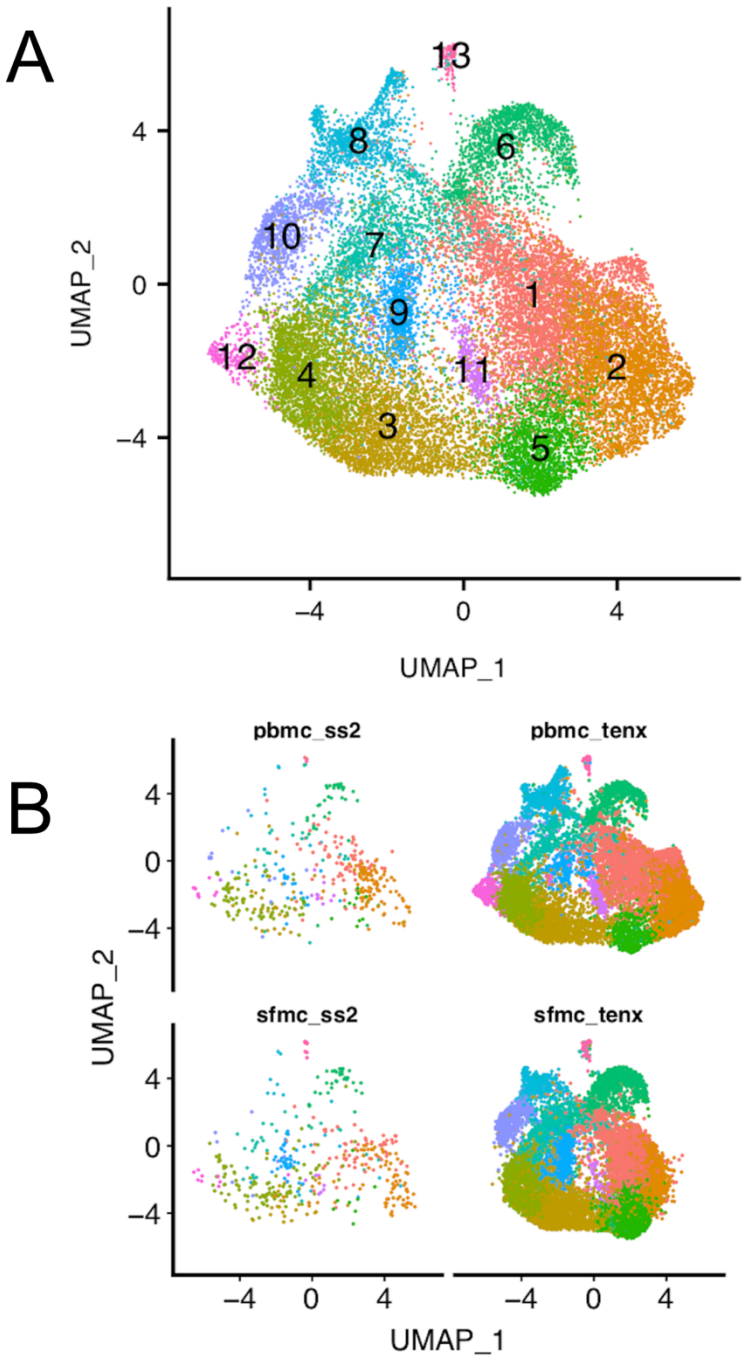


**Figure 5-13 Integrated analysis of 10x and SS2 datasets from patient PSA1607 identifies 13 clusters represented by cells sequenced across both platforms.**

(A): UMAP plot of PSA1607-integrated dataset combining *ex vivo* memory CD4 and CD8 T cells from the peripheral blood and synovial fluid of patient PSA1607, sequenced using both 10x and SS2 platforms. Seurat 3.0 integration was used in combination with the Seurat SCTransform function for normalisation of datasets. 13 phenotypic clusters were identified numbered from largest to smallest by number of cells in cluster. (B): Plot from (A) split to show distribution of *ex vivo* memory CD4 and CD8 T cells originating from PBMC or SFMC and from each platform. Each dot represents a cell.

