Investigating the genetics and immunopathogenesis of ankylosing spondylitis



Feng Liu

Wolfson College

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences

University of Oxford

A thesis submitted for the degree of D.Phil

Abstract

Ankylosing spondylitis is an immune-mediated inflammatory arthritis mainly affecting the spine, although the peripheral joints, skin, intestine, and eyes could also be affected. I investigated the genetic risks and immune dysfunction of AS to achieve a better understanding of the underlying pathogenic mechanisms.

Genetic association studies have made significant progress in AS. I hypothesized that studying the common genetic risks among AS and its associated conditions could help prioritize risk genes as potential drug targets and reveal potential opportunities for drug repurposing. I assessed the genetic risks overlapping between AS and 30 other diseases and identified 9 conditions that are genetically close to AS. Utilizing these shared genetic risks, I found a risk locus at the intron of *ZMIZ1* shared by AS and IBD, which regulates the expression of *PPIF* through an enhancer-enhancer-promoter interaction and transcriptional factors RELB and SP1. PPIF has a higher expression in CD4+ T cells from synovial fluid compared to blood, particularly in the proliferating population. Interestingly, PPIF, as a cyclophilin, can bind to Cyclosporin and the latter has been approved to treat IBD. The genetic risk overlapping suggested the opportunity of repurposing Cyclosporin to treat AS.

Additionally, I found that activated T cells can instruct monocytes from AS blood to produce IL-1β and IL-23 in the absence of pathogen-associated molecular patterns (PAMPs). Monocytes in AS patients are known to be hyper-proinflammatory, the cause of which is not clear. T-cell activated monocytes were heterogeneous but included a subset of “hyperinflammatory monocytes” not induced by LPS activation and enriched for expression of AS risk genes, including *ERN1*. This subset of monocytes exhibited a gene expression profile highly similar to a major population of CD14+ myeloid cells identified in AS synovial fluid (SF), strongly suggesting that T cell-induced myeloid cell activation occurs *in vivo* in inflamed joints. In line with this, PD1+IFN-γ+ T-cells were enriched in AS SF compared to blood. T-cell instructed activation of primary patient-derived

monocytes *in vitro* was inhibited by silencing the *ERN1* gene or by chemical inhibition of its gene product IRE1α.

Lastly, I identified a CXCR6+ Th17/Tph polyfunctional CD4+ T cell subset with the enriched expression of AS GWAS genes through integrating scRNA-seq profiling of AS synovial fluid T cells and genetics approaches. Trajectory analysis revealed these CXCR6+ Th17 / Tph cells were derived from effector Th17 cells, activated by IL-1 and IL-23, followed by antigen engagement. Through demonstrating Tph cells in breast cancer were generated from Tfh cells, I proposed Tph cells are generally derived from different types of effector helper T cells depending on the environment.

My work integrated genetics and immunology approaches to investigate AS pathology and has provided new insights into potential therapeutic targets with genetic support, as well as novel mechanisms of monocyte dysfunctions.

Declaration

I hereby declare that the thesis titled ‘Investigating the genetics and immunopathogenesis of ankylosing spondylitis' is entirely my own work, except where contributions from my collaborators have been explicitly acknowledged. None of the material presented in this thesis has been submitted for any degree or other qualification, either at this University or elsewhere.



Feng Liu

2023/03/31

Acknowledgments

I would like to express my sincere gratitude to my supervisors Liye Chen and Paul Bowness for their generous support of my D.Phil study, for their patience, and their immense knowledge. I would also like to greatly thank my parents for funding my living and training course. Additionally, a huge thank you to my friends and colleagues in my group and facilities.

Contents

Abstract **2**

Declaration **4**

Acknowledgments **5**

List of Figures **9**

List of Tables **11**

List of Abbreviations **12**

Chapter 1 Introduction **14**

1. Clinical features and diagnosis of Ankylosing spondylitis **14**
2. Management of Ankylosing spondylitis **17**
3. Genetics of Ankylosing spondylitis **20**
3.1 HLA-B27 ..... **20**
3.2 From monogenic to polygenic: Genome-wide association studies **22**
4. Immunopathogenesis of Ankylosing spondylitis **37**
4.1 Type 3 immunity **37**
4.2 Peripheral helper T cells and tertiary lymphoid structures **48**
5. Project aims **50**

Chapter 2: Material and Methods **51**

1. Laboratory experiments **51**
1.1 Patient and Control Recruitment **51**
1.2 Mononuclear cell separation from peripheral blood **52**
1.3 Mononuclear cell separation from synovial fluids **52**
1.4 Magnetic isolation of cells from peripheral blood mononuclear cells **52**
1.5 Tissue digestions **54**
1.6 Cellular culture and stimulation **54**
1.7 siRNA silencing of ERN1 gene in primary monocytes **55**
1.8 Intracellular cytokine staining (ICS) **56**
1.9 Enzyme-linked immunosorbent assays (ELISA) **57**
1.10 Single cell RNA sequencing using 10X platform **57**
2. Computational analysis **58**
2.1 GWAS Database **58**
2.2 Immune eQTL Database **61**
2.3 Drug target analysis **62**

2.4 Genetic risk overlapping and clustering analysis 62
2.5 ATAC-seq and ChIP-seq 62
2.6 H3K27ac HiChIP 62
2.7 Single-cell RNA-seq analysis 63

Chapter 3: Studying common genetic risks shared across AS and its associated diseases 65

1. Introduction 66
2. Aims 67
3. Results and discussion 68
3.1 Overlapping of genetic risk between AS and 30 other common diseases 68
3.2 Drug target discovery analysis of commonly shared risk genes 73
3.3 Common risk gene PPIF suggests a drug repurposing opportunity 74
3.4 Specific targeting of JAK2 and TYK2 in AS suggested by genetic risk 86
3.5 A common risk locus at 12q24.12 regulates a transcriptional hub 87
4. Conclusion 92

Chapter 4: T-cell-instructed monocyte activation is a key feature in ankylosing spondylitis and provides novel therapeutic opportunities 94

1. Introduction 94
2. Special Acknowledgements 96
3. Results 97
3.1 Activated T-cells induce monocyte IL-1β, IL-23 and TNF-α production 97
3.2 T-cell instructed monocytes produce IL-1β and IL-23 at levels that are adequate to promote Th17 response in AS 98
3.3 T-cell activated monocytes exhibit unique transcriptional features 103
3.4 CD14+ myeloid cells in AS synovial fluid are transcriptionally similar to T-cell induced IL-23+CXCL2+ C2 monocytes 110
3.5 Activated T-cells are present in AS SF and are composed of both CXCR6+ tissue resident and CXCR6- circulating T-cells 114
3.6 T-cell induced IL-23+CXCL2+ monocytes express high levels of AS-associated risk genes and are regulated by the AS risk gene ERN1 118
4. Discussion 123

Chapter 5: Integration of genetics and single-cell transcriptomics identified CXCR6+ Th17/Tph cells as a pathogenic subset in ankylosing spondylitis 126

1. Introduction 126
2. Results 127

2.1 Genetic risk scoring of CD4+ T cells at single cell level prioritized potential pathological populations **127**
2.2 Transcriptional features and lineage development of CXCR6+ Th17 ..... **131**
2.3 Investigating the developmental route of Tph in different chronic immune responses **138**
3. Discussion **144**

Chapter 6: Overall discussion and future directions **147**

1. Overall discussion **147**
2. Future directions **150**

List of Figures

Chapter 1

Figure 1.1 Algorithm based on the ASAS-EULAR recommendations for the management of axial spondyloarthritis (axSpA). 19

Figure 1.2 Type 3 immunity-driven chronic inflammation in AS 45

Chapter 2

Figure 2.1 Purity of CD14+ monocytes after magnetic isolation. 53

Figure 2.2 Purity of pan-T cells after magnetic isolation. 53

Chapter 3

Figure 3.1 Exploring the translational potential of studying genetic risks shared across diverse diseases 69

Figure 3.2 Linkage disequilibrium of SNPS associated with PPIF (R2 is shown) 79

Figure 3.3 Genomic regulation of AS risk variant on *PPIF* expression 80

Figure 3.4 *PPIF* was uniquely expressed in proliferating T cells at joints of AS 84

Figure 3.5 Common risk variants at 12q24.12 regulate a transcriptional hub 91

Chapter 4

Figure 4.1 Activated T-cells induce monocyte activation that acts back to T-cells and supports Th17 response 100

Figure 4.2 T-cell activation beads or LPS induced immune cell activation in PBMCs from patients with AS. 101

Figure 4.3 Gating strategy for monocyte cytokine production and Th17 cells. 102

Figure 4.4 T-cell-activated monocytes exhibit unique transcriptional features 105

Figure 4.5 CD14+ myeloid cells in AS synovial fluid are transcriptionally similar to T-cell-induced IL-23+CXCL2+ C2 monocytes 111

Figure 4.6 Feature genes of *in vitro* generated calprotectin+ C0 are not enriched in CD14+ myeloid cells from AS synovial fluid 113

Figure 4.7 Activated T-cells are present in AS synovial fluids 115

Figure 4.8 Annotation of T-cell subsets from matched blood and synovial fluid from patients with AS....117
Figure 4.9 T-cell-induced IL-23+CXCL2+ monocytes express high levels of AS-associated risk genes and are regulated by the AS risk gene *ERN1* 120
Figure 4.10 Confirmation of IRE1α suppression by siRNA at the protein level.... 122
Figure 4.11 The inhibition of IRE1α suppresses the secretion of IL-1β, IL-23 and CXCL2 by T-cell-stimulated monocytes.... 122

Chapter 5

Figure 5.1 GWAS scoring identified CXCR6+ Th17 cells as a pathogenic subset in AS.... 130
Figure 5.2 Transcriptional feature and lineage development of CXCR6+ Th17.... 133
Figure 5.3 Developmental trajectory from CXCR3-TCF+ Th17 to CXCR6+ Th17/Tph.... 137
Figure 5.4 A scRNA-seq map of ileum biopsy in Crohn's disease.... 139
Figure 5.5 A single cell transcriptional map of CXC13+ T cells in breast cancer.... 142
Figure 5.6 Tph cells were derived from Tph in breast cancer....143

List of Tables

Chapter 1

Table 1.1 ASAS classification criteria 16

Table 1.2 Modified New York criteria 16

Table 1.3 ASDAS-CRP formula 18

Table 1.4 BASDAI formula 18

Table 1.5 Major genome-wide association studies in AS 27

Table 1.6 List of lead SNPs for genomics loci associated with AS 28

Chapter 2

Table 2.1 Demographics of AS patients 51

Table 2.2 Demographics of AS patients 58

Table 2.3 GWAS result data sources 59

Table 2.4 eQTL data sources 61

Chapter 3

Table 3.1 36 traits selected for genetic risk overlapping analysis 70

Table 3.2 25 AS risk variants associated with PPIF expression 77

Table 3.3 Linkage disequilibrium of SNPS associated with PPIF (R2 is shown 78

Table 3.4 Association of JAK family with different diseases 86

Table 3.5 Four AS risk variants at 12q24.12 89

Chapter 4

Table 4.1 Top 20 gene up regulated in clusters defined in Fig 4.4D 106

Table 4.2 Enrichment of AS risk genes in monocyte clusters induced by *in vitro* stimulation 119

List of Abbreviations

Abbreviation	Full name
ADermatitis	Atopic dermatitis
AITD	Autoimmune thyroid diseases
AlzheimerD	Alzheimer disease
AMD	Age-related Macular Degeneration
AR	Allergic rhinitis
AS	Ankylosing spondylarthritis
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
axSpA	Axial spondyloarthritis
BA	Brain aneurysm
BASDAI	Bath ankylosing spondylitis disease activity index
BMI	Body mass index
CAD	Coronary artery disease
CD	Coeliac disease
ChIP-seq	Chromatin Immunoprecipitation sequencing
COPD	Chronic obstructive pulmonary disease
CRE	Cis-regulatory Element
CTL	Cytotoxic T lymphocyte
ERN1	Endoplasmic reticulum to nucleus signaling 1
ELISA	Enzyme-linked Immunosorbent Assays
eQTL	Expression quantitative trait locus
ER	Endoplasmic reticulum
GWAS	Genome-wide Association Studies
H3K27ac	Histone 3 lysine 27 acetylation
HCV	Hepatitis C virus
hg19	The human genome reference hg19
hg38	The human genome reference hg38
HIV	Human immunodeficiency virus
IBD	Inflammatory bowel diseases
ICS	Intracellular cytokine staining
ILC	Innate lymphoid cell
ILC3	Group 3 innate lymphoid cell
IRE1α	Inositol-requiring enzyme 1
IStroke	Ischemic stroke

JAK	Janus kinase
JIA	Juvenile idiopathic arthritis
LPS	Lipopolysaccharides
MAIT	Mucosal-associated invariant T
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MIF	Macrophage migration inhibitory factor
nr-axSpA	Non-radiographic Axial Spondyloarthritis
NSAIDs	Non-steroidal anti-inflammatory drugs
PAD	Peripheral arterial disease
PAMPs	Pathogen-associated molecular pattern molecules
ParkinsonD	Parkinson disease
PBC	Primary biliary cholangitis
PBMCs	Peripheral blood mononuclear cells
PPIF	Peptidyl-prolyl cis-trans isomerase, mitochondrial
PSC	Primary sclerosing cholangitis
RA	Rheumatoid arthritis
RT	Room temperature
RIPA	Radioimmunoprecipitation assay
scRNA-seq	Single-cell RNA Sequencing
scRNA-seq	Single cell RNA sequencing
SF	Synovial fluid
SH2B3	SH2B adapter protein 3
siRNA	Small inference RNA
SLE	Systemic lupus erythematosus
SNP	Single-nucleotide Polymorphisms
SNVs	Single-nucleotide variants
SPP1	Secreted phosphoprotein 1
T1D	Diabetes mellitus type 1
T2D	Diabetes mellitus type 2
TCR	T-cell Receptor
TLS	Tertiary lymphoid structures
TYK2	Tyrosine kinase 2

Chapter 1: Introduction

Ankylosing spondylitis (AS) is a subtype of axial spondyloarthritis (axSpA) which primarily affects the spine and is frequently accompanied by peripheral arthritis, gut inflammation, uveitis, and psoriasis. While the term ankylosing spondylitis was coined in the early 1900s, it was not differentiated from rheumatoid arthritis (RA) until the 1970s[1]. The prevalence of AS varies, ranging from 0.02% in Sub-Saharan Africa to 0.35% in the Northern Arctic population[2,3]. This variation is likely influenced by the prevalence of human leukocyte antigen (HLA)-B27, a major genetic risk factor associated with AS[2].

This introduction chapter will discuss:

1) the clinical features, diagnosis criteria, and management strategies of ankylosing spondylitis
2) the genetics of ankylosing spondylitis
3) the immunopathogenesis of ankylosing spondylitis

1. Clinical features and diagnosis of ankylosing spondylitis

Radiographic sacroiliitis is the hallmark feature of AS, which differentiates it from non-radiographic axial spondyloarthritis (nr-axSpA)[4]. Non-radiographic axSpA can be diagnosed by the presence of MRI-demonstrated sacroiliitis, and these patients often subsequently develop radiographic damage. Anatomical changes in the affected joints include synovitis, enthesitis, new bone formation, and bone erosion, which result in back pain and stiffness. These symptoms can be temporarily relieved by physical activity and non-steroidal anti-inflammatory drugs (NSAIDs)[5]. AS typically presents in the third decade of life with a male-to-female ratio of 2:1[4]. In contrast, an equal sex distribution has been reported among patients with non-radiographic axial spondyloarthritis[4].

Many AS patients develop concomitant extra-musculoskeletal inflammatory conditions. Approximately 25% of AS patients have uveitis, and the frequency is associated with HLA-B27 positivity[6]. Psoriasis and inflammatory bowel diseases are less frequent comorbidities, occurring in 10% and 7% of the patients respectively[7]. It is noteworthy that gut involvement could be far more frequent, over half of the patients show evidence of macro or microscopic inflammation on endoscopy[8,9]. H. Cypers et al. from the D. Elewaut group[10] have documented that fecal calprotectin can serve as a robust marker for assessing gut inflammation in SpA patients. Given the widespread accessibility of fecal calprotectin testing, the screening of AS patients for concurrent gut inflammation in expansive cohorts could yield a more comprehensive understanding of its prevalence. This information could be particularly instructive in guiding therapeutic choices, as patients exhibiting gut inflammation may benefit from avoiding medications associated with the potential exacerbation of gastrointestinal symptoms, such as IL-17A inhibitors.

Ankylosing Spondylitis (AS) is diagnosed using two widely used guidelines: Assessment of SpondyloArthritis international Society-European League Against Rheumatism (ASAS-EULAR, 2009[11]) (**Table 1.1**) and the Modified New York criteria[12] (**Table 1.2**). The ASAS criteria were developed to improve early-stage diagnosis and sensitivity, which the Modified New York criteria lack. The Modified New York criteria consider radiographic sacroiliitis as the essential criterion, reflecting bone structure damage caused by years of chronic inflammation. The ASAS criteria make use of magnetic resonance imaging (MRI), which allows doctors to directly see inflammation at an early stage, even before structural damage has occurred. ASAS also introduced HLA-B27 and familial disease history, which could further improve the diagnosing sensitivity and specificity given that AS is a highly inheritable disease. It is noteworthy that although ASAS may offer greater sensitivity than the Modified New York criteria, resulting in improved patient outcomes, it may be associated with lower specificity and, therefore, not optimal for research purposes.

Table 1.1 ASAS classification criteria (for people with back pain for 3 months or longer who were < 45 years old at the onset of back pain)

Sacroilitis on imaging* plus 1 or more spondyloarthritis features†
or
HLA-B27 plus 2 or more other spondyloarthritis features†
*Sacroiliitis on imaging: active (acute) inflammation on MRI highly suggestive of sacroiliitis associated with spondyloarthritis, or definite radiographic sacroiliitis according to the modified New York critieria. †Spondyloarthritis features: inflammatory back pain; arthritis; enthesitis (heel); uveitis; dactylitis; psoriasis; Crohn's disease/ulcerative colitis; good response to NSAIDs; family history of spondyloarthritis; HLA-B27; elevated C-reactive protein

Table 1.2 Modified New York criteria

Clinical criteria: Low back pain; present for more than 3 months; improved by exercise but not relieved by rest. Limitation of lumbar spine motion in both the sagittal and frontal planes. Limitation of chest expansion relative to normal values for age and sex.
Radiological criterion: Sacroiliitis on X-ray.
Diagnose: Definite ankylosing spondylitis if the radiological criterion is present plus at least one clinical criterion. Probable ankylosing spondylitis if three clinical criteria are present alone, or if the radiological criterion is present but no clinical criteria are present.

2. Management of Ankylosing spondylitis

AS can have a profound negative impact on patients' quality of life and often leads to disability. The proactive management of AS has multiple aims, including: 1) improving the signs and symptoms of the disease; 2) controlling inflammation; 3) retarding radiographic progression of structural damage; 4) preventing complications associated with AS; 5) maintaining physical function; and 6) promoting work and social participation[13].

The pharmacotherapy of AS has significantly improved over the past few decades, owing to breakthroughs in pathogenesis research. These new therapies were mainly developed from the angle of containing the excessive immune response. Typically, non-steroidal anti-inflammatory drugs (NSAIDs) are the first-line treatment and can effectively relieve pain in at least half of patients[14]. However, the use of NSAIDs is often limited by inefficacy and intolerable side effects, such as headache and abdominal pain[14]. For patients who are not eligible or fail to NSAIDs, TNF blockade (first USA FDA approval in 2003) and IL-17 blockade (first USA FDA approval in 2016) are the next choices. Approximately half of the patients taking anti-TNF therapy can achieve satisfactory remissions and reduced inflammation in the spinal and peripheral joints. However, some patients may develop tolerance to a specific antibody drug over time and require switching to a different TNF inhibitor or other therapies[15]. IL-17 blockade therapy also has an approximately 50% remission rate[16], while it has been associated to an increased risk of *Candida spp.* Infection and induction/exacerbation of IBD. More recently, Janus kinase (JAK) inhibitors have been added to the therapeutic toolbox. The remission rate of JAK inhibitors is similar to that of TNF and IL-17 blockade therapy[13]. The FDA recommends considering JAK inhibitors after anti-TNF therapy due to the increased risk of cardiovascular events with the former[17].

Along with the development and clinal practice of new therapies, the experience of tailoring management strategies for patients having different clinical presentations has been accumulated. The selection of therapeutic tools is mainly based on disease activity, treatment history,

comorbidities, and psychosocial factors. Currently, Ankylosing Spondylitis Disease Activity Score (ASDAS)[18] (**Table 1.3**) and the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (**Table 1.4**) are most widely used to index the disease activities. Last year, ASAS-EULAR published updated management recommendations using ASDAS (**Figure 1.1**)[5].

Table 1.3 ASDAS-CRP formula

0.12 x Back Pain + 0.06 x Duration of Morning Stiffness + 0.11 x Patient Global + 0.07 x Peripheral Pain/Swelling + 0.58 x Ln(CRP+1)
ASDAS, Ankylosing Spondylitis Disease Activity Score; Ln(CRP+1), natural logarithm of the C-reactive protein (mg/L) + 1. Back pain, patient global, duration of morning stiffness and peripheral pain/swelling are all assessed on a visual analogue scale (from 0 to 10cm) or on a numerical rating scale (from 0 to 10). Back pain, BASDAI question 2: "How would you describe the overall level of AS neck, back or hip pain you have had?". Duration of morning stiffness, BASDAI question 6: "How long does your morning stiffness last from the time you wake up?". Patient global: "How active was your spondylitis on average during the last week?" Peripheral pain/swelling, BASDAI question 3: "How would you describe the overall level of pain/swelling in joints other than neck, back or hips you have had?"

Table 1.4 BASDAI formula

BASDAI = ((Q1 + Q2 + Q3 + Q4) + ((Q5 + Q6) / 2)) / 5
The Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) disease activity questionnaire contains six questions regarding subjective symptoms during the week prior to answering the questions. 1. How would you describe the overall level of fatigue/tiredness you have experienced? 2. How would you describe the overall level of AS neck, back, or hip pain you have had? 3. How would you describe the overall level of pain/swelling in joints other than neck, back, hips you have had? 4. How would you describe the level of discomfort you have had from an area tender to touch or pressure? 5. How would you describe the level of morning stiffness you have had from the time you wake up? 6. How long does your morning stiffness last from the time you wake up? Each question is scored on a scale of 0 to 10. Aside from the last question, 0 indicates none and 10 indicate very severe. For the last question, 0 is 0 hours, 5 is one hour, and 10 is two or more hours.

Despite exciting progress, current management options remain limited. There is still a proportion of patients not responding to all the therapies and there is a lack of precise predictors to prioritize treatments for individuals. More importantly, none of the existing therapies can eliminate inflammation permanently or halt radiographic progression. To achieve the ultimate goal of curing AS, we must advance our understanding of AS pathogenesis.

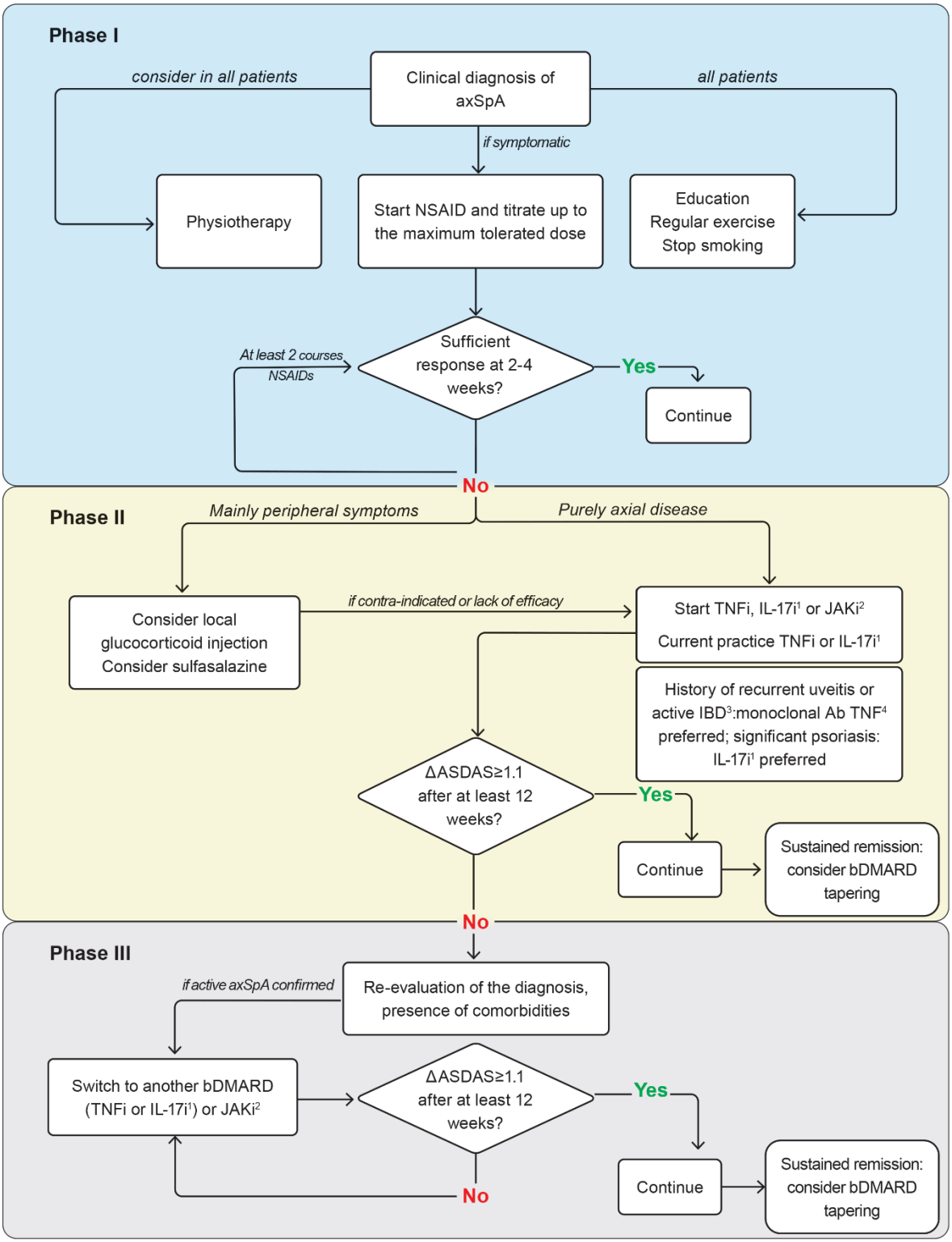


Figure 1.1 (taken from ASAS-EULAR) **Algorithm based on the ASAS-EULAR recommendations for the management of axial spondyloarthritis (axSpA)**. Ab, antibody; ASAS, Assessment of SpondyloArthritis international Society; ASDAS, Ankylosing Spondylitis Disease Activity Score; bDMARD, biological disease-modifying antirheumatic drug; IBD, inflammatory bowel disease; IL-17i, interleukin-17 inhibitors; JAKi, Janus kinase inhibitors; NSAID, non-steroidal anti-inflammatory drug; TNFi, tumour necrosis factor inhibitors.

3. Genetics of Ankylosing spondylitis

Ankylosing spondylitis has long been recognized as a highly inheritable disease[19]. Twin studies have demonstrated a recurrence rate of 63% in monozygotic twins[20] and a heritability of over 90%[21]. However, a later study utilizing SNP genotype data from the UK Biobank cohort of unrelated individuals estimated the heritability of AS to be 40%, compared to 24.1% for Crohn's disease and 7.4% for rheumatoid arthritis (UKB Heritability by Neale lab, https://nealelab.github.io/UKBB_ldsc/). While accurately estimating heritability has proven to be challenging due to the complex impact of environmental factors, it is undeniable that genetic backgrounds strongly influence the development of AS. This provides important clues for studying the etiology of AS.

3.1 HLA-B27

Around ~94% of AS patients carry HLA-B27, while this frequency is estimated as 6-9% in the whole US and European population[22]. HLA-B27 is a subtype of the HLA-B molecule. Together with HLA-A and HLA-C, these three molecules constitute the classical MHC class 1 molecule family, whose primary function is to present peptide fragments of intracellular proteins to cytotoxic T cells. MHC class 1 molecules are expressed by all kinds of human cells. When a "non-self" antigen is presented, cytotoxic T cells recognize such cells as being infected or abnormal and eliminate them. Cells that attempt to evade this surveillance by not presenting internal peptides will be eliminated by NK cells. This mechanism ensures an effective defense against external invaders and internal corruptions.

In a 2015 review, Paul Bowness effectively summarized three major hypotheses regarding the etiology of HLA-B27, namely[23]: 1) the arthritogenic peptide hypothesis, which posits that self- and/or pathogen-derived peptides are presented by properly folded forms of HLA-B27 for recognition by the T cell receptor of autoreactive CD8+ T cells. 2) the B27 misfolding hypothesis,

which suggests that misfolding of B27 within the endoplasmic reticulum leads to ER stress, unfolded protein response, or autophagy, with downstream effects on cellular function that promote the development of spondyloarthritides (e.g., IL-23 release). 3) the cell surface B27 free heavy chain expression and immune recognition hypothesis, which proposes that B27 free heavy chains, including dimers, are expressed at the cell surface, where they are recognized by cells bearing killer immunoglobulin-like receptors and/or leukocyte immunoglobulin-like receptors to trigger inflammation. Despite significant efforts, none of these hypotheses has yet accumulated sufficient evidence.

A new hypothesis has recently emerged, suggesting that HLA-B27 may shape the immune system by affecting the gut microbiota. The intestine serves as an interface where our immune system constantly interacts with environmental microbes, and this interaction can modulate the systematic immune response. Variations in gut microbiome composition have been linked to the induction of inflammatory disease[24,25] and response to immune blockade therapy when treating cancers[26]. Given that ankylosing spondylitis (AS) is often accompanied by gut inflammation, it is reasonable to suspect that HLA-B27 may affect host-microbe interactions in the gut. Mary-Ellen Costello and colleagues from the Matthew A. Brown group were the first to report differences in the gut microbiome composition between AS patients and healthy individuals, revealing a higher abundance of 5 families of bacteria (*Lachnospiraceae*, *Ruminococcaceae*, *Rikenellaceae*, *Porphyromonadaceae*, and *Bacteroidaceae*) and a decrease in the abundance of 2 families of bacteria (*Veillonellaceae* and *Prevotellaceae*) in the terminal ileum of patients with AS compared with healthy control subjects[27]. The same group later associated this microbiome composition change with the presence of the HLA-B27 allele in a healthy individual cohort. These pieces of evidence support the hypothesis that HLA-B27 causes or increases the risk of AS induction through interaction with the intestinal microbiome[28]. However, the mechanism underlying these associations has not been fully elucidated, and it is yet unknown whether HLA-B27 can

preferentially recognize the peptides of certain microbes or whether it has intrinsic immune regulatory functions that modulate cellular programs.

In addition to its implications for spondyloarthritides, HLA-B27 also confers protective effects against hepatitis C virus (HCV)[29] and human immunodeficiency virus (HIV)[30]. Given that HLA-B27 affects diseases involving different human tissues and pathogens, it is reasonable to suspect that it does not recognize certain autoantigens. It is possible that the presence of HLA-B27 may broaden the repertoire of TCRs through a unidentified mechanism, thereby augmenting the probability of recognizing both autoantigens and pathogens (it is important to note that this hypothesis currently lacks evidence).

3.2 From monogenic to polygenic: Genome-wide association studies

Although HLA-B27 renders a strong genetic risk to develop AS, only ~1-5% of the HLA-B27 carriers eventually develop AS[31]. Also, diagnosed AS was only found in 21% of HLA-B27 positive relatives of HLA-B27 positive AS patients[32]. It is estimated that HLA-B27 only accounts for ~20% of AS heritability[21,33]. This implies that additional genetic risks inherited together with HLA-B27 could contribute to the HLA-B27-mediated immune variation. The advent of the genome-wide association studies (GWAS) era enabled the illustration of the polygenetic structure hidden behind HLA-B27.

3.2.1 The era of genome-wide association studies

Approximately thirty years ago, proponents of the Human Genome Project (HGP) proposed that sequencing the human genome would have significant implications for human life science and clinical medicine. Risch and Merikangas, in 1996, highlighted the potential of the HGP by stating that it could lead to the identification of polymorphisms for all the genes in the human genome and the diseases to which they contribute[34]. The completion of the HGP in 2003 marked the

beginning of the paradigm shift in genetic studies towards Genome-wide association studies (GWAS) to fulfil this goal.

GWAS represents a hypothesis-free approach to screen the genome for detecting associations between genetic variants and traits in the population. The core promise of GWAS was the ability to identify polymorphisms associated with disease risk at the population level, thereby explaining a proportion of familial risk. This experimental design was expected to enable the detection of previously unknown genetic associations, leading to an improved understanding of the genetic basis of complex diseases[35]. The publication of the first Genome-wide association study (GWAS) was in 2005 by Robert J Klein et al[36]. The authors profiled 116,204 single-nucleotide polymorphisms (SNPs) of 96 age-related macular degeneration (AMD) patients and 50 healthy individuals. This study only identified one genetic risk locus sitting around gene CFH. It was the Wellcome Trust Case Control Consortium (WTCCC) that set the standard for the study design of GWASs[37]. The WTCCC demonstrated that reproducible discoveries in GWASs required combining forces through large sample sizes, a rigorous study design involving discovery and replication stages, and the use of stringent criteria such as multiple testing corrected significance level[38]. Subsequent studies have increasingly recognized the significance of large-scale cohorts in Genome-wide association studies (GWAS). Recent GWASs, such as those for height[39] and blood pressure[40], have included over one million individuals. This remarkable achievement is a result of the collaborative efforts of smaller cohorts and the establishment of large-scale biobanks. Examples of such biobanks include deCODE Genetics, the Estonian Biobank, Biobank Japan, China Kadoori Biobank, FinnGen in Finland, Lifelines in the Netherlands, the Million Veteran Program in the USA, and more recently, the All of Us Research Program in the US[35]. These biobanks have played a crucial role in enabling the generation of large and diverse datasets necessary for detecting subtle genetic variants associated with diseases and traits. Simultaneously, despite the increasing emphasis on large-scale cohorts, the expansion of cohorts

specialized in certain diseases remains necessary, particularly for diseases with rare prevalence. For example, a cohort with 1 million individuals may only have ~5000 AS patients, which will not enable the discovery of all generic risk signals.

3.2.2 Benefits of GWAS

GWAS has been proven very successful of identifying novel and subtle disease-associated genetic variants. To date, GWAS have identified tens to hundreds of genetic regions affecting the risk of developing each form of the condition in the field of immune-mediated inflammatory diseases [41,42]. These GWAS studies have already helped in drug target discovery. It is estimated that selecting targets with genetic support could double the success rate in clinical development[43]. Among 50 FDA-approved drugs in 2021, two-thirds of them were supported by genetic evidence[44]. A good example is that the association of genetic risks altering IL-23/IL-17 axis with multiple inflammatory diseases supported the development of new treatments for IBD, inflammatory arthritis, and psoriasis by inhibiting the IL-23 pathway[45].

GWASs have also facilitated the discovery of novel biological mechanisms. The most well-known example is probably the association between the *CFH* gene and age-related macular degeneration (AMD)[36], which implicated the complement system of innate immunity in this most common cause of blindness[46]. The second example is the rs1421085 T-to-C single-nucleotide variant in obesity. This SNP disrupts the binding of ARID5B on an enhancer regulating IRX3 and IRX5 expression. ARID5B is a repressor for this enhancer, and such disruption led to the doubling of IRX3 and IRX5 expression. This genomic dysregulation can result in a reduction in mitochondrial thermogenesis and the development of excessive lipid deposits. Another example is the discovery of single-nucleotide polymorphisms (SNPs) associated with IBD risk in the genes *ATG16L1* (rs2241880)[47] through GWAS. The rs2241880 SNP results in a missense mutation (p.Thr300Ala) that increases caspase-3-mediated cleavage of ATG16L1, leading to decreased

autophagy under cellular stress[48]. Consequently, this reduction impairs pathogen clearance and enhances inflammatory cytokine production, thus establishing chronic inflammation.

GWASs can also directly improve clinical practice. 1) Genetic biomarkers can help distinguish conditions that have similar clinical presentations. For example, the increasing prevalence of obesity has resulted in a greater challenge to differentiate between type 1 diabetes (T1D) and type 2 diabetes (T2D), especially among young adults. Despite some overlapping, these two diseases have distinct genetic risk backgrounds[49]. It has been demonstrated that this difference can aid the discrimination of these two conditions. 2) Early identification of individuals highly susceptible to certain diseases can instruct the use of preventative measures before disease induction. For instance, AMD has been found to have a strong genetic component. Individuals with a high genetic risk of AMD may be advised to quit smoking and increase their intake of certain nutrients, which could reduce the potential for disease development. 3) GWASs could inform drug selection. SNPs near the *IL28B* gene (also known as *IFNL3*) have been associated with the response to pegylated interferon-α and ribavirin therapy for hepatitis C virus infection[50–52]. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has published guidelines for pegylated interferon-α-based treatment regimens based on the *IL28B* genotype[53].

GWASs also revealed the ethnic variation of complex traits. Given the substantial differences in physiological characteristics among ethnic populations, it is reasonable to expect differences in disease pathology. Large-scale GWASs have demonstrated the extent to which such differences exist. On the one hand, the majority of risk loci are shared among different ethnic groups, providing reassurance that humans share a common ancestry and that cultural and environmental selection throughout history has not resulted in significant differences[54]. On the other hand, unique risk loci have been identified, and there are substantial differences in association strength at most loci[54]. These differences highlight the need for large-scale GWASs within diverse

populations to facilitate precision medicine beyond the western world. However, economic difficulties are a clear obstacle to achieving this goal.