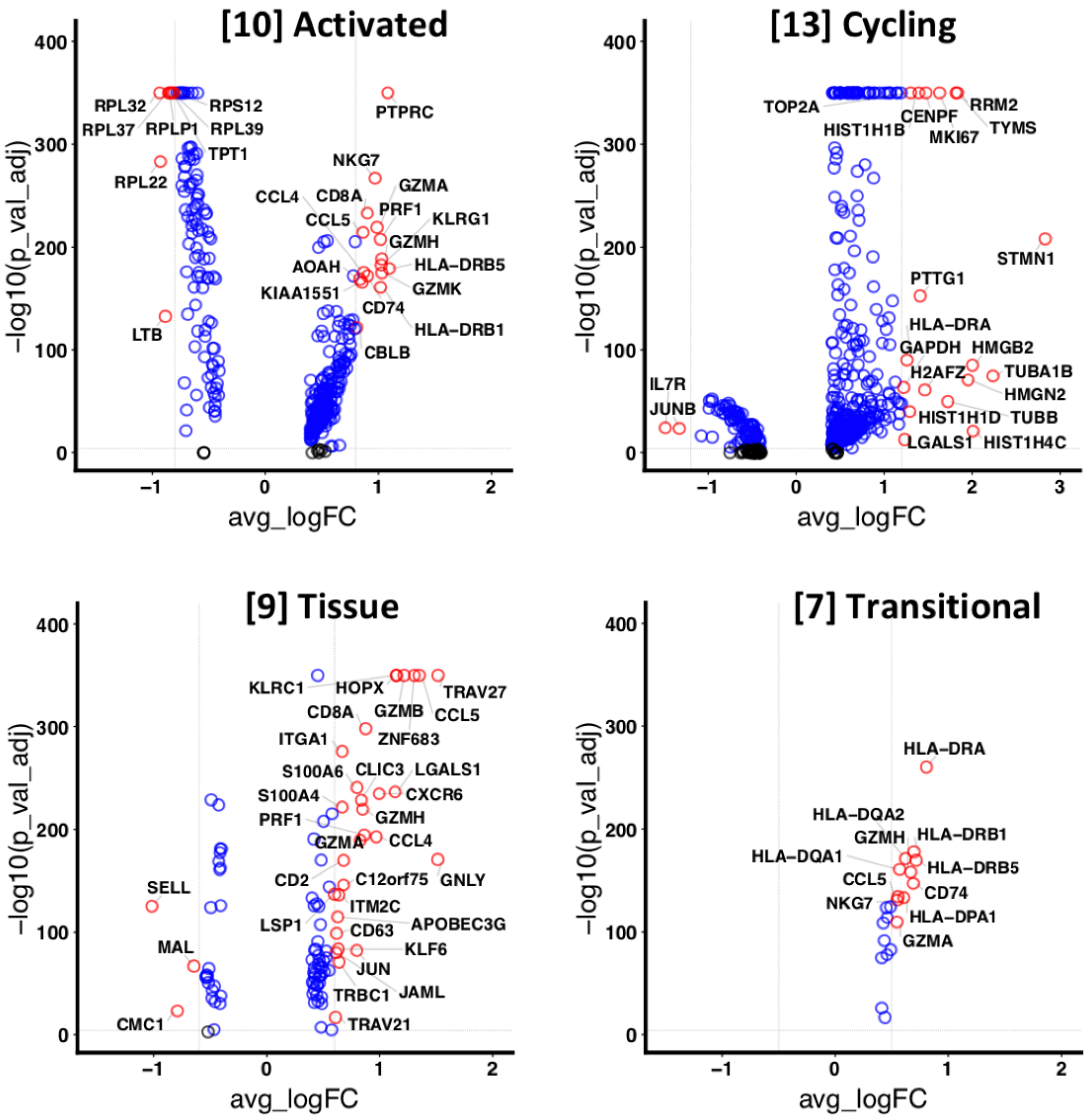


**Figure 5-14 The Activated, Tissue, Cycling and Transitional clusters in both PSA1607-integrated and PSA-integrated datasets are of a similar phenotype.**

Volcano plots showing differential gene expression of Activated, Tissue, Cycling and Transitional clusters from Figure 5-13A versus all other cells in remaining clusters. Statistical significance calculated using Wilcoxon rank sum test. To help visualisation, only genes significantly overexpressed by adjusted p-value < 0.5 and high or low average logFC are labelled with the gene name and represented by red circles. Each circle represents a gene. Genes with an average logFC > -0.4 or < 0.4 are not represented.

### *5.3.8 CD8 memory T cells from the most enriched clone in the synovial fluid of patient PSA1607 are present in multiple distinct phenotypic clusters.*

One key question which the large PSA1607-integrated dataset helps to answer is whether the synovial enriched clones present in PSA synovial fluid are each of a distinct phenotype, or whether multiple phenotypes are adopted by a single clone either before trafficking to the synovial fluid or at the site of tissue inflammation. Figure 5-15 clearly shows that cells from the most synovial enriched clone of patient PSA1607 are present in Memory, Transitional, Activated, Cycling and (predominantly) Tissue clusters (blue dots). The smaller number of cells belonging to the same clone and sequenced using the SS2 platform also cluster with Tissue cluster 9 (red dots), but are not seen in the other clusters. This highlights the importance of sequencing a sufficient number of cells to make valid conclusions regarding scRNA-seq data.

### *5.3.9 CD8 T cells from the most synovial enriched clone of patient PSA1607 follow two pseudotime trajectories, bifurcating at a transitional cell state towards activated and cycling cell clusters.*

Pseudotime analysis of the most synovial enriched clone (top clone) from patient PSA1607 revealed two similar but distinct trajectories to those identified by pseudotime analysis of CRG-1 cells (Figure 5-16 ). In both cases trajectories begin in the Central memory cluster, progressing through the Tissue cluster and into the Transitional cluster. At this point both trajectories identified for CRG-1 cells continue to the Activated cluster before bifurcating towards cycling or exhausted phenotypes. Some cells belonging to the top clone also continue to the Activated cluster but do not progress to an exhausted phenotype, whereas cells following the second trajectory identified for the top clone bypass an activated phase, directly entering a cycling cell state.