3.2.3 Limitations of GWAS

Although GWAS have yielded significant insights biomedicine, the study design has faced criticism. Critics have argued that single-nucleotide variants (SNVs) identified in GWAS only account for a small fraction of complex trait heritability and most traits are influenced by thousands of causal variants that individually confer very little risk and are often associated with many other traits. Also, GWASs usually do not directly provide the causal genetic variants. Linkage disequilibrium is a double-edged sword: On the one hand, it allows the genome-wide genotype through imputation when whole genome sequencing is unaffordable. On the other hand, it makes it difficult to discern the causal variant or variants. Furthermore, interpreting the biological and genomic context of these associations poses a significant challenge. These limitations make it difficult to draw unambiguous conclusions about the biological meaning of GWAS results, sometimes limiting their utility to produce mechanistic insights or serve as starting points for drug development.

3.2.4 GWAS of ankylosing spondylitis

To date, there are five major genome-wide association studies published for AS, with the largest cohort comprised of > 10,000 cases (Table 1.5). Totally, 113 non-MHC loci have been identified at genome-wide significance levels (Table 1.6).

Table 1.5 Major genome-wide association studies in AS

Studies	year	Discovery cohort (case/control)	Ethinicity	No. of non-MHC loci reaching genome-wide significance
WTCCC & TASC[55]	2007	1000/1500	European	2
TASC[56]	2010	2053/5140	European	4
TASC[31]	2011	3023/8779	European	8
IGAS[57]	2013	10,619/15,145	European, East Asian, and Latin American	24
Ellinghaus et al.[33]	2016	8726/34,213	European and East Asian	113

Table 1.6 List of lead SNPs for genomics loci associated with AS (adapted from Matthew Brown et al.[58])

rsID*	Chromosome	Position	Nearby Gene	Allele (Effect/reference)	OR**
rs183686347	1	67703442	*IL23R*	A/G	1.78
rs469758	5	96121715	*ERAP1*	C/T	1.34
rs10865331	2	62551472	*B3GNT2*	A/G	1.3
rs72871627	2	163136942	*IFIH1*	G/A	1.28
rs7115956	11	109956346	*ZC3H12C/RDX*	C/T	1.23
rs11624293	14	88488821	*GPR65*	C/T	1.23
rs11363316	16	30171017	*CORO1A*	D/I	1.23
rs10750899	11	58284951	*LPXM*	G/A	1.23
rs2910686	5	96252589	*ERAP2*	C/T	1.21
rs67025039	6	167415404	*FGFR1OP*	I/D	1.21
rs3124998	9	139389432	*NOTCH1*	T/C	1.21
rs6908425	6	20728731	*CDKAL1*	C/T	1.2
rs11249215	1	25297184	*RUNX3*	A/G	1.19
rs3749171	2	241569692	*GPR35*	T/C	1.18
rs6556416	5	158818745	*IL12B*	C/A	1.17
rs1525735	7	17197571	*AHR*	C/T	1.17
rs26528	16	28517709	*IL27*	C/T	1.17
rs10889676	1	67722567	*IL23R*	A/C	1.16
rs12943464	17	45690351	*NPEPPS*	A/T	1.15
rs17622517	5	131803537	*C5orf56*	C/T	1.15
rs11306716	2	204707771	*CD28/CTLA4*	I/D	1.14
rs6781808	3	137441372	*SOX14*	C/T	1.14
rs9797244	17	26097131	*NOS2*	C/T	1.14
rs196941	17	62147192	*ERN1/ICAM2*	G/C	1.14
rs11221322	11	128346793	*ETS1*	C/T	1.14
rs56167332	5	158827769	*IL12B*	A/C	1.13
rs11749391	5	150229066	*IRGM*	C/T	1.13
rs35874463	15	67457698	*SMAD3*	G/A	1.13
rs2666218	2	9402988	*ASAP2*	G/A	1.12
rs3774937	4	103434253	*NFKB1*	C/T	1.12
rs7933433	11	128194450	*ETS1*	T/G	1.12
rs12968719	18	12879466	*PTPN2*	A/G	1.12
rs11750385	5	10521668	*LINC02213*	G/T	1.11

rs10748781	10	101283330	*NKX2-3*	C/A	1.11
rs11616188	12	6502742	*LTBR*	A/G	1.11
rs2779255	17	26137540	*NOS2*	C/T	1.11
rs61802846	1	161473873	*FCGR2A*	C/A	1.11
rs11236797	11	76299649	*LRRC32*	A/C	1.105
rs3024493	1	206943968	*IL10/IL19*	A/C	1.1
rs8006884	14	35563211	*PPP2R3C*	C/T	1.1
rs1292035	17	57989557	*RPS6KB1*	A/G	1.1
rs2266961	22	21928597	*UBE2L3*	G/C	1.1
rs3197999	3	49721532	*MST1*	A/G	1.09
rs2451258	6	159506600	*TAGAP*	C/T	1.09
rs2812378	9	34710260	*CCL21*	G/A	1.09
rs10870077	9	139263891	*CARD9*	G/C	1.09
rs10761648	10	64354262	*ZNF365*	T/C	1.09
rs11221332	11	128380974	*ETS1*	T/C	1.09
rs12720356	19	10469975	*TYK2*	C/A	1.09
rs6058869	20	31348750	*DNMT3B*	T/C	1.09
rs4845604	1	151801680	*RORC*	A/G	1.08
rs12694846	2	231148128	*SP140*	G/A	1.08
rs72928038	6	90976768	*BACH2*	A/G	1.08
rs11574938	16	30485393	*ITGAL*	G/C	1.08
rs4456788	21	45616324	*ICOSLG/DNMT3L*	G/A	1.08
rs4971079	1	155130391	*THBS3*	A/G	1.076
rs10800314	1	161472789	*FCGR2A*	C/A	1.07
rs201014116	2	25500905	*DNMT3A*	C/A	1.07
rs4728142	7	128573967	*IRF5*	A/G	1.07
rs2236379	10	6527143	*PRKCQ*	A/G	1.07
rs35736272	17	38032680	*ZPBP2*	C/T	1.07
rs679574	19	49206108	*FUT2*	G/C	1.07
rs13407913	2	25097644	*ADCY3*	G/A	1.06
rs7426056	2	204612058	*CD28/CTLA4*	A/G	1.06
rs11676348	2	219010146	*CXCR2*	T/C	1.06
rs582757	6	138197824	*TNFAIP3*	C/T	1.06
rs10758669	9	4981602	*JAK2*	C/A	1.06
rs2823288	21	16820888	*USP25*	A/G	1.06

rs6426833	1	20171860	*OTUD3*	G/A	1.05
rs12718244	7	50175654	*C7orf72*	A/G	1.05
rs9687958	5	40496423	*PTGER4*	G/T	0.95
rs4917129	7	50323174	*IKZF1*	T/C	0.95
rs12369214	12	107198611	*TMEM263*	A/G	0.95
rs367569	16	11365500	*TNP2*	T/C	0.95
rs587259	19	34656406	*LSM14*	T/C	0.95
rs4676406	2	241579108	*GPR35*	G/T	0.94
rs1001007	3	46428658	*CCR5*	A/G	0.94
rs71624119	5	55440730	*ANKRD55*	A/G	0.94
rs2145623	14	35839236	*NFKBIA*	C/G	0.94
rs1569328	14	75741751	*FOS*	T/C	0.94
rs7517847	1	67681669	*IL23R*	G/T	0.93
rs12075255	1	206961628	*IL10/IL19*	A/G	0.93
rs1992661	5	40414989	*PTGER4*	G/A	0.93
rs7915475	10	64381668	*ZNF365*	G/A	0.93
rs9889296	17	32570547	*CCL2/7/8/11*	A/G	0.93
rs1267499	6	14715882	*6p23 Intergenic*	T/C	0.92
rs726657	9	117696336	*TNFSF8*	T/C	0.92
rs3184504	12	111884608	*SH2B3*	T/C	0.92
rs4243971	20	30849517	*KIF3B/POFUT1*	T/G	0.92
rs1333062	1	160846284	*LY9/CD244*	T/G	0.91
rs2301436	6	167437988	*FGFR1OP*	T/C	0.91
rs4986790	9	120475302	*TLR4*	G/A	0.9
rs2143178	22	39660829	*PDFGFB*	C/T	0.9
rs1250573	10	81042475	*ZMIZ1*	A/G	0.89
rs1800682	10	90749963	*ACTA2*	G/A	0.89
rs4851529	2	102647300	*IL1R2*	A/G	0.88
rs12504282	4	80927001	*ANTXR2*	C/T	0.88
rs2328530	6	20643727	*CDKAL1*	G/A	0.88
rs11190133	10	101278725	*NKX2-3*	T/C	0.88
rs871656	2	102771282	*IL1R1*	A/T	0.87
rs10510607	3	28286261	*CMC1*	T/C	0.87
rs2549803	5	96174929	*ERAP1*	T/C	0.87
rs1569414	22	45727565	*FAM118A*	G/T	0.87

rs4129267	1	154426264	*IL6R*	T/C	0.85
rs1860545	12	6446777	*TNFRSF1A*	A/G	0.85
rs6600247	1	25305114	*RUNX3*	T/C	0.83
rs12131796	1	200878727	*C1orf106/GPR25*	A/G	0.83
rs9977672	21	40463283	*21q22 Intergenic*	A/G	0.83
rs5743293	16	50763781	*NOD2*	I/D	0.82
rs74956615	19	10427721	*RAVER1*	A/T	0.81
rs72796367	16	50762771	*NOD2*	C/T	0.79
rs2066845	16	50756540	*NOD2*	C/G	0.78
rs80174646	1	67708155	*IL23R*	T/G	0.6
rs35018800	19	10464843	*TYK2*	A/G	0.6
rs141992399	9	139259592	*CARD9*	G/C	0.48

*Note: *rsID: Reference SNP cluster ID defined by dbSNP; **OR: odds ratio*

Despite the scarcity of information pertaining to the function of these GWAS candidate genes, several cellular programs appear to exhibit enrichment for these genetic elements.

IL-23/IL-17 axis

One of the first two non-MHC signals identified in early AS GWAS exits at the *IL23R* locus. *IL23R* encodes the unique subunit of the interleukin 23 (IL-23) receptor, in complex with the interleukin-12 receptor β1 subunit (IL-12Rβ1). This provided the first evidence implicating IL-23/IL-17 axis into AS pathology. This locus was also associated with psoriasis[59] and IBD[60], presenting an early example of genetic risk pleiotropy in inflammatory diseases. Functional studies demonstrated that the protective allele at rs11209026 impairs the IL-23 function and leads to less Th17 response[61].

Multiple other genes in the IL-23/IL-17 axis have been associated with AS, including *IL1R1*[33], *IL6R*[62], *RORC*[33], and *CCR6*[33], *TYK2*[62]. IL-1, IL-6, and IL-23 provide the Th17 polarization signals. RORC encodes the master transcriptional factor RORγt for Th17 cells and controls IL-17 production. CCR6 is the chemokine receptor uniquely expressed on the Th17 cell surface. TYK2 is required to transduce the signal of IL-6 and IL-23 receptors to STAT3.

Antigen presentation and ER stress

The *ERAP1* locus constitutes another robust signal identified in early investigations and has been found to be exclusively associated with ankylosing spondylitis (AS) in HLA-B27 positive patients[31]. *ERAP1* is instrumental in the trimming of peptides presented to MHC class I molecules, which underscores the significance of the antigen-presenting program in the pathogenesis of AS. AS risk variants were associated with increased activity and expression of *ERAP1*[63]. A recent study by Tran et al. elucidated the mechanism through which this enhanced ERAP1 function interacts with HLA-B27 in the pathogenesis of AS[64]. The researchers utilized an HLA-B27 transgenic mouse model that spontaneously develops spondyloarthritis symptoms[65]. They discovered that

the absence of ERAP1 resulted in an increase in HLA-B27 folding and a reduced incidence of disease, which was concomitant with a diminished unfolded protein response. Conversely, ERAP1 ablation led to increased misfolding of HLA-B7 and ER stress in HLA-B7 transgenic mice. They proposed a plausible explanation that the absence of ERAP1 could contribute to the accumulation of peptides longer than 8-mers, which selectively bind to HLA-B27 rather than HLA-B7. This binding to peptides stabilizes HLA-B27 folding and alleviates the ER stress induced by its inherent propensity to misfold.

3.2.5 Challenges of translating GWAS discoveries to new therapies

Despite the remarkable advancements achieved through genome-wide association studies (GWAS), a mere fraction of the findings has been successfully translated into clinical medicine[42]. One significant obstacle lies in elucidating the molecular mechanisms that underpin the associations identified in GWAS. The functional examination of genetic associations currently faces several challenges: 1) Complex genomic regulation renders difficulty in identifying causal genes and determining the appropriate cellular contexts related to specific disease. More than 90% of disease-associated genetic variants genetic variants reside in non-coding regions of the genome[66]. Furthermore, a cis-regulatory element (CRE) influenced by a variant can regulate genes situated millions of base pairs away[67,68]. In addition, a high degree of cell-type specificity in genetic regulation by non-coding regions has been extensively documented[69,70]; 2) Current limitations exist in recapitulating the human biological system *in vitro* and conducting gene perturbation in disease-relevant cellular assays utilizing human primary immune cells, particularly in a high-throughput manner.

3.2.6 From genetic variants to genes in immune cells: eQTL study and 3D genomic mapping

Drawing on a rationale akin to that of GWAS, expression quantitative trait locus (eQTL) studies endeavor to uncover associations between genetic variants and gene expression. In 2015, the Genotype-Tissue Expression (GTEx) project published the first comprehensive eQTL atlas, encompassing 44 types of tissues obtained from 175 individuals[71]. This groundbreaking study showcased the cell and tissue-specificity of eQTLs and emphasized the importance of cohort size in such investigations. Although the study included blood and EBV-transformed lymphocytes, it did not incorporate primary immune cells. Subsequently, in 2016, the BLUEPRINT project published an eQTL study featuring monocytes, CD4+ T cells, and neutrophils, which revealed distinct cell-type specificities[72]. This latter study catalyzed further endeavors to generate eQTL datasets for additional immune cell types. In 2018, the DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs), and Epigenomics) project broadened the scope to 13 cell types, albeit with a reduced cohort size of 91 donors[73]. Most recently, the ImmuNexuUT consortium unveiled the most comprehensive eQTL database to date, encompassing 28 cell types derived from 416 donors[69].

The eQTL approach exhibits limitations in pinpointing causal variants due to the presence of linkage disequilibrium. Additionally, differentiating between cis-eQTL and trans-eQTL presents a considerable challenge. Most studies conventionally employ a 1-million-bp boundary; however, the protein product of a causal gene may regulate the transcription of adjacent genes, leading to two eQTL signals and complicating the identification of the true causal gene. This issue could potentially be addressed by generating enhancer connectome maps, which provide information on the physical interactions between putative enhancers and gene promoters. Organizing a 2-meter-long genome is one of the key programs regulating transcriptional activity in human cells. The advent of technologies like Hi-C[74,75], Micro-C[76–78], ChIA-PET[79–81], and HiChIP[82] has enabled

researchers to elucidate this regulatory program by capturing the three-dimensional interactions of genomic DNA. In 2016, Javierre and colleagues published the first comprehensive chromatin interaction atlas of immune cells, profiling 17 primary human hematopoietic cell types using promoter-capture Hi-C[83], achieving a resolution of up to 25 kb. Although this resolution was insufficient to pinpoint causal genetic variants, it facilitated the linking of GWAS SNPs to their putative target genes. In 2017, Mumbach et al. employed H3K27ac (Histone 3 lysine 27 acetylation, and active enhancer/promoter histone marker) HiChIP to profile the enhancer connectome of naïve CD4+ T cells, Th17, and Treg, obtaining a 2-5 kb resolution[84]. In 2020, researchers from the DICE project published their results using the same technology for monocytes, NK cells, naïve B cells, naïve CD4+ T cells, and naïve CD8+ T cells[85]. Notably, the DICE project developed a method called promoter-interacting QTL to integrate eQTL and chromatin interaction information, identifying a dataset of eQTLs supported by physical interactions. Despite the substantial contributions of these datasets, their resolutions typically range from 2 to 25 kb, encompassing multiple genetic variants and CREs, which complicates the identification of causal variants. In 2015, Hsieh et al. developed Micro-C, in which they replaced restriction enzymes with micrococcal nuclease in Hi-C. This modification enabled chromatin interaction mapping to achieve a nucleosome resolution (approximately 200 bp)[76]. Generating a Micro-C landscape of immune cells could potentially assist in pinpointing causal genetic variants in numerous complex loci.

eQTL analysis and 3D genomic mapping constitute the most effective methods for associating genes with disease-linked genetic variants. Additionally, cell-type-specific profiling can provide insight into the potential cellular contexts in which genetic variations exert their effects. Nevertheless, eQTL studies are constrained by cohort size, while 3D genomic mapping is hindered by limited resolution. Despite these limitations, the aforementioned valuable databases

have proven instrumental in assisting the research community in uncovering the functional basis of numerous GWAS findings.

4. Immunopathogenesis of Ankylosing spondylitis

4.1 Type 3 immunity

4.1.1 Three types of immune reponse

The human immune system has tailored its effector functions to respond optimally to specific pathogens. In 2014, Francesco Annunziato and his colleagues proposed a three-type immune response classification system[86]. In this theory, the innate and adaptive immune systems intersect to form three primary categories of cell-mediated effector immunity: type 1, type 2, and type 3. Type 1 immunity specializes in eliminating intracellular infections and dysfunctional cells. It consists of T-bet+ T cells (Th1, CD8+ CTLs, and ILC1s), and mononuclear phagocytes. This system scans MHC class I-presented peptides and purges infected cells and aberrant cells such as tumor cells. This system can also cause autoimmune diseases like type 1 diabetes when attacking auto-antigen. Type 2 immunity is mainly mediated by GATA-3+ T cells (Th2 and ILC2s), mast cells, basophils, and eosinophils, and confers protection against helminths and venoms. Nevertheless, dysregulation of type 2 immunity can lead to allergic disorders such as asthma. Type 3 immunity encompasses RORγt+ T cells (Th17 and ILC3s), mononuclear phagocytes, and neutrophils, and stimulates epithelial antimicrobial responses, thus safeguarding against extracellular fungi and bacteria. This immune response has been linked to a broad spectrum of chronic inflammatory disorders including inflammatory bowel diseases, multiple sclerosis, among others. This three type immunity framework aids in comprehending how different components of the immune system synergize to respond to distinct invading stimuli. However, it is not all-encompassing, and the above description is an oversimplification. Other cell types such as γδT cells and MAIT cells have also been reported to participate in type 3 immunity. Some cell types, such as B cells, may potentially be involved in all three types of immunity.

The following discussion will focus on type 3 immunity in consideration of its driving role in ankylosing spondylitis.

4.1.2 IL-17 and Th17 cells

Helper T cells are traditionally regarded as the primary orchestrators of distinct types of immune responses. Upon recognition of pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns, innate immune cells produce specific cytokines that guide naïve CD4+ T cells towards particular lineages. IL-12 induces Th1 differentiation that generates IFN-γ; IL-4 polarizes Th2 towards IL-4, IL-5, and IL-13 production, while IL-1 and IL-23 promote the Th17 lineage that secretes IL-17.

Discovery of IL-17 and Th17 cells

Th17 cells were first recognized as a unique lineage distinct from Th1 and Th2 cells in 2005[87–89]. However, the discovery journey of Type 3 immunity dates back to 1993 when the IL17A transcript was detected in T cells[90]. In pursuit of identifying novel molecules involved in immune functions, Eric Rouvier and his colleagues screened the cDNA of murine T cell hybridoma, leading to the identification of a novel transcript, which they named Cytotoxic T Lymphocyte-Associated Protein 8 (CTLA-8). In 1995, Zhengbin Yao et al. cloned the orthologous gene of CTLA-8 from human CD4+ T cells and renamed it IL-17[91,92]. They also demonstrated that this cytokine could induce pro-inflammatory phenotypes (IL-6 and IL8 production) in fibroblasts. Subsequently, F Fossiez et al. reported a similar impact on other types of stromal cells, including synoviocytes, and epithelial and endothelial cells[93,94]. A sustained innate immune response requires activation signaling from T cells. In 1998, Dragan V. Jovanovic et al. examined if IL-17 is one of such mediators and found that IL-17 indeed controls the production of IL-1β and TNF-α by macrophages[95]. In the same year, Marcel Teunissen et al reported the IL-17-producing cells are present in lesional psoriatic skin but not in nonlesional control biopsies, suggesting the pathogenic roles of IL-17 and IL-17-producing cells in inflammatory diseases for the first time[96].

One year later, Martine Chabaud and his colleagues implicated IL-17 into arthritis pathology[97]. They found the presence of IL-17 in the synovium of rheumatoid arthritis (RA) but not in

osteoarthritis (OA) and healthy individuals. Additional IL-17 into *in vitro* cultured RA synovium increased IL-6 production while anti-IL17 antibodies had the opposite effect. The author also tried to fit these IL-17-producing cells into the established Th1-Th2 paradigm. It was known that effector cytokines of a Th lineage would suppress the commitment of naïve T cells to other lineages, restricting the immune system from producing weapons that are not optimal for recognized invaders and causing unnecessary tissue damage. With this knowledge and assumption, the authors tested if Th2 effector cytokines can suppress IL-17 production and demonstrated it is true. Therefore, they concluded that "Through the contribution of IL-17, some Th1-like T cells appear to mediate synovial inflammation". Now we know these Th1-like cells are actually Th17 cells. Besides inflammation, bone destruction is another major manifestation of arthritis. In 1999, Shigeru Kotake et al. showed that IL-17 contributed to this as well[98]. In their experiments using murine cells, IL-17 promoted the differentiation of osteoclasts from their progenitors (osteoblast and hematopoietic cells) by inducing the expression of *RANKL* (Receptor activator of nuclear factor kappa-B ligand, used to be known as osteoclast differentiation factor (ODF)). Blockade of IL-17 significantly inhibited osteoclast formation induced by culture media of RA synovial tissues. In 2000 and 2001, two *in vivo* studies further demonstrated the injection of IL-17 could enhance bone resorption, while IL-17 blockade had protective effects[99,100]. In 2002, Katherine A. Bush et al. presented the first comprehensive assessment of the potential use of IL-17 blockade to treat arthritis using animal models[101]. They induced arthritis symptoms in rats by injecting heat-killed and dried *Mycobacterium butyricum* and subsequently treated them with anti-IL-17-receptor antibodies. A dose-dependent reduction in disease activity was observed. Following this study, Susumu Nakae et al reported that IL-17$^{-/-}$ mice had markedly reduced incidence rate of arthritis under collagen challenge[102]. These two studies, in conjunction with previous phenotyping and functional analyses, laid the foundation for using IL-17 as a therapeutic target.

During the early stages of exploration, researchers also discovered that IL-17 is involved in several other inflammatory diseases. In a study by Kazuhiro Kurasawa et al., it was found that systemic sclerosis patients have higher concentrations of IL-17 in their serum and higher levels of IL-17 mRNA in peripheral blood mononuclear cells (PBMCs) when compared to healthy individuals[103]. The study also revealed the enrichment of IL-17-producing cells from skin lesions and lung liquid of systemic sclerosis patients. Sophie Molet et al. found that in asthma, IL-17-producing cells are significantly increased in sputum and bronchoalveolar lavage fluids of asthma patients compared to healthy individuals[104]. IL-17 was also implicated in inflammatory bowel diseases in a study conducted by S Fujino et al.[105], wherein they detected IL-17 in the mucosa of ulcerative colitis (UC) and Crohn's disease (CD) patients but not in the normal colonic mucosa, infectious colitis, or ischemic colitis. The number of IL-17-producing cells correlated with the disease activity. Recent studies using single-cell transcriptional atlas show that healthy gut tissues actually have some Th17 cells, and the lack of detection of IL-17 in normal mucosa could be due to limitations in the technique[106].

So far, IL-17 had been recognized as a pro-inflammatory molecule derived from T cells that has a profound impact on cells in inflamed tissues in various inflammatory diseases. However, the origin of IL-17-producing CD4+ T cells remained unclear. Studies attempting to fit IL-17-producing cells into the Th1/Th2 paradigm have been unable to provide clear categorization[107,108]. and some IL-17-producing CD4+ T cells do not produce either IFN-γ or IL-4[107]. In fact, Th2 was excluded as the source of IL-17 at the first place as Th2 polarization condition suppress IL-17 production. Also, an apparent problem was that neither Th1 nor Th2 differentiation conditions can efficiently generate IL-17-producing cells from naïve CD4+ T cells. In a review paper published by Austin L. Gurney et al. in 2002[109], the authors proposed "… these data suggest a new paradigm for the consideration of IL-17 within the immune system.". Only one year later, these authors themselves published an original research article providing evidence that IL-23, a newly-identified cytokine,

but not IL-12 (Th1 condition) can promote IL-17-producing cell generation. However, they found that IL-23 primarily affected memory T cells and can only induce a few IL-17-producing cells from naïve cells, therefore not concluding Th17 is a distinct lineage like Th1 and Th2.

IL-23 was discovered by Birgit Oppmann et al in 2000[110]. IL-23 is a heterodimeric protein composed of two subunits: p19 and p40, coded by *IL23A* and *IL12B*. The p19 subunit is unique to IL-23, while the p40 subunit is shared with IL-12. The discovery of IL-23 could be considered as the start of a real paradigm shift as its functional studies soon resolved the puzzles of immunopathology under the old Th1/Th2 paradigm. For a long time before IL-23 was discovered, people believed that IL-12 is essential for T-cell mediated pathology in chronic inflammatory diseases by promoting IFN-γ+ Th1 cell differentiation. This sparked the exploration of IFN-γ blockade therapy in chronic inflammatory diseases. However, conversely, a series of *in vivo* studies showed that IFN-γ+ inhibition/genomic ablation could exacerbate disease presentations. These models include collagen-induced arthritis[111,112], experimental autoimmune encephalomyelitis (EAE)[113,114], nonobese diabetes (NOD)[115], and experimental autoimmune uveitis (EAU)[116]. These findings challenged people's understanding regarding the role of IL-12/IFN-γ axis in the pathology of inflammatory diseases. Previous studies that implicated IL-12 in inflammatory disease diseases primarily used IL-12 p40 gene-targeted mice and neutralizing antibodies against p40. However, when people found IL-23 shares p40 with IL-12 in 2000, an identity crisis was raised. In 2003, a seminal study by Daniel J. Cua et al.[117] demonstrated that it is IL-23, rather than IL-12, that triggers neuroinflammation. The authors compared p19-deficient mice (lack interleukin-23) and p35-deficient mice (lack interleukin-12) using the model of experimental autoimmune encephalomyelitis. They found that although p19-deficient mice generate IFN-γ+ Th1 cells, they do not develop neuroinflammation upon induction, and administration of IL-23 alone can reverse the KO effects exogenous IL-23 could reconstitute EAE susceptibility in the $p19^{-/-}$ and $p40^{-/-}$ mice. The authors also examined the synergistic effect of IL-

12 and IL-23 in disease development. They found that administration of IL-23 to p40-deficient mice did not induce experimental autoimmune encephalomyelitis as strong as the wild-type, while administration of IL-12 followed by IL-23 did result in more intense disease symptoms.

In 2005, two legendary papers published by Harrington et al.[89] and Park et al[88]. definitively established that interleukin-17 (IL-17)-producing CD4+ T cells represent a distinct lineage from Th1 and Th2. The secret to Th17 differentiation, they found, is to inhibit the effector cytokines of Th1 and Th2. As mentioned previously, effector cytokines of one lineage can suppress other lineages. It seemed that Th17 lineage is different from Th1 and Th2 that only need one cytokine for potent differentiation (IL-12 and IL-4 respectively). Th17 differentiation needs additional signals on top of IL-23, and the search for such signaling sparked numerous studies in the following five years. Park et al. additionally showed that IL-17 provokes fibroblasts and lung endothelial cells to produce chemokines and matrix metalloproteinases, ultimately leading to macrophage and neutrophil infiltration into tissue in experimental autoimmune encephalomyelitis. Neutralizing IL-17 with antibodies successfully suppressed disease symptoms, further supporting the therapeutic modulation of IL-17-producing effector T cells as an important method for treating T cell-mediated chronic inflammatory diseases.

A lineage of T cells usually harbors a distinct master transcriptional factor that commits them to a dedicated fate. Th1 has T-bet (also known as T-box expressed in T cells) that controls the expression of *IFNG* and *CXCR3*. Th2 uses GATA3 to regulate expression of IL-4, IL-5, and IL-13. After Th17 was determined as a distinct subset of helper T cells, a missing piece of its biology was which TF is its fate-determining transcriptional factor. This missing piece was found at amazing speed. In 2006, researchers from Daniel J. Cua and Dan R. Littman groups reported that RORγt (RAR-related orphan receptor gamma t) is the master transcriptional factor for Th17[118]. They performed an Affymetrix gene chip analysis to compare the transcriptome of *in vitro* differentiated Th17 and Th1 cells. This analysis showed the enriched expression of RORγt mRNA

in Th17 cells. At the same time, they found a distinct RORγt+ T cell population in the lamina propria when studying the role of RORγt in T cell development. These two pieces of evidence suggested the possibility that these RORγt+ T cells in gut were Th17 cells and RORγt is potentially an important regulator of Th17 development and function. Thus, they examined if Rorγt $^{-/-}$ naïve CD4+ T cells were able to generate Th17 cells and found that only limited IL-17-producing cells were generated. They also showed that over-expression of RORγt in naïve CD4+ T cells can induce IL-17 production. These data strongly supported RORγt as a critical transcriptional factor of Th17 cells. Now it is recognized that RORγt is the central regulator and robust marker for these cells.

Th17 plasticity and heterogeneity

Th17 cells exhibit plastic changes in their effector programs under inflammatory conditions, therefore forming functionally and transcriptionally distinct subsets. In 2007, Francesco Annunziato et al. reported the presence of IFN-γ–producing Th17 as a subset distinct from the IFN-γ-negative Th17 in the gut of subjects with Crohn's disease[119]. Later studies showed IFN-γ–producing Th17 harbors strong Th1 features such as the expression of T-bet and CXCR3[120]. Under an inflammatory environment, these IFN-γ–producing Th17 cells can further lose their RORγt and IL-17 production and become Th1-like “ex-Th17”[120]. The fate of human Th17 cells towards IFN-γ-IL10+ Th17 or IFN-γ+ Th17 can be determined by the type of pathogens challenging the immune system. For example, Christina E. Zielinski et al. reported that *Candida albicans* could induce IFN-γ+ Th17 while *Staphylococcus aureus* induced IL-10+ Th17 cells. However, when IL-1β is present, IFN-γ+ Th17 will become dominant when under *Staphylococcus aureus* challenge[121]. In animal models of multiple sclerosis and inflammatory bowel diseases, the presence of IFN-γ+ Th17 cells were found to be required for disease induction and IL-23 was attributed to the generation of these cells[122]. Thus, IFN-γ+ Th17 were called “pathogenetic” Th17 cells. A recent study from Kuchroo lab used single-cell RNA/TCR sequencing technique to

intensively profile Th17 cells from IL-17 fate mapping mice at health and EAE condition[123]. They found that these phenotypically different Th17 subsets represent different states of Th17 cells. They proposed a model explaining how Th17 mediates chronic inflammatory disease: a repertoire of SLAMF6+ stem-like Th17 cells is maintained in the gut, which can circulate and be converted to CXCR6+IFN-γ+GM-CSF+ effector Th17 cells. These CXCR6+ Th17 cells can be recruited to inflammatory tissue through CXCL16 and contribute to sustained inflammation.

4.1.3 Type 3 immunity in ankylosing spondylitis

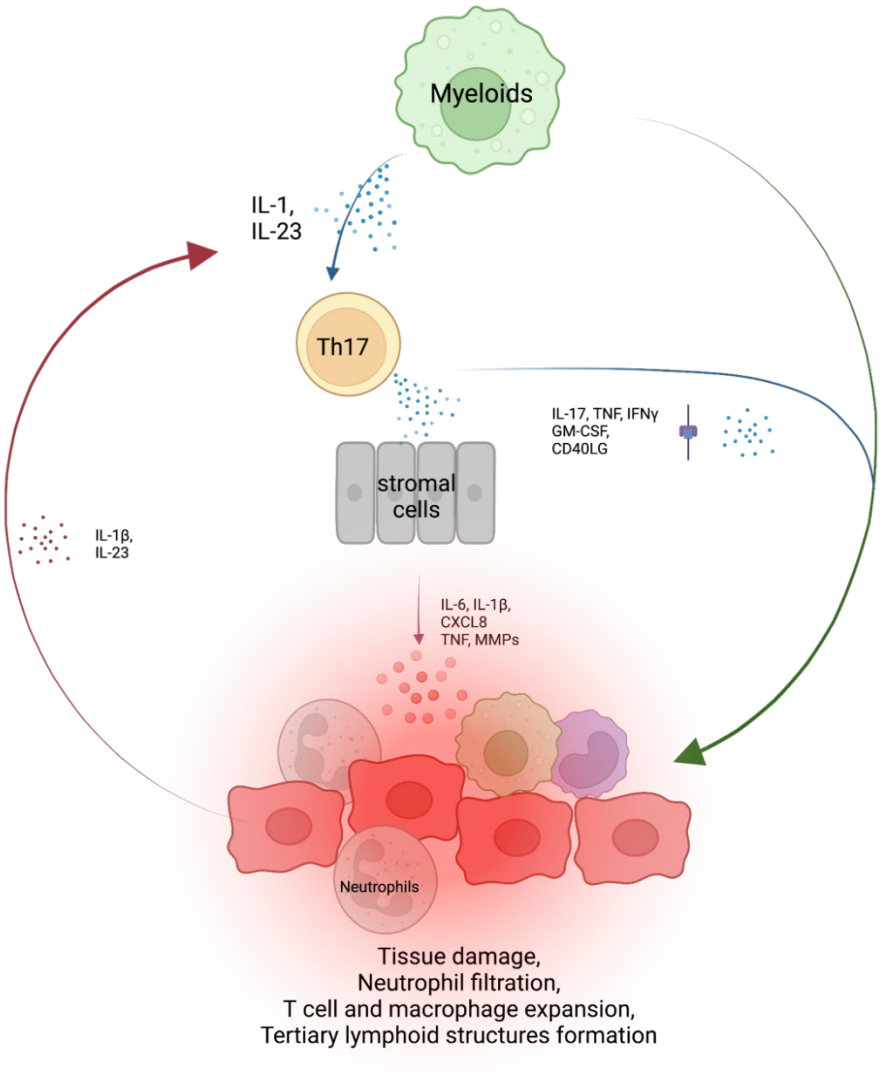


Figure 1.2 Type 3 immunity-driven chronic inflammation in AS (generated using biorender).

Type 3 immunity has been identified as a key driver of inflammation in ankylosing spondylitis (Figure 1.2). Genetic studies provided evidence implicating Type 3 immunity in AS at the early stage. In 2007, the first genome-wide association study in AS identified rs11209032 at *IL23R* locus affects susceptibility to AS, plays a major role in the generation and/or maintenance of Th17 cells[124]. Subsequent studies have associated several other functional or marker genes of Th17 cells with AS susceptibility, including *IL1R1*[33], *IL6R*[62], *RORC*[33], *CCR6*[33], and *TYK2*[62]. Later

phenotyping studies further supported the involvement of the IL-23/IL-17 axis in AS. Patients with AS have been found to exhibit higher levels of IL-17 in their blood compared to healthy individuals, with this concentration linked to disease activity[125,126]. In the inflammation site, more IL-17 and IL-17-producing cells were enriched in the synovial fluid of AS patients compared to their blood[127]. Finally, the success of IL-17 blockade therapy in AS provided conclusive evidence supporting the role of IL-17 and IL-17-producing cells in inflammation maintenance in AS.

Despite the considerable success of IL-17 blockade therapy, a number of unanswered questions persist with respect to the pathological mechanism of the IL-23/IL-17 axis in AS. These include: (1) whether Th17 cells are the primary source of IL-17 or if other cell types contribute; (2) the precise triggers that initiate excessive type 3 immunity; and (3) the environmental factors that drive Th17 cells from a homeostatic state, where they function to protect against infection, to a pathogenic state that promotes inflammation and disease. Further investigation is needed to address these outstanding questions and to gain a better understanding of the complex interplay between IL-23, IL-17, and the various cell types involved in the pathogenesis of AS. Solving those problems can potentially help improve current therapy by avoiding the increased infection risk caused by IL-17 blockade therapy and providing therapies for patients who are now not responding to this treatment.

Various studies have identified alternative cellular sources of IL-17 apart from Th17 cells, including γδ T cells, MAIT cells, and CD8+ T cells. γδ T cells, although scarce in secondary lymphoid organs, are prevalent in diverse peripheral tissues such as skin, intestines, and lungs, distinguishing them as a unique T cell subpopulation. Similar to αβ T cells, γδ T cells originate in the thymus and undergo T cell receptor gene somatic mutation, while employing TRD and TRC instead of TRA and TRB. Notably, the TCRs of γδ T cells recognize a wide range of antigens, unlike αβ T cells. A 2004 study identified γδ T cells as a primary source of IL-17+ and mediators of neutrophil infiltration in response to FasL+ tumor cells[128], sparking the interest of examining

their role as a crucial cellular source of IL-17 in health and disease. In 2019, Koen Venken et al. demonstrated that γδ T cells are the main IL-17 producers among synovial T cells in *in vitro* restimulation experiments[129]. Mucosal-associated invariant T (MAIT) cells are innate-like T cells characterized by CD8 and CD161 expressions. These cells utilize TCR α and β chains while being restricted to Vα7.2-Jα33. In 2016, Eric Gracey et al. reported that IL-17+ MAIT cells occur more frequently in the blood of AS patients than in healthy individuals, although the overall proportion of MAIT cells remained constant[130]. Additionally, they discovered that MAIT cells are enriched in the synovial fluids of AS patients and can produce IL-17 under IL-7 stimulation[130]. The presence of IL17+CD8+ T cells in AS blood and synovial fluid was also reported, but it remains uncertain whether these cells are innate-like T cells or conventional cytotoxic T cells[131]. There are also non-T cells that were suggested as source of IL-17A in SpA, including group 3 innate lymphoid cells (ILC3s) and NK cells[132]. It is important to note that these findings were based on *in vitro* reactivation experiments, which assess the capacity for IL-17 production but not the actual cellular activity *in vivo*. To conclusively determine the primary cellular sources at the site of inflammation, a sensitive assay must be developed to directly measure IL-17 production by all cell types.

As to the second question, mechanical stress is a plausible cause initiating chronic inflammation at and around joints. Repetitive or increased strain can cause microdamage and the repair process can involve homeostatic immune response[133]. Athletes and women who have given birth could develop inflammatory change in their sacroiliac joints while this inflammation is usually going to resolved and not develop into ankylosing spondylitis[134,135]. An immune system polarized by genetic components and environmental factors in certain individuals might eventually break this homeostatic immune response and amplify the inflammation. However, this hypothesis lacks supporting evidence and it is not clear which cells and which molecules participate in this progress.

In terms of the third question, the model proposed by Kuchroo lab (mentioned in the Th17 plasticity and heterogeneity section) could, in my opinion, be close to the truth. Homeostatic Th17

cells are maintained by gut microbes and turned to pathogenic states when being recruited to inflamed joints.

4.2 Peripheral helper T cells and tertiary lymphoid structures

Peripheral helper T cells (Tph), a recently identified T cell subset defined as CD4+PD-1^{hi}CXCR5-, have garnered increasing attention in the contexts of autoimmunity and cancer[136,137]. No published studies have fully examined the role of Tph cell in AS yet, but a thorough investigation is worthwhile given the substantial pathology overlapping between AS and other inflammatory conditions.

Tph possesses comparable functions to follicular helper T cells (Tfh) (CD4+PD-1^{hi}CXCR5+), which support survival and differentiation of plasmablasts [137]. However, unlike Tfh, which primarily localize to and function within secondary lymphoid organs through CXCR5-mediated chemotaxis, Tph reside in the periphery at inflammatory sites[137]. This is due to their lack of *CXCR5* expression and instead, their expression of chemokine receptors that facilitate migration to inflamed tissues[138]. At these sites, Tph promotes the formation of tertiary or ectopic lymphoid structures (TLS or ELS, respectively) by producing the chemokine CXCL13, analogous to the role played by follicular dendritic cells (fDCs) in organized secondary lymphoid organs[139]. The expression of CXCL13 and subsequent TLS (tertiary lymphoid structures) formation by Tph recruit B cells, and Tph then support B cells similarly to Tfh by secreting interleukin-21, interacting with SLAMF5, and expressing *ICOS*, *MAF*, and other factors[136]. This peripheral interaction facilitates antibody production by B cells by promoting plasma cell differentiation.

The peripheral functions of Tph cells have broad implications for various disease contexts. In healthy individuals, Tph cells are present in circulation in limited numbers, while autoimmune patients, particularly those with autoantibody-mediated conditions, exhibit increased Tph cell counts[136]. Tph cells have been associated with pathological roles in several autoimmune diseases,

such as RA[136,140], UC[141], SLE[142], and Sjogren's syndrome[143]. These observations are most likely attributable to their capacity to promote the formation of tertiary lymphoid structures (TLS) and facilitate autoantibody production. Indeed, seropositive rheumatoid arthritis patients display higher Tph cell counts compared to seronegative individuals[136]. In contrast, Tph cells have been linked with improved prognosis in diverse types of cancer, which may be linked to the formation of TLS at the tumor site and the potential generation of tumor-specific antibodies by B cells[139]. Additional research is necessary to clarify the mechanistic roles of Tph cells in different disease contexts and determine whether they can serve as a therapeutic target for both autoimmune diseases and cancer.

Furthermore, the current comprehension of Tph generation harbors some uncertainty. *In vitro* studies have indicated that both IL-2 neutralizing and addition of TGF-β are necessary for Tph generation from naïve CD4+ T cells, with Sox4 recognized as a crucial transcription factor in their development[140,144]. Comparable conditions are believed to be essential for Tfh generation, making it somewhat indistinct what prompts the Tph vs Tfh developmental divergence[137]. The possibility of Tfh to Tph (or vice versa) interconversion has also been inadequately researched and is worth further examination, given their differing expression levels of *Bcl6* (the Tfh master transcription factor) and *Blimp1* (a transcription factor upregulated in Tph but downregulated in Tfh), which reciprocally repress each other[137,145]. Uncertainty also persists as to whether pro-inflammatory cytokines, such as TNF-α and IL-1β, enhance or suppress Tph generation *in vitro*[140,144]. Different conditions may promote Tph generation in disease-specific contexts. Lastly, the cellular interactions that could support Tph generation are yet to be explored. Given the suggested TGFβ and IL-2 limited conditions *in vitro*, it is likely that Treg cells play a significant role.

5. Project aims

As mentioned in the *Management of ankylosing spondylitis* section, a better understanding into AS pathology is clearly needed for the development of more therapeutic options and the eventual achievement of a cure for this disease. This DPhil project endeavors to combine genetics approaches and immunology to elucidate the cellular and molecular mechanisms responsible for the immune dysfunction observed in AS. Specifically, the proposed research plan includes (a dedicated introduction section justifying the rationale and the use of methods is included in each separated experimental chapter):

1) Examining the genomic regulation of risk variants shared by AS and its associated conditions, as well as exploring the biological functions of the corresponding risk genes. This investigation may shed light on the common pathological features underlying AS and related inflammatory conditions and reveal potential avenues for drug repurposing.
2) Investigating the mechanisms underlying monocyte activation in AS, and identifying genetic risk factors that regulate these cellular processes.
3) Combining genetic analysis with single-cell transcriptomics to pinpoint pathogenic T cell subsets in AS.

Chapter 2: Material and Methods

1. Laboratory experiments

1.1 Patient and Control Recruitment

AS synovial fluid and blood samples collected in Oxford: Samples were collected from patients attending the Oxford University Hospitals National Health Service Foundation Trust (OUH) with appropriate consent and ethical approval (Ethics reference number 06/Q1606/139, National Health Service, Health Research Authority, South Central - Oxford C Research Ethics Committee). All patients included in this study met the criteria of the Assessment of Spondyloarthritis International Society (ASAS) for AxSpA (which we henceforth term AS). Demographics of 51 AS patients recruited for functional experiments (excluding scRNA-seq experiments) were described in Table 2.1. Demographics of AS patients recruited for scRNA-seq experiments were described in related method sections.

Table 2.1 Demographics of AS patients (n = 51)

Age, mean (range) years	45 (25-89)
Sex, male/female	32/19
HLA-B27+ no. (%)	34* (82.9%)
ESR, mean (range)	11.8 (2-49)
CRP, mean (range)	10.76 (0.2-87.5)
BASDAI, mean (range)	4.3 (0.6-8.5)
NSAIDs therapy current	28 (54.9%)
Biologic DMARD therapy current	15 (29.4%)
Synthetic DMARD therapy current	2 (4%)
* Unknown HLA-B27 status in 10 AS patients	
ESR: Erythrocyte sedimentation rate; CRP: C-reactive protein; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; NSAIDs: Non-steroidal anti-inflammatory drugs; DMARD: Disease-modifying antirheumatic drugs.	

AS synovial fluid and blood samples collected in Gwangju, Korea: All patients consent to the use of their biological samples under approved ethics (CNUH-2011-199). All AS patients fulfilled the modified New York criteria.

Crohn's disease ileum biopsies collected in Guangzhou, China: Research was approved by the ethics committee in Guangdong Provincial People's Hospital. The written consent form was obtained from all patients. Crohn's disease patients fulfilled the criteria of Crohn's Disease Activity Index (CDAI) >220 and <450.

1.2 Mononuclear cell separation from peripheral blood

Blood was stored in BD Vacutainer® Heparin Tubes during transfer. Blood was dilute use RPMI 1640 (Sigma-Aldrich; cat.no. R8758) at a 1:1 ratio. Diluted samples were then layered over 15 ml of Histopaque in 50ml tubes. Tubes were then centrifuged at 850g at RT without brake for stopping. The white cells interface was then collected using Pasteur pipettes. Collected cells was then washed using RPMI 1640 and counted using a hemacytometer.

1.3 Mononuclear cell separation from synovial fluids

Synovial fluids were stored in BD Vacutainer® Heparin Tubes during transfer. Heparin Tubes containing synovial fluids were centrifuged for 10 minutes at 850g at room temperature. After the removal of the supernatant, pelleted cells were then resuspended in 25ml RPMI 1640 (Sigma-Aldrich, R8758). Cellular suspensions were then layered over 15ml of Histopaque (Sigma-Aldrich, 10771-500ML) in 50ml tubes. Tubes were then centrifuged at 850g for 30 mins at RT without brake for stopping. The white cells interface was then collected using Pasteur pipettes. Collected cells were then washed using RPMI 1640 (Sigma-Aldrich, R8758) with 10% FBS (Sigma-Aldrich, F9665-500ML) and counted using a hemocytometer.

1.4 Magnetic isolation of cells from peripheral blood mononuclear cells

This study involved CD14+ monocyte isolation and T-cell isolation. CD14+ monocyte isolation used CD14 MicroBeads (positive isolation) from Miltenyi Biotec (130-050-201) (In-house testing demonstrated this kit can yield CD14+ monocytes of ~95% purity (n = 1, Figure 2.1).). T cells

isolation used Pan T Cell Isolation Kit (negative isolation), human from Miltenyi Biotec (130-096-535) (In-house testing demonstrated this kit can yield CD3+ cells of 94% purity (n = 1, Figure 2.2).). Operations were referred to the manufacturer's protocol.

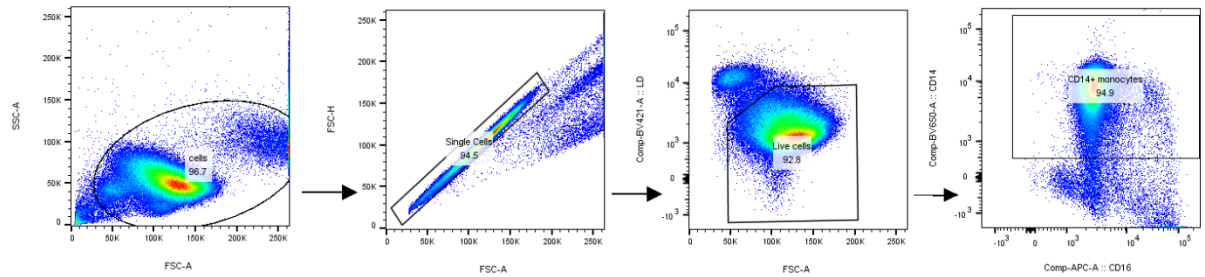


Figure 2.1 Purity of CD14+ monocytes after magnetic isolation. PBMCs from a health donor were used for magnetic isolation using CD14 MicroBeads from Miltenyi Biotec (130-050-201). The figure was made using Flowjo.

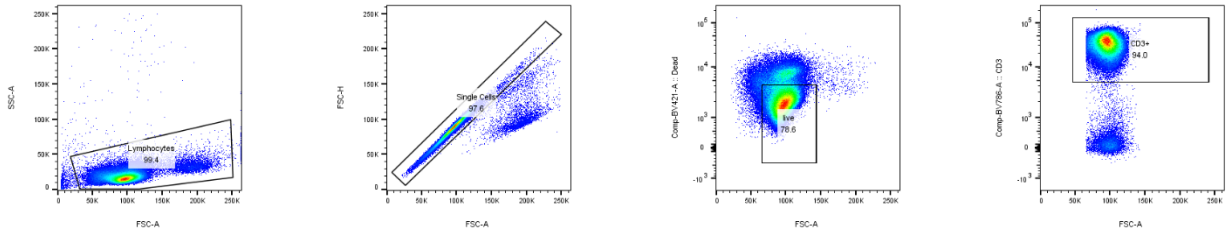


Figure 2.2 Purity of pan-T cells after magnetic isolation. PBMCs from a health donor were used for magnetic isolation using Pan T Cell Isolation Kit (negative isolation), human from Miltenyi Biotec (130-096-535). The figure was made using Flowjo.

1.5 Tissue digestions

Tissues were freshly collected and placed on Petri dishes. RPMI 1640 (room temperature) was used to wash the biopsies. Next, tissues were minced into small pieces and put into a prewarmed digestion buffer (RPMI 1640 containing 10% FBS, 100 µg/ml Liberase TL, and 10 mg/ml DNAse I). During digestion, magnetic stirring was applied, and tubes were incubated at 37 °C. After enzymatic digestion, cells were collected from the suspension through centrifuge and debris were removed through 70um cellular strainers.

1.6 Cellular culture and stimulation

Assay of monocytes activation in PBMCs: $1x10^6$ PBMCs were cultured in 200 µl medium in non-tissue culture treated 96-well round bottom plates (Corning). The use of non-tissue culture treated is critical here, as monocytes can adhere to tissue culture-treated plates. as monocytes have the propensity to adhere to tissue culture-treated plates. The retention of these cells on such plates poses a challenge for their subsequent transfer to the next container for subsequent treatments. The cells were either left unstimulated or treated with 200 ng/ml LPS (Enzo Bioscience) or anti-CD2/CD3/CD28-coated T-cell activation/expansion beads (Miltenyi Biotec) for 16 hours.

Assay of T cells expansion in PBMCs: PBMCs and CD14-depleted PBMCs were cultured in 96-well u-bottom plates at a density of 5×10^5 per well and stimulated with anti-CD2/CD3/CD28-coated T-cell activation/expansion beads (Miltenyi Biotec) in the presence of IL-2 (20 ng/ml, PeproTech) for 6 days. For Th17 polarization condition, recombinant human IL-23 (20 ng/mL, PeproTech) and IL-1β (20 ng/mL, PeproTech) were added at the initiation of culture. For Th0 polarization condition, cells were stimulated as above in the absence of IL-1β and IL-23. On day 3, half of the culture medium was replaced with the fresh medium containing the same concentration of IL-2, IL-23 and IL-1β. On day 6, cells were harvested for flow cytometry staining.

1.7 siRNA silencing of ERN1 gene in primary monocytes (conducted by Jinyi Zhao)

Primary monocytes were cultured in 24-well plates and then transfected with control non-targeting siRNA (Dharmacon) or siRNA targeting *ERN1* (endoplasmic reticulum to nucleus signaling 1, which codes a protein named inositol-requiring enzyme 1 α (IRE1α)) (Dharmacon) using HiPerFect reagent (QIAGEN) following the instructions provided by the manufacturer. After 3 days, the cells were collected for subsequent experiments. The knockdown efficiency was determined by quantitative PCR (qPCR) and western blotting. For qPCR, primary monocytes were pelleted and resuspended in TRIzol. Total RNA was extracted using the Direct-zol RNA MiniPrep Kit according to the manufacturer's instructions (Zymo Research, R2052). The RNA was then reverse transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 4368813). Using Taqman Gene Expression Assay probe Hs00980095_m1 for ERN1 and Hs1060665_g1 for ACTB (Applied Biosystems), quantitative real-time PCR was performed with an Applied Biosystems Prism 7500 Fast Sequence Detection System using the TaqMan fast universal PCR master mix reagents (Applied Biosystems, 4352042). The mRNA expression levels were analysed using the 2–ΔΔCt method.

For western blotting, primary monocytes were lysed using RIPA (Radioimmunoprecipitation assay) buffer supplemented with protease inhibitors (Sigma). Protein concentrations were quantified by bicinchoninic acid (BCA) assay. Total proteins (18 μg) were electrophoresed through a 12% sodium dodecyl sulfate (SDS)–polyacrylamide gels, transferred to the polyvinylidene difluoride membranes and analyzed by immunoblotting. The primary antibodies used were rabbit monoclonal anti-IRE1α (1:1000 dilution; Cell Signalling; 14C10) and mouse anti-α-Tubulin (1:1000 dilution; Cell Signalling; DM1A). The secondary antibodies used were goat anti-rabbit IgG (1:35000, Bethyl Laboratories, A120-201P) or goat anti-mouse IgG (1:25000, Bethyl Laboratories, A90-516P).

1.8 Intracellular cytokine staining (ICS)

For T-cell, cells were stimulated with 100ng/mL PMA (Sigma-Aldrich) and 1ug/mL Ionomycin (Sigma-Aldrich) for 2.5 h followed by 2.5 h incubation with 5 µg/ml Brefeldin-A (Sigma-Aldrich) before intracellular cytokine staining. The following FACS ((Fluorescence-activated cell sorting)) anti-human antibodies were used: CD3 (clone OKT3; Biolegend); CD4 (clone RPA-T4; BioLegend) and IL-17A (clone eBio64DEC17; eBioscience). For monocyte, cells were stimulated with either LPS or T cell activation beads for 4 h followed by 12 h incubation with Brefeldin-A (Sigma-Aldrich). Following FACS (Fluorescence-activated cell sorting) anti-human antibodies were used for analysis: CD3 (clone OKT3; Biolegend); CD14 (clone M5E2; Biolegend); IL-1 beta (clone CRM56 eBioscience); IL-6 (MQ2-13A5; eBioscience); XBP1s (Q3-695; BD Biosciences) and TNF-a (clone Mab11; BioLegend). To identify living cells and exclude dead cells, samples were stained with LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit (L34955, Invitrogen). For intracellular cytokines staining experiments, BD Cytofix/Cytoperm™ Fixation/Permeabilization Kit (554714) was used for cell fixation and permealization. For ICS experiments involving transcription factor staining, BD Pharmingen™ Transcription Factor Buffer Set (562574) was used instead. BD LSRFortessa™ Cell Analyzer was used for the flow cytometry analysis. Plate mode was used for sample injection and events from 50ul cell solutions were recorded. Compensation were performed using OneComp eBeads™ Compensation Beads (Thermo Fisher, 01-1111-41). Data analysis was performed using FlowJo software.

A detailed intracellular staining protocol using 96-well round bottom plates was described below:

a. Remove supernatant (culture medium usually) from pelleted cells sample.
b. Pipette 50ul surface staining antibody solutions (0.1-10 µg/ml in FACS buffer (1% FBS in PBS)) to each well, mix by pipetting.
c. Put 96-well plate on ice box with cover, incubating for 20 minutes.
d. Wash X2: Centrifuge at 500g for 3 minutes, wash with 200ul FACS buffer
e. Fix/Perm: aspirate residual stain buffer and add 100 ul of freshly prepared 1x Fix/Perm Buffer working solution to each well and resuspend cell pellets by pipetting

f. Put 96-well plate on ice box with cover, incubating for 20 minutes.
g. Wash X2: Centrifuge at 500g for 3 minutes, wash with 200ul FACS buffer
h. Pipette 50ul intracellular staining antibody solutions to each well, mix by pipetting.
i. Put 96-well plate on ice box with cover, incubating for 20 minutes.
j. Wash X2: Centrifuge at 500g for 3 minutes, wash with 200ul FACS buffer
k. Resuspend cells with 100ul FACS buffer for subsequent flow cytometry analysis.

1.9 Enzyme-linked immunosorbent assays (ELISA) (conducted by Jinyi Zhao)

The concentrations of IL-23, IL-1β, TNF-α and CXCL2 in cell culture supernatants were assayed using standard ELISA kits (Invitrogen or R&D) in accordance with the protocol set out by the manufacturer. The Multiskan Ascent ELISA reader was employed to read the ELISA plates at a test wavelength of 450 nm and a background wavelength of 570 nm. Then, the standard curve of ELISA was plotted using Microsoft Excel software, while the corresponding concentrations of samples were calculated using the absorbance value and standard curve.

1.10 Single-cell RNA sequencing using 10X platform

In vitro monocytes in PBMCs experiment: Blood samples from donor AS1802, AS1830, and AS2311 were used in this experiment, whose demographics was described in Table 2.2. For AS1802, we loaded 20,000 cells from each condition into a channel of Chromium Next GEM Chip K of the 10x Genomics platform. For AS1830 and AS2311, we multiplexed cells from the same conditions at a 1:1 ratio using Total-seq-C hashtag antibodies (C0251 and C0252) and loaded 20,000 multiplexed samples from each reaction onto the chip separately. Single-cell RNA-sequencing libraries were generated using the Chromium Next GEM Single Cell 5' v2 kit according to the manufacturer's instructions. The resulting DNA libraries were sequenced at >50,000 reads/cell on a Novoseq 6000 using PE150 mode.

Table 2.2 Demographics of AS patients (n = 3)

Donor ID	AS1802	AS1830	AS2311
Age	48	36	33
Gender	Female	Female	Male
BASDAI score	6.2	0.8	2.9
CRP level	1 mg/L	Not available	Not available
HLA-B27-positive	Positive	Positive	Positive
Current treatments	Meloxicam	Methotrexate	Naproxen

Ex vivo profiling of paired PBMCs and SFMCs: For two pairs of AS blood and synovial fluid samples, isolated PBMCs and SFMCs were subject to fluorescence-activated cell sorting to enrich for CD3- live cells. Library construction and sequencing strategies were similar to those used in the *in vitro* stimulation experiment, except the use of the Chromium Next GEM Single Cell 5' v1 kit instead of v2. For the additional six donors, whole PBMCs and SFMCs were directly sequenced using a similar workflow.

2. Computational analysis:

2.1 GWAS Database

GWAS data sourced from 68 studies for 31 conditions were selected for GWAS data curation (**Table 2.3**). Each condition has at least one published study with available GWAS summary statistics and >2000 individuals (not including the control cohort) in its discovery cohort. Genetic variants passing P-value < 1×10^{-6} were accepted for the following analysis. Data were either directly downloaded as summary statistics/set of lead SNPs or request through email. When data were only available as hg19 coordinates, Liftover tool from UCSC was used to convert the coordinates to hg38.

Table 2.3 GWAS result data sources

Author/Consortium	Traits	Published Date	Pubmed ID
Lars G Fritsche, et al.	Age-related macular degeneration	21 December 2015	26691988
Johannes Waage, et al.	Allergic Rhinitis	16 July 2018	30013184
Brian W. Kunkle, et al.	Alzheimer's disease	28 February 2019	30820047
eremy Schwartzentruber, et al.	Alzheimer's disease	15 February 2021	33589840
Itziar de Rojas, et al.	Alzheimer's disease	7 July 2021	34099642
IGAS	Ankylosing Spondyloarthrtis	9 June 2013	23749187
David Ellinghaus, et al.	Ankylosing Spondyloarthrtis	14 March 2016	26974007
Carolina Roselli, et al.	Arrhythmia	11 June 2018	29892015
Jonas B Nielsen, et al.	Arrhythmia	30 July 2018	30061737
Florence Demenais, et al.	Asthma	22 December 2017	29273806
Michelle Daya, et al.	Asthma	20 February 2019	30787307
Thorunn A. Olafsdottir, et al.	Asthma	20 January 2020	31959851
Yi Han, et al.	Asthma	15 April 2020	32296059
David Ellinghaus, et al.	Atopic dermatitis	2 June 2013	23727859
Lavinia Paternoster, et al.	Atopic dermatitis	19 October 2015	26482879
Nao Tanaka, et al.	Atopic dermatitis	8 June 2021	34116867
Saedis Saevarsdottir, et al.	Autoimmune thyroid diseases	24 June 2020	32581359
Mark K Bakker, et al.	Brain aneurysm	16 November 2020	33199917
Hélène Choquet, et al.	Cataract	14 June 2021	34127677
ICGC	Chronic obstructive pulmonary disease	6 February 2017	28166215
ICGC	Chronic obstructive pulmonary disease	25 February 2019	30804561
Stephan Buch, et al.	Cirrhosis	19 October 2015	26482880
Patrick C A Dubois, et al.	Coeliac disease	28 February 2010	20190752
Gosia Trynka, et al.	Coeliac disease	6 November 2011	22057235
UKBB	Coronary artery disease	17 July 2017	28714975
Pim van der Harst, et al.	Coronary artery disease	6 December 2017	29212778
Satoshi Koyama, et al.	Coronary artery disease	5 October 2020	33020668
Joshua Chiou, et al.	Diabetes mellitus type I	19 May 2021	34012112
Jamie R J Inshaw, et al.	Diabetes mellitus type I & II	8 April 2021	33830302
Anubha Mahajan, et al.	Diabetes mellitus type II	8 October 2018	30297969
Ken Suzuki, et al.	Diabetes mellitus type II	4 February 2019	30718926
Cassandra N. Spracklen, et al.	Diabetes mellitus type II	6 May 2020	32499647
Akiyoshi Nakayama, et al.	Gout	1 April 2020	32238385
UKBB	Gout	2 October 2019	31578528
Katrina M de Lange, et al.	Inflammatory Bowel Diseases	9 January 2017	28067908
Rainer Malik, et al.	Ischemic Stroke	12 March 2018	29531354
Anne Hinks, et al.	Juvenile idiopathic arthritis	21 April 2013	23603761
Elena López-Isac, et al.	Juvenile idiopathic arthritis	26 October 2020	33106285
IMSGC	Multiple sclerosis	27 September 2019	31604244
MSKKP	Osteoarthritis	24 November 2021	34822786
Mike A Nalls, et al.	Parkinson's disease	1 September 2014	25064009
Mike A Nalls, et al.	Parkinson's disease	December 2019	31701892
Cornelis Blauwendraat, et al.	Parkinson's disease	22 November 2019	31755958

Author/Consortium	Traits	Published Date	Pubmed ID
Derek Klarin, et al.	Peripheral arterial disease	8 July 2019	31285632
Jimmy Z Liu, et al.	Primary biliary cholangitis	9 September 2012	22961000
Heather J. Cordell, et al.	Primary biliary cholangitis	22 September 2015	26394269
Jimmy Z Liu, et al.	Primary sclerosing cholangitis	21 April 2013	23603763
Sun-Gou Ji, et al.	Primary sclerosing cholangitis	19 December 2016	27992413
Liang-Dan Sun, et al.	Psoriasis	17 October 2010	20953187
Eva Ellinghaus, et al.	Psoriasis	17 October 2010	20953188
Philip E Stuart, et al.	Psoriasis	17 October 2010	20953189
GAPC & Wellcome Trust...	Psoriasis	17 October 2010	20953190
Lam C Tsoi, et al.	Psoriasis	11 November 2012	23143594
Lam C. Tsoi, et al.	Psoriasis	24 May 2017	28537254
Yukinori Okada, et al.	Rheumatoid arthritis	25 December 2013	24390342
Eunji Ha, et al.	Rheumatoid arthritis	11 December 2020	33310728
Timothy R D J Radstake, et al.	Scleroderma	11 April 2010	20383147
Elena López-Isac, et al.	Scleroderma	31 October 2019	31672989
James Bentham, et al.	Systemic Lupus Erythematosus	26 October 2015	26502338
David L Morris, et al.	Systemic Lupus Erythematosus	11 July 2016	27399966
Carl D. Langefeld, et al.	Systemic Lupus Erythematosus	17 July 2017	28714469
Yong-Fei Wang, et al.	Systemic Lupus Erythematosus	3 February 2021	33536424
Ying Jin, et al.	Vitiligo	23 January 2019	30674883
UKBB-Neale	7,228 traits in UK biobank	15 June 2020	
Saori Sakaue, et al.	220 traits	30 October 2021	34594039
BioBank Japan Project	42 diseases	8 June 2020	32514122
David Ellinghaus, et al.	5 chronic inflammatory diseases	14 March 2016	26974007

2.2 Immune eQTL Database

eQTL (Expression quantitative trait loci) data were curated from 12 independent studies and significant thresholds defined by the original authors were used (**Table 2.4**). When data were available as hg19 (The human genome reference hg19, completed in 2009, also known as GRCh37 (Genome Reference Consortium human genome build 37), is a version of the human genome widely used in genomic research, https://www.ncbi.nlm.nih.gov/genome/guide/human/) coordinates, Liftover tool from UCSC was used to convert the coordinates to hg38 (The human genome reference 38, completed in 2013, also known as GRCh38 (Genome Reference Consortium human genome build 38).

Table 2.4 eQTL data sources

Author/Consortium	Cohort size	Cell types	Stimulation	Published Date	Pubmed ID
Benjamin Fairfax et al.	432	Monocytes	LPS, IFNγ	2014	24604202
Vivek Naranbhai et al.	101	Neutrophils	No	2015	26151758
Hélène Quach et al.	200	Monocytes	LPS, Pam3CSK4, R848, Influenza A	2016	27768888
Yohann Nédélec et al.	175	Macrophages	Listeria monocytogene, Salmonella typhimurium	2016	27768889
Blueprint	197	T cells, Monocytes, Neutrophils	No	2016	27863251
JapanBB	105	CD4+ T cells, CD8+ T cells, B cells, NK cells, Monocytes, PBMCs	No	2017	28553958
DICE	91	14 types of blood cells	No	2018	30449622
Cedar	100	CD4+ T cells, CD8+ T cells,B cells, Monocytes, Neutrophils	No	2018	29930244
ImmuNexUT	337	28 types of blood cells	No	2021	33930287
eQTLgen	31,684	PBMCs	No	2021	34475573

2.3 Drug target analysis

The Drug-target database was requested from the DrugBank (https://go.drugbank.com/). Only targets active on the market (pharmacologically active) were included as approved targets.

2.4 Genetic risk overlapping and clustering analysis

To quantify the correlation between each pair of conditions, an enrichment score was calculated based on a rationale similar to over-representation analysis (**Formula 1**). The enrichment score was then transformed to a scale from 0 to 1(**Formula 2**).

$$Enrichment\ Score = \frac{N_{overlap} \times N_{all_traits}}{N_{trait1} \times N_{trait2}} \quad [1]$$

$N_{overlap}$: Number of genetic variants overlapping between trait 1 and trait 2
N_{all_traits}: Number of genetic variants associated at least one condition out of 31 included
N_{traitn}: Number of genetic variants associated with trait n.

$$Genetic\ distance\ index = 1 - \frac{1}{Enrichment\ Score + 1} \quad [2]$$

Using the genetic distance index for each pair of conditions, a genetic risk overlapping matrix was formed. A hierarchical clustering was then applied to this matrix to group conditions with similar genetic risk structure.

2.5 ATAC-seq and ChIP-seq

Fastq data was downloaded and aligned to hg38 reference genome using bwa (Burrows-Wheeler Aligner, a software package for mapping low-divergent sequences against a large reference genome.). Reads aligned to blacklist defined by Encode were removed. PCR duplicates were removed using Picard tools. Signal tracks were then generated using MAC2.

2.6 H3K27ac HiChIP

Fastq data were downloaded and aligned to the hg38 reference genome using bwa. Valid pairs were then recorded and PCR duplicates were removed using pairtools. FitHiChIP was used to call significant loops from the valid pairs. For FitHiChIP, significant peaks identified using H3K27ac ChIP-seq of CD4+ T cells from ENCODE were used as 1D input. Only loops with at

least one anchor mapped to H3K27ac ChIP-seq were considered. Coverage was used as bias factor. Resolution was set to 1000bp.

2.7 Single-cell RNA-seq analysis

The paired reads obtained were mapped to the hg38 reference genome to generate gene expression matrices using CellRanger v7.0.0 (https://www.10xgenomics.com/cn/support/software/cell-ranger). The raw matrices were then analyzed using the Seurat R package (v4)[146]. Cells with low-quality profiles were excluded based on the number of detected genes, the percentage of mitochondrial RNA among total UMIs, and the total number of UMIs. The raw read counts were normalized using the NormalizeData function, and variable genes were identified for each sample. Data across different conditions were then integrated using the SCTtransform function. To select candidate anchor genes, genes were first selected using the SelectIntegrationFeatures function, and TCR genes were then further excluded.

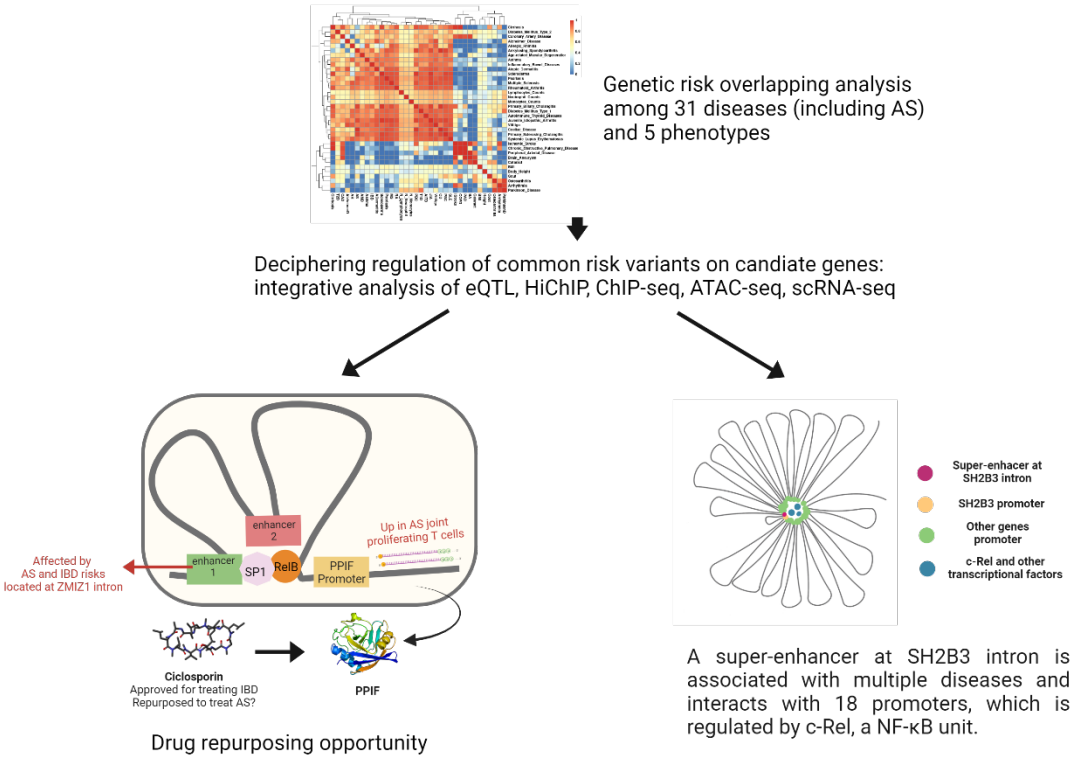
To identify cell types, I performed principal component analysis (PCA) on integrated data with the RunPCA function to generate principal components (PCs) used for dimensionality reduction. PCs representing >90% variance were used for Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction. The FindNeighbors and FindCluster functions were used to cluster cells based on global transcriptional profile. Clusters in the 2-dimensional UMAP were used to identify cell types based on marker genes. When re-clustering CD14+ myeloids, CD14+ populations were extracted, and dendritic cells were further excluded based on the expression of *FLT3*, *CLEC9A*, and *CLEC10A*. Pure CD14+ myeloids were then split by donors and re-integrated to capture the variance within this population. For differential expression analysis, the FindMarkers function was used to test the normalized data with the default Wald test method. Unless specified, default parameters were used for each function.

Feature scores calculation: the normalized and scaled counts of feature were summed and divided by the number of genes used.

AS GWAS candidate gene enrichment analysis: SNPs associated with AS susceptibility reaching $p < 1x10^{-5}$ reported by the International Genetics of Ankylosing Spondylitis Consortium (IGAS)[57] in 2013 and David Ellinghaus in 2016[33] were used as AS risk variants. Cell-type eQTL data were subet from the eQTL database described above. By combining GWAS data and eQTL data, a list of AS GWAS candidate genes was defined, and their enrichment in marker genes of diverse monocyte subsets was tested using clusterProfiler.

Trajectory analysis: I employed Slingshot algorithm[147] for the trajectory analysis in Chapter 5. This algorithm presupposes the existence of a continuum of cellular states within the context of single-cell transcriptional profiling experiments. Subsequently, a developmental trajectory can be deduced from the low-dimensional features of the cell transcriptome (positions on the UMAP). Starting or ending clusters were assigned as indicated. In practice, target cells were extracted and subject to dimension reduction. Principle components and UMAP coordinates were then used for the Slingshot algorithm to infer the developmental trajectory. To interrogate the transcriptional programs mediating cellular development, dynamic changes of gene expression over the trajectory were tested using tradeSeq[148] (fitGAM and associationTest function). Genes that have over two-fold significant changes (p-value < 0.05) were then used to get smooth expression shifting over the trajectory and be plotted as a heatmap. Cluster of genes with similar dynamics patterned with grouped through cutting at certain distance in hierarchical clustering dendrogram.

Chapter 3: Studying common genetic risks shared across AS and its associated diseases



Graphic abstract of major findings

1. Introduction

Ankylosing spondylitis (AS) frequently presents together with other inflammatory conditions, including uveitis, psoriasis, and IBD[7] with some AS patients also developing cardiovascular disease[149]. These conditions have significant overlap in genetic risks (e.g., overlapping risk genes in IL-23/IL-17 axis) and therapies (e.g., TNF inhibition), suggesting a shared pathology[45]. Understanding the shared etiology of AS and its associated conditions could provide a unique perspective on disease development and may reveal further opportunities for drug repurposing. Here, I explored this direction from a genetic perspective.

Genome-wide association studies (GWAS) have identified >100 genetic loci associated with AS (see detail in Chapter 1), while there is no high-throughput approach to decipher the contribution of these loci to the pathology. Prioritizing loci that impact both AS and multiple conditions associated with AS may increase the likelihood of discovering pathogenic genes with fundamental roles in the disease. If proven as an effective and safe therapeutic target, developed therapeutics may treat a variety of different conditions.

The assessment of shared genetic risks in diseases requires the identification of conditions with potential etiological overlap with Ankylosing Spondylitis (AS). In 2015, David Ellinghaus and his colleagues reported a cross-trait genome-wide association study (GWAS) analysis to investigate the genetic risk pleiotropy in AS and other clinically closely related conditions, including psoriasis, Crohn's disease, ulcerative colitis, and primary sclerosing cholangitis[33]. To date, there have been no systematic studies that screen for conditions that are closely related to AS based on overlapping genetic risks. In this study, a curated GWAS database for AS and 30 other conditions was used to identify conditions that share the most genetic risk with AS, using an unbiased approach.