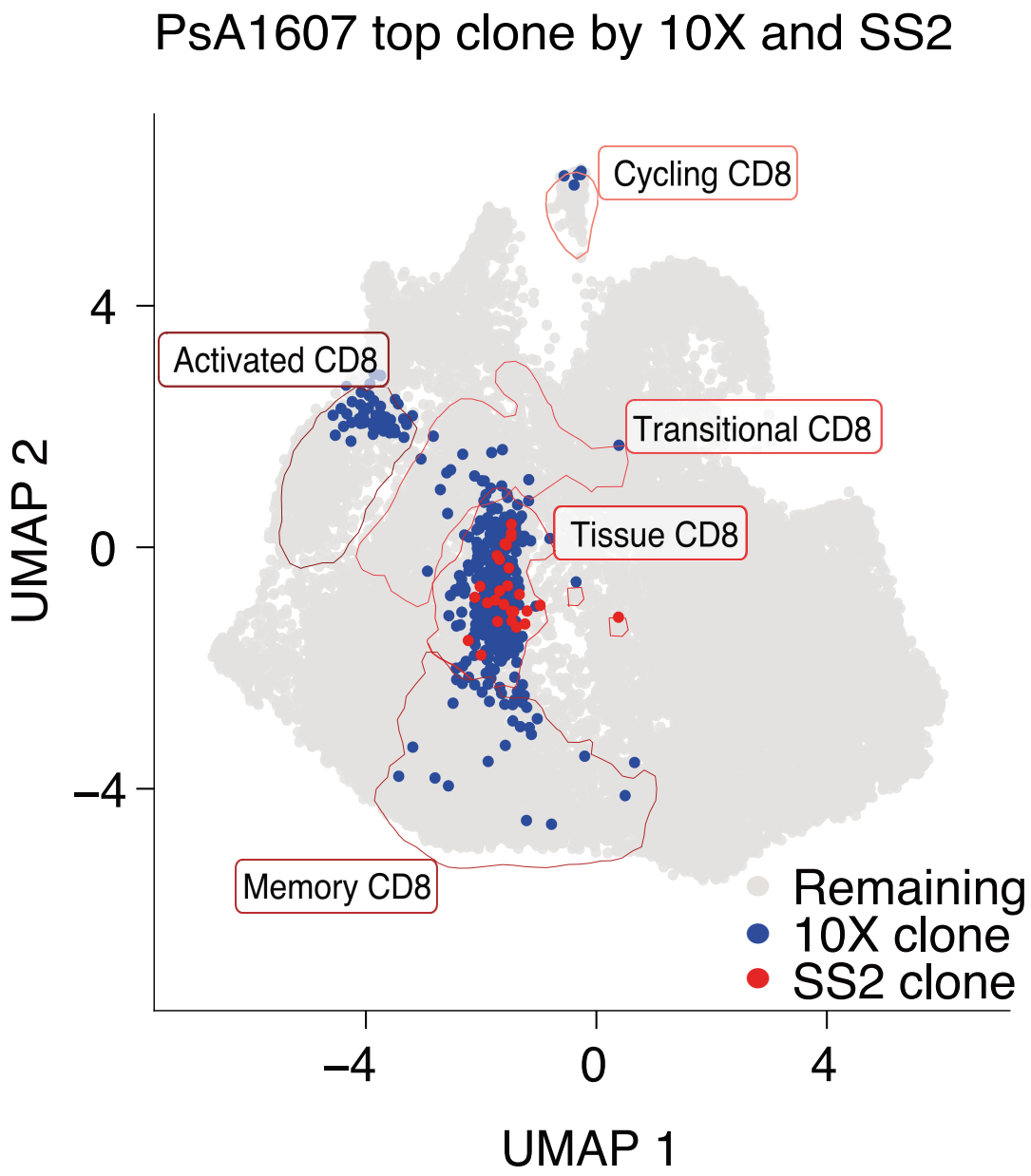


**Figure 5-15 CD8 memory T cells from the most enriched clone in the synovial fluid of patient PSA1607 are present in multiple distinct phenotypic clusters.**

UMAP plot of integrated memory CD4 and CD8 T cells from one donor (PSA1607) including cells from 5' 10x and SS2 data sets. Synovial CD8 T cells from the clone most enriched in synovial fluid for this patient is highlighted in blue for the 10x data set and red for the SS2 data set. Annotated clusters had upregulation of key markers shared with clusters identified in the main 10x integrated dataset representing 3 patients.