2. Aims

AIM-1: Identify diseases genetically close to AS

AIM-2: Deciphering the genomic regulation of risk variants on genes at shared risk loci

AIM-3: Exploring opportunities for drug repurposing using shared risk genes

3. Results and discussion

3.1 Overlapping of genetic risk between AS and 30 other common diseases

To systematically screen for traits genetically close to AS, I established a curated genome-wide association study (GWAS) database that included AS and a broad range of common diseases. As AS is a member of the immune-mediated inflammatory diseases (IMIDs) family, 21 other IMIDs were included in the analysis as associated diseases (**Table 3.1**). The increased risk of cardiovascular disease in individuals with AS (discussed in Chapter 1) led to the inclusion of six cardiovascular diseases as AS-associated comorbidities. Additionally, despite the controversy, evidence of immune involvement in some neurodegenerative diseases, including Alzheimer's disease[150] and Parkinson's disease[151], led me to include them in the database. The GWAS data for these 31 conditions was sourced from 68 studies, and 295,584 genetic variants reached suggestive genome-wide significance (1×10^{-6}) for at least one condition. 31.8% (94,099) of these variants were associated with two or more conditions, with a single variant potentially being associated with up to 16 conditions, demonstrating the frequent occurrence of genetic risk pleiotropy (**Figure 3.1B**).

To determine the relative genetic proximity of these traits, I designed a clustering technique that blends genetic risk overlap assessment and unsupervised clustering (refer to the methodology section). Briefly, a genetic risk overlap score was computed for each pair of conditions using all significant SNPs curated (including those in MHC region), and then the overlapping score matrix was subjected to Hierarchical clustering. As a means of evaluating the relationship of a trait with the immune system, three classic immune-related traits (blood lymphocyte, monocyte, and neutrophil counts) and two immune-irrelevant traits (Body Mass Index (BMI) and height) were included as positive and negative controls, respectively (**Figure 3.1A)**. The results show that classical IMIDs have a close relationship with lymphocyte, monocyte, and neutrophil counts, but

not with BMI and height. Conversely, cardiovascular diseases, excluding coronary artery diseases (CAD) and age-related macular degeneration (AMD), form a separate cluster and demonstrate a distant relationship to immune traits. The clustering result is in line with current understanding, thus validating the approach.

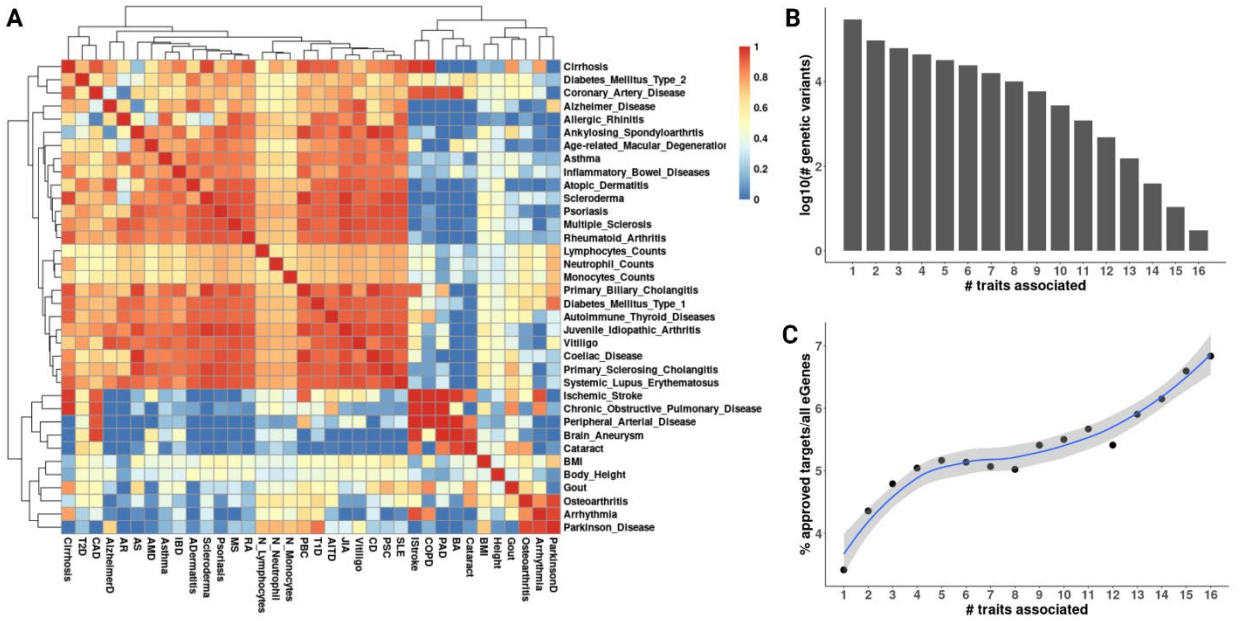


Figure 3.1. Exploring the translational potential of studying genetic risks shared across diverse diseases. A. Heatmap of genetic correlation among diverse traits. All abbreviations used are listed in the section of LIST OF ABBREVIATIONS. B. Distribution of genetic variants with different levels of commonality (the number of diseases influenced by a single shared variant). C. Correlation between the number of conditions (shown in x-axis) that a risk gene associated with and the percentage of gene targets (shown in y-axis) with approved therapies targeting their protein products. e.g.: for all genes (1533) where each one is associated with at least 8 diseases, 5% (77) of them have been used as drug targets in existing therapies.

Table 3.1 36 traits selected for genetic risk overlapping analysis

Traits	Classification
Ankylosing Spondylarthritis	IMIDs
Rheumatoid Arthritis	
Juvenile Idiopathic Arthritis	
Osteoarthritis	
Gout	
Psoriosis	
Scleroderma	
Vitiligo	
Atopic Dermatitis	
Inflammatory Bowel Diseases	
Coeliac Disease	
Primary Biliary Cholangitis	
Primary Sclerosing Cholangitis	
Cirrhosis	
Type 1 Diabetes Mellitus	
Type 2 Diabetes Mellitus	
Multiple Sclerosis	
Systemic Lupus Erythematosus	
Chronic Obstructive Pulmonary Disease	
Asthma	
Autoimmune Thyroid Diseases	
Allergic Rhinitis	
Ischemic Stroke	Cardiovascular diseases
Coronary Artery Disease	
Peripheral Arterial Disease	
Brain Aneurysm	
Cataract	
Age-related Macular Degeneration	
Alzheimer Disease	Neurodegenerative diseases
Parkinson Disease	
Blood neutrophil count	Immune-related traits
Blood monocyte count	
Blood lymphocyte count	
Body mass index	Non-Immune-related traits
Height	

The results of the analysis revealed that nine conditions clustered with AS, including IBD, psoriasis, RA, AMD, asthma, atopic dermatitis, scleroderma, and MS. The observed genetic overlap between AS and IBD, psoriasis, and RA aligns with the common clinical presentation of AS. Although a study of the risk overlaps between AS and uveitis, which is commonly associated with HLA-B27, would be informative, it could not be included in this analysis due to the lack of large-scale GWAS studies of uveitis. Notably, AMD was found to be most closely associated with AS, likely due to their shared association with HLA-B27. This analysis also suggests genetic proximity between AS and MS, which is in line with previous reports on the overlapping of genetics and pathology of the IL-23/IL17 axis. Secukinumab, which has good efficacy in treating AS, has also been tested in treating MS in a phase 2 clinical trial[152]. The preliminary findings demonstrated promising efficacy in mitigating MS symptoms. However, subsequent trials were prematurely discontinued in accordance with corporate policy decisions[153]. Further studies are needed to assess if IL-17A blockade could benefit the majority of MS patients or a subgroup of them. Surprisingly, AS was found to have closer genetic proximity to allergy-driven diseases such as asthma and atopic dermatitis than to JIA, a condition traditionally considered clinically similar to AS. JIA is a heterogeneous condition including many subtypes including oligoarticular JIA, polyarticular JIA, enthesitis-related JIA, psoriatic JIA, systemic JIA, and undifferentiated JIA[154]. Enthesitis-related JIA is commonly perceived to exhibit a closer association with ankylosing spondylitis AS. However, it was not encompassed in the genome-wide association studies (GWAS) that I curated for this analysis, potentially due to the relatively low frequency of this sub-form[154]. This omission may, in part, account for the observed absence of risk overlap in my analysis.

An unexpected discovery was made in the clustering analysis, which showed that coronary artery diseases (CAD) clustered closely with AS, classic immune-related traits (blood lymphocyte, monocyte, and neutrophil counts), and other immune-mediated inflammatory diseases (IMIDs).

Notably, the strongest association between CAD and these traits was observed with monocyte counts. This observation aligns with previous research findings indicating the role of foamy macrophages in the development of atherosclerosis[155]. This example supports the potential of using genetics to guide the discovery of unexpected yet important biology in the pathology of human diseases.

Several limitations exist in this analysis. 1) a comprehensive genetic risk profile is not accessible for any of the conditions included, largely due to the cost of genotyping and the challenges in organizing large-scale cohort studies. 2) The study scale, genotyping technology platform (which affects the genome coverage), statistical methods, and data availability vary across different studies, which could introduce confounding factors into this analysis. 3) The inclusion of the major histocompatibility complex (MHC) region could introduce confounding factors into the genetic correlation analysis. The MHC region is characterized by high linkage disequilibrium, posing a challenge in distinguishing causally associated single nucleotide polymorphisms (SNPs) identified in genome-wide association studies (GWAS). These SNPs may contribute to the observed genetic risk overlap among conditions without necessarily reflecting shared pathology. Conversely, the MHC region harbours numerous immune-related genes, and the exclusion of SNPs within this region may inadvertently overlook crucial pathological associations. 4) The consideration of association directionality was omitted in the analysis, primarily due to constrained access to these data in some of the curated studies. The inclusion of such information in future studies has the potential to enhance the precision of the analysis. 5) Further tests such as Mendelian randomization and linkage disequilibrium scoring could be used to investigate the causal relationships and quantify the genetic correlations between conditions, which was not included in this thesis.

3.2 Drug target discovery analysis of commonly shared risk genes

In recognition of the pivotal role played by immune cells in Ankylosing Spondylitis (AS), I utilized immune cell expression quantitative trait loci (eQTL) data obtained from 12 consortia/studies (refer to the methods section for further details) to identify gene candidates that are regulated by genetic variants associated with the disease. Approximately 78% of the disease-associated genetic variants were found to have eQTL signals in immune cells. It is widely acknowledged that drugs targeting proteins with genetic evidence of involvement have a higher chance of success in clinical trials, with a twofold increased likelihood[43]. Based on this premise, I hypothesized that drug targets associated with multiple conditions would have an even higher success rate to be developed for therapy. Using DrugBank as the source of approved drug targets, I evaluated the correlation between the successful drug development rate and the number of diseases a target is associated with (**Figure 3.1C**). I found that the percentage of approved drug targets for eGenes (Genes identified by eQTL) positively correlates to the number of diseases that it is associated with. Notably, genes associated with 16 conditions had double the chances of having an approved drug compared to those associated with only 1 condition. This highlights the significance and benefit of investigating the functional basis of common genetic risks.

It is pertinent to emphasize that the aforementioned analysis did not account for the directional aspect of expression quantitative trait loci (eQTL). In this context, there exists the possibility of erroneously identifying an antagonist for a gene linked to the upregulation associated with a protective allele as a prospective therapeutic candidate for a specific medical condition.

In 2019, Hai Fang, a member of the Julian Knight research group, introduced a priority index (Pi) methodology designed to prioritize genes with genetic support as potential drug targets[156]. This approach centers on the prioritization of genes primarily through considerations of informativeness, encompassing the strength of evidence derived from genetics and functional

annotation, along with their connectivity within biological pathways. The approach I have undertaken shares the same overarching objective — prioritization. However, it diverges by examining the issue from a distinct perspective, specifically focusing on the overlap of genetic risks across various conditions. Comparing the merits of these two methodologies is challenging, given that they are grounded in disparate rationales, and the effectiveness of these approaches is difficult to assess.

3.3 Common risk gene PPIF suggests a drug repurposing opportunity

Based on the integration of GWAS, eQTL, and drug-target database analyses, I identified risk genes for AS and its comorbidities that have existing drugs known to target their protein products. One particularly interesting risk AS eGene I found with potential for drug repurposing is *PPIF* (Peptidyl-prolyl cis-trans isomerase, mitochondrial). PPIF is an enzyme as part of the mitochondrial permeability transition pore, and it catalyzes the cis-trans isomerization of proline imidic peptide bonds[157]. I identified 25 risk SNPs associated with the expression of *PPIF* in various immune cells (**Table 3.2**), which were not previously thought to affect *PPIF* as they are located in the intron of *ZMIZ1* (Zinc Finger MIZ-Type Containing 1) and approximately 50,000 base pairs away from the *PPIF* promoter region (**Figure 3.3A**). Moreover, these SNPs are associated with three of the nine diseases genetically linked to AS, including IBD, psoriasis, and MS, as well as with lymphocyte and neutrophil counts, implying a crucial role in immune regulation (**Table 3.2**). PPIF is primarily known as a cyclophilin and is capable of binding Ciclosporin, an approved drug for IBD, RA, atopic dermatitis, psoriatic arthritis, and psoriasis. Ciclosporin can form a complex with PPIF to inhibit calcineurin, which regulates T-cell activation[158]. The association of *PPIF* with AS suggests a potential for repurposing Ciclosporin to treat AS. A recent case study reported the successful control of peripheral arthritis in an AS patient treated with Ciclosporin[159]. Additionally, Colombo and colleagues reported that Ciclosporin improved spinal symptoms in psoriatic arthritis patients[160].

Despite the association between risk variants and the expression of *PPIF*, as revealed by eQTL analysis, the underlying mechanism of regulation remains unclear. It is possible that distal genetic variants exert their effects by modulating the activity of cis-regulatory elements (CREs), which regulate gene expression through long-range chromatin interactions. The binding of transcription factors to DNA in eukaryotic cells is a fleeting process, lasting for mere seconds even in the presence of assisting DNA or protein elements[161]. The presence of CREs, such as enhancers, can increase the concentration of certain transcriptional factors by binding to the factors themselves or to cofactors[162]. DNAse-seq or ATAC-seq are commonly used to identify active CREs in the genome by mapping accessible non-coding regions. I analyzed DNAse-seq data from primary CD4+ and CD8+ T cells provided by the Encyclopedia of DNA Elements (ENCODE) Consortium to identify candidate cis-regulatory elements (CREs) in the region of interest (**Figure 3.3A**). My analysis revealed that AS risk variants specifically flank an open chromatin region in the *ZMIZ1* intron, suggesting it as the putative causal CRE (**Figure 3.3A**). To ascertain the functional role of this open region as an enhancer, I analyzed H3K27ac ChIP-seq data from ENCODE and found that this region is marked by the H3K27ac modification in both CD4+ and CD8+ T cells, indicating its identity as an enhancer. Herein, I refer to this enhancer as enhancer 1 (**Figure 3.3A**). A recent functional genomic study by Joseph Nasser et al. investigated this region to decode the risk variants for inflammatory bowel disease (IBD) and found that silencing enhancer 1 through CRISPR interference significantly downregulated *PPIF* expression in a human cell line of T cells and monocytes[66]. This finding consolidates the regulatory function of enhancer 1 on the *PPIF* expression. To determine whether the regulation is mediated by direct physical interactions, I analyzed Th17 H3K27ac HiChIP data published by Maxwell Mumbach and colleagues[84] and found that direct interactions between enhancer 1 and the PPIF promoter were not present (**Figure 3.3A**). However, an intermediate enhancer, named enhancer 2, is located between enhancer 1 and the *PPIF* promoter and interacts with both (**Figure 3.3A**). In Joseph Nasser's study, enhancer 2 was also shown to regulate *PPIF* expression. Taken together, these

findings suggest that the causal enhancer 1 indirectly regulates PPIF expression through enhancer 2 (**Figure 3.3B**). Further studies are required to confirm this mechanism, which is beyond the scope of this thesis.

Table 3.2 25 AS risk variants associated with PPIF expression.

(Note: Linkage disequilibrium among these variants were shown in Table 3.3 and Figure 3.3)

rsID	Chromosome	Position	Effector Allele	Reference Allele	P value	Other trait associated	cell type/tissue with significant eQTL of PPIF
rs1250544	chr10	79273128	A	G	2.10E-07	IBD,MS,CD,Neutrophil counts	DC
rs1250542	chr10	79274913	A	G	1.23E-07	IBD,MS,Neutrophil counts	NK
rs1250540	chr10	79276250	G	A	2.77E-07	IBD,MS,CD,Neutrophil counts	NK
rs1250538	chr10	79278043	T	G	1.18E-07	IBD,MS,CD,Neutrophil counts	NK
rs1250537	chr10	79278312	C	T	3.12E-07	IBD,MS,CD,Neutrophil counts	CD4+ T cells
rs1250573	chr10	79282718	T	C	1.00E-08	IBD,MS,AITD	Treg
rs1108617	chr10	79283816	C	T	3.56E-07	IBD,MS,Neutrophil counts, Lymphocytes counts	Treg
rs1892497	chr10	79283950	T	C	9.42E-09	IBD,MS,CD,AITD,Neutrophil counts	Treg
rs1108618	chr10	79283986	C	T	1.48E-07	IBD,MS,Psoriasis,CD, Neutrophil counts,Lymphocytes counts	Treg
rs1250566	chr10	79286696	A	G	1.96E-08	IBD,MS,CD,AITD,Neutrophil counts	Treg
rs1250565	chr10	79287258	A	G	1.15E-07	IBD,MS,CD,Neutrophil counts	CD8+ T cells
rs1250564	chr10	79287585	C	A	1.22E-07	IBD,MS,CD,Neutrophil counts	CD8+ T cells,Treg, Blood
rs2802372	chr10	79287818	G	T	5.75E-07	IBD,MS,CD,AITD, Neutrophil counts,Lymphocytes counts	CD8+ T cells,Blood
rs1782645	chr10	79288854	A	G	5.95E-07	IBD,MS,CD, Neutrophil counts,Lymphocytes counts	Blood
rs1250560	chr10	79293718	T	C	2.79E-07	IBD,MS,CD,Neutrophil counts	Blood
rs1250559	chr10	79293803	T	C	2.26E-07	IBD,MS,CD,Neutrophil counts	Blood
rs1250550	chr10	79300560	A	C	1.46E-09	IBD,MS,CD	Blood
rs1250549	chr10	79300717	G	A	1.05E-07	IBD,MS,CD, Neutrophil counts,Lymphocytes counts	Blood
rs1782648	chr10	79301072	T	C	1.27E-08	IBD,T1D,MS,CD,AITD	Blood
rs1250558	chr10	79305444	A	G	7.90E-08	IBD,MS,CD	Blood
rs1272633	chr10	79305751	A	G	1.17E-07	IBD,MS,CD	Blood
rs1250556	chr10	79306820	T	C	1.10E-07	IBD,MS,CD	Blood
rs1250555	chr10	79307225	T	C	4.98E-08	IBD,MS,CD	Blood
rs1250554	chr10	79307283	A	G	1.14E-07	IBD,MS,CD	Blood
rs1972346	chr10	79307723	C	G	1.30E-07	IBD,MS,CD,AITD	Blood

Table 3.3 Linkage disequilibrium of SNPS associated with PPIF (R2 is shown)

Note: Genome data of European population from 1000 genome project was used for the calculation.

SNP_ID	rs1250544	rs1250542	rs1250540	rs1250538	rs1250537	rs1250573	rs1108617	rs1892497	rs1108618	rs1250566	rs1250565	rs1250564	rs2802372	rs1782645	rs1250560	rs1250559	rs1250550	rs1250549	rs1782648	rs1250558	rs1272633	rs1250556	rs1250555	rs1250554	rs1972346
rs1250544	1	1	1	0.954	0.954	0.705	0.705	0.668	0.741	0.668	0.909	0.909	0.688	0.643	0.597	0.597	0.292	0.383	0.323	0.349	0.349	0.349	0.412	0.336	0.254
rs1250542	1	1	1	0.954	0.954	0.705	0.705	0.668	0.741	0.668	0.909	0.909	0.688	0.643	0.597	0.597	0.292	0.383	0.323	0.349	0.349	0.349	0.412	0.336	0.254
rs1250540	1	1	1	0.954	0.954	0.705	0.705	0.668	0.741	0.668	0.909	0.909	0.688	0.643	0.597	0.597	0.292	0.383	0.323	0.349	0.349	0.349	0.412	0.336	0.254
rs1250538	0.954	0.954	0.954	1	1	0.74	0.746	0.701	0.783	0.701	0.953	0.953	0.729	0.681	0.635	0.635	0.313	0.415	0.346	0.373	0.373	0.373	0.444	0.359	0.275
rs1250537	0.954	0.954	0.954	1	1	0.74	0.746	0.701	0.783	0.701	0.953	0.953	0.729	0.681	0.635	0.635	0.313	0.415	0.346	0.373	0.373	0.373	0.444	0.359	0.275
rs1250573	0.705	0.705	0.705	0.74	0.74	1	0.539	0.948	0.565	0.948	0.689	0.689	0.526	0.471	0.445	0.445	0.474	0.299	0.523	0.554	0.554	0.554	0.315	0.539	0.441
rs1108617	0.705	0.705	0.705	0.746	0.746	0.539	1	0.581	0.955	0.581	0.79	0.79	0.889	0.846	0.496	0.496	0.237	0.496	0.312	0.286	0.286	0.286	0.47	0.274	0.314
rs1892497	0.668	0.668	0.668	0.701	0.701	0.948	0.581	1	0.609	1	0.735	0.735	0.568	0.51	0.484	0.484	0.508	0.332	0.561	0.593	0.593	0.593	0.347	0.577	0.477
rs1108618	0.741	0.741	0.741	0.783	0.783	0.565	0.955	0.609	1	0.609	0.828	0.828	0.934	0.888	0.528	0.528	0.254	0.467	0.281	0.306	0.306	0.306	0.502	0.293	0.337
rs1250566	0.668	0.668	0.668	0.701	0.701	0.948	0.581	1	0.609	1	0.735	0.735	0.568	0.51	0.484	0.484	0.508	0.332	0.561	0.593	0.593	0.593	0.347	0.577	0.477
rs1250565	0.909	0.909	0.909	0.953	0.953	0.689	0.79	0.735	0.828	0.735	1	1	0.773	0.721	0.676	0.676	0.336	0.449	0.371	0.398	0.398	0.398	0.477	0.384	0.298
rs1250564	0.909	0.909	0.909	0.953	0.953	0.689	0.79	0.735	0.828	0.735	1	1	0.773	0.721	0.676	0.676	0.336	0.449	0.371	0.398	0.398	0.398	0.477	0.384	0.298
rs2802372	0.688	0.688	0.688	0.729	0.729	0.526	0.889	0.568	0.934	0.568	0.773	0.773	1	0.912	0.481	0.481	0.229	0.479	0.252	0.277	0.277	0.277	0.518	0.264	0.357
rs1782645	0.643	0.643	0.643	0.681	0.681	0.471	0.846	0.51	0.888	0.51	0.721	0.721	0.912	1	0.511	0.511	0.239	0.516	0.266	0.24	0.24	0.24	0.487	0.229	0.322
rs1250560	0.597	0.597	0.597	0.635	0.635	0.445	0.496	0.484	0.528	0.484	0.676	0.676	0.481	0.511	1	1	0.348	0.433	0.374	0.347	0.347	0.347	0.41	0.332	0.245
rs1250559	0.597	0.597	0.597	0.635	0.635	0.445	0.496	0.484	0.528	0.484	0.676	0.676	0.481	0.511	1	1	0.348	0.433	0.374	0.347	0.347	0.347	0.41	0.332	0.245
rs1250550	0.292	0.292	0.292	0.313	0.313	0.474	0.237	0.508	0.254	0.508	0.336	0.336	0.229	0.239	0.348	0.348	1	0.532	0.812	0.763	0.763	0.763	0.456	0.743	0.618
rs1250549	0.383	0.383	0.383	0.415	0.415	0.299	0.496	0.332	0.467	0.332	0.449	0.449	0.479	0.516	0.433	0.433	0.532	1	0.588	0.491	0.491	0.491	0.773	0.472	0.571
rs1782648	0.323	0.323	0.323	0.346	0.346	0.523	0.312	0.561	0.281	0.561	0.371	0.371	0.252	0.266	0.374	0.374	0.812	0.588	1	0.812	0.812	0.812	0.481	0.79	0.655
rs1250558	0.349	0.349	0.349	0.373	0.373	0.554	0.286	0.593	0.306	0.593	0.398	0.398	0.277	0.24	0.347	0.347	0.763	0.491	0.812	1	1	1	0.588	0.975	0.823
rs1272633	0.349	0.349	0.349	0.373	0.373	0.554	0.286	0.593	0.306	0.593	0.398	0.398	0.277	0.24	0.347	0.347	0.763	0.491	0.812	1	1	1	0.588	0.975	0.823
rs1250556	0.349	0.349	0.349	0.373	0.373	0.554	0.286	0.593	0.306	0.593	0.398	0.398	0.277	0.24	0.347	0.347	0.763	0.491	0.812	1	1	1	0.588	0.975	0.823
rs1250555	0.412	0.412	0.412	0.444	0.444	0.315	0.47	0.347	0.502	0.347	0.477	0.477	0.518	0.487	0.41	0.41	0.456	0.773	0.481	0.588	0.588	0.588	1	0.604	0.72
rs1250554	0.336	0.336	0.336	0.359	0.359	0.539	0.274	0.577	0.293	0.577	0.384	0.384	0.264	0.229	0.332	0.332	0.743	0.472	0.79	0.975	0.975	0.975	0.604	1	0.844
rs1972346	0.254	0.254	0.254	0.275	0.275	0.441	0.314	0.477	0.337	0.477	0.298	0.298	0.357	0.322	0.245	0.245	0.618	0.571	0.655	0.823	0.823	0.823	0.72	0.844	1

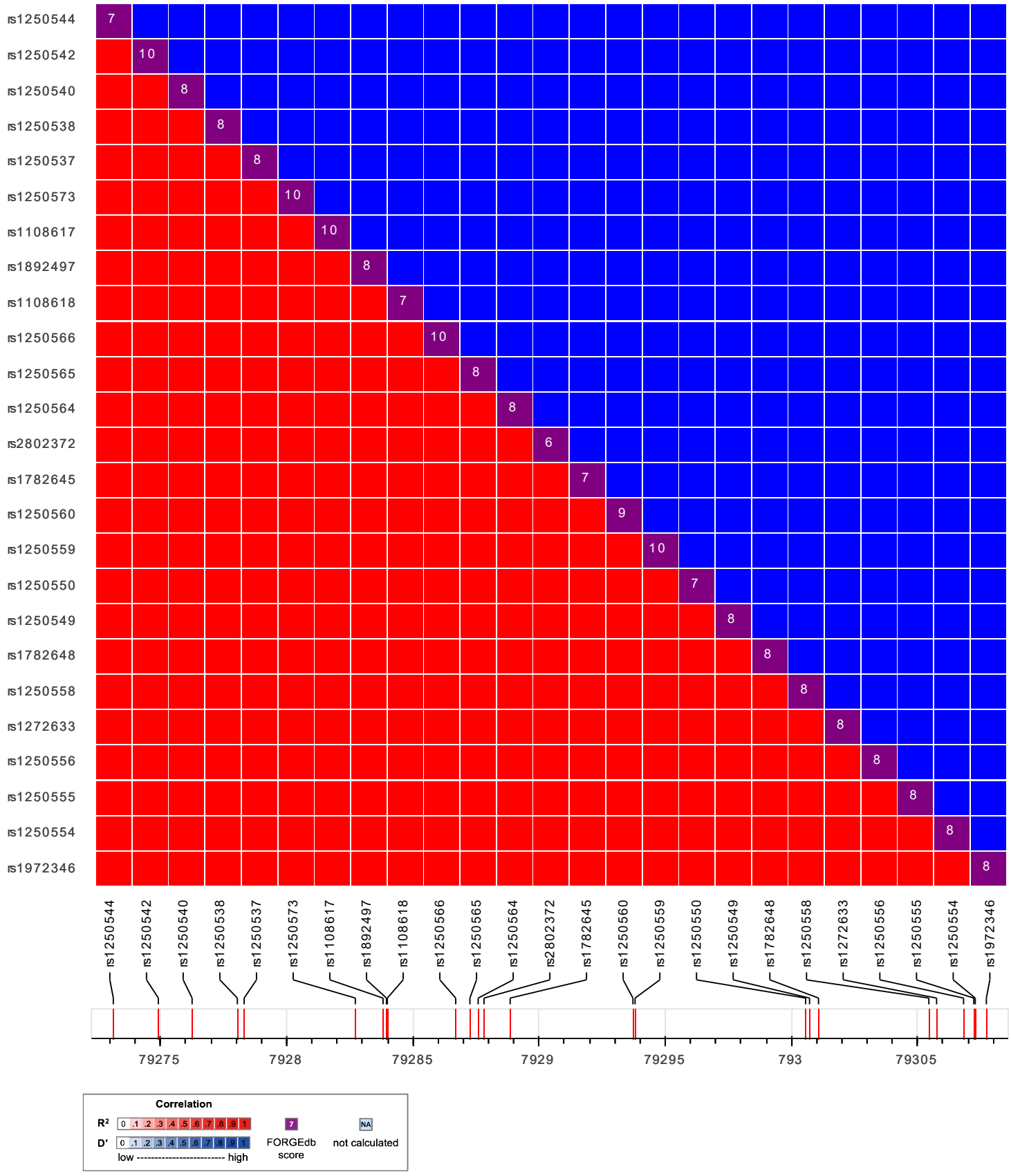


Figure 3.2 Linkage disequilibrium of SNPS associated with PPIF (R^2 is shown)
(Note: Genome data of European population from 1000 genome project was used for the calculation.)



Figure 3.3 Genomic regulation of AS risk variant on *PPIF* expression. A. Genomic view of ZMIZ1-PPIF locus. Tracks show (from top to bottom): gene body, chromosome coordinates of hg38, AS risk variants (height indicates the number of studies supporting the corresponding SNPs), enhancer annotation, DNAse-seq signals in CD4+ T cells, H3K27ac ChIP-seq signals in CD4+ T cells, DNAse-seq signals in CD8+ T cells, H3K27ac ChIP-seq signals in CD8+ T cells, chromatin interactions identified from H3K27ac HiChIP in Th17 cells[84]).

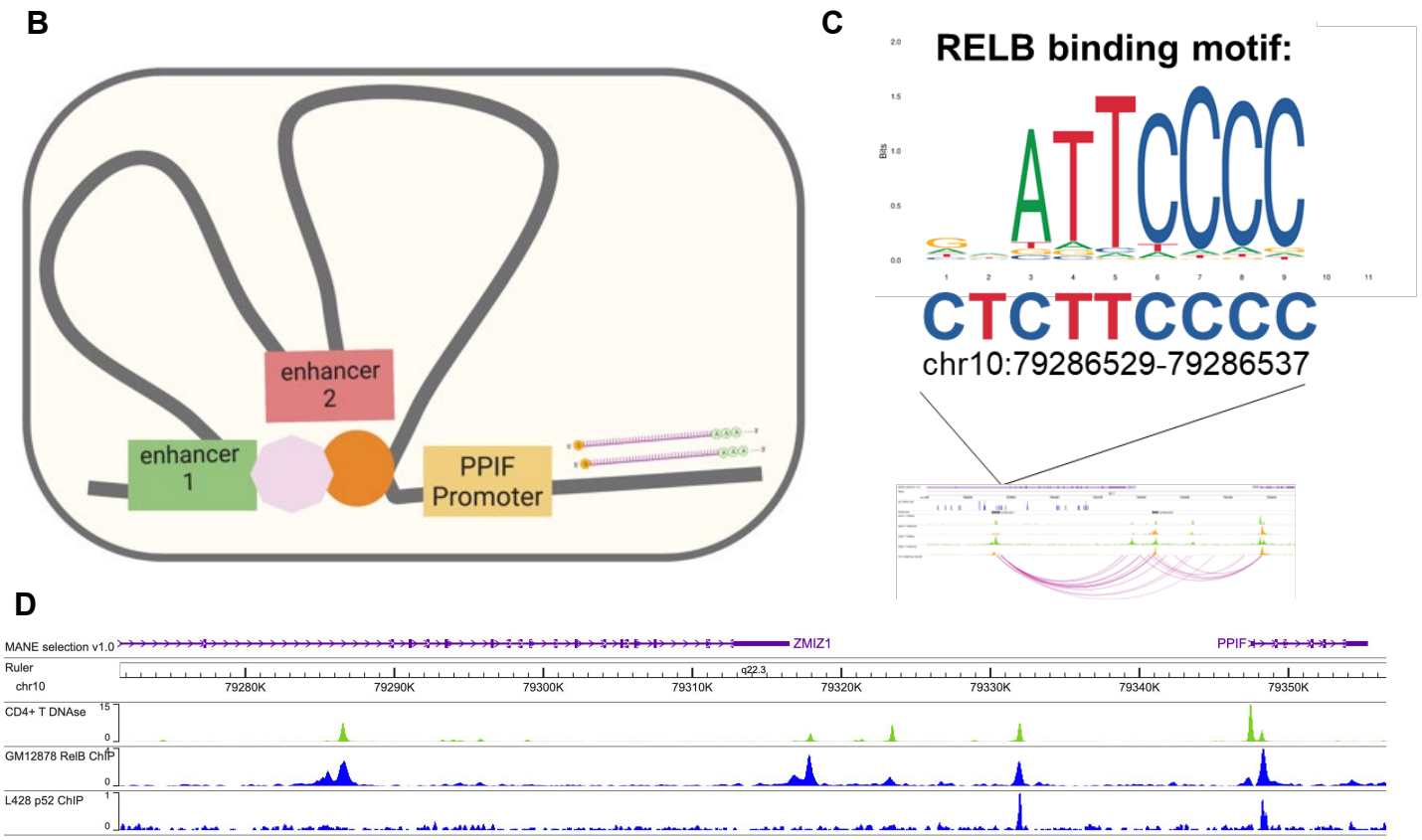


Figure 3.3 Genomic regulation of AS risk variant on *PPIF* expression. B. A proposed model of regulation of AS risk variants on PPIF expression. C. RELB motif on the casual enhancer 1. The sequence below the motif is the reference genome sequence. D. ChIP-seq signal of RelB on ZMIZ1-PPIF locus. Tracks show (from top to bottom): gene body, chromosome coordinates of hg38, DNAse-seq signals in CD4+ T cells (ENCODE), RelB ChIP-seq signal in GM12878 cells (ENCODE), p52 ChIP-seq signal in Hodgkin lymphoma cell line L428 (ENCODE).

PPIF not only acts as a cyclophilin but also serves as a key component of the mitochondrial permeability transition pore, in addition to regulating mitochondrial metabolism and apoptosis. However, its role in the immune response remains unclear. Identifying the transcription factors (TFs) that control *PPIF* expression can help understand the upstream signaling pathways and biological programs involving PPIF. Using the JASPAR database, I scanned enhancer 1 for known motifs of human TFs and found RelB transcriptional factor motifs at the peak summit area (chr10:79286507-79286545) (**Figure 3.3C**). RelB is a crucial member of the NF-κB family of transcription factors that play essential roles in the immune system. Typically, RelB pairs with p52 to activate the noncanonical NF-κB pathway, which differs from the canonical NF-κB pathway in the stages of T cell development it affects[163]. The canonical NF-κB pathway is mainly triggered by T-cell receptor signaling and co-stimulatory signals, promoting T-cell activation and clonal expansion. On the other hand, the noncanonical NF-κB pathway is typically activated by a subset of the TNF receptor family, such as OX40 and CD27, and regulates the function of effector cells[163]. The noncanonical NF-κB pathway is usually stimulated by a subset of the TNF receptor family (such as OX40 and CD27) and regulates the function of effector cells[163]. ChIP-seq of RelB in GM12878 cells from the ENCODE consortium showed signals in enhancer 1, enhancer 2, and the *PPIF* promoter, indicating that RelB is involved in the transcriptional regulation of *PPIF* (**Figure 3.3D**). Interestingly, the signal in enhancer 2 was weaker than that of enhancer 1 and PPIF (**Figure 3.3D**). It is possible that this signal is a result of the proximity of enhancer 2 to the enhancer 1/RelB/*PPIF* promoter interaction, rather than a genuine binding interaction of enhancer 2. These findings could be confirmed by ChIP-seq in primary human T cells, but it is not covered here. To examine if p52 worked as a co-factor with RelB in regulation at this locus, I analyzed ChIP-seq data of p52 in the Hodgkin lymphoma cell line L428 provided by ENCODE consortium. The result demonstrated the p52 binding to the PPIF promoter but not enhancer 1 or enhancer 2 (**Figure 3.3D**). A plausible model is that enhancer 1 recruits the RelB-p52 complex by binding to RelB, while the p52 component of the complex binds to the *PPIF* promoter.

To further investigate PPIF's role in AS, I conducted a re-analysis of a single-cell RNA sequencing (scRNA-seq) dataset of peripheral blood mononuclear cells (PBMCs) and synovial fluid mononuclear cells (SFMCs) from six AS patients, recently published by Kijong Yi et al from Kim group [127]. First, T cells from clusters 1 and 9 were extracted based on the authors' clustering and annotation (**Figure 3.4A and B**). A total of 6,603 and 6,026 T cells were obtained from PBMCs and SFMCs, respectively (**Figure 3.4C**). Re-clustering these cells identified nine distinct populations, including both adaptive (Th, Treg, and CTL) and innate (γδ T cells and MAIT) arms (**Figure 3.4D and G**). *PPIF* expression was mostly upregulated in the proliferating cells, which were enriched in SFMCs (**Figure 3.4E and F**). These findings suggest that PPIF may promote the T cell proliferation program in AS joints and treating patients with Ciclosporin might inhibit this program.