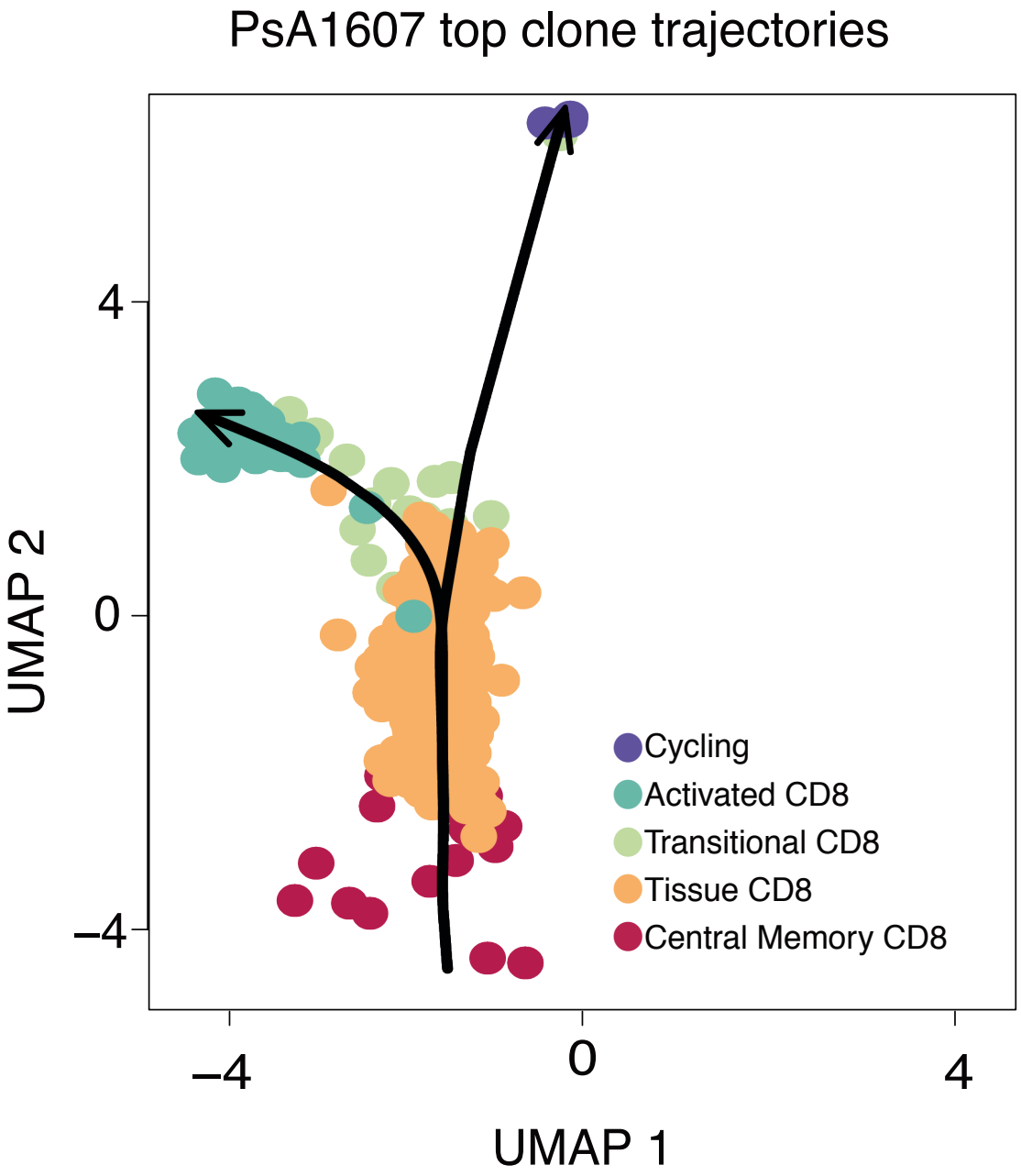


**Figure 5-16 Cells from the most synovial enriched clone of patient PSA1607 follow two pseudotime trajectories, bifurcating at a transitional cell state towards activated and cycling cell clusters.**

Pseudotime analysis of the most synovially enriched clone from donor PSA1607 showing shared central memory and tissue trajectory ending in either activated or cycling cell state. Cells are coloured by phenotypic cluster to which they belong. Pseudotime was calculated using the Seurat Slingshot algorithm on all (10x derived) cells belonging to the most synovial enriched clone of patient PSA1607 using the integrated dataset from Figure 5-15. Only the top 3000 genes identified by the Seurat SCTransform function as being able to explain the most variation between cells, further filtered for robustness to include only those genes represented by more than 3 transcripts in at least 10 cells within the integrated dataset were included in the analysis.

### *5.3.10 Ex vivo CD4 and CD8 memory T cell clusters express distinct pro-inflammatory and cell trafficking genes in peripheral blood and synovial fluid of PsA patients.*

The heatmap in Figure 5-17 summarises the similarities and differences in trafficking and pro-inflammatory gene expression between subsets of the 16 clusters identified in the PSA-integrated dataset. CRG-1 cells (red labels) were subsetted from their respective clusters to see whether they were driving a particular pattern of gene expression. In most cases CRG-1 subsets conformed to a very similar phenotype of the cluster they were subsetted from. Exceptions to this were the increased expression of CCR1 and CSF1 in the CRG-1 cells of Activated and Cycling synovial clusters; the increased expression of TNF for CRG-1 cells from the Activated synovial cluster; increased CX3CR1 and notably CSF1 expression in Tissue CRG-1 cells from blood; and increased CXCR6 expression in Tissue CRG-1 cells from synovial fluid.

When comparing the expression of cell trafficking genes between peripheral blood (PB) and synovial fluid (SF) subsets, chemokine receptors CCR1, CCR5, CXCR3 and CXCR6 were generally upregulated in SF by CD8 subsets and to a lesser extent CD4 subsets. However, MAIT cells maintained comparable expression of these genes in both blood and synovial fluid, and particularly overexpressed CCR1 in blood relative to other blood subsets. In contrast to the above mentioned chemokine receptors, gene expression of CX3CR1 was strongly overexpressed in PB relative to SF, especially for Activated, Transitional, Cycling and to a lesser extent Tissue CD8 subsets. High gene expression of the integrin subunit ITGA1 (encoding CD49a) was maintained for the Tissue cluster in both PB and SF but otherwise increased only in the SF subsets of Transitional and Cycling CD8 clusters, suggesting that these clusters may share a degree of tissue residency phenotype or common ancestry in SF. ITGA4 was most strongly expressed in the PB and SF Transitional cluster subsets and average expression

was also higher in PB and SF ITGB1+ subsets. Gene expression of the chemokines listed in Figure 5-17 was primarily restricted to CD8 subsets, with the exception of CXCL13 which was overexpressed in the mixed ITGB1+ Central Memory and Effector CD4 clusters in SF. Strikingly CCL3 and CCL23 were strongly upregulated only in Activated and Cycling SF subsets compared to all other subsets, and within Activated and Cycling clusters CXCL13 was also only upregulated in SF.

Although cytokine expression was typically restricted to a smaller number of cells within clusters, the increased average gene expression of cytokines in SF was striking compared to PB. In particular, IFNG gene expression was increased in the SF Activated, Transitional, Tissue and Cycling CD8 cluster subsets, in addition to the ITGB1+ Central Memory and Effector CD4 SF cluster subsets. TNF expression was also generally higher in the SF CD8 and CD4 cluster subsets, but less so in Treg subsets. In contrast, both Treg subsets had high average gene expression of CSF1 in SF, a trait shared with the Cycling and Activated CRG-1 subsets already mentioned.

The strongest increase in SF granzyme expression by clusters compared with PB was seen for GZMB, however among PB subsets, GZMB was also strongly expressed by the CRG-1 Tissue subset. The S100A4, S100A6, S100A10 and S100A11 genes were all strongly upregulated in the majority of SF subsets. Interestingly, within PB, S100A4 and S100A6 were most strongly upregulated among the Tissue CRG-1 subset, whereas in SF both the CRG-1 Tissue subset and remaining Tissue cluster subset upregulated these genes. Upregulation of HLA-DRA and HLA-DMA associated most strongly with the Activated and Cycling clusters as expected, however it was interesting that this association was maintained, particularly for Cycling cells in PB.

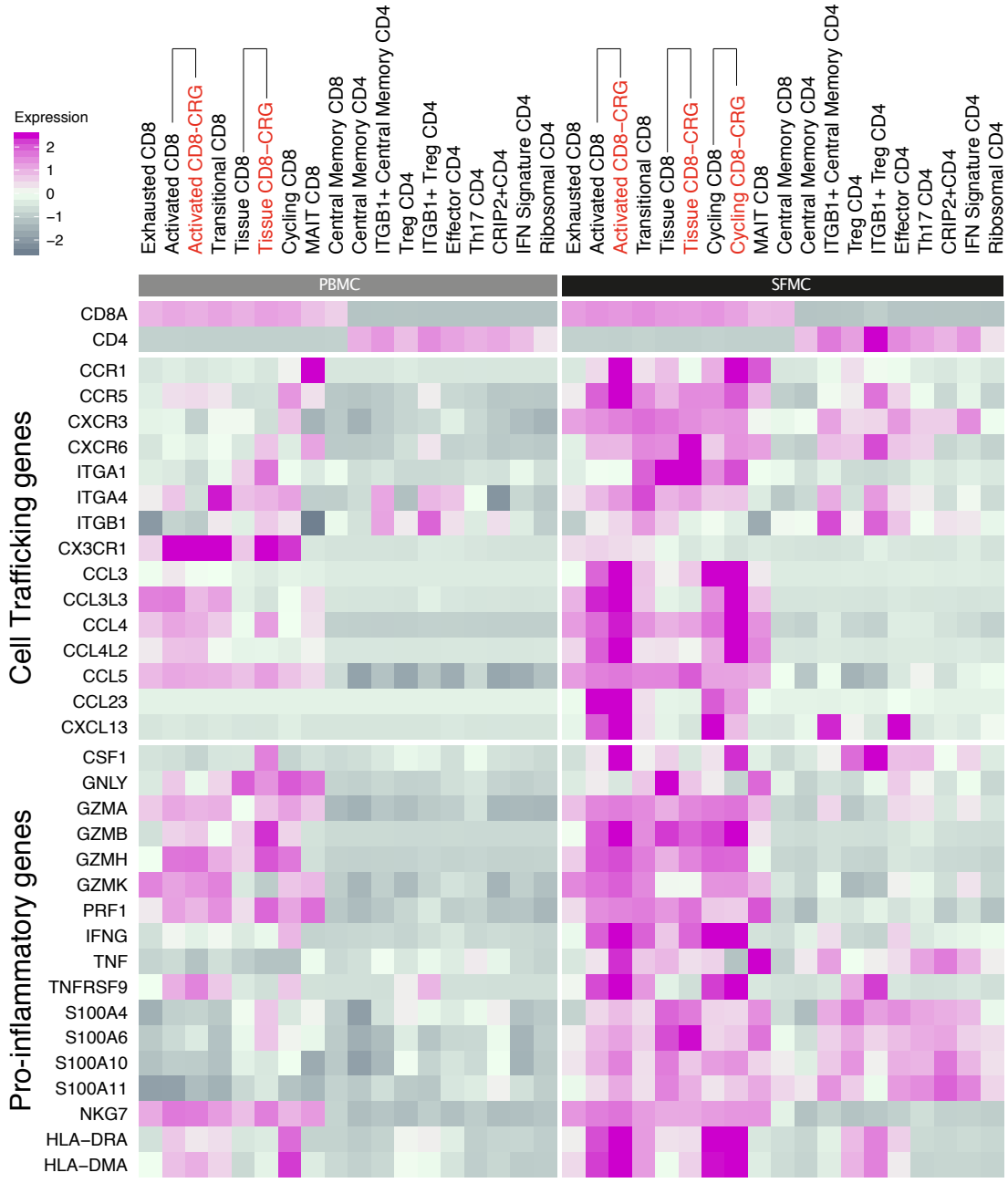


**Figure 5-17 *Ex vivo* CD4 and CD8 memory T cell clusters express distinct pro-inflammatory and cell trafficking genes in peripheral blood and synovial fluid of PsA patients.**

A heatmap of average pro-inflammatory and cell-trafficking gene expression from *ex vivo* CD4 and CD8 memory T cells subsetted according to the groups listed along the top of the figure. Cells were first subsetted according to the 16 clusters identified during the integrated analysis of 10x datasets from 3 PsA patients (Figure 5-1), then further subsetted by peripheral or synovial origin, and finally by membership in the CRG-1 convergence group where indicated. CRG-1 cells subsetted from Activated, Tissue and Cycling clusters are compared only with remaining cells from those clusters (no cells are represented more than once). No CD8 cycling cells belonging to CRG-1 originated from PBMC. The intensity of purple colour indicates level of gene expression according to the scale shown, based on SCTransformed, normalised and scaled average gene expression. Either CD4 or CD8 cells were selected to represent each of the 16 phenotypic clusters based on the majority of cells classified as either of these 2 cell types within each cluster. CD4 or CD8 T cells forming a minority within each cluster were excluded from analysis. Whether CD4 or CD8 cells were included for each cluster is indicated by the cluster label.

## 5.4 Discussion

In contrast to chapter 4, this chapter dealt entirely with unstimulated *ex vivo* CD4 and CD8 memory T cells from the blood and synovial fluid of multiple PSA patients. In addition, multiple scRNA-seq platforms (10x and SS2) were used to generate datasets. Two datasets (for blood and synovial fluid) were generated for each of the 6 patients, in addition to 1 patient sequenced using both platforms, resulting in a total of 14 gene expression matrices requiring analysis. Within each of these expression matrices, there was an additional implied subdivision of memory T cells into CD4+ and CD8+, increasing the theoretical number of datasets to keep track of during the analysis to 28. In addition, the acquisition of V(D)J sequencing data which could be mapped back to each of these 28 datasets effectively increased the number of distinct datasets to 56. This illustrates how quickly scRNA-seq experiments can result in very complex datasets for analysis. Finding the most logical way to integrate and analyse these already high-dimensional datasets to glean the most biologically relevant information proved to be the most challenging aspect of this chapter. Luckily, dedicated labs are working to provide software pipelines to assist with scRNA-seq analysis, and this chapter made heavy use of the Seurat[10] pipeline (which continues to evolve rapidly). However, these tools mainly deal with gene expression analysis. As our lab were very early adopters of TCR sequencing using the 10x platform, no tools existed to specifically aide with this aspect of the analysis. The code required to import the TCR output of Cell Ranger into R and convert it to a workable format had to be developed, and tools such as GLIPH[2] and VDJ tools, which were designed to accommodate bulk sequencing of TCR beta chains were adapted for integration into the workflow.