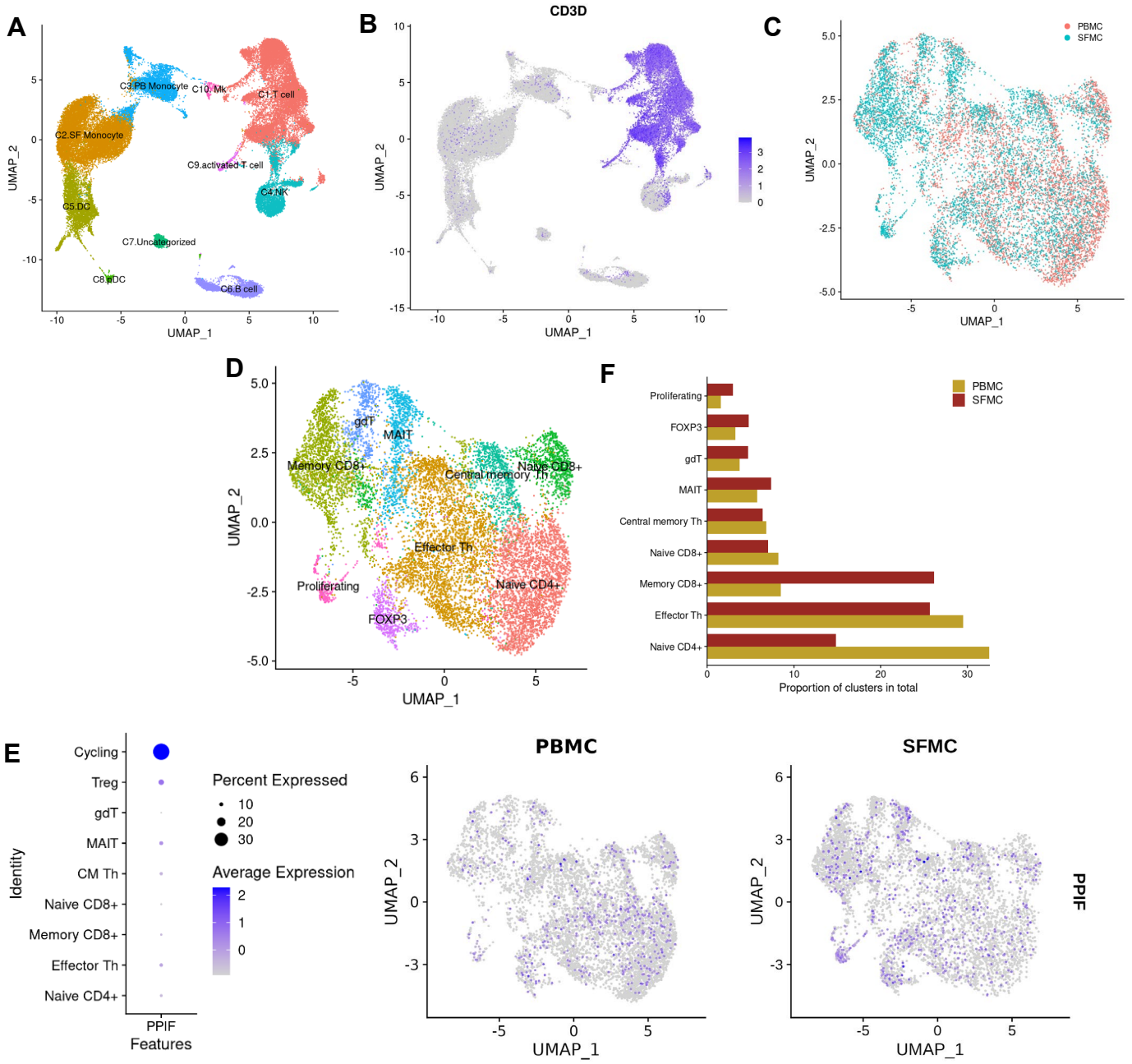


Figure 3.4 PPIF was uniquely expressed in proliferating T cells at joints of AS. A. UMAP plot of cells and cell type annotations from PBMCs and SFMCs of 6 AS donors. B. UMAP plot of CD3D expression to identify T cell compartment. C. UMAP plot of T cells from PBMC and SFMC of 6 AS donors. D. UMAP plot of cells and cell type/state annotations of T cells from AS PBMCs and SFMCs. E. PPIF expression in different clusters of T cells. (left: expression in cells mixed with SF and PB sourced; Right: UMAP of PPIF expression in SF and PB T cells separately.) F. Proportion of each cluster in PBMCs and SFMCs.

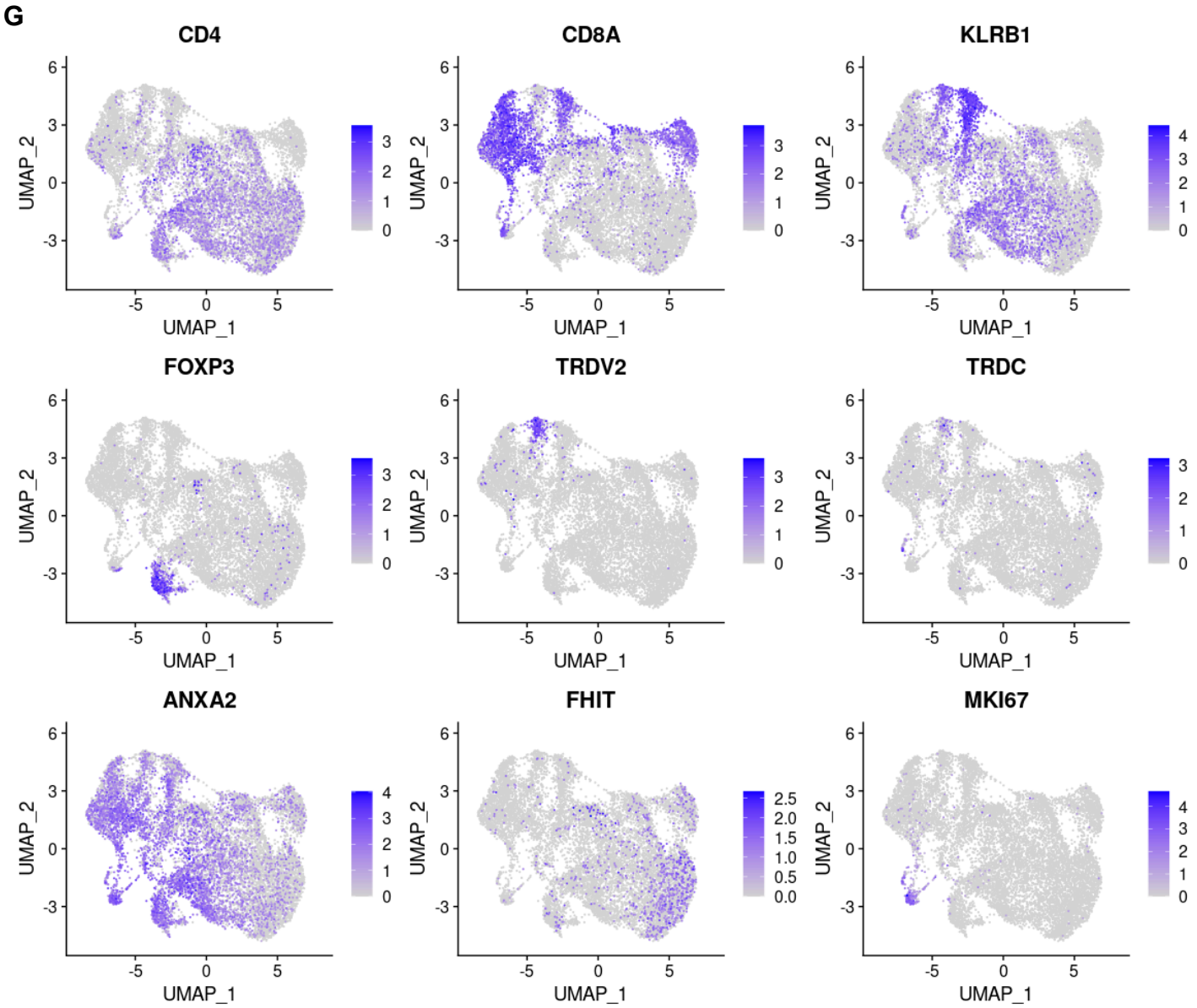


Figure 3.4 PPIF was uniquely expressed in proliferating T cells at joints of AS. G. UMAP visualization of expression of marker genes of different clusters. (KLRB1: Killer Cell Lectin Like Receptor B1, coding CD161; FOXP3: forkhead box P3; TRDV2: T cell receptor delta variable 2; TRDC: T cell receptor delta constant; ANXA2: annexin A2; FHIT: fragile histidine triad diadenosine triphosphatase; MKI67: marker of proliferation Ki-67)

3.4 Specific targeting of JAK2 and TYK2 in AS suggested by genetic risk

JAK inhibitors have emerged as promising therapies for treating a wide range of immune-mediated inflammatory diseases (IMIDs), such as Ankylosing Spondylitis (AS). Due to their oral delivery and cost-effectiveness, these drugs have the potential to become a major therapy option. However, the US Food and Drug Administration (FDA) has raised concerns about the safety of these drugs, citing potential side effects including serious cardiovascular events, cancer, blood clots, and death[17]. Current JAK inhibitors lack specificity and act on all JAK family members including JAK1, JAK2, JAK3, and TYK2 (tyrosine kinase 2). By selectively targeting the JAK family members that are more relevant to specific diseases, the side effects of these drugs may be reduced.

According to my database analysis, only *JAK2* and *TYK2* were associated with AS and its genetically linked diseases (**Table 3.4**). These proteins are involved in transducing signaling of IL-12 family receptors such as IL-12 and IL-23, which mediate the generation of pathogenic Th17 cells. Notably, the two JAK inhibitors approved for AS treatment, Tofacitinib (JAK1/3) and Upadacitinib (JAK1), are more effective against other JAK family members than JAK2 and TYK2.

Given the genetic evidence and the crucial role of IL-23 in AS pathology, developing JAK2/TYK2 specific inhibitors may improve treatment efficacy and safety. A recent phase 3 clinical trial showed that Deucravacitinib, a selective TYK2 inhibitor, successfully treated PsA, and it has been approved by US FDA for treating psoriasis. Since AS and psoriasis share a genetic risk overlap regarding *TYK2*, repurposing Deucravacitinib for AS treatment is a promising avenue to explore.

Table 3.4 Association of JAK family with different diseases

JAK family member	Diseases associated
JAK1	AITD
JAK2	AS, IBD, Asthma, MS, T1D, PSC
JAK3	None
TYK2	AS, IBD, RA, MS, Psoriasis, T1D, PBC, AITD, SLE, CAD

Note:1) these association were identified in my analysis and some associations have been reported from other studies; 2) Red labelled conditions were genetically associated to AS

3.5 A common risk locus at 12q24.12 regulates a transcriptional hub

Efforts have been focused on identifying the singular causal gene responsible for disease-associated risk loci based on the assumption that genetic variants at a locus primarily affect a single gene. However, advancements in functional genomics have recently revealed that gene regulation networks do not always function in isolation[164]. Instead, the promoters of multiple genes can be driven proximal to each other and regulated by the same enhancers, forming a transcriptional hub. In this study, I identified a risk locus at 12q24.12 that interacts with the promoters of up to 18 genes, providing an example of a risk locus that can affect multiple genes.

This locus is associated with susceptibility to AS through four SNPs that also affect the incidence of 15 other diseases (**Table 3.5**). One of these SNPs is a missense mutation (rs3184504) of *SH2B3* (SH2B adapter protein 3), while the other three are located at introns of *SH2B3*. *SH2B3* encodes a protein named LNK, which is a negative regulator of JAK-STAT signaling. Prior research suggests that loss-of-function of *SH2B3* due to rs3184504 leads to hematopoietic stem cell expansion, NOD2 recognition activation, and enhanced bacterial clearance[165–168]. Along with *SH2B3*, eQTL data indicates that genetic variants at this locus affect the gene expression of 17 additional genes (**Table 3.5**). A colocalization analysis could potentially help identify the causal SNPs at this local, which is beyond this thesis. These gene expression changes could be due to the functional alteration of the cis-regulatory element or trans-regulation through SH2B3.

To investigate whether the risk locus directly regulates multiple genes, I employed the Th17 H3K27ac HiChIP data that mentioned in the PPIF analysis. Interestingly, significant interactions among this risk locus and promoters of 18 genes were identified (**Figure 3.5A**). This presents an example that a risk locus may regulate not only one causal gene but a transcriptional hub involving multiple genes, and that the true causal genes could be a few of the affected genes or a combination of them. Genes that are regulated together may suggest that they function in the

same cellular program, as exemplified by the production of an intact CD3 complex requiring CD3γ, CD3δ, and two CD3ε, which are produced by *CD3G*, *CD3D*, and *CD3E* respectively. These genes are located at the same genomic locus, with their promoters showing strong interaction, indicative of a transcription hub organizing their expression (**Figure 3.5B**). However, due to limited knowledge of gene functions at the *SH2B3* locus, the functional relevance of these co-regulated genes is difficult to determine. Next, I looked for transcription factors participating in transcriptional regulation at this locus. I found the presence of a binding motif of a transcriptional factor named c-Rel (NF-κB subunit), with its binding on promoters of all 18 genes further supported by ChIP-seq data in primary human naïve CD4+ T cells published by Romain Lévy et al[169] (**Figure 3.5C**). The binding signal was found only in the PMA-stimulated state, indicating that the transcription activity of this hub is induced by T-cell activation through the NF-κB pathway.

Table 3.5 Four AS risk variants at 12q24.12

rsID	Chromosome	Position	Effector Allele	Reference Allele	P value	Other trait associated	eGenes
rs3184504	chr12	111446804	T	C	5.32E-07	IBD, T1D, T2D, RA, JIA, PBC, PSC, CD, AITD, Gout, Vitiligo, SLE, Asthma, Ischemic Stroke, CAD	SH2B3, ALDH2, NAA25, RAD9B, PHETA1, MAPKAPK5-AS1,HSPA8P14, LINC02356, ADAM1B, GPN3, TMEM116, TRAFD1, TCTN1, VPS29, FAM216A, ACAD10, PPTC7, HVCN1, HECTD4
rs12580300	chr12	111411467	A	G	8.39E-07	T1D, PBC, AITD, CAD	SH2B3, ALDH2, MAPKAPK5-AS1, NAA25, ADAM1B, TMEM116, ACAD10, PHETA1, HSPA8P14, LINC02356, ERP29, ATXN2, MAPKAPK5
rs11065898	chr12	111424771	A	G	4.71E-08	T1D, AITD, CAD	SH2B3, NAA25, ALDH2, MAPKAPK5-AS1, ADAM1B, TMEM116, ACAD10, PHETA1, HSPA8P14, LINC02356, ERP29, PTPN11, FAM216A, ATXN2, MAPKAPK5
rs2283358	chr12	111427761	A	G	6.01E-07	T1D, PBC, AITD, CAD	SH2B3, NAA25, ALDH2, MAPKAPK5-AS1, ADAM1B, TMEM116, ACAD10, PHETA1, HSPA8P14, LINC02356, ERP29, PTPN11, ATXN2, MAPKAPK5

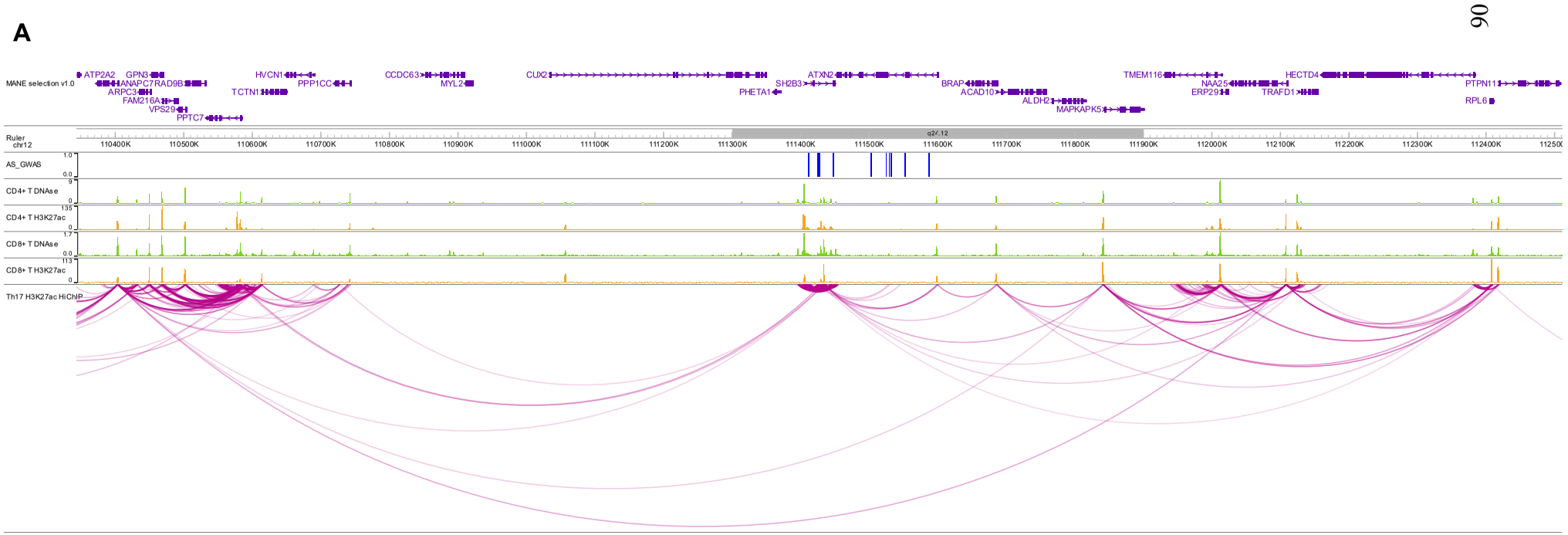


Figure 3.5 Common risk variants at 12q24.12 regulate a transcriptional hub. A. Genomic view of the risk variants and the genes they interact with. Tracks show (from top to bottom): gene body, chromosome coordinates of hg38, AS risk variants variants (height indicates the number of studies supporting the corresponding SNPs), enhancer annotation, DNAse-seq signals in CD4+ T cells, H3K27ac ChIP-seq signals in CD4+ T cells (ENCODE), DNAse-seq signals in CD8+ T cells (ENCODE), H3K27ac ChIP-seq signals in CD8+ T cells (ENCODE), chromatin interactions identified from H3K27ac HiChIP in Th17 cells[84]).

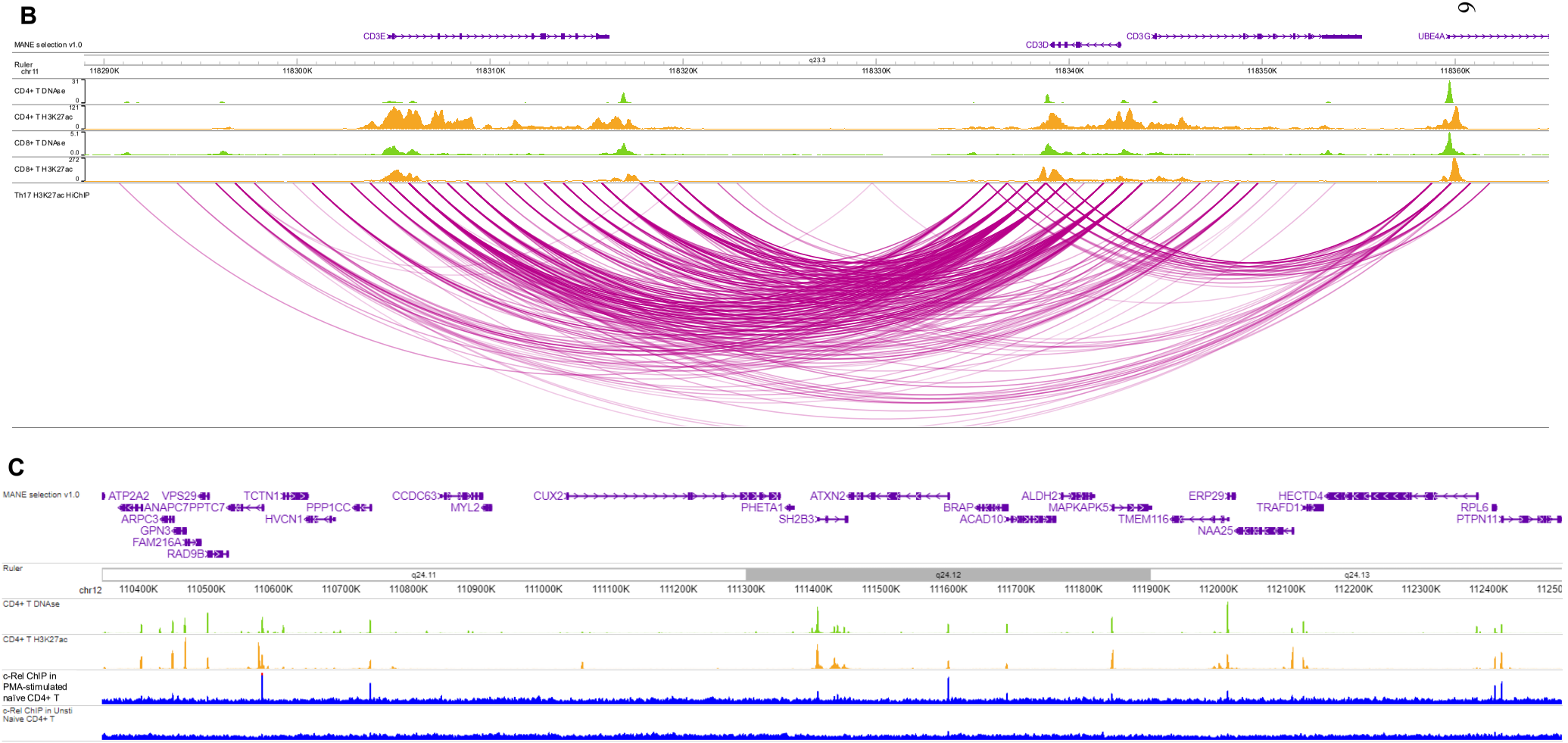


Figure 3.5 Common risk variants at 12q24.12 regulate a transcriptional hub. B. Genomic view of the transcriptional hub at CD3 region. Tracks show (from top to bottom): gene body, chromosome coordinates of hg38, enhancer annotation, DNAse-seq signals in CD4+ T cells (ENCODE), H3K27ac ChIP-seq signals in CD4+ T cells (ENCODE), DNAse-seq signals in CD8+ T cells (ENCODE), H3K27ac ChIP-seq signals in CD8+ T cells (ENCODE, chromatin interactions identified from H3K27ac HiChIP in Th17 cells[84]). C. ChIP-seq signal of c-Rel in naïve CD4+ T cells at unstimulated and PMA-stimulated state[168]. Tracks show (from top to bottom): gene body, chromosome coordinates, DNAse-seq signals in CD4+ T cells (ENCODE), H3K27ac ChIP-seq signals in CD4+ T cells (ENCODE), c-Rel ChIP-seq signals in untreated primary naïve T cells, c-Rel ChIP-seq signals in PMA-stimulated primary naïve T cells.

4. Conclusion

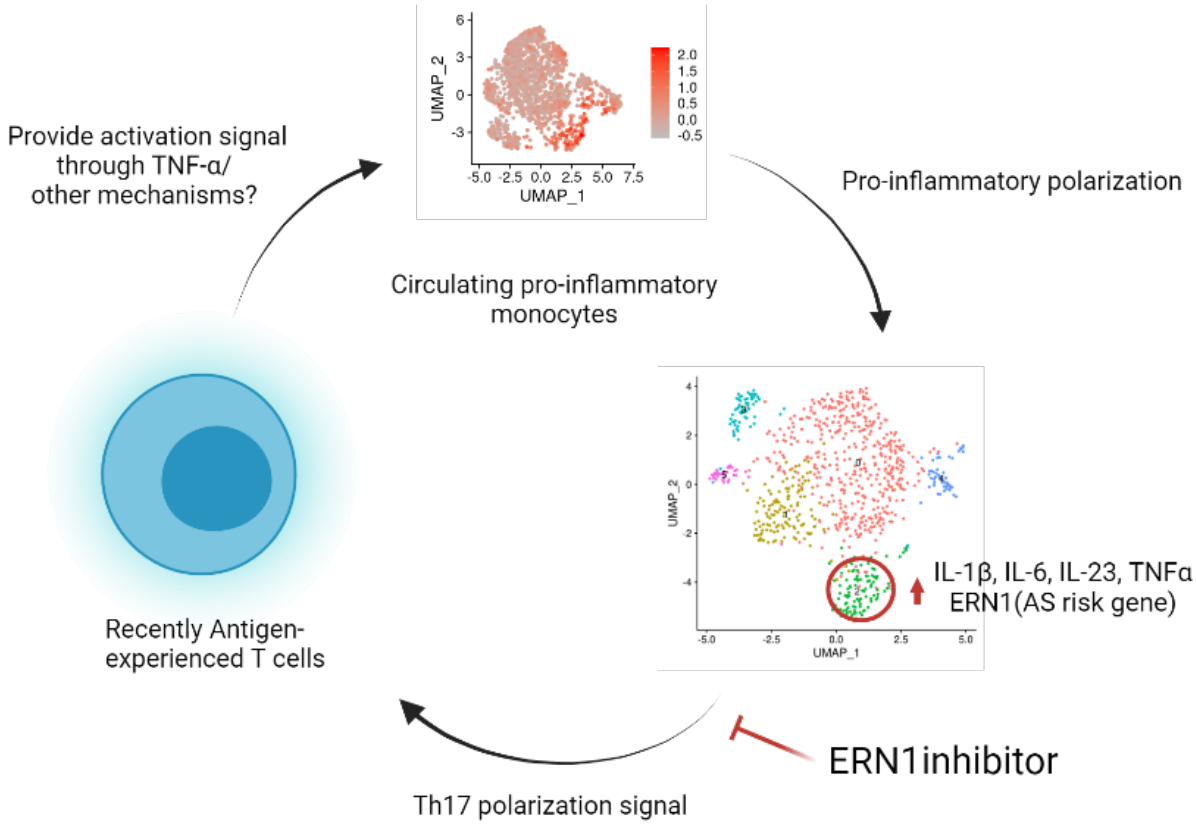
In this study, I quantified the genetic risk overlaps between AS and 30 other conditions and examined the genomic regulation and translational potential in several loci. A comprehensive cross-trait GWAS database was established, accompanied by an atlas of immune cell eQTL. Analysis based on this database revealed the widespread existence of genetic risk pleiotropy. Approximately one-third of the curated risk variants were associated with more than one condition, and a single genetic variant could affect up to 16 conditions. Clustering diseases based on their genetic background identified 9 conditions mostly closely related to AS, including both clinically related and unexpected ones. Previous studies have highlighted that therapeutic targets with genetic evidence have a higher chance of being successfully developed for approved therapies. In this study, it was demonstrated that prioritizing risk genes shared by multiple conditions could further increase this chance.

This study showed that PPIF (Peptidyl-prolyl cis-trans isomerase, mitochondrial) could be a good example of using shared genetic risk for identifying drug repurposing opportunities. Through the combination of 3D genome and epigenetic modification mapping, it was shown that the expression of *PPIF* is affected by AS risk variants located at *ZMIZ1* introns through enhancer-enhancer-promoter interactions. RelB was identified as one of the mediators of this genomic regulation, providing clues to the upstream signaling of PPIF. Examining *PPIF* expression in AS T cells profiled by scRNA-seq techniques revealed that it is upregulated in joint T cells compared to blood and mainly operates in proliferating T cells. The AS risk variants regulating *PPIF* expression were also associated with susceptibility to IBD. An old drug named Ciclosporin, which has been approved for treating IBD, was found to suppress inflammation through targeting PPIF. Such genetic risk overlap suggests the possibility of repurposing Ciclosporin to treat AS.

The JAK families were associated with a wide range of inflammatory conditions. Through specificity analysis of genetic risk across AS and other conditions, evidence supporting the repurposing/development of drugs targeting JAK2 and TYK2 was provided. Such drugs could potentially give better efficacy and fewer side effects for treating AS, compared to the approved drugs either targeting all JAK family members or primarily inhibiting JAK1/3.

A multiple-disease-associated risk locus at *SH2B3* intron was used as an example demonstrating a risk locus that could directly regulate the expression of multiple genes. 3D chromatin interaction mapping and eQTL association results show this risk locus affects an enhancer interacting with the promoter of 18 genes. Additionally, the transcriptional factor c-Rel was found to regulate the transcriptional hub upon PMA stimulation in naïve T cells using comparative ChIP-seq.

Chapter 4: T-cell-instructed monocyte activation is a key feature in Ankylosing Spondylitis and provides novel therapeutic opportunities



Graphic abstract of major findings

1. Introduction

IL-1β and IL-23 are key cytokines for initiating type 3 immunity, and their production by myeloid cells was reported to be elevated in patients with AS[170,171]. Monocytes in AS patients show enhanced spontaneous (and stimulation-induced) production of IL-1β compared with healthy controls[170]. In addition, the level of serum IL-23 and lipopolysaccharides (LPS)-induced macrophage IL-23 production was significantly higher in AS patients in comparison with healthy controls[171–173]. Thus, monocyte activation is likely a key driver of Th17 pathology in AS and offers opportunities for better clinical treatments.

Inositol-requiring enzyme 1 (IRE1α), encoded by the AS risk gene *ERN1*, is an endoplasmic reticulum (ER) stress sensor that initiates the adaptive unfolded protein response (UPR) to maintain cellular homeostasis[174,175]. IRE1α is activated under ER stress by self-dimerization and phosphorylation, activated IRE1α then cleaves the mRNA of the inactive transcription factor X-box binding protein 1 (XBP1) to generate the mature mRNA that encodes the cleaved XBP-1 (XBP1s) protein, thus upregulating expression of UPR target genes involved in protein folding, processing, and degradation[176,177]. The IRE1α /XBP1 pathway has been shown to be important for the activation of several types of immune cells including macrophages. It has been proposed that the IRE1α /XBP1 pathway is required for toll-like receptor (TLR)-mediated production of pro-inflammatory cytokines IL-6 and TNF-α in macrophages[178].

How monocytes are activated in AS is unclear. Gut inflammation and leakage have been found in a proportion of patients, accompanied with the upregulation of IL-1β and IL-23 in inflamed gut tissue[179–181]. Accordingly, a model in which monocytes are activated by bacteria in the gut or by bacterial derivatives in the bloodstream has been considered[179]. We here provide evidence that T cell activation of monocytes stimulates the production of pro-inflammatory cytokines of known

pathological relevance to AS. We identify a population of monocytes in AS joints that closely resemble T cell-activated monocytes and identify IRE1α as a (novel) therapeutic target within this pathway.

2. Special Acknowledgements

Despite my significant contribution, this chapter describes a study that involved extensive collaboration with my colleagues, particularly Jinyi Zhao. Specifically, I generated and analyzed single-cell RNA sequencing (scRNA-seq) data profiling *in vitro* stimulated monocytes and identified a unique monocyte subset that is induced by T cells and expresses IL-23 and CXCL2. I also analyzed scRNA-seq data from AS PBMCs and SFMCs generated by Tae-Jong Kim group[127] and Davide Simone and demonstrated that CD14+ myeloid cells in AS synovial fluid share transcriptional similarities with the T-cell-induced IL-23+CXCL2+ monocytes. Furthermore, I performed gene set enrichment analysis of AS-associated risk genes and identified ERN1 as a potential regulator of the function and development of IL-23+CXCL2+ monocytes.

Meanwhile, my colleague Jinyi Zhao utilized large-scale AS blood samples to confirm, at the protein level, the production of IL-1β, IL-23, and TNF-α by monocytes when activated by T cells. She also demonstrated that this monocyte activation can promote Th17 polarization. Based on the hypothesis that the activated T cells providing stimulation signals to monocytes could originate from the gut, she performed experiments to demonstrate that T cells can still activate monocytes even 48 hours after antigen engagement. Finally, she validated that ERN1 can indeed regulate the production of IL-1β, IL-23, and CXCL2 by monocytes upon T cell activation. Additionally, Claudia Worth, Shristi Lamma, and Davide Simon contributed to patient recruitment. This study was supervised by my supervisor, Dr. Liye Chen, and Professor Paul Bowness. I would like to extend our gratitude to the patient participants and healthy donors without whom this study would not have been possible.

3. Results

3.1 Activated T-cells induce monocyte IL-1β, IL-23 and TNF-α production

We utilized 10X single-cell RNA sequencing (scRNA-seq) to profile the cellular heterogeneity and transcriptome of PBMCs from 3 ankylosing spondylitis (AS) patients comparing three different conditions: unstimulated control, stimulation with T-cell activation beads (TAB) or lipopolysaccharides (LPS) for 16 hours (**Figure 4.1A**). A total of 59,840 cells, representing all typical blood cell types, were obtained (**Figure 4.1B left and 4.2 A**). As expected, TAB and LPS activated T cells and monocyte respectively (**Figure 4.2B and C**). To our surprise, TAB also induced strong monocyte activation, characterized by the upregulation of *IL23A*, *IL1B*, and *TNF* (**Figure 4.1B right**).

To confirm this finding at the protein level, we performed intracellular cytokine staining (ICS) for these cytokines in PBMCs from 13 AS patients stimulated with LPS or TAB. In line with the scRNA-seq findings, the production of IL-1β and TNF-α by monocytes was significantly increased after both LPS and TAB stimulation (**Figure 4.3A and Figure 4.1C**). Of note, TAB induced a level of cytokine production either comparable to (IL-1β) or even higher (TNF-α) than that of LPS. Since ICS was unable to measure the mature form of IL-1β or quantify IL-23 (very low percentage in ICS), we utilized the enzyme-linked immunosorbent assay (ELISA) assays to detect the secreted form of IL-1β and IL-23 in the culture supernatant and confirmed the production of these two cytokines in TAB condition (**Figure 4.1D**). To completely exclude the potential secretion of IL-1β and IL-23 from other cells in PBMCs, we carried out a similar experiment with isolated monocytes and CD3+ T-cells from the same donors with AS. TAB only induced IL-1β and IL-23 secretion in T-cell/monocyte coculture but not in T-cell alone conditions (**Figure 4.1E**). As expected, TNF-α, known to be produced by activated T-cells, was detected in both conditions. Taken together, these

data reveal a novel mechanism of T-cell instructed monocyte activation leading to the production of IL-1β, IL-23 and TNF-α by monocytes.

3.2 T-cell instructed monocytes produce IL-1β and IL-23 at levels that are adequate to promote Th17 response in AS

We then asked if the T-cell instructed monocyte activation could play a downstream role in AS pathology. Given the clear pathological role of Th17 cells in AS and the well-established role of IL-1β and IL-23 in driving Th17 responses, we asked if T-cell-induced monocyte activation fed back to T-cells to support Th17 expansion. To this end, we examined the effect of monocyte depletion on the Th17 response in AS PBMCs stimulated with Th0 condition (TAB and IL-2). IL-17A production by CD4+ T-cells from patients with AS was significantly downregulated in monocyte-depleted PBMCs compared with whole PBMCs (**Figure 4.1F, Figure 4.3B and 4.3C**). Strikingly, the Th17-enhancing effect driven by exogenous IL-1β and IL-23 was only observed in monocyte-depleted PBMCs and not whole PBMCs from AS patients (**Figure 4.1G, Figure 4.3B and 4.3C**), suggesting that the monocytes secreted sufficient levels of IL-1β and IL-23 that are adequate to enhance Th17 response.

(see figure in the next pages)

Figure 4.1. Activated T-cells induce monocyte activation that acts back to T-cells and supports Th17 response. (A) Overview of experimental workflow profiling *in vitro* stimulated AS PBMCs using the scRNA-seq technique. (B) Left: UMAP visualization of the major populations of all cells integrated from unstimulated, LPS-, and TAB-stimulated conditions. Right: Expression of IL23A, IL1B, and TNF projected to UMAP of monocytes from three conditions. (C) PBMCs from AS patients (n=13) were stimulated with T-cell activation beads or LPS for 16 hours. Left: representative flow cytometry plots and gating strategy of IL-1β- and TNF-α-producing monocytes in response to stimulation. Right: summary graphs showing the percentage of monocytes producing IL-1β or TNF-α in response to stimulation. (D) ELISA results for the production of inflammatory cytokine IL-1β, IL-23 and TNF-α in AS PBMCs stimulated with T-cell activation beads or LPS for 16 hours. (E) Isolated CD3+ T-cells from PBMCs from AS patients (n=9) were cultured alone or co-cultured with autologous monocytes in the absence/presence of TAB for 16 hours. The levels of cytokines IL-1β, IL-23 and TNF-α in the culture supernatant were measured using ELISA. (F) Left: Representative flow cytometry plots showing the percentage of IL-17A-producing CD4+ T cells within the whole PBMCs or monocyte-depleted PBMCs from AS patients (n=9) after 6 days *in vitro* culture under Th0 condition. Right: a summary graph for the frequency of IL-17A+ cells in CD4. (G) Summary graphs showing the frequency of IL-17A+ CD4 within the whole PBMCs (left) or monocyte-depleted PBMCs (right) from AS patients (n=9) after *in vitro* culture under Th0 or Th17 polarizing conditions for 6 days. The P-value was assessed using the paired two-tailed Student's t-test (* $P \leqslant 0.05$; ** $P \leqslant 0.01$; *** $P \leqslant 0.001$).