One of the aims of this chapter was to validate the presence of clonally expanded CD8 T cells in the synovial fluid of PsA patients, first seen in patient PSA1607. The V(D)J

sequencing of 2 additional patients using the 10x platform supported this finding. The V(D)J sequencing of a limited number of cells (approximately 500 synovial and 500 blood CD8 T cells) from 3 more patients using the SS2 platform was considered underpowered in detecting clonal expansion and could therefore not be used to support or refute the finding. Although a clonal expansion was detected in synovial CD8 T cells using SS2 for patient PSA1607, this patient was already known to have a particularly strong expansion of a single clone relative to other patients from the 10x cohort, and therefore this clone had a higher probability of being classified as enriched in the SS2 dataset. SS2 TCR clonal information for patient PSA1607 was instead used to complement the 10x dataset for this patient with full length sequencing of the same clone. It also provided a matching alpha chain which had not previously been detected by 10x sequencing. Although the possibility of the 10x clone having a different alpha chain or being associated with multiple alpha chains cannot be excluded, it is reasonable to infer from the similar enrichment between both platforms and also differential gene expression of TRAV27 in the 10x dataset for this clone (data not shown) that V(D)J and CDR3 sequences discovered by both platforms were of the same clone (or that the SS2 paired sequences at least represented the majority of cells belonging to the 10x clone). In addition the PSA1607-integrated dataset showed that cells of this clonotype sequenced by both platforms associated most strongly with the Tissue cluster (Figure 5-15)

Clonal expansions in the synovial fluid of PsA patients have previously been reported[1]. However, to the best of my knowledge these expansions have never before been explored using scRNA-seq to determine both the precise paired-chain sequences of expanded clones and attempt to define their phenotype. This chapter first described an integrated approach to analysing the phenotype of cells belonging to the maximally

expanded clones of patients PSA1505 and PSA1607 together with cells having related TCR specificity towards a common antigen as determined by the GLIPH algorithm (collectively termed CRG-1). CRG-1 included a similar proportion of cells from all three patients in the 10x cohort, and although no synovial enriched clones from patient PSA1801 were associated with CRG-1, it may be that a more diverse TCR repertoire exists for this patient cognate for the same antigen. This integrated approach relied to some extent on findings of the GLIPH analysis, however given that a large fraction of the cells belonging to CRG-1 were also members of expanded clones, it should be largely representative of an analysis of only expanded clones (an analysis I am currently undertaking). The association made by GLIPH also provides grounds to argue that a particular antigen, potentially an auto-antigen, may be driving clonal expansion in the joints of these patients. To specifically address clonal phenotype in isolation, the maximally enriched clone from patient PSA1607 was also mapped back to a separately integrated dataset of the total (filtered) gene expression data available for this patient. The results of pseudotime analysis on cells from both the top clone of PSA1607 and the CRG-1 convergence group showed similar trajectories, sharing a common path segment between Central Memory, Tissue and Transitional clusters. Interestingly the Tissue cluster (and in particular the CRG-1 subset of cells from the tissue cluster) had similar average expression of many of the cell trafficking and pro-inflammatory genes shown in (Figure 5-17) in both blood and synovial fluid, including ITGA1 (encoding CD49a). Psoriasis typically precedes large joint involvement in PsA, and although not possible to conclude from the data available in this chapter, it is tempting to speculate that upon activation, a portion of CD49a+ skin resident CD8 T cells[11], sharing many markers in common with the "Tissue" cluster identified in this chapter, could traffic to joints in PsA patients. There, upon encountering cognate antigen, these cells could then

be driven towards a phenotype represented by the Activated and Cycling clusters identified in this chapter, leading to inflammation and tissue destruction.

The focus of this chapter has been on characterising *ex vivo* memory CD8 T cells in PsA patients, however analysis of the memory CD4 T cell clusters identified in this chapter, and their potential relationship to the Th17.1 phenotype of *Salmonella* stimulated cells discussed in chapter 4 is ongoing.

# I. References


1. Costello, P.J., Winchester, R.J., Curran, S.A., Peterson, K.S., Kane, D.J., Bresnihan, B., and FitzGerald, O.M. (2001). Psoriatic Arthritis Joint Fluids Are Characterized by CD8 and CD4 T Cell Clonal Expansions that Appear Antigen Driven. The Journal of Immunology *166*, 2878–2886.

2. Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., *et al.* (2017). Identifying specificity groups in the T cell receptor repertoire. Nature *547*, 94–98.

3. Penkava, F., Velasco-Herrera, M.D.C., Young, M.D., Yager, N., Lara, A.L., Guzzo, C., Maroof, A., Mamanova, L., Cole, S., Efremova, M., *et al.* (2019). Single-cell sequencing reveals a clonal expansion of pro-inflammatory synovial CD8 T cells expressing tissue homing receptors in psoriatic arthritis. bioRxiv, 704494.

4. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. bioRxiv, 576827.

5. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21.

6. Ciucci, T., Vacchio, M.S., Gao, Y., Ardori, F.T., Candia, J., Mehta, M., Zhao, Y., Tran, B., Pepper, M., Tessarollo, L., *et al.* (2019). The Emergence and Functional Fitness of Memory CD4+ T Cells Require the Transcription Factor Thpok. Immunity *50*, 91-105.e4.

7. Huang, H., Sikora, M.J., Islam, S., Chowdhury, R.R., Chien, Y., Scriba, T.J., Davis, M.M., and Steinmetz, L.M. (2019). Select sequencing of clonally expanded CD8+ T cells reveals limits to clonal expansion. PNAS *116*, 8995–9001.

8. Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J.Y., Zhang, Q., *et al.* (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell *169*, 1342-1356.e16.

9. Park, D., Kim, H.G., Kim, M., Park, T., Ha, H.-H., Lee, D.H., Park, K.-S., Park, S.J., Lim, H.J., and Lee, C.H. (2019). Differences in the molecular signatures of mucosal-associated invariant T cells and conventional T cells. Sci Rep *9*, 1–10.

10. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495–502.

11. Cheuk, S., Schlums, H., Gallais Sérézal, I., Martini, E., Chiang, S.C., Marquardt, N., Gibbs, A., Detlofsson, E., Introini, A., Forkel, M., *et al.* (2017). CD49a Expression Defines Tissue-Resident CD8+ T Cells Poised for Cytotoxic Function in Human Skin. Immunity *46*, 287–300.

# Chapter 6: Conclusion and future directions

This thesis made use of 3 single-cell technologies to characterise T lymphocyte immune responses in SpA. The first, fluorescence activated flow cytometry, has a long history with the first commercially available FACS machines appearing in the early 1970's[1]. This technology has been refined over many years, continues to be developed, and is widely considered to be the gold standard for rapid detection and enumeration of cellular populations [2]. It is a very valuable tool for both stand-alone phenotyping of cells in addition to the isolation and enrichment of cells for downstream experiments. A FACS machine was used throughout this study, firstly to phenotype the functional response of rare microbe-reactive Th memory cells, and later to isolate both microbe-reactive Th memory cells and *ex vivo* memory CD4 and CD8 T cells. The high-throughput capacity of FACS to analyse individual cells showed utility in detecting rare microbe-reactive populations of Th cells in chapter 3, however this technique was not without its limitations. The most obvious limiting factor in terms of analysis is the relatively small number of parameters which can be tested (comfortably 10, at most 30). To some extent this is being addressed by new technologies such as spectral flow cytometry, however this form of analysis is still very much hypothesis driven and still requires antibodies to identify proteins of interest, which can themselves be expensive, fluctuate between lots and bind non-specifically. Additionally, when used to isolate cells, FACS does so serially, one cell at a time. This generally provides very good purity of the sorted population of interest, and is flexible in terms of gating strategies which can be used to identify cells. However, if a larger number of cells is desired for downstream work this may also create a bottleneck. Alternate techniques such as magnetic bead separation could for example be used to separate 100 000, rather than 10 000 *Salmonella*-reactive CD154+Th memory cells

from SFMC for downstream scRNA-seq within an hour. Isolating the same number of cells using FACS would take 5 - 10 hours. For the experiments set out in this thesis however, FACS was an appropriate tool and served its purpose well.

The other 2 techniques used to analyse single cells in this thesis were the 10x droplet based scRNA-seq platform, and the plate-based SS2 platform. Both of these techniques were used to determine the relative level of mRNA gene transcripts and the TCR alpha and beta chain paired CDR3 sequences within T cells. These techniques benefited from not requiring any specific labelling of features and provided a hypothesis-free approach to characterise the gene expression of target cells. Disadvantages relating to both techniques were primarily cost, especially the per cell cost of SS2 scRNA-seq, and the possible disconnect between mRNA expression and protein expression, as seemed to be the case for both IL-22 and CD4 in chapter 4 of this thesis. There is also a risk that sparsely expressed gene transcripts can go undetected due to a lack of read depth, which is more likely using the 10x platform. This could lead to the false conclusion that a gene is not being expressed if no transcripts are detected, or to an under-appreciation of the frequency of cells expressing a gene. An example of the later point which occurred in chapter 4 was the increased sensitivity to cytokine gene expression when a read depth of 90 0070 reads / cell was applied to the *Salmonella*-reactive Th memory dataset (Table 4-3). However, it is also possible that the Cell Ranger aggr function was too aggressive in its normalisation of the combined stimulated and unstimulated datasets for patient PSA1607. Although the high read depth used in SS2 is likely to avoid many of these issues, the higher cost SS2 and typically lower cell numbers sequenced as a result, can lead to false conclusions if not enough cells are sequenced to begin with. In chapter 5 for example, if only relying on SS2 data, one might have mistakenly surmised that clonality of the maximally enriched clone was restricted to a

specific cell phenotype. This study made use of both SS2 and 10x, and in doing so helped mitigate the chance of drawing false conclusions from an overreliance on any one scRNA-seq platform.

The overall aim of this study was to characterise T lymphocyte immune responses in SpA. Chapter 3 addressed this task by testing the hypothesis that an aberrant or enhanced immune response existed in SpA patients towards known commensal or pathogenic microbes typically found in the human gut. The Th memory compartment within both blood and synovial fluid was targeted for investigation. It was expected that either the type of cytokines produced by SpA Th memory cells in response to stimulation with gut microbial antigen would differ from healthy controls and/or RA patients (an aberrant response), or that an increased frequency of cells reactive to some or all of these microbes would be detected in SpA (enhanced response). The main findings of this chapter were that Th memory immune responses against microbes capable of inducing IL17A+ and particularly IL17A+IFNg+ production were enhanced in the synovial fluid of SpA patients relative to their blood; that SpA patients had an impaired (aberrant)Th memory IL-2 response against *Salmonella typhimurium* and *Mycobacterium tuberculosis* in blood compared to both RA patients and healthy controls; and that a pre-existing population of CD154+ Th memory cells were present in the synovial fluid of SpA patients without any *in vitro* stimulation. These findings warranted further characterisation of the SpA Th17/Th17.1 response towards IL17A inducing gut microbes, and *S. Typhimurium*, already having a known association with reactive arthritis, was chosen as a model Th17.1 inducing stimulus to explore these responses further. The main limitation of these findings was the inability to compare SpA synovial fluid responses to RA controls due to a lack of available samples. The CD154 assay was not compatible with frozen samples, making it difficult to plan for

the availability of sufficient synovial fluid samples ahead of time. However investigations are ongoing.