Figure 4.2. **T-cell activation beads or LPS-induced immune cell activation in PBMCs from patients with AS.** (A) Expression of a selected set of marker genes by different populations of PBMCs shown in Figure 4.1 A. (KLRF1: killer cell lectin-like receptor F1, a well-known maker for NK cells; FLT3: fms-related receptor tyrosine kinase 3, a gene uniquely expressed in DCs.) (B) UMAP visualization of PBMCs from different conditions, coloured by treatment conditions. (C) Expression of activation markers projected to UMAP of PBMCs integrated from three conditions.

Figure 4.3. **Gating strategy for monocyte cytokine production and Th17 cells.** (A) Gating strategy for identifying CD14+ monocytes in PBMCs of AS patients. Myeloid cells were gated on singlet cells and monocytes were defined as CD3−CD14+ cells. (B) Gating strategy for identifying CD4+ T-cells in PBMCs of AS patients. Lymphocytes were gated on singlet cells and CD4+ T cells were defined as CD3+CD4+CD8- cells. (C) Representative flow cytometry plots of IL-17A-producing CD4+ T cells from the whole PBMCs (left) or monocyte-depleted PBMCs (right) from AS patients after 6 days *in vitro* culture under Th0 (TAB+IL-2) or Th17 (TAB+IL-2/IL-1β/IL-23) polarizing conditions.

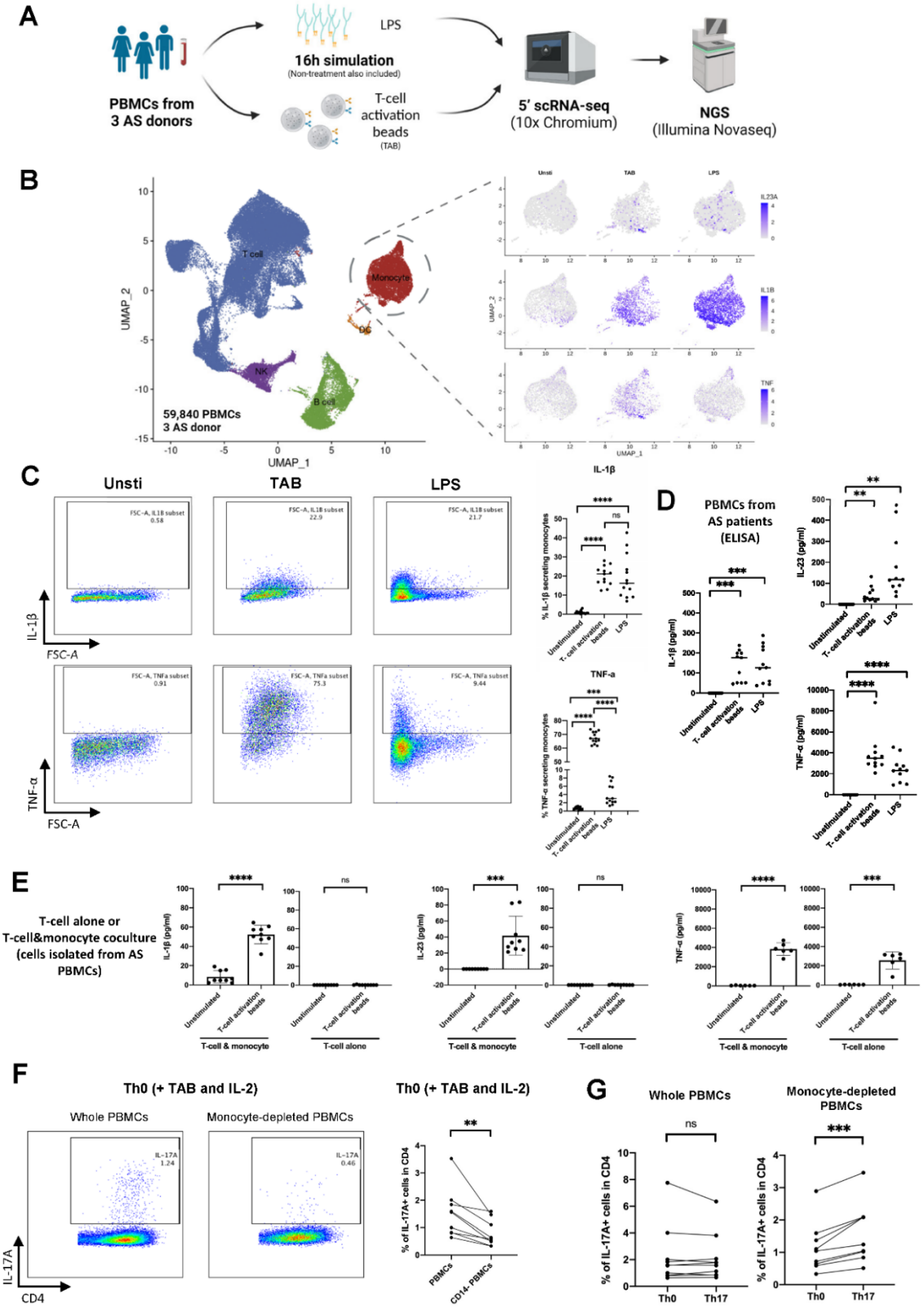


Figure 4.1. Activated T-cells induce monocyte activation that acts back to T-cells and supports Th17 response.

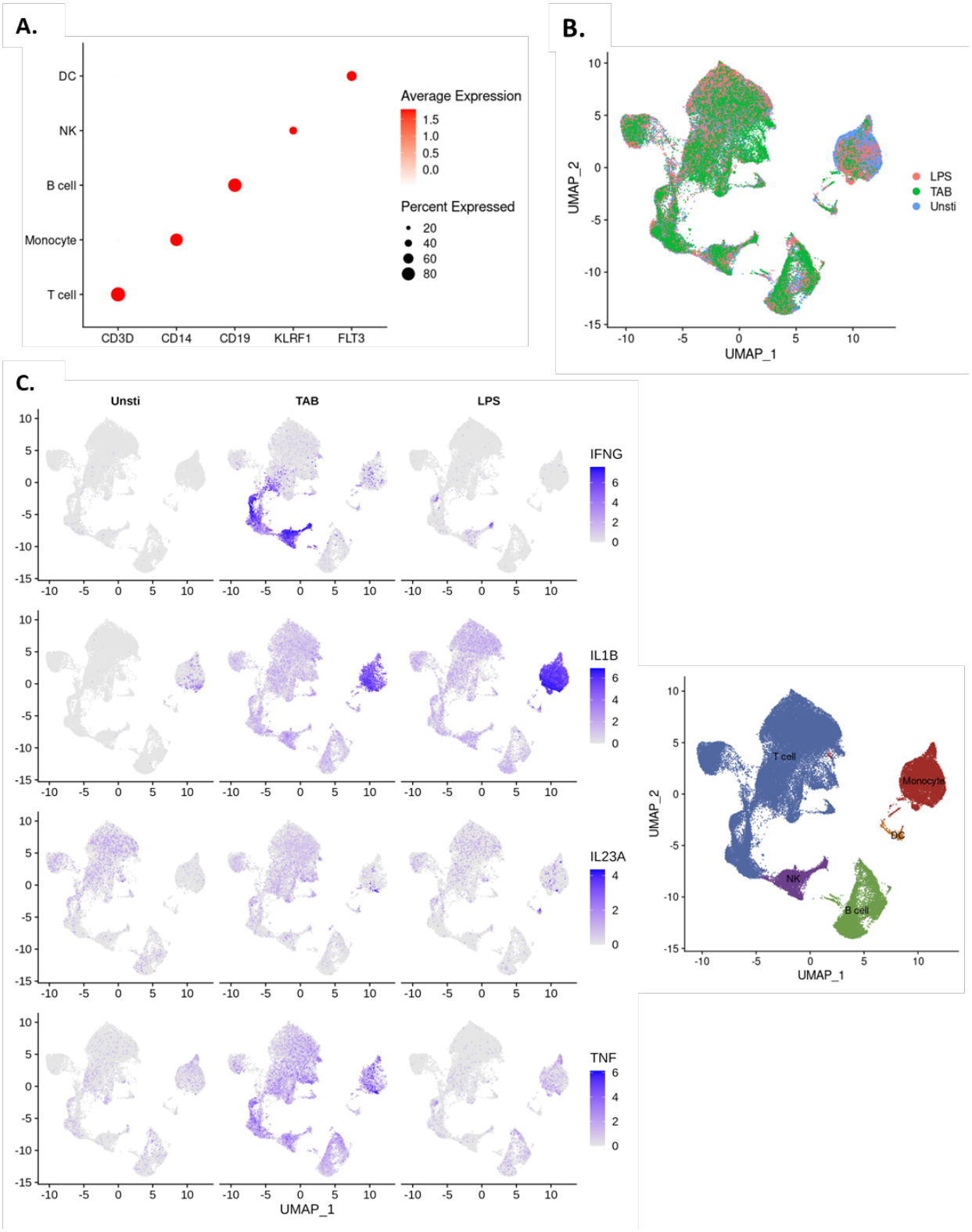


Figure 4.2. T-cell activation beads or LPS induced immune cell activation in PBMCs from patients with AS.

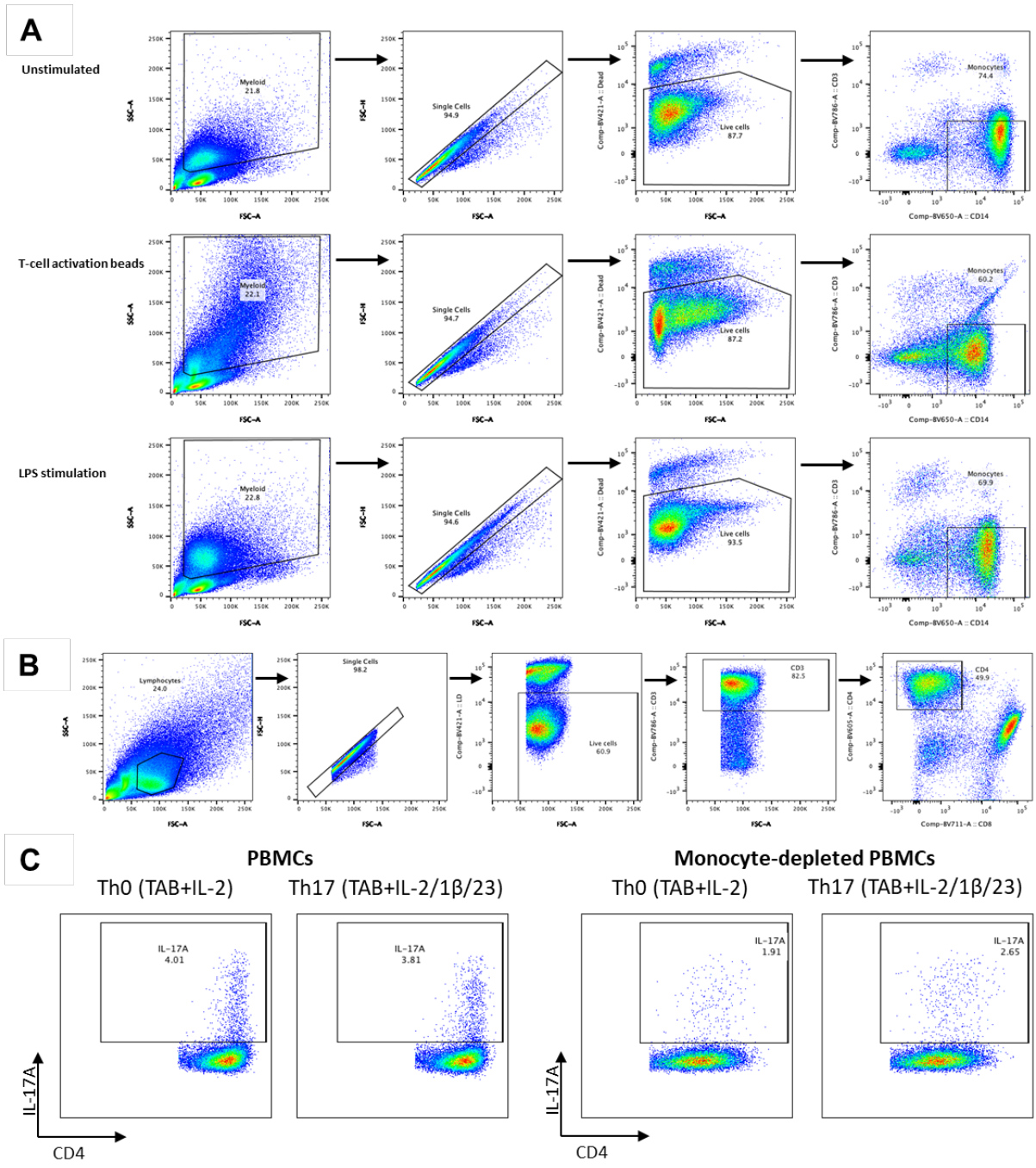


Figure 4.3. Gating strategy for monocyte cytokine production and Th17 cells.

3.3 T-cell activated monocytes exhibit unique transcriptional features

We then asked if T-cell instructed and LPS-induced monocyte activation involve similar or distinct transcriptional programs. Compared to unstimulated condition, TAB induced substantial transcriptional changes in monocyte, with 1638 up- and 4097 genes down-regulated upon TAB activation (adjusted p-value < 0.05). In addition to TNF, IL1B, and IL23A, multiple inflammation-related genes were induced including NFKB1 (encoding the NF-κB subunit p105), JUN (encoding c-Jun, an AP-1 subunit), CXCL2 (a chemokine recruiting neutrophils), CD40 (mediating interaction with T cells) and interferon-stimulated genes (RSAD2, GBP5, ISG20, and ISG15) (**Figure 4.4A**). 2263 and 3540 genes were found to be up- or down-regulated respectively following LPS stimulation (compared to the unstimulated control) (**Figure 4.4B**). Of note, around one third of upregulated genes (547) in TAB condition were not induced by LPS (**Figure 4.4C**), strongly suggesting the involvement of unique transcriptional programs.

To better evaluate the transcriptional differences between monocytes following T-cell and LPS-stimulation, we pooled and re-clustered monocytes from these two conditions and identified seven populations with distinct gene expression profiles (**Figure 4.4D, E and G, Table 4.1**). The largest cluster (Calprotectin+C0) comprised the majority of LPS-stimulated cells and displayed elevated expression of *S100A8/9* (calprotectin is a heterodimer of S100A8 and A9) (**Figure 4.4F**). With the exception of *CCR7*+ C4 that was similar in abundance between LPS and TAB, the remaining clusters were dominated by cells from TAB condition (**Figure 4.4F**). The CD16+ (encoded by *FCGR3A* gene) non-classical monocytes were found in cluster 1. Cluster 2 expressed a high level of *JAK2* but was not enriched for pro-inflammatory cytokines or chemokines. Cluster C3 (*IL23A*+*CXCL2*+) expressed high levels of pro-inflammatory cytokines (*IL23A*, *TNF* and *IL1B*) and genes associated with neutrophil recruitment (*JUN*/*CXCL2*) (**Figure 4.4G**). The *LGALS9*+ C5 cluster expressed high levels of *LGALS9*, *SPP1* (secreted phosphoprotein 1, which encodes a protein named Osteopontin involving bone remodelling and immune response.), *MIF*

((macrophage migration inhibitory factor). Cluster C6 exhibited high levels of multiple chemoattractants (*CCL8*, *CCL2* and *CCL7*), suggesting a function in recruiting inflammatory leucocytes. Overall, these results show that activated T-cells induce a heterogeneous monocyte activation profile with populations distinct to those seen following LPS challenge.

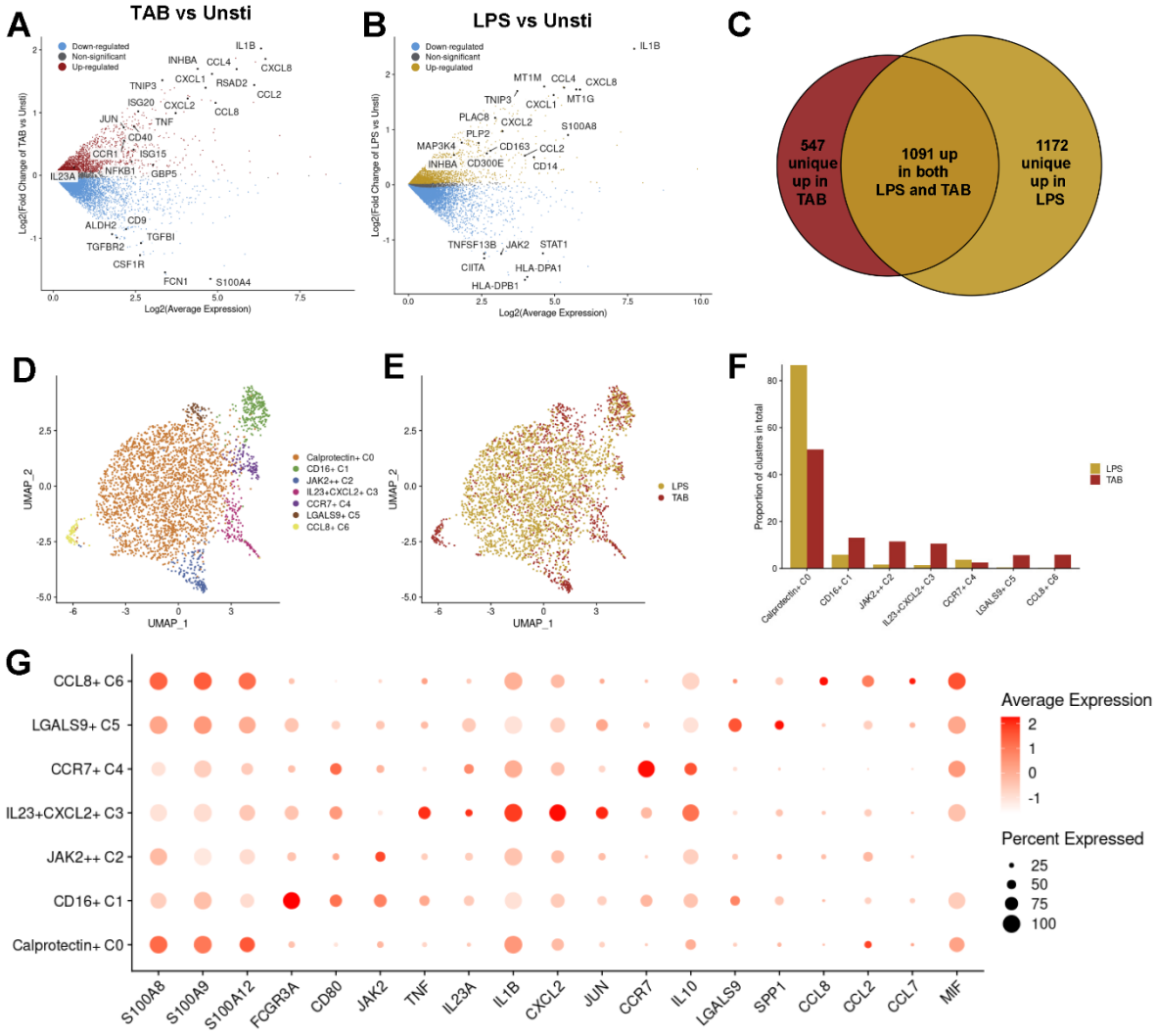


Figure 4.4. T-cell-activated monocytes exhibit unique transcriptional features. (A) MA plot of differentially expressed genes between TAB and unstimulated conditions (FDR <0.05). The average expression levels are shown for TAB when the fold change is greater than 0, and for Unsti when the fold change is less than 0. (Note: MA plot shows the log average (A) on the x -axis and the log ratio (M) on the y -axis.) (B) MA plot of differentially expressed genes between LPS and unstimulated conditions (FDR <0.05). The average expression levels are shown for LPS when the fold change is greater than 0, and for Unsti when the fold change is less than 0. (C) Venn plot depicting the overlap of up-regulated genes between LPS and TAB conditions. (D) UMAP visualization of transcriptionally distinct populations of monocytes integrated from LPS- and TAB-stimulated conditions. (E) UMAP visualization of monocytes from different conditions, coloured by treatment conditions. (F) Frequency of cells from various monocyte subsets in LPS and TAB conditions, respectively. (G) Expression of a selected set of cluster markers. Top 20 up-regulated genes in each cluster were listed in Table 4.1.

Table 4.1 Top 20 gene up regulated in clusters defined in Fig 4.4D (order by fold change)

Cluster ID	Gene	log2(average fold change)	Adjusted P value
C1 IL-23+	CALHM6	2.672214997	2.02E-62
C1 IL-23+	FCGR3A	2.51526704	1.61E-152
C1 IL-23+	CXCL10	2.375020669	3.29E-31
C1 IL-23+	APOBEC3A	2.281205265	9.35E-53
C1 IL-23+	CXCL9	2.258332863	6.87E-18
C1 IL-23+	IFITM1	1.933350939	4.97E-47
C1 IL-23+	GBP1	1.894078702	1.78E-32
C1 IL-23+	CCL4L2	1.884048787	1.16E-08
C1 IL-23+	LAMP3	1.864274239	1.06E-69
C1 IL-23+	IFITM3	1.737511764	4.08E-43
C1 IL-23+	C1QB	1.725767052	3.19E-76
C1 IL-23+	EBI3	1.630242471	1.25E-37
C1 IL-23+	ACOD1	1.540964074	5.90E-20
C1 IL-23+	GBP4	1.520563577	2.18E-35
C1 IL-23+	LILRA5	1.511671333	1.50E-42
C1 IL-23+	CD40	1.455388996	9.13E-23
C1 IL-23+	CD83	1.455122113	1.83E-34
C1 IL-23+	DAPK1	1.409170644	1.67E-39
C1 IL-23+	ADA	1.405846978	8.73E-21
C1 IL-23+	PPA1	1.380284191	7.85E-31
C0 S100+	S100A8	2.693005461	2.61E-173
C0 S100+	CXCL5	2.21932849	3.86E-74
C0 S100+	MT1G	1.916238718	2.59E-75
C0 S100+	CD14	1.842302156	3.61E-131
C0 S100+	MT1M	1.811566334	1.66E-84
C0 S100+	CXCL1	1.722784326	2.31E-69
C0 S100+	S100A9	1.717115092	4.28E-141
C0 S100+	RETN	1.582463863	1.54E-49
C0 S100+	MT1F	1.54733468	4.90E-85
C0 S100+	CD163	1.537221697	2.31E-80
C0 S100+	MT1X	1.535914017	4.21E-78
C0 S100+	MT2A	1.50842732	9.13E-92
C0 S100+	MCEMP1	1.481395231	1.22E-129
C0 S100+	VCAN	1.456353283	5.15E-82
C0 S100+	S100A12	1.445205806	1.28E-81
C0 S100+	MT1H	1.358956235	4.41E-39
C0 S100+	PID1	1.241432132	1.07E-74
C0 S100+	LGALS1	1.232658597	1.11E-92
C0 S100+	SLC11A1	1.228758051	2.84E-117
C0 S100+	MT1E	1.116157719	3.49E-59
C2 CD16+	TNF	4.126183215	3.21E-46

C2 CD16+	MIR155HG	2.699682538	1.01E-47
C2 CD16+	CXCL2	2.687799558	6.23E-21
C2 CD16+	SNX9	2.660000577	3.26E-45
C2 CD16+	CD83	2.639518575	1.11E-48
C2 CD16+	HIF1A-AS3	2.606364071	2.60E-06
C2 CD16+	INHBA	2.517899306	9.51E-36
C2 CD16+	PPP1R15A	2.384596327	1.13E-46
C2 CD16+	AC015912.3	2.356645591	1.09E-56
C2 CD16+	HSPA1B	2.331323744	4.37E-05
C2 CD16+	RASGEF1B	2.329400373	4.67E-43
C2 CD16+	SGK1	2.31533653	7.77E-34
C2 CD16+	IL23A	2.248224652	1.11E-26
C2 CD16+	JUN	2.204882431	4.13E-28
C2 CD16+	SNAPC1	2.185197592	3.35E-10
C2 CD16+	SPINK1	2.1508161	3.90E-27
C2 CD16+	CDC42EP3	2.100424158	9.47E-33
C2 CD16+	GADD45B	2.091610795	2.44E-10
C2 CD16+	LINC02154	2.057057942	4.81E-16
C2 CD16+	REL	1.98681885	1.06E-30
C3 IL10+	THBS1	3.79588912	8.08E-29
C3 IL10+	CCR7	3.193857893	9.21E-161
C3 IL10+	EBI3	2.738323457	9.86E-63
C3 IL10+	CCL22	2.60982637	2.10E-25
C3 IL10+	LAMP3	2.554173098	1.85E-130
C3 IL10+	ADAM19	2.308898201	4.06E-190
C3 IL10+	SLC1A2	2.258626857	2.45E-69
C3 IL10+	PPA1	2.236882617	1.73E-66
C3 IL10+	MMP7	2.116030207	4.30E-34
C3 IL10+	CCL19	2.025225816	3.76E-67
C3 IL10+	FSCN1	1.885561591	3.21E-89
C3 IL10+	BLVRA	1.877958119	3.43E-56
C3 IL10+	LNCAROD	1.814962804	7.68E-92
C3 IL10+	CXCL9	1.814588235	0.008266123
C3 IL10+	CKB	1.779895744	5.21E-13
C3 IL10+	IL2RA	1.72419743	2.61E-48
C3 IL10+	TMSB4X	1.673002088	7.65E-21
C3 IL10+	ABTB2	1.65467421	6.10E-98
C3 IL10+	LAD1	1.577436606	2.82E-69
C3 IL10+	MCOLN2	1.561919677	3.49E-59
C4 SPP1+	CCL24	4.676107805	5.65E-15
C4 SPP1+	RSAD2	2.36603501	2.88E-21
C4 SPP1+	CIR1	2.293996937	3.51E-27
C4 SPP1+	RGCC	2.242785917	1.25E-24

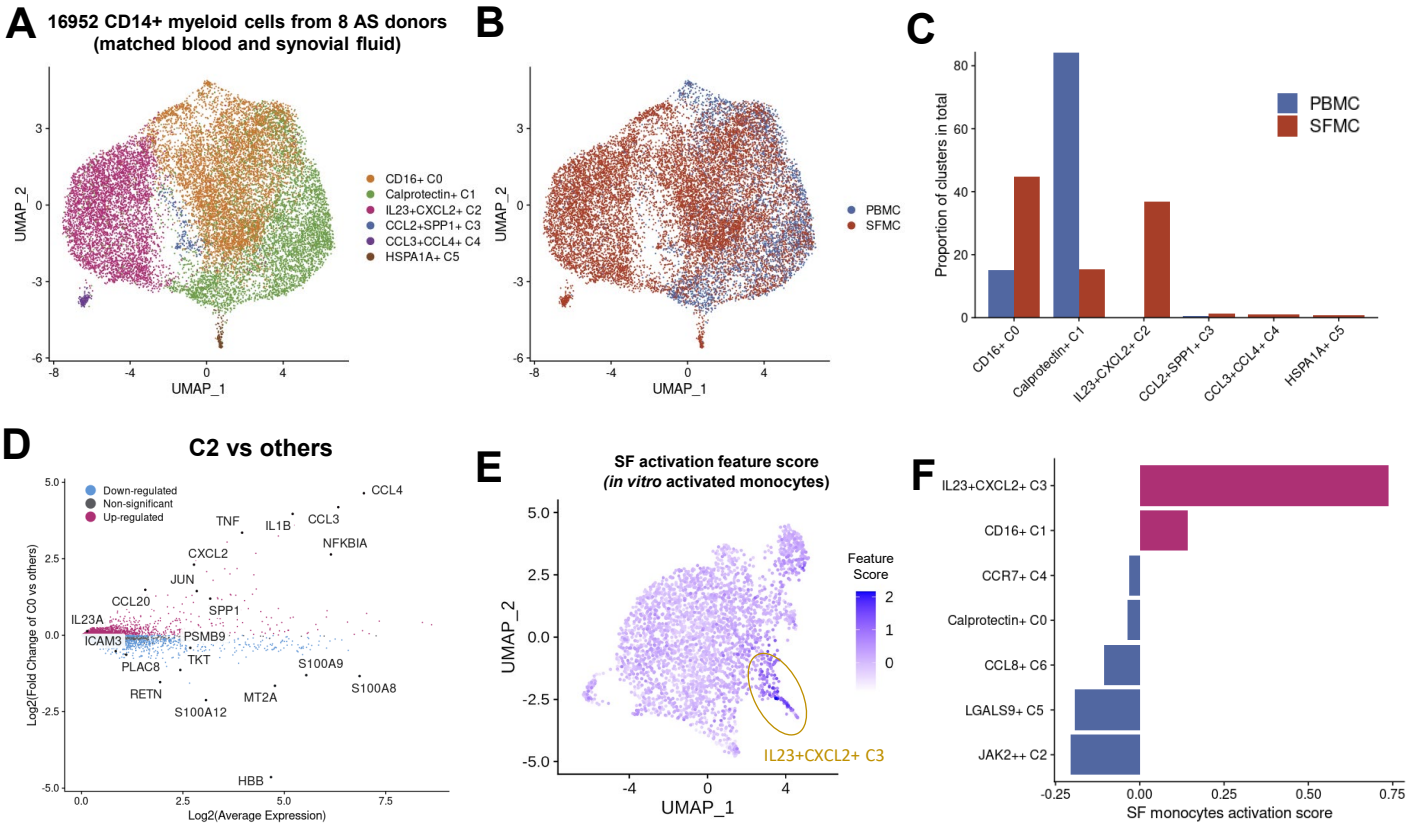
C4 SPP1+	CHI3L1	2.209576998	1.47E-12
C4 SPP1+	GM2A	2.19148904	3.40E-28
C4 SPP1+	SPINK1	2.128573768	2.44E-15
C4 SPP1+	SPP1	2.078199101	1.96E-11
C4 SPP1+	LSP1	1.984277433	6.33E-36
C4 SPP1+	OCSTAMP	1.886418502	1.11E-67
C4 SPP1+	AC015660.2	1.876660505	1.54E-97
C4 SPP1+	HLA-DRB1	1.866886175	2.82E-18
C4 SPP1+	CYP1B1	1.854387792	3.67E-10
C4 SPP1+	MFSD12	1.850707459	2.19E-30
C4 SPP1+	ATF3	1.817159172	1.36E-19
C4 SPP1+	CSF1	1.770118683	1.14E-43
C4 SPP1+	HLA-DPA1	1.670502566	9.19E-15
C4 SPP1+	HLA-DRA	1.642680037	1.63E-15
C4 SPP1+	CD74	1.623787128	1.09E-11
C4 SPP1+	MATK	1.609336265	3.37E-62
C5 JAK2++	NEAT1	2.343087212	1.77E-33
C5 JAK2++	ALCAM	2.189935374	1.65E-13
C5 JAK2++	CHST11	2.121728353	2.30E-11
C5 JAK2++	LINC01705	2.100632349	0.007126536
C5 JAK2++	SLC8A1	2.062172974	9.88E-08
C5 JAK2++	PRR16	1.990986965	0.045691795
C5 JAK2++	JAK2	1.984088352	4.01E-14
C5 JAK2++	TPRG1	1.944455908	2.36E-08
C5 JAK2++	ARHGAP24	1.941375957	1.27E-06
C5 JAK2++	ARL15	1.94109114	3.63E-23
C5 JAK2++	MALAT1	1.917634612	7.23E-29
C5 JAK2++	PARP14	1.892018587	6.05E-18
C5 JAK2++	ASPH	1.865735266	3.20E-04
C5 JAK2++	PTPRJ	1.833068683	1.47E-27
C5 JAK2++	DOCK4	1.826698473	3.26E-37
C5 JAK2++	SLC39A11	1.818237284	6.45E-10
C5 JAK2++	HIVEP2	1.795671622	2.72E-16
C5 JAK2++	DNAJC6	1.792075692	1.63E-11
C5 JAK2++	SIPA1L1	1.790801829	3.07E-17
C5 JAK2++	P2RX7	1.78894511	1.10E-15
C6 Hemoglobin+	HBB	4.641210175	5.19E-04
C6 Hemoglobin+	HBA2	3.024530254	1.66E-28
C6 Hemoglobin+	TFPI2	2.505542217	2.57E-10
C6 Hemoglobin+	HBA1	2.404339789	6.57E-42
C6 Hemoglobin+	WARS	1.803611464	5.67E-11
C6 Hemoglobin+	PSME2	1.719976752	7.68E-06
C6 Hemoglobin+	CCL8	1.466492031	7.35E-13

C6 Hemoglobin+	LAP3	1.409139802	1.66E-05
C6 Hemoglobin+	RSAD2	1.327542801	3.43E-08
C6 Hemoglobin+	MX1	1.314335915	0.017979135
C6 Hemoglobin+	ISG20	1.287050565	0.006456135
C6 Hemoglobin+	TXN	1.209667678	7.42E-07
C6 Hemoglobin+	HSP90AB1	1.194223398	0.007385345
C6 Hemoglobin+	SPINK1	1.191671565	1.48E-05
C6 Hemoglobin+	IRF1	1.177672272	1.73E-04
C6 Hemoglobin+	CD74	1.16754113	2.55E-04
C6 Hemoglobin+	CCL2	1.165481856	0.017021894
C6 Hemoglobin+	GBP1	1.078076074	7.10E-10
C6 Hemoglobin+	MED13L	0.339605103	0.008521223
C6 Hemoglobin+	CSNK2B	0.337945043	0.004571072

3.4 CD14+ myeloid cells in AS synovial fluid are transcriptionally similar to T-cell induced IL-23+CXCL2+ C2 monocytes

Next, we asked if T-cell instructed monocyte activation is relevant to the pathology of inflammatory synovium in patients with AS. To this end, we examined the expression profile of CD14+ myeloid cells from the synovial fluids of eight Ankylosing Spondylitis (AS) patients and their corresponding blood. A total of approximately 16952 CD14+ myeloid cells with high-quality single-cell RNA sequencing (scRNA-seq) transcriptomic profiles were obtained (**Figure 4.5A and B**). We identified six clusters but decided to focus on cluster C0-2 as the remaining clusters were scarce in both blood and SF (**Figure 4.5C**). CD16+ non-classical monocytes (cluster C0) and Calprotectin+ cluster C1 were more enriched in SF and blood respectively. Notably, cluster C2, shown in purple in **Figure 4.5A**, was dominated by SF-sourced monocytes (**Figures 4.5B and C**), and was highly enriched for pro-inflammatory genes (**Figure 4.5D**). The upregulation of *IL1B* and *IL23A*, as well as *CCL20* (a known Th17 recruiter through CCR6), suggests that cluster C2 could enhance both enrichment and expansion of Th17 cells. Built on this finding, we used the feature genes from the cluster C2 to generate a "SF activation score" and applied it to the different clusters of *in vitro* TAB- and LPS-stimulated monocytes. Remarkably, cells from the *IL23A+CXCL2+* C3 were significantly enriched for the SF activation score (**Figure 4.5E and F**). Interestingly, in addition to well-established AS-related cytokines (*TNF* and *IL23A*), the *JUN*/*CXCL2* pathway, known to recruit neutrophils, was also upregulated in cluster C2 from AS SF and cluster C3 from *in vitro* activated monocytes (**Figure 4.5G**). Notably, the key feature genes (*S100A8* and *S100A9*) in the Calprotectin+ C0 from *in vitro* activated monocytes, representing for almost 90% monocytes following LPS stimulation (**Figure 4.4F**), were not enriched in cluster C2 from AS SF (**Figure 4.6**). This suggests the hyperinflammatory C2 unique to AS SF is unlikely induced by activation mediated via pathogen-associated molecular pattern molecules (PAMPs). Taken together, pro-inflammatory CD14+ myeloid cells in SF have gene signatures of that in T-cell activated monocytes.