scRNA-seq seemed an ideal fit for characterising the SpA Th17.1 response to *Salmonella* further, which was the focus of chapter 4. This revealed very distinct subsets of CD154+ Th helper cells responding to stimulation, including a Th1 subset expressing high levels of not only TNF and IFNG, but also, almost exclusively, CSF2 (encoding GM-CSF) and IL-2. In addition, a portion of the cells within this Th1 subset also expressed IL17A, thus specifically identifying the cells behind an SpA Th17.1 response. Two other subsets shared expression of IL-17A, one of which produced high IL-22 and IL-17F while the other took on a more T regulatory phenotype producing IL-10. Another subset expressing high PLAC8, CXCR5, TCF7, and CD27 resembled a "stem-like" memory subset and shared these characteristics with a neighbouring proliferative subset of Th memory cells. These last two clusters would be hidden from functional assays only measuring cytokine output, as their cytokine producing program was completely switched off, even though they expressed CD154 on the cell surface in response to stimulation with *Salmonella* lysate. Pseudotime and clonal analysis both indicated a relationship linking these clusters together, unveiling a trajectory along which Th17.1 cells may develop into either regulatory or stem-like memory cells with proliferative capacity. Although further experiments would be required to validate these findings, for this patient, scRNA-seq provided a high resolution snapshot of the varied but related Th immune subsets responding towards stimulation with *S. typhimurium,* a known causative agent of reactive arthritis. Furthermore, all microbial stimulations carried out as part of this thesis used microbial lysate, not live microbes. Therefore Th memory cells reactive to antigens within *Salmonella* lysate, and potentially cross-reactive to other gut microbes (as has been demonstrated for *E. coli*[3]) were found in

the synovial fluid of SpA patients, and have the potential to be activated, if for example immunogenic peptides were to find their way from the gut to the joints. Alternatively, given the transient nature of flares in PsA, another possibility is that microbe-reactive Th17 cells stimulated at barrier sites such as the gut, traffic to the joints during periods of increased inflammation in these patients.

In addition to characterising the memory Th17 response to *Salmonella* stimulation in patient PSA1607, chapter 4 also found a strong clonal enrichment of CD8 memory T cells in the *ex vivo* synovial fluid of this patient compared with their peripheral blood. Chapter 5 built on these findings, confirming that an enrichment of CD8 memory T cell clones existed in the synovial fluid of 3 PsA patients. Chapter 5 then characterised the phenotype of both CD4 and CD8 ex vivo memory T cells, with an emphasis on clonally expanded CD8 T cells which potentially shared specificity towards a common antigen. This analysis showed that pro-inflammatory and clonally expanded CD8 memory T cells were present in the synovial fluid of PsA patients. The analysis also revealed pseudotime trajectories implicating CD8 T cells with an ITGA1/ZNF683 "Tissue resident" phenotype representing an intermediate developmental state giving rise to activated, cycling and exhausted CD8 T cell populations[4].

The conclusions reached in this thesis support a role for both CD4 and CD8 T cells in the pathogenesis of SpA. The enrichment of Th17.1 cells reactive to the gut microbes *S. typhimurium* and *E. coli* in synovial fluid of PsA patients is especially of interest in the context of the strong evidence implicating Type 17 immune responses with SpA. It was also interesting that lysate from *C. trachomatis*, which is not a gut microbe but (in common with *S. typhimurium*) is a known causative agent of reactive arthritis, induced a very similar type 17 immune response to *S. typhimurium*. The finding by Breban et al. that CD4 T cells are able to transfer disease phenotype in an animal model

independently of CD8 T cells[5], and the clinical success of anti-IL17 drugs in treating SpA[6] has placed a great deal of emphasis on CD4 T cells as drivers of disease in SpA. However, the well-established association of MHC-I genes with SpA has long implicated CD8 T cells, and it is hoped that the findings from chapter 5 will encourage further research into the role of clonally expanded CD8 T cells in the pathogenesis of SpA. In particular, the identification of paired-chain TCR CDR3 sequences belonging to clones enriched in the synovial fluid of PsA patients can be used to help identify cognate antigens towards these clones using yeast display[7], which may shed light on autoantigen(s) associated with disease.

# I. References


1. Herzenberg, L.A., Parks, D., Sahaf, B., Perez, O., Roederer, M., and Herzenberg, L.A. (2002). The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. Clinical Chemistry *48*, 1819–1827.

2. Rane, A.S., Rutkauskaite, J., deMello, A., and Stavrakis, S. (2017). High-Throughput Multi-parametric Imaging Flow Cytometry. Chem *3*, 588–602.

3. Hegazy, A.N., West, N.R., Stubbington, M.J.T., Wendt, E., Suijker, K.I.M., Datsi, A., This, S., Danne, C., Campion, S., Duncan, S.H., *et al.* (2017). Circulating and Tissue-Resident CD4+ T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. Gastroenterology. Available at: http://www.sciencedirect.com/science/article/pii/S0016508517359796 [Accessed October 13, 2017].

4. Penkava, F., Velasco-Herrera, M.D.C., Young, M.D., Yager, N., Lara, A.L., Guzzo, C., Maroof, A., Mamanova, L., Cole, S., Efremova, M., *et al.* (2019). Single-cell sequencing reveals a clonal expansion of pro-inflammatory synovial CD8 T cells expressing tissue homing receptors in psoriatic arthritis. bioRxiv, 704494.

5. Breban, M., Fernández-Sueiro, J.L., Richardson, J.A., Hadavand, R.R., Maika, S.D., Hammer, R.E., and Taurog, J.D. (1996). T cells, but not thymic exposure to HLA-B27, are required for the inflammatory disease of HLA-B27 transgenic rats. Journal of Immunology *156*, 794–803.

6. Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., van der Heijde, D., McInnes, I., van Laar, J.M., Landewé, R., Wordsworth, P., *et al.* (2013). Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. The Lancet *382*, 1705–1713.

7. Birnbaum, M.E., Dong, S., and Garcia, K.C. (2012). Diversity-oriented approaches for interrogating T-cell receptor repertoire, ligand recognition, and function. Immunological Reviews *250*, 82–101.