(see figure legends in the next page)



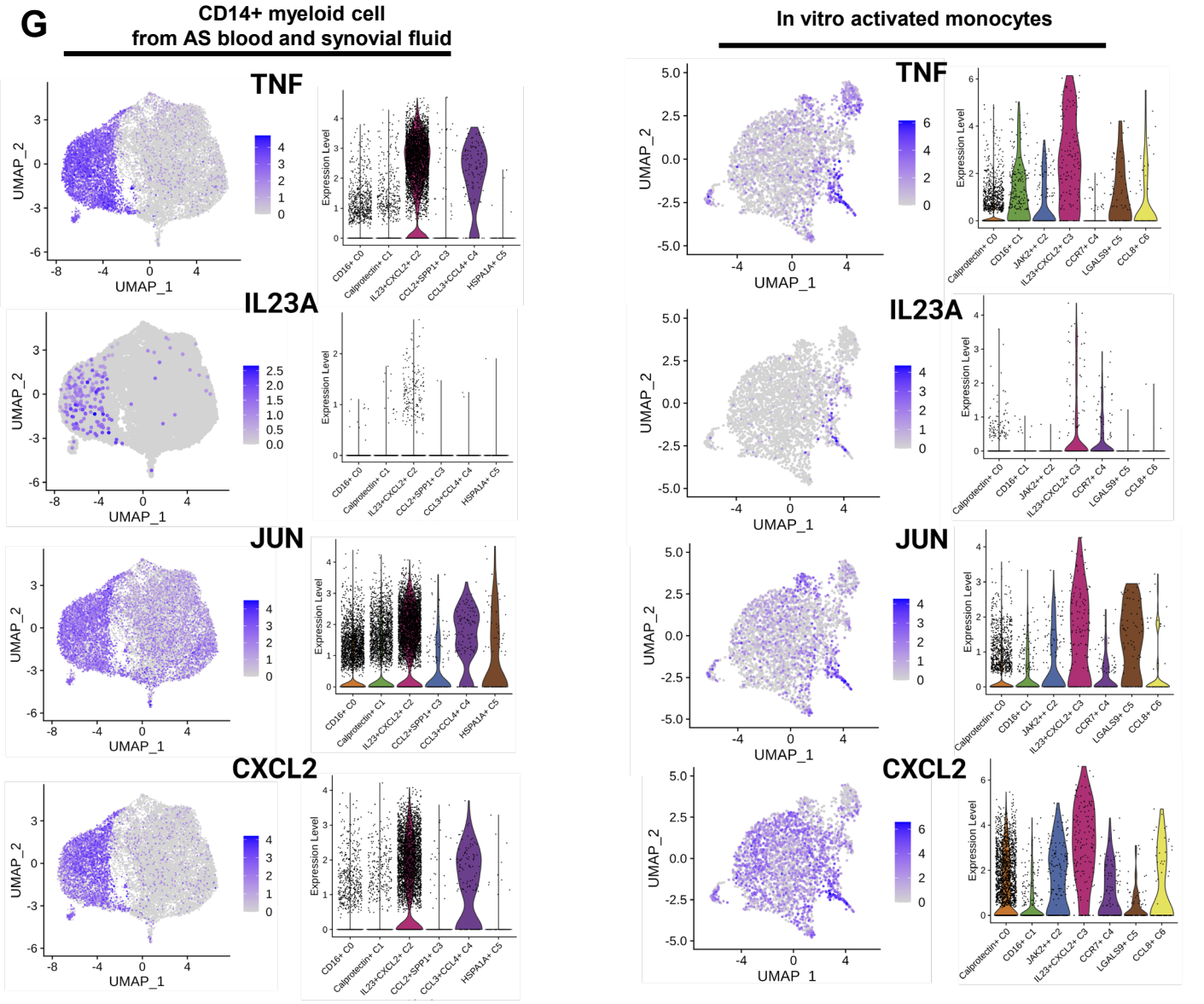


Figure 4.5. CD14+ myeloid cells in AS synovial fluid are transcriptionally similar to T-cell-induced *IL23A+CXCL2+* C2 monocytes. (A) UMAP visualization of transcriptionally distinct populations of CD14+ myeloid cells from 8 pairs of blood and synovial fluid samples from AS donors. (Data of six donors were obtained through collaboration with Kim group[127], and the T cells of these donors have been used in Chapter 3. Data of another 2 samples were obtained in-house and the method was described in Chapter 2. (B) UMAP visualization of CD14+ myeloid cell from blood and synovial fluids, coloured by source of cells. (C) Frequency of cells from various clusters in PBMCs and SFMCs, respectively. (D) MA plot of differentially expressed genes in *IL23A+CXCL2+* C2 cells versus the rest (FDR <0.05). The average expression levels are shown for C2 when the fold change is greater than 0, and for the others when the fold change is less than 0. (E) Summed expression of synovial C2 marker genes in monocytes stimulated by LPS or T cells. (F) Synovial activation feature scores by monocyte subsets identified from *in vitro* stimulated monocytes, matching figure E. (G) Expression of *TNF*, *IL23A*, *JUN*, and *CXCL2* projected to UMAP of CD14+ myeloid cells from AS blood and synovial fluids or *in vitro* stimulated PBMCs.

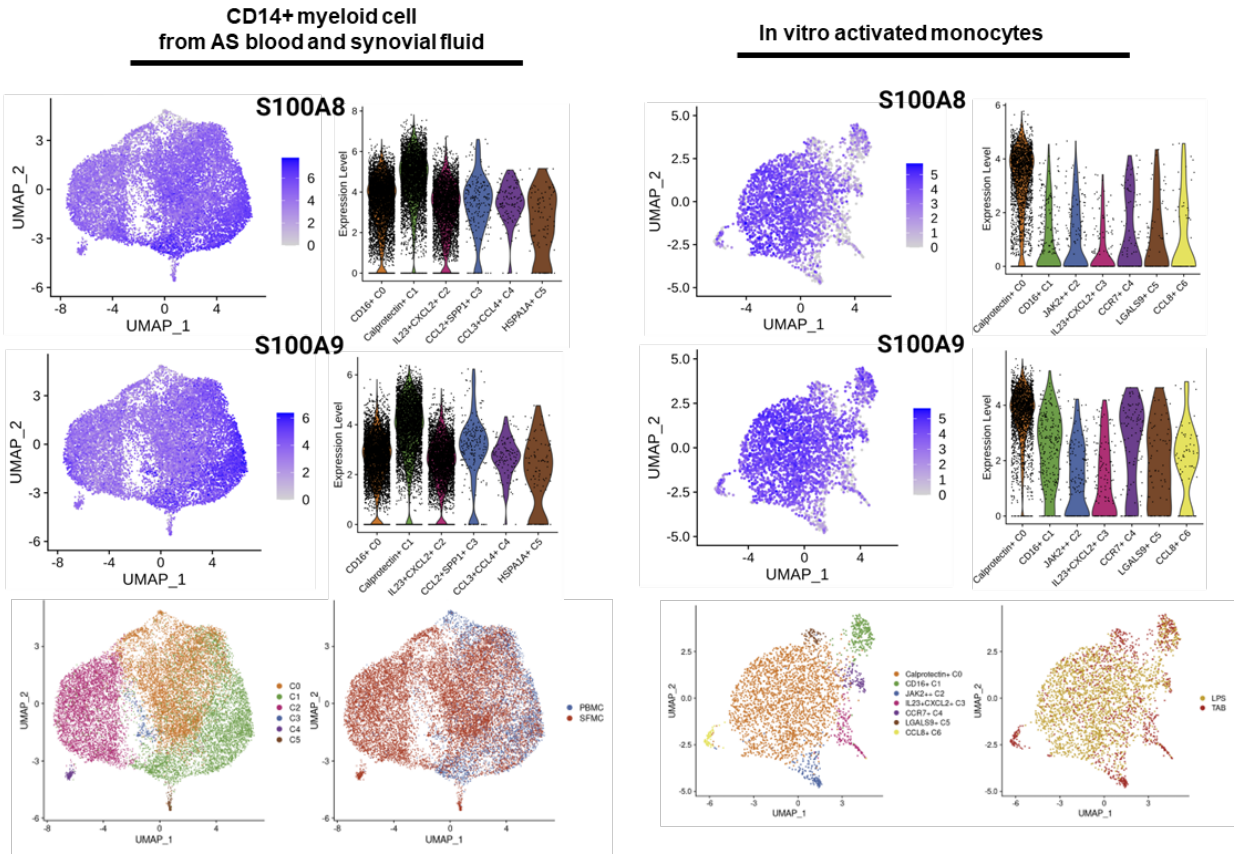
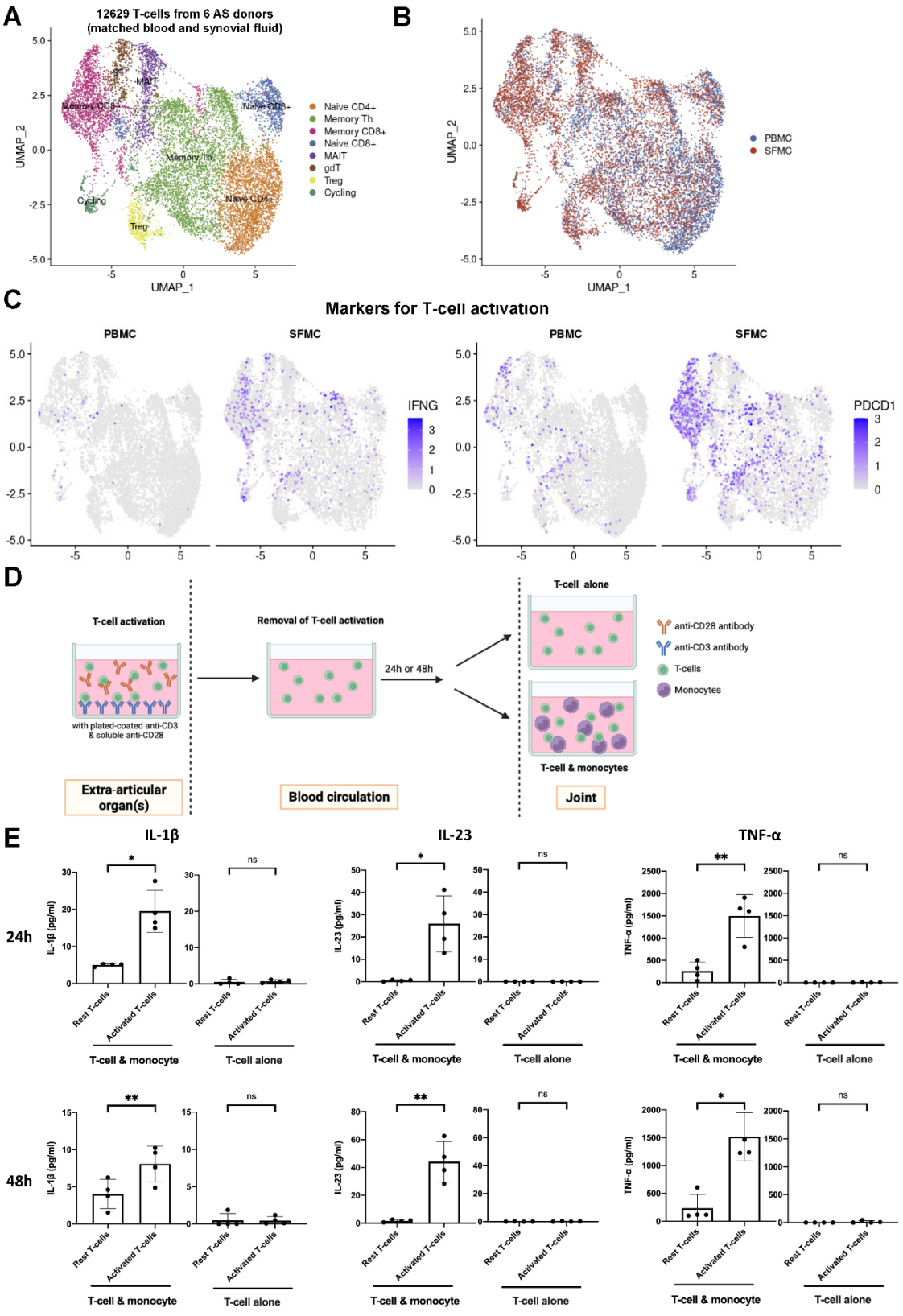


Figure 4.6. Feature genes of *in vitro* generated calprotectin+ C0 are not enriched in CD14+ myeloid cells from AS synovial fluid. UMAP and violin plot of *S100A8* and *S100A9* expression of CD14+ myeloid cells from AS blood and synovial fluids or *in vitro* stimulated PBMCs.

3.5 Activated T-cells are present in AS SF and are composed of both CXCR6+ tissue resident and CXCR6- circulating T-cells

We next looked for evidence of T-cell activation in SF using the same AS synovial and blood datasets from which CD14+ myeloid cells were extracte. A total number of 12629 T-cells were acquired revealing 8 major populations (**Figure 4.7A and B, Figure 4.8A**). Clear enrichment of *PDCD1*+*IFNG*+ activated T-cell was observed in memory CD4+ and CD8+ T-cells from SFMCs compared to those from PBMCs (**Figure 4.7C,** *PDCD1* encodes the PD-1 protein). Notably, these activated T-cells were composed of *CXCR6*+ tissue resident and *CXCR6*- circulating T-cells (**Figure 4.8B**), suggesting that these effector cells could have been both locally generated in joints and migrated from other organs (such as gut). Overall, these findings suggest that activated T-cells were present in AS SF that could provide stimulation to CD14+ myeloid cells in SF.

Multiple pieces of evidence have supported the enrichment of gut-derived T-cells in inflamed AS synovium[182,183]. Thus, the *CXCR6*-*PDCD1*+*IFNG*+ activated T-cells that we observed in AS SF were likely generated in the gut. To test whether gut-derived and activated T-cells could maintain their ability to activate monocytes in the joints, we designed an *in vitro* assay to model this biological process (**Figure 4.7D**). We found that T-cells pre-activated either 24 or 48 hours before co-culture preserved their ability to induce monocytes to produce pro-inflammatory cytokines IL-1β, IL-23 and TNF-α (**Figure 4.7E**). These data, together with data shown in **Figure 4.1**, strongly suggests that T-cells either locally activated in synovium or pre-activated in the gut are capable of inducing monocyte activation in joints.



(see figure legends in the next page)

Figure 4.7. Activated T-cells are present in AS synovial fluids. (A) UMAP visualization of T cells subsets from 6 pairs of blood and synovial fluid samples from AS patients. (B) UMAP visualization of T cells from blood and synovial fluids, coloured by source of cells. (C) Expression of T cell activation marker genes (*IFNG* and *PDCD1*) projected to UMAP of T cells from AS blood and synovial fluids. (D) Schematic representation of the *in vitro* assay simulating the activation of monocytes in joints by T-cells activated in extra-articular organs. CD3+ T-cells (green) were isolated from PBMCs of AS patients and cultured overnight in the presence or absence of anti-CD3 (blue, plate-coated) and anti-CD28 (yellow, soluble). Different from previous experiments, this experiment used plate-bound anti-CD3/CD28 antibodies instead of beads-coated anti-CD2/CD3/CD28 antibodies as this allow the removal of the stimuli afterwards. Antibodies. Unstimulated and stimulated $CD3^+$ T-cells were then cultured in the fresh culture medium for 24 or 48 hours before being co-cultured with the monocytes (purple) from the same donors for 16 hours. The levels of cytokines IL-1β, IL-23 and TNF-α in the culture supernatant were then measured by ELISA (E, n=4). The P-value was assessed using the paired two-tailed Student's t-test (* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$).

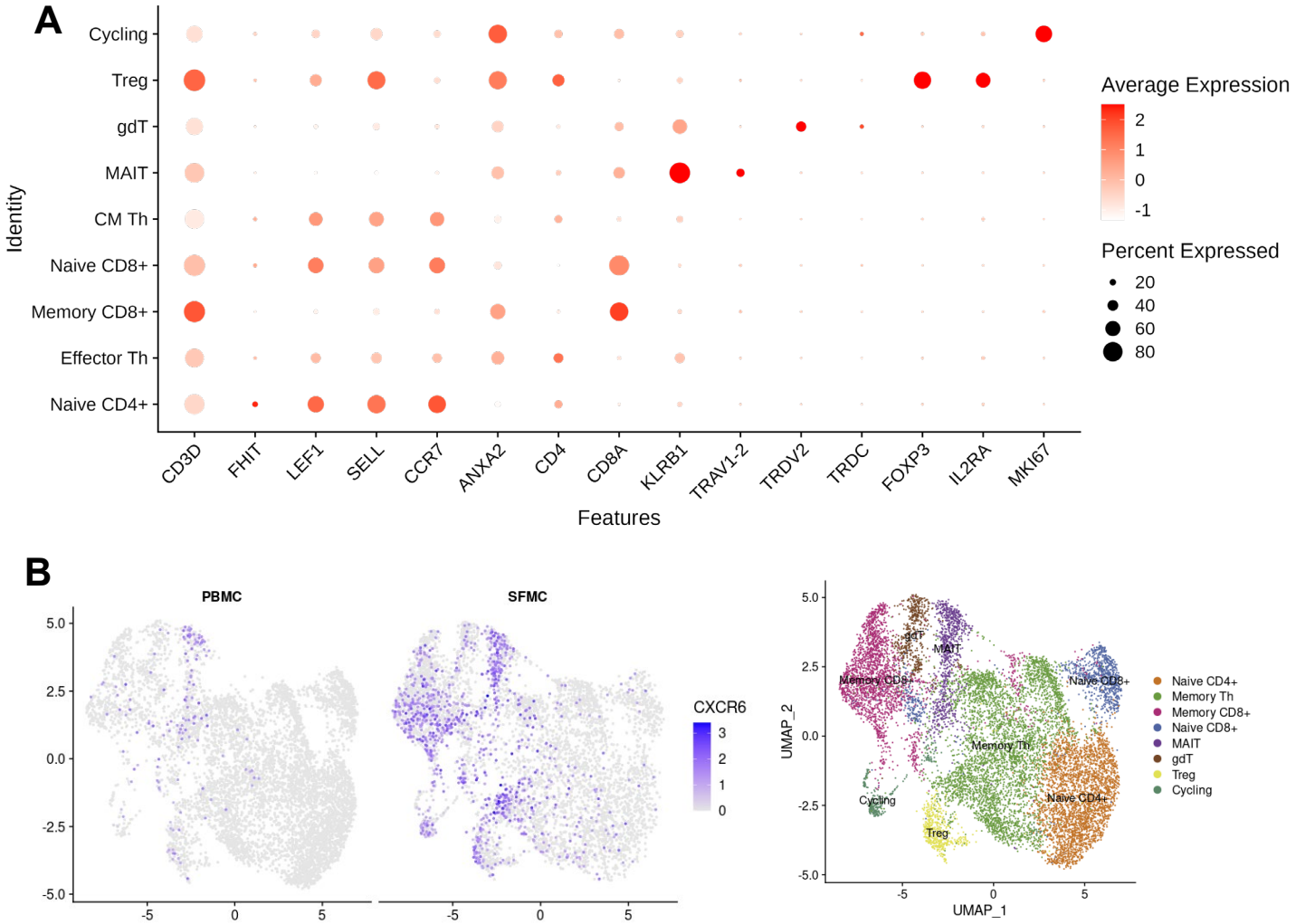


Figure 4.8. A. Expression of a selected set of cluster markers by T-cell subsets from matched blood and synovial fluid from patients with AS (Shown on Figure 4.7) B. Expression of CXCR6 projected to UMAP of T cells from AS blood and synovial fluid.

3.6 T-cell induced IL-23+CXCL2+ monocytes express high levels of AS-associated risk genes and are regulated by the AS risk gene ERN1

Genome-wide association studies (GWASs) have identified over 100 genomic loci affecting the risk of developing AS. Notably, the TNF-α and IL-17A pathways are enriched for AS risk genes, highlighting the potential value of GWASs for the discovery of relevant drug target(s). We performed an AS GWAS candidate gene enrichment analysis using all subsets of monocytes induced by T-cell and LPS stimulation (**Figure 4.4D-F**), and found only the *IL23A+CXCL2+* C3 cluster reached significance (**Figure 4.9A and Table 4.2**). This provides genetic evidence supporting the relevance of T-cell induced monocyte activation to AS pathology.

ERN1 was one of the AS risk genes upregulated in T-cell induced *IL23A+CXCL2+* C3 monocytes (**Figure 4.9B**). It encodes the IRE1α protein, which plays a central role in regulating the unfolded protein response (UPR), a key hypothesized cause of AS. In addition, using a publicly available dataset (ImmuNexUT), we found that the AS risk allele (G) is associated with an increased *ERN1* expression level in human primary monocytes (**Figure 4.9C**). These pieces of evidence encouraged us to hypothesize that *ERN1* might contribute to T-cell instructed monocyte activation. We firstly asked if T-cell-induced cytokine production was associated with the function of IRE1α measuring spliced X-box binding protein 1 (XBP1s), a classic marker of IRE1α activation. In AS PBMCs stimulated with T-cell activation beads (TABs), we observed a significant increase of XBP1s+ cells in monocytes that produced IL-1β (**Figure 4.9D**). To test if IRE1α has a causal role in T-cell-induced monocyte activation, we used siRNA to suppress endogenous IRE1α expression in isolated primary monocytes from patients with AS and observed significant reduction in the production of IL-1β and IL-23 (**Figure 4.9E**). The suppression of IRE1α protein expression by siRNA was confirmed using western blot (**Figure 4.10**). In addition to IL-1β and IL-23, CXCL2 was significantly enriched the T-cell induced cluster 3 (**Figure 4.4G**). We confirmed this finding at protein level using ELISA (**Figure 4.9F**) and found that *ERN1* siRNA also inhibited the secretion

of CXCL2 by T-cell-activated monocytes (**Figure 4.9G**). Lastly, we found 4u8C, a widely used inhibitor specific for IRE1α, significantly reduced the production of IL-1β, IL-23 and CXCL2 using the same assay (**Figure 4.11**). Taken together, these results confirm the activation of IRE1α /XBP1s pathway in T-cell stimulated monocytes and reveal a key role of IRE1α in the production of proinflammatory cytokines and chemokine in this setting.

Table 4.2 Enrichment of AS risk genes in monocyte clusters induced by *in vitro* stimulation

Cluster ID	Number of Maker genes	Number of overlapping risk genes	Enrichment Score	Adjusted p-value	Overlapping genes
IL23+CXCL2+ C3	177	9	1.773	0.032	CCRL2, CCR5AS, ERN1, NFKB1, ZC3H12C, ETS2, TNFSF15, IL27, IL6R
Calprotectin+ C0	26	1	1.341	0.540	SLC11A1
CCR7+ C4	63	2	1.107	0.558	ALDH2, IL7R
JAK2++ C2	141	4	0.989	0.610	JAK2, ITGAX, ZC3H12C, TNFSF15
LGALS9+ C5	110	2	0.634	0.859	LGALS9, ITGAX
CCL8+ C6	37	0	0.000	1.000	
CD16+ C1	66	0	0.000	1.000	

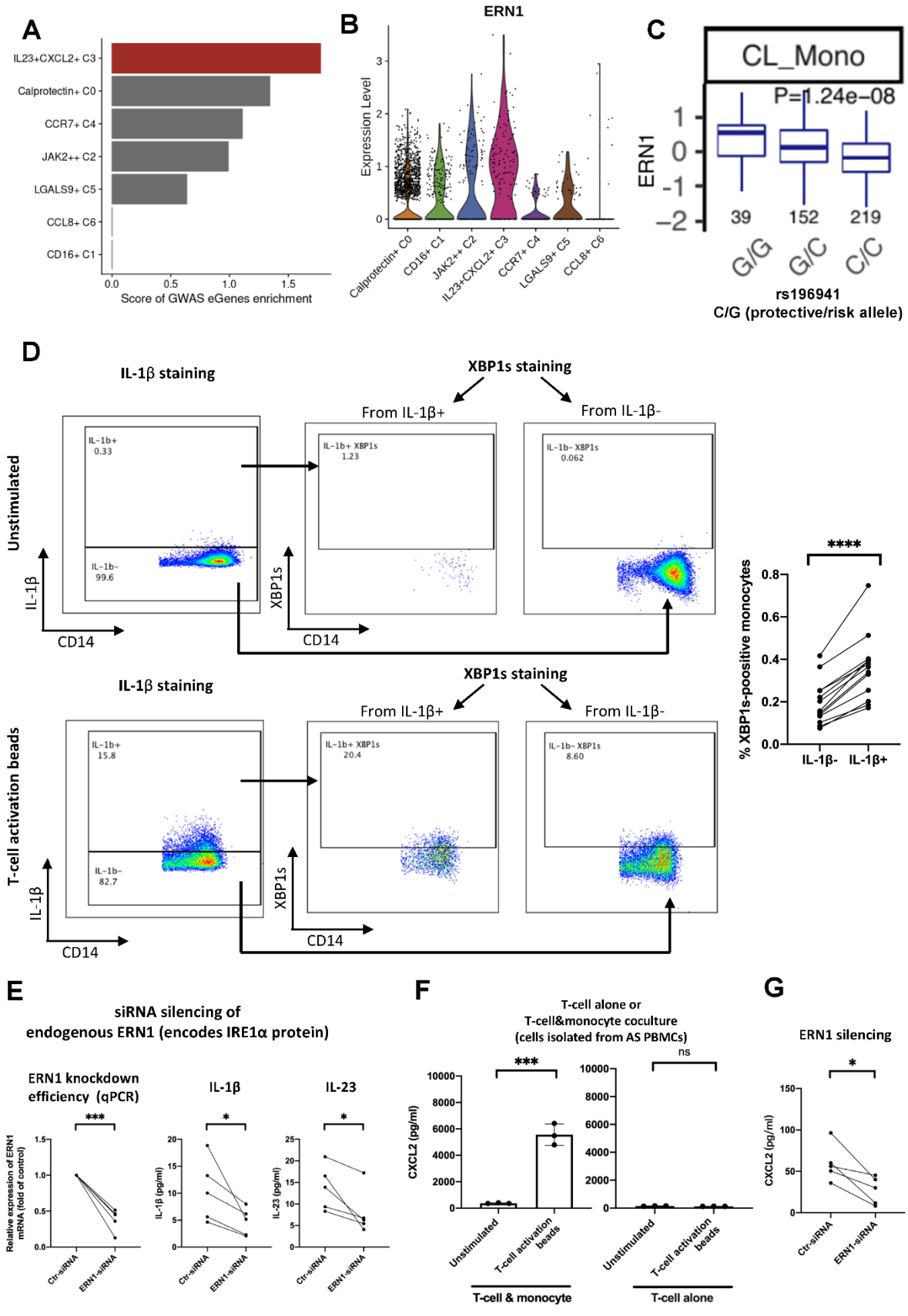

A
IL23+CXCL2+ C3
Calprotectin+ C0
CCR7+ C4
JAK2++ C2
LGALS9+ C5
CCL8+ C6
CD16+ C1
Score of GWAS eGenes enrichment
B
ERN1
Expression Level
Calprotectin+ C0
CD16+ C1
JAK2++ C2
IL23+CXCL2+ C3
CCR7+ C4
LGALS9+ C5
CCL8+ C6
C
CL_Mono
P=1.24e-08
ERN1
39
152
219
G/G
G/C
C/C
rs196941
C/G (protective/risk allele)
D
IL-1β staining
XBP1s staining
From IL-1β+
From IL-1β-
Unstimulated
IL-1b+ 0.33
IL-1b- 99.6
IL-1b+ XBP1s 1.23
IL-1b- XBP1s 0.062
IL-1β
CD14
XBP1s
T-cell activation beads
IL-1b+ 15.8
IL-1b- 82.7
IL-1b+ XBP1s 20.4
IL-1b- XBP1s 8.60
% XBP1s-poositive monocytes
****
IL-1β-
IL-1β+
E
siRNA silencing of
endogenous ERN1 (encodes IRE1α protein)
ERN1 knockdown
efficiency (qPCR)
IL-1β
IL-23
Relative expression of ERN1 mRNA (fold of control)
IL-1β (pg/ml)
IL-23 (pg/ml)
Ctr-siRNA
ERN1-siRNA
F
T-cell alone or
T-cell&monocyte coculture
(cells isolated from AS PBMCs)
CXCL2 (pg/ml)
***
ns
Unstimulated
T-cell activation beads
T-cell & monocyte
T-cell alone
G
ERN1 silencing
*

Figure 4.9. T-cell-induced *IL23A*+*CXCL2*+ monocytes express high levels of AS-associated risk genes and are regulated by the AS risk gene *ERN1*. (A) Enrichment of AS risk genes (identified through monocyte eQTL) in monocyte clusters induced by *in vitro* stimulation. (B) Expression of ERN1, an AS risk gene, in different monocyte subsets. (C) A higher expression level of *ERN1* in monocytes is associated with AS risk allele G at SNP rs196941. (D) PBMCs from AS patients (n=13) were stimulated with T-cell activation beads for 16 hours. Left: representative flow cytometry plots and gating strategy for XBP1s expression in IL-1β+ and - monocytes in response to stimulation. Right: a summary graph showing the percentage of XBP1s+ monocytes. This analysis is limited by the sensitivity of XBP1s detection, which makes it challenging for the gating strategy. Additional work using PCR analysis on cDNA can potentially further validate this finding (Unspliced XBP1 transcript is 26bp longer than the spliced form that encodes XBP1s). (E and G) Monocytes from AS patients (n=5) were transfected with non-targeting control or ERN1 siRNA and then co-cultured with autologous $CD3^+$ T cells and T-cell activation beads for 16 hours. The levels of cytokines IL-1β, IL-23 and CXCL2 in the culture supernatant were measured by ELISA. (F) CD3+ T-cells from AS patients (n=3) were cultured alone or co-cultured with monocytes from the same donors and stimulated with TAB for 16 hours. The level of CXCL2 in the culture supernatant was measured by ELISA.

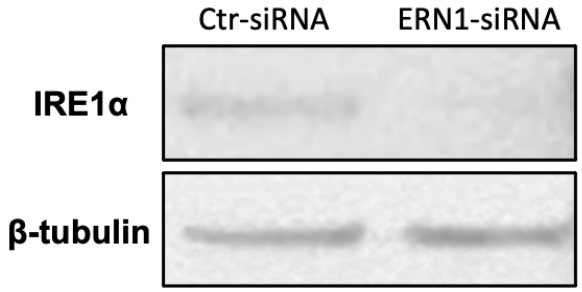


Figure 4.10. Confirmation of IRE1α suppression by siRNA at the protein level. Primary human monocytes were transfected using control non-targeting siRNA (Ctr-siRNA) or siRNA targeting ERN1 (ERN1-siRNA). 3 days later, cells were collected and lysed for western blot to measure the protein level IRE1α (encoded by ERN1 gene). Western blotting analysis of IRE1α (130 kDa) expression in monocytes； α-Tubulin (52 kDa) was used as an internal control.

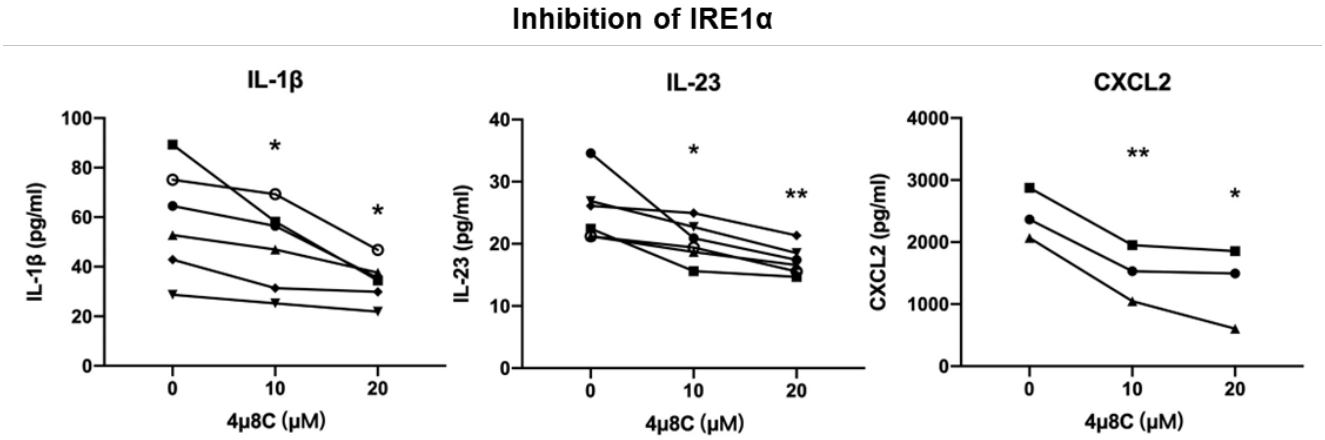


Figure 4.11. The inhibition of IRE1α suppresses the secretion of IL-1β, IL-23 and CXCL2 by T-cell-stimulated monocytes. Isolated CD14+ monocytes from AS patients were pre-treated with the IRE1α inhibitor 4μ8C (0, 10 and 20uM) for 17 h. The 4 μ8C was then washed away before the 4 h co-culture with autologous CD14-depleted PBMCs (for IL-1β and IL-23 experiment, n=6)/isolated T-cell (for CXCL2 experiment, n=3) and anti-CD2/CD3/CD28-coated beads. The levels of IL-1β, IL-23 and CXCL2 in the culture supernatant were measured by ELISA. The P-value was assessed using the paired two-tailed Student's t-test (* P≤ 0.05; ** P≤ 0.01; *** P≤ 0.001).

4. Discussion

In this study, we show that activated T-cells potently stimulate monocytes from patients with AS leading to the production of IL-1β and IL-23. These in turn act back on T cells to enhance Th17 responses. Using single-cell RNA sequencing (scRNA-seq), we find that T-cell activated monocytes include a subset of hyperinflammatory monocytes that is not induced by LPS activation and enriched for expression of AS risk genes including ERN1. Importantly, this monocyte subset is transcriptionally very similar to a population of CD14+ myeloid cells abundant in AS patients' synovial fluid (but not blood). Additionally, both genetic and chemical suppression of IRE1α, the protein encoded by the ERN1 gene, inhibits T-cell-instructed monocyte activation and production of IL-1β, IL-23 and CXCL2. We propose that this is a novel and important mechanism underlying monocyte activation in AS, advancing our knowledge of AS pathology and providing opportunities for the development of new drugs for this condition. Significantly, it is imperative to recognize that this mechanism may not be restricted solely to cells from AS patients. Further investigations employing samples from healthy donors could provide valuable insights in this regard. Additionally, it would be worthwhile to explore the impact of HLA-B27 on the responsiveness of monocytes to T-cell activation by conducting comparative analyses between HLA-B27-positive and HLA-B27-negative AS patients.

We observed that the presence of T cell-activated monocytes was as efficient as exogenous IL-1β and IL-23 in driving Th17 response in AS peripheral blood. This suggests that monocytes in AS have capacity to provide Th17-supporting cytokines in the absence of pathogen-associated molecular patterns (PAMPs) such as LPS. This mechanism of monocyte activation is likely more relevant to the sterile inflammation in synovium of patients with AS. In line with this, we found a CD14+ myeloid cell population present in synovial fluid that was highly similar to *IL23A+CXCL2+* monocyte cluster induced by T-cell activation (but not the Calprotectin+ monocyte cluster constituting the majority of cells stimulated by LPS). Additionally, in contrast to the lack of

evidence supporting the abundance of PAMPs in AS joints, we observed the enrichment of *PDCD1*+*IFNG*+ T-cells in synovial fluid from patients with AS. Our data thus suggest that T-cell-induced monocyte activation is occurring in inflamed AS joints and likely contributes to pathology. We also show that T cells activated 24 to 48 hours previously are capable of activating monocytes - thus they would have time to migrate to the joints following activation at other sites (e.g. gut). These data are consistent with models of joint or distant T cell activation with subsequent effector function in the joint. Indeed, we here provide evidence that T-cell-induced monocyte activation may indeed be a key part of this effector function.

Interestingly, around half of the activated T-cells we observed in AS SF were CD8+ supporting the previous studies proposing that antigen-experienced CD8+ T cells travel from gut to joint and then undergo local expansion[182,183]. In addition, a recent study reports the presence of HLA-B*27-restricted peptides that activate T cell receptors present in AS joint CD8+ T cells[184]. Our data along with these published findings imply a role of gut-derived HLA-B*27-restricted CD8+ T-cells in the activation of CD14+ myeloid cells in joints, which lead to the upregulation of proinflammatory cytokines such as IL-1β and IL-23. These Th17-supporting cytokines could be the upstream of IL-17 production in the synovium. Future work needs to be carried out to test this hypothesis.

Drug targets with genetic linkage are more likely to reach late-stage clinical trials[43,185]. We found that the *IL23A*+*CXCL2*+ monocyte subset induced by T-cell activation was significantly enriched for AS risk genes, including ERN1. Notably, silencing of ERN1 or inhibiting the function of its gene product (IRE1α) both reduced the production of IL-1β, IL-23 and CXCL2 by T-cell-activated monocytes from patients with AS. Thus, the inhibition of IRE1α or other protein(s) contributing to T-cell induced monocyte activation is an attractive approach for AS treatment.

In summary we here identify a novel pathway of T-cell-instructed monocyte activation that is likely present in AS and offers therapeutic potential.

Chapter 5: Integration of genetics and single-cell transcriptomics identified CXCR6+ Th17/Tph cells as a pathogenic subset in ankylosing spondylitis

1. Introduction

In recent years, the application of single-cell transcriptome profiling techniques has revolutionized the study of disease pathology. Through the use of cellular maps, researchers can identify unique cell subsets and cellular networks that may serve as potential pathogenic components in disease conditions. However, these approaches often generate lengthy candidate lists, as the initial drivers of disease may impose systemic impacts on the tissue. Consequently, it can be challenging to distinguish critical cell populations from those that are a consequence of abnormality resulting from disease progression.

To address these challenges, genetic analysis can potentially provide valuable insights, as genetic risk genes are more likely to take effect in key drivers of pathology. A few studies have explored the integration of genetics and single-cell transcriptomics and have revealed important insights in various diseases, including IBD, neurodegenerative diseases, and cardiovascular diseases[186–189]. More recently, researchers from the Alkes Price and Aviv Regev groups published two comprehensive frameworks for integrating genetics and single-cell transcriptomics, named scDRS and sc-linker, which enable the identification of disease-critical cell types and cellular processes[190,191]. These two methods first define a candidate risk gene set using information of genome proximity, eQTL, chromatin looping, epigenome modification, etc. Then they assess the enrichment of the expression of these candidate genes in the transcriptome at single-cell resolution/disease-enriched gene programs identified from scRNA-seq data.

Despite these advances, these methods have not been applied to the investigation of ankylosing spondylitis (AS) pathology. In this study, I used this strategy to interrogate the CD4+ T cells from AS joints, with an aim to identify the disease-critical subset of this heterogeneous population.

2. Results

2.1 Genetic risk scoring of CD4+ T cells at single cell level prioritized potential pathological populations

I first investigated the heterogeneity of CD4+ T cells in ankylosing spondylitis (AS) utilizing a single-cell transcriptional map of paired peripheral blood mononuclear cells (PBMCs) and synovial fluid mononuclear cells (SFMCs) obtained from AS patients (The scRNA-seq data were obtained through collaboration with Kim group[127], and it is the same data used in Chapter 3 and 4). A total of 12,629 T cells were characterized, derived from six donors (**Figure 4.7A and B**). Following dimensionality reduction and clustering, eight canonical T cell subsets were discerned, including naïve CD4+ T cells (characterized by *FHIT*+*ANXA2*-CD4+), memory helper T cells (characterized by *FHIT*-*ANXA2*-*FOXP3*-*CD4*+), regulatory T cells (Treg; *CD4*+*FOXP3*+), naïve CD8+ T cells (characterized by *FHIT*-*ANXA2*+*CD8A*+), memory CD8+ T cells (characterized by *FHIT*-*ANXA2*-*CD8A*+), mucosal-associated invariant T (MAIT) cells (characterized by *CD8A*+*KLRB1*^{high}, where KLRB1 encodes protein CD161), γδ T cells (*TRDC*+, encoding the constant region of the T cell receptor δ chain), and cycling cells (characterized by *MKI67*+) (**Figure 4.8A**).

I then looked for evidence of SF-specific T-cell activation (**Figure 4.7 C**) and found it to be present mostly in CD4+ and CD8+ T subsets. Despite the recent data showing the joint expansion of CD8+ T-cells which were trained in the gut[182], the pathogenic role(s) of CD8+ T-cells in AS has not been established. In contrast, multiple lines of evidence support the key role of CD4+ T-cells in AS, especially in the context of Th17 pathology (see detail in Chapter 1). Thus, I chose to focus on CD4+ T-cells. Naïve CD4+, memory Th, and Treg cells, along with bioinformatically isolated

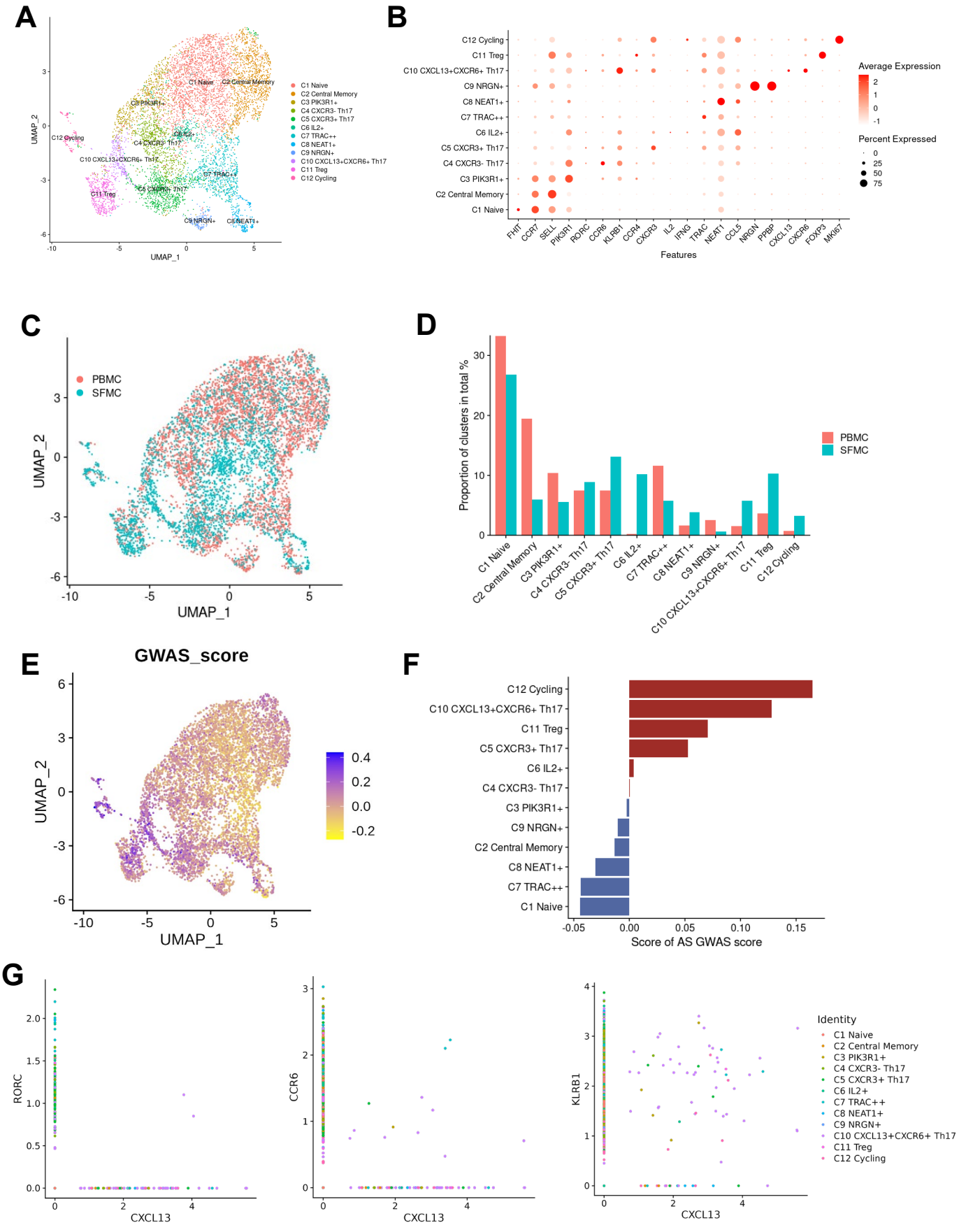
CD4+ cycling cells (CD4+ and CD8A-), were extracted and re-clustered into 12 transcriptionally distinct populations (**Figure 5.1 A, B**). These populations can be roughly classified into 1) naïve and central memory (C1, C2), 2) *RORC*+ effector (C4, C5, C6, C10), 3) *RORC*- effector (C3, C7, C8, C9), 4) Treg, 5) Cycling cells (marked by *MKI67* expression). Cells from synovial fluid (SF) and blood exhibited varying distributions across these clusters (**Figure 5.1 B**). Consistent with current knowledge, the *RORC*+ population constituted the majority of effector cells, indicating a local expansion of Th17 cells. IL17A was scarcely detected in this dataset, potentially due to stringent regulation of this cytokine. Previous studies have further subdivided Th17 cells into IL-17+IFNγ- "good" Th17, IL-17+IFNγ+ “pathogenic” Th17.1, and IL-17-IFNγ+RORγt^{low} ex-Th17[192]. Interestingly, despite the expression of *CXCR3*, the surface marker of Th17.1, C5 did not exhibit *IFNG* expression and thus did not completely fit the Th17.1 notion. This emphasizes the limitations of using surface markers to define Th17 subsets. Both C10 and C6 expressed *RORC* and could correspond to the Th17.1 subset and ex-Th17, respectively. C6 was exclusively present in joints (**Figure 5.1 B**) and exhibited high expression of *IL2* and *IFNG*, suggesting local activation. C10 was also highly enriched in synovial fluids and uniquely expressed *CXCR6*, a key marker for tissue residence. A recent study by Alexandra Schnell et al. from the Kuchroo and Regev groups proposed that the recruitment of pathogenic Th17 cells from the gut to inflammatory sites in a neuroinflammation model is mediated through CXCR6-CXCL16. Here this CXCR6+ Th17 population might contribute to AS pathology through this mechanism[123].

Multiple clusters with pro-inflammatory phenotypes were enriched in synovial fluid (SF), raising the question of which subset is more critical for disease development. To utilize genetics in aiding the identification of disease-critical subsets, I defined a set of 232 CD4-Tcells-specific ankylosing spondylitis (AS) genetic risk genes using published AS genome-wide association study (GWAS) data and CD4+ T cell expression quantitative trait locus (eQTL) data (see methods). Subsequently, I scored all CD4+ T cells based on the cumulative expression of this gene set (see methods)

(**Figure 5.1E**). The results reveal the enrichment of AS risk genes in cycling cells, *CXCR6*+ Th17 cells, regulatory T cells (Tregs), and *CXCR3*+ Th17 (Th17.1).

(See Figure in the next page)

Figure 5.1. GWAS scoring identified *CXCR6*+ Th17 cells as a pathogenic subset in AS. (A) UMAP visualization of transcriptionally distinct populations of CD4+ T cells from 6 pairs of blood and synovial fluid samples from AS donors. (B) Expression of a selected set of cluster markers by different populations of CD4+ T cells shown in Figure 5.1 A. (C) UMAP visualization of CD4+ T cells from blood and synovial fluids, colored by source of cells. (D) Frequency of cells from various clusters in PBMCs and SFMCs, respectively. (E) UMAP plot of Summed expression of GWAS candidate genes in monocytes stimulated by LPS or T cells. (F) Average GWAS scores of cell in different T-cells subsets identified from AS SFMCs and PBMCs, matching figure (E). (G) Co-expression of CXCL13 (Tph feature) with RORC, CCR6, and KLRB1 (Th17 feature).


A
UMAP_2
UMAP_1
C1 Naive
C2 Central Memory
C3 PIK3R1+
C4 CXCR3- Th17
C5 CXCR3+ Th17
C6 IL2+
C7 TRAC++
C8 NEAT1+
C9 NRGN+
C10 CXCL13+CXCR6+ Th17
C11 Treg
C12 Cycling
B
Average Expression
Percent Expressed
Features
FHIT
CCR7
SELL
PIK3R1
RORC
CCR6
KLRB1
CCR4
CXCR3
IL2
IFNG
TRAC
NEAT1
CCL5
NRGN
PPBP
CXCL13
CXCR6
FOXP3
MKI67
C
PBMC
SFMC
D
Proportion of clusters in total %
E
GWAS_score
F
Score of AS GWAS score
G
RORC
CCR6
KLRB1
CXCL13
Identity

2.2 Transcriptional features and lineage development of CXCR6+ Th17

I then focused on the *CXCR6*+ Th17 cells because this cluster has a high GWAS score and tissue-resident features but has not been reported or studied in AS. A particularly intriguing phenotype of these *CXCR6*+ Th17 cells is that they exhibit both peripheral helper T (Tph; characterized by *CXCL13*+*PDCD1*+*CXCR5*-) and Th17 features (**Figure 5.1 A and G**). Recent findings have revealed the critical role of Tph cells in rheumatoid arthritis (RA) pathology by promoting tertiary lymphoid structure formation. These cells were not RORC+ and shared TCR with effector CD4+ cells[193], implying that Tph found in RA tissue might have raised from effector CD4+. Thus, I asked if the *CXCR6*+ Th17 cells in AS SF could be derived from effector Th17 cells. To answer this question, I systematically examined the unique cellular program operating in these *CXCR6*+ Th17/Tph cells and compared their transcriptome with that of other CD4+ T cells. In total, 2,876 and 121 genes were found to be significantly upregulated and downregulated in this population, respectively (**Figure 5.2 B**), suggesting that activation rather than regulatory programs dominate their cellular activities. These cells highly express *GZMA*, which encodes granzyme A and typically functions in cytotoxicity programs. Nevertheless, the absence of expression of the classical markers *GZMB* (encodes granzyme B) and *GNLY* (encodes granulolysin) precluded classifying them as cytotoxic CD4+ T cells (**Figure 5.2 C**). Two key checkpoint molecules, *PDCD1* (encoding PD-1) and *LAG3*, are also enriched in this population, indicating that they might experience continuous antigen engagement (**Figure 5.2 B**). Another upregulated negative regulator is *LGALS1*, which encodes Galectin-1 and can induce T cell apoptosis (**Figure 5.2 B**). Conversely, some negative regulators are downregulated in these cells, such as *SOCS3* (**Figure 5.2 B**). SOCS3 can suppress STAT1 and STAT3 signaling, thereby reducing cytokine production in Th17 cells[194]. Its downregulation here might pave the way for IL-17 production. A more strikingly diminished cellular activity in these cells is the production of ribosomal proteins (**Figure 5.2 B**). Among the 121 significantly downregulated genes, 74 belong to the ribosomal protein family. This

is likely regulated by a metabolic regulator called c-MYC, which is among the top downregulated genes (**Figure 5.2 B**).

(see the figure in the next page)

Figure 5.2 Transcriptional feature of CXCR6+ Th17. (A) Expression of Th17 and Tph markers projected to UMAP of CD4+ T cells from AS blood and synovial fluids. (B) MA plot of differentially expressed genes in *CXCR6*+ Th17/Tph cells versus the rest (FDR <0.05). The average expression levels are shown for *CXCR6*+ Th17/Tph cells when the fold change is greater than 0, and for the others when the fold change is less than 0. (C) Expression of cytotoxic CD4+ T cell markers in different populations. (D) Expression of marker genes for “stem-like” and “pathogenic” Th17 cells in CD4+ T cells from AS and Th17 cells from central nervous system in mouse EAE model.

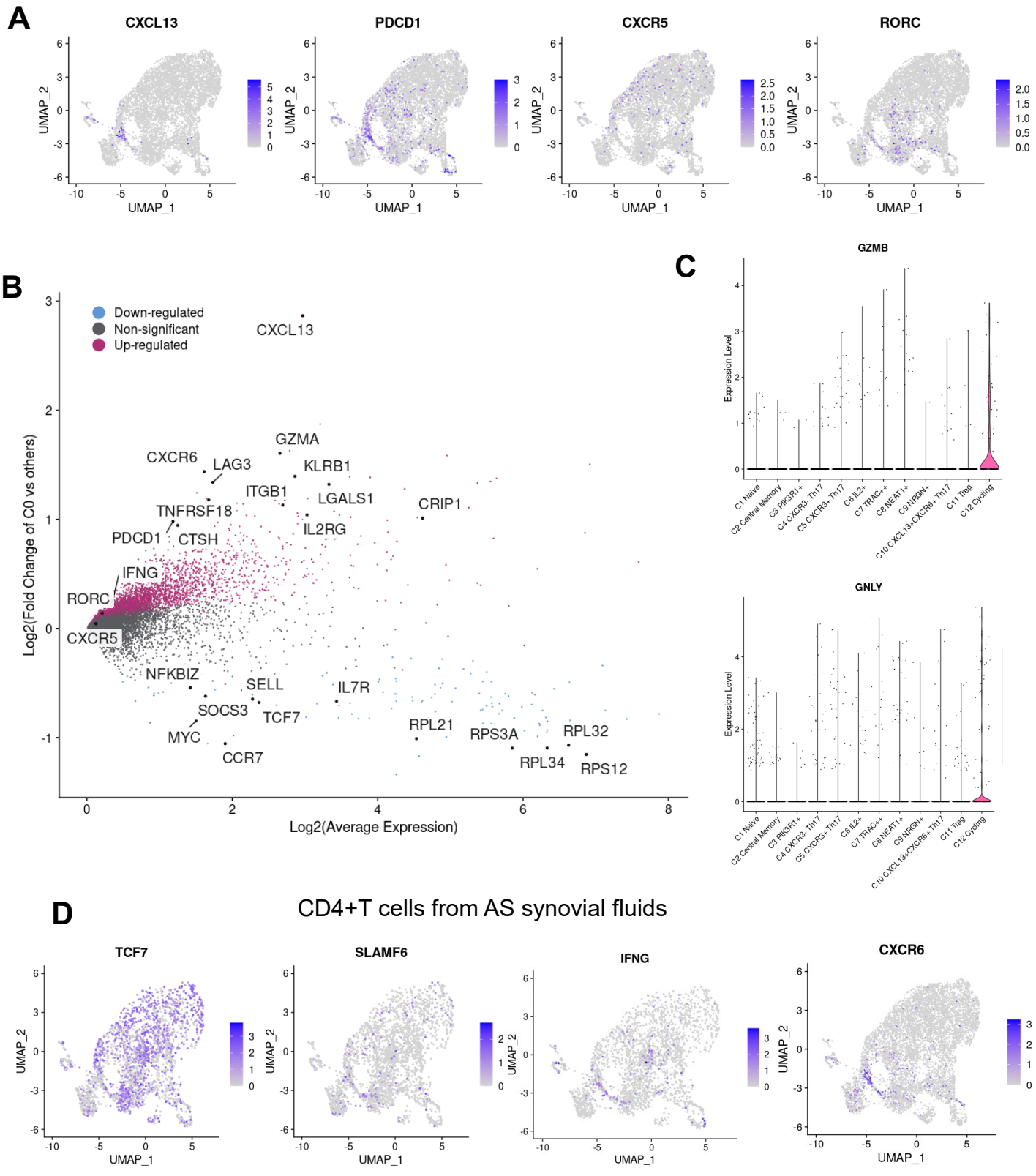


Figure 5.2 Transcriptional feature and lineage development of *CXCR6*+ Th17.

Subsequently, I sought to understand how this pathogenic Th17 subset was developed in ankylosing spondylitis (AS). Answering this question could elucidate the context and signaling that mediate their generation and provide insights for the development of new therapeutics. According to the model proposed by the Kuchroo group[123], the progenitors of these cells could be SLAMF6+TCF7+ stem-like Th17 (TCF7: transcription factor 7, which encodes a protein named TCF-1.). I noticed the presence of *TCF7*+ Th17 cells in AS joints, which suggests the presence of stem-like subsets (**Figure 5.2D**). However, SLAMF6 was not enriched in any specific population, which might be attributed to differences between humans and mice. TCF7 has been shown to functionally control the stemness of T cells in both humans and mice[195], making it a reliable marker for stem-like subsets in conjunction with the absence of IFN-γ production. Kuchroo group's study demonstrated the developmental relationship by showing the TCR repertoire sharing between the TCF7+ and CXCR6+ subsets and laboratory modelling. Here, I hypothesized this biology also stands true in AS joints, and investigated the developmental lineage of CXCR6+ Th17/Tph cells under this hypothesized scenario.

Both C4 *CXCR3*- Th17 and C5 *CXCR3*+ Th17 were *TCF*+ thus could be potential progenitors of *CXCR6*+ Th17 cells. I extracted cells from these three clusters from synovial fluid mononuclear cells (SFMCs) and utilized trajectory analysis to infer the developmental relationships among them. Following the dimensionality reduction (**Figure 5.3 A**), the Slingshot algorithm was employed to infer the developmental trajectory in two separate runs, where C4 and C5 were set as starting points (with no assigned endpoints) separately (**Figure 5.3 B and C**). Intriguingly, in either setting, *CXCR3*+ Th17 cells but not *CXCR3*- Th17 cells could give rise to *CXCR6*+ Th17 cells/Tph, suggesting *CXCR3*+ Th17 cells as an intermediate state between *CXCR3*- Th17 cells and *CXCR6*+ Th17/Tph. When setting C10 (*CXCR6*+ Th17) cells as endpoints (with no assigned starting points), the inferred trajectory aligned with the result using *CXCR3*- C4 cells as starting points (**Figure 5.3 D**). This further supported the C4->C5->C10 developmental lineage. Based on

this inferred trajectory (inferred using C10 as the ending cluster), I constructed a pseudo-time-course transcriptional dynamics map. A total of 7,478 genes were found to be significantly changed over this trajectory and were clustered into eight groups based on the similarity of their expression patterns (**Figure 5.3 E**). *RORC* exhibited limited dynamics over this trajectory and hence not shown in the graph, suggesting a robust Th17 identity feature throughout this development. These transcriptional activities could be divided into a series of waves. At the initial stage, IL-1 and IL-23 receptors were highly expressed, equipping cell ability to respond to stimulation by Th17-expanding cytokine (clusters 1 and 2) (**Figure 5.3 E**). Interestingly, in line with the work from the Kuchroo group's study, IL-23 receptor production diminished over development, while the IL-1 receptor exhibited only temporary downregulation. Clusters 7 and 8 might represent the second wave of transcriptional programs, where *JAK2* was upregulated, reflecting cytokine activation (**Figure 5.3 E**). This activation could lead to the acquisition of Th1 features in the third wave, such as the up-regulation of *EOMES*, *PVR*, and *CXCR3* (cluster 6) (**Figure 5.3 E**). Prior to the final stage of *IFNG* and *CXCL13* production (cluster 4), there was an intriguing stage marked by the up-regulation of *IL2*, *CD69*, and *ICOS* (cluster 3) (**Figure 5.3 E**). This implies that TCR activation, likely by antigen in the tissue, might be the signal that triggered Tph and tissue-resident phenotypes. This could also explain the strong checkpoint features (*PDCD1*, *LAG3*, *HAVCR2* [coding for Tim-3]), which are generally thought to arise after multiple rounds of antigen engagement (**Figure 5.3 E**). However, it must be noted that these observations were based on a bioinformatically inferred developmental trajectory, which requires thorough laboratory validation before being used for further hypothesis generation. A TCR sharing analysis could also test this theory while the single-cell experiments described in this work did not obtain high-quality TCR data.

(see the figure in the next page)

Figure 5.3 Developmental trajectory from *CXCR3-TCF+* Th17 to *CXCR6+* Th17/Tph. (A) UMAP visualization of C4, C5, and C10 after a new run of a dimensional reduction. (B) Developmental trajectory inferred by slingshot algorithm (C4 was set as starting point). (C) Developmental trajectory inferred by slingshot algorithm (C5 was set as starting point). (D) Developmental trajectory inferred by slingshot algorithm (C10 was set as the ending point). (E) Gene expression dynamics along the trajectory infer using C10 as ending cluster. Genes cluster into 8 gene sets, each characterized by specific expression profiles, as depicted by a selection of marker genes characteristic for each cluster (right).

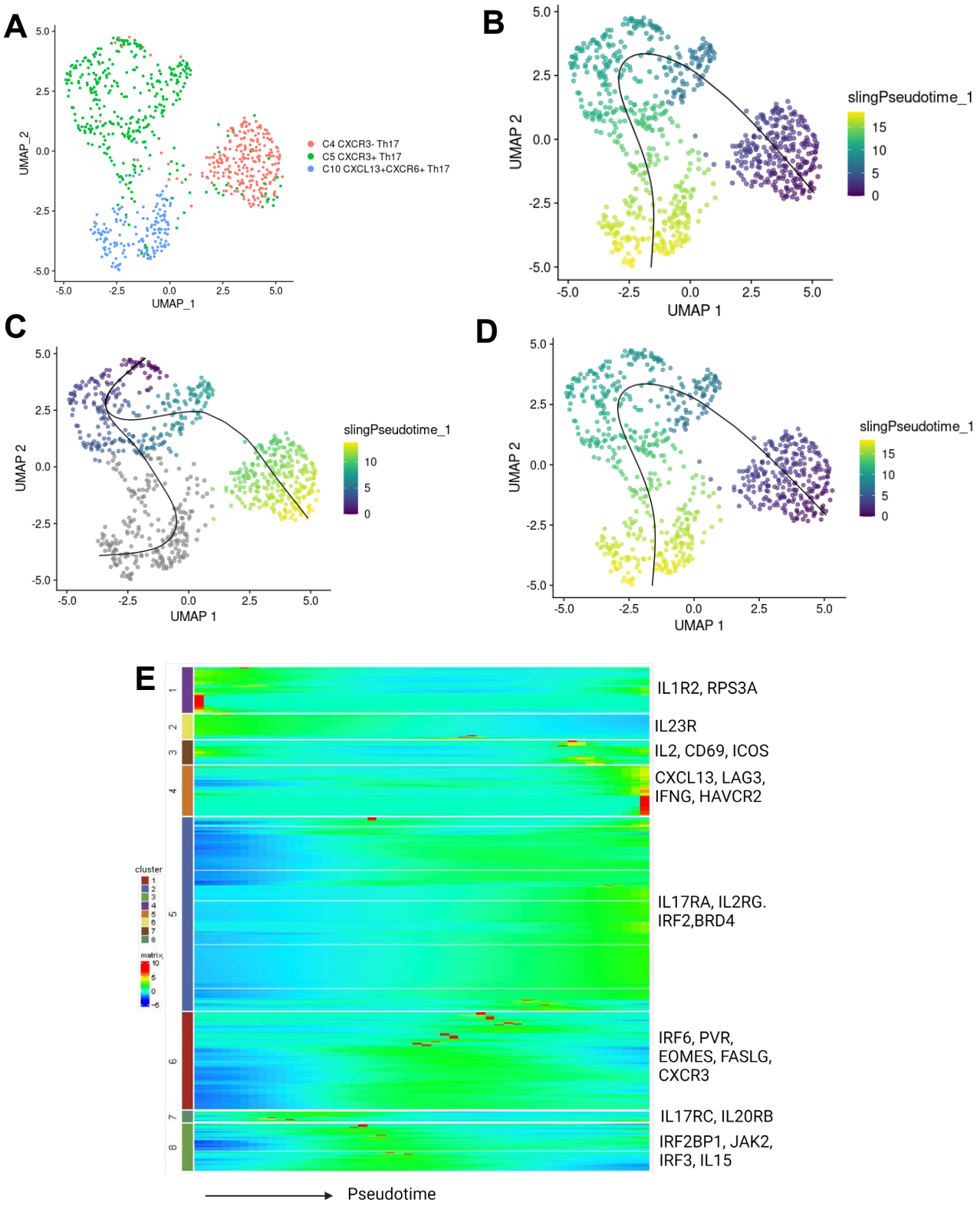


Figure 5.3 Developmental trajectory from CXCR3-TCF+ Th17 to CXCR6+ Th17/Tph.

2.3 Investigating the developmental route of Tph in different chronic immune responses

The analyses presented above suggest the presence of Tph in ankylosing spondylitis and Th17 cells as their precursors. This developmental route of Tph could be attributed to the type 3 immunity polarization environment in AS. I hypothesized that Tph could be generated from different types of memory helper T cells depending on the biological context. To examine this hypothesis, I analyzed the tissue data from patients with other clinical conditions.

Given the overlap in pathology between AS and inflammatory bowel disease (IBD), I first examined the single-cell RNA sequencing (scRNA-seq) profiles of immune cells in one type of IBD, Crohn's disease. I employed the 10x 5' scRNA-seq technique to profile ileum biopsies from two newly diagnosed Crohn's disease patients. After removing cells with low-quality transcriptional profiles, 18,643 cells were obtained, encompassing both immune (CD45+) and stromal (CD45-) compartments (**Figure 5.4 A and B**). Subsequently, 2,404 T cells were extracted and re-clustered into 13 transcriptionally distinct populations (**Figure 5.4 C, D, E, and F**). The CD4+ compartment consisted of Th17, Tfh, Treg, and cytotoxic populations. Both *CXCR3*- and *CXCR3*+ Th17 cells were identified. Unlike AS synovial fluids, *CXCR3*+ Th17 cells in Crohn's disease highly expressed IFNG. The cytotoxic CD4+ cells also exhibited Tph (*PDCD1*+*CXCR5*-*CXCL13*+) and Th17 (*RORC*+*IL17A*+) phenotypes. However, their strong cytotoxic features (*GZMB*+*GNLY*+) make it debatable whether they represent a subset similar to the *CXCR6*+ Th17/Tph identified in AS synovial fluids. Consequently, examination of the immune environment in Crohn's disease cannot provide conclusive evidence supporting the hypothesis regarding context-dependent Tph development. Additional work to profile more tissue samples and assess whether these *CXCL13*+ Th17 cells were derived from *CXCR3*+ Th17 cells or cytotoxic T cells are needed.

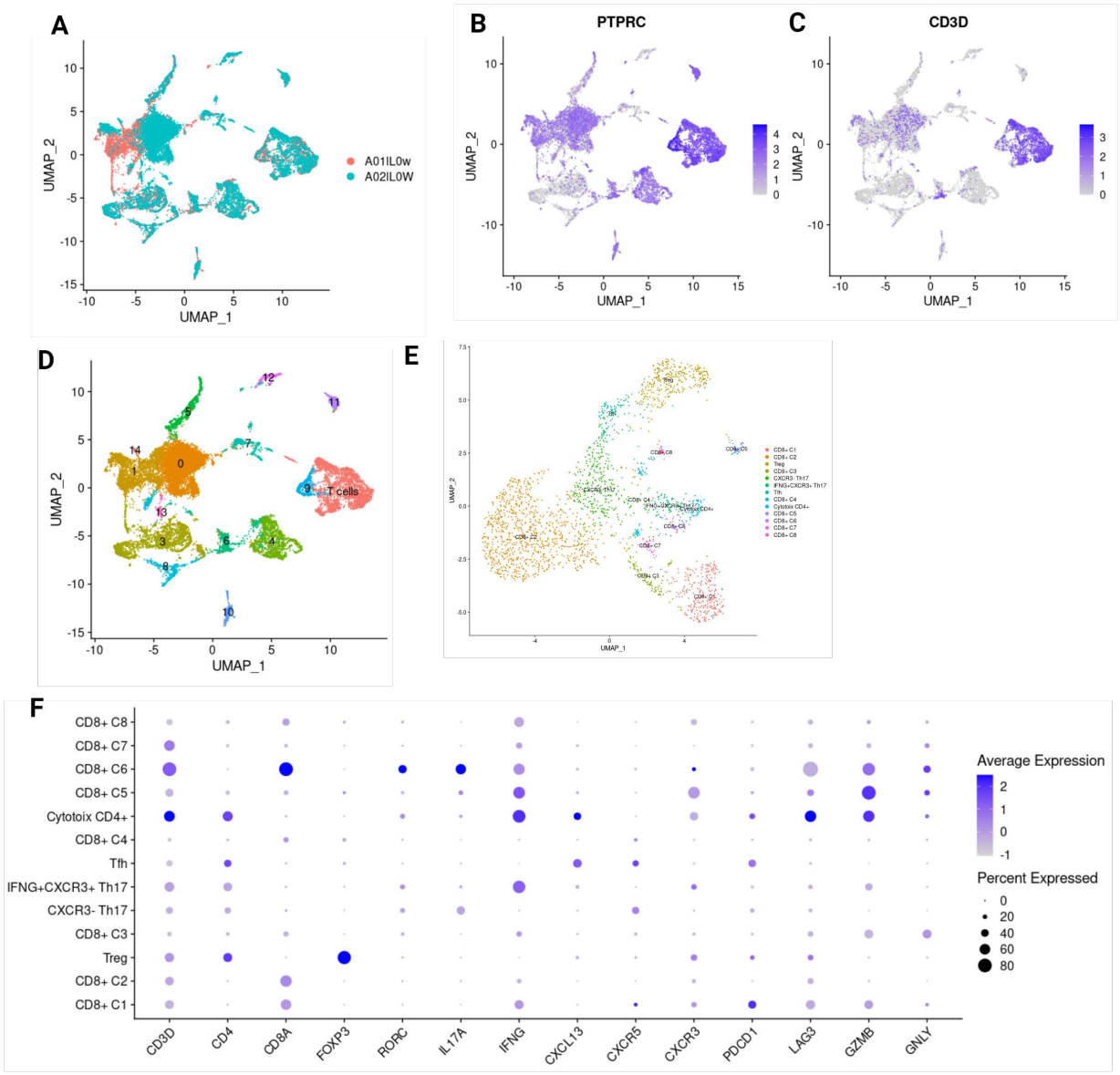
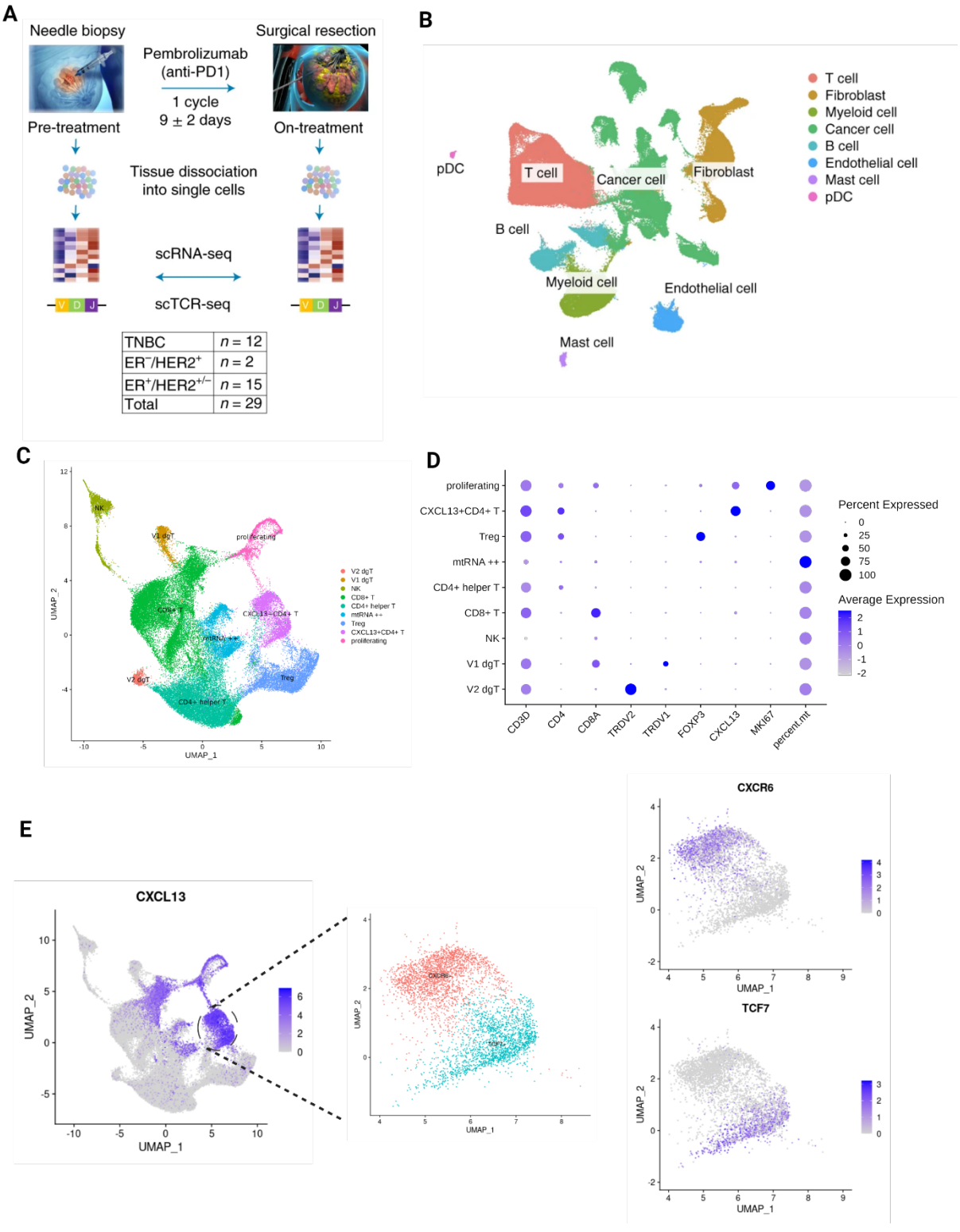


Figure 5.4 A scRNA-seq map of ileum biopsy in Crohn's disease. (A) UMAP visualization of cells from ileum biopsies from two Crohn's disease patients, colored by sample ID. (B) UMAP plot of *PTPRC* (encodes CD45) expression to distinguish immune and stromal compartments. (C) UMAP plot of *CD3D* expression to distinguish T cells. (D)UMAP visualization of cells and cell type/states annotation from ileum biopsies. (E) UMAP visualization of transcriptionally distinct populations of T cells from ileum biopsies after re-clustering. (F) Expression of a selected set of cluster markers by different populations of T cells shown in Figure 5.4 D.

Next, to examine Tph development in cancer, I utilized a scRNA-seq dataset of breast cancer published by Ayse Bassez and colleagues[196]. The authors obtained paired pre- and on-treatment biopsies from 29 treatment-naive patients receiving anti-PD1 therapy and applied scRNA/TCR sequencing to characterize the tissue cells (**Figure 5.5 A**). They classified 175,942 well-profiled cells into T cells, fibroblasts, myeloid cells, cancer cells, B cells, endothelial cells, mast cells, and pDCs (plasmacytoid dendritic cells) (**Figure 5.5 B**). Based on this annotation, I extracted the T cell compartment and re-clustered them into Vδ2+ γδ T cells, Vδ1+ γδ T cells, NK cells (mistakenly classified as T cells by the original authors), CD8+ T cells, helper T cells, Treg cells, proliferating T cells, *CXCL13*+*CD4*+ T cells, and an mtRNA^{high} population (**Figure 5.5 C and D**). Interestingly, these *CXCL13*+*CD4*+ T cells comprise two distinct subsets, which were exclusively marked with *CXCR6* and *TCF7* expression respectively (**Figure 5.5E**). This is similar to the stem-like-to-effector development scheme identified in AS. However, neither of these two populations expresses *RORC*. Instead, the *TCF7*+ population exhibited a classic Tfh feature marked by *CXCR5* and *BCL6* expression (**Figure 5.6A**). This suggests that the *CXCR6*+*CXCL13*+ subset is the main Tph subset in breast cancer and is likely derived from the Tfh population present in breast cancer. This developmental relationship was then supported by the substantial TCR repertoire sharing between the Tfh and Tph populations (**Figure 5.6B**). This observation, along with my findings in AS, supported the hypothesis that Tph could derive from various memory helper T cells depending on the context. In breast cancer, the Tfh cells that were activated by tumor antigen in the tissue seem to be the main progenitor of Tph. Thus, the tumor reactive T-cells could be most enriched in the Tph population. In line with this, I found that the pre-treatment frequencies of Tph, not Tfh, were associated with T cell expansion following PD-1 blockade.

(see the figure in the next page)

Figure 5.5 A single cell transcriptional map of *CXC13*+ T cells in breast cancer. (A) Design of sample collections (the figure is sourced from Ayse Bassez, et al.[196]) (B) UMAP visualization of cells from tumor tissue, colored by cell type annotation. (the figure is sourced from Ayse Bassez, et al.[196]) (C) UMAP visualization of transcriptionally distinct populations of T cells and NK cells from tumor tissues. (D) Expression of a selected set of marker genes by different populations of T cells and NK cells shown in Figure 5.5 C. (E) Left: UMAP plot of CXCL13 expression by T cells and NK cells from tumor tissues. Middle: UMAP visualization of CXCL13+ cell subsets annotation. Right: UMAP plot of marker genes expression by CXCL13+ cell subsets.


A
Needle biopsy
Surgical resection
Pembrolizumab
(anti-PD1)
1 cycle
9 ± 2 days
Pre-treatment
On-treatment
Tissue dissociation
into single cells
scRNA-seq
scTCR-seq
V D J
TNBC n = 12
ER⁻/HER2⁺ n = 2
ER⁺/HER2⁺/⁻ n = 15
Total n = 29
B
T cell
Fibroblast
Myeloid cell
Cancer cell
B cell
Endothelial cell
Mast cell
pDC
C
UMAP_1
UMAP_2
D
proliferating
CXCL13+CD4+ T
Treg
mtRNA ++
CD4+ helper T
CD8+ T
NK
V1 dgT
V2 dgT
CD3D
CD4
CD8A
TRDV2
TRDV1
FOXP3
CXCL13
MKI67
percent.mt
Percent Expressed
Average Expression
E
CXCL13
CXCR6
TCF7
UMAP_1
UMAP_2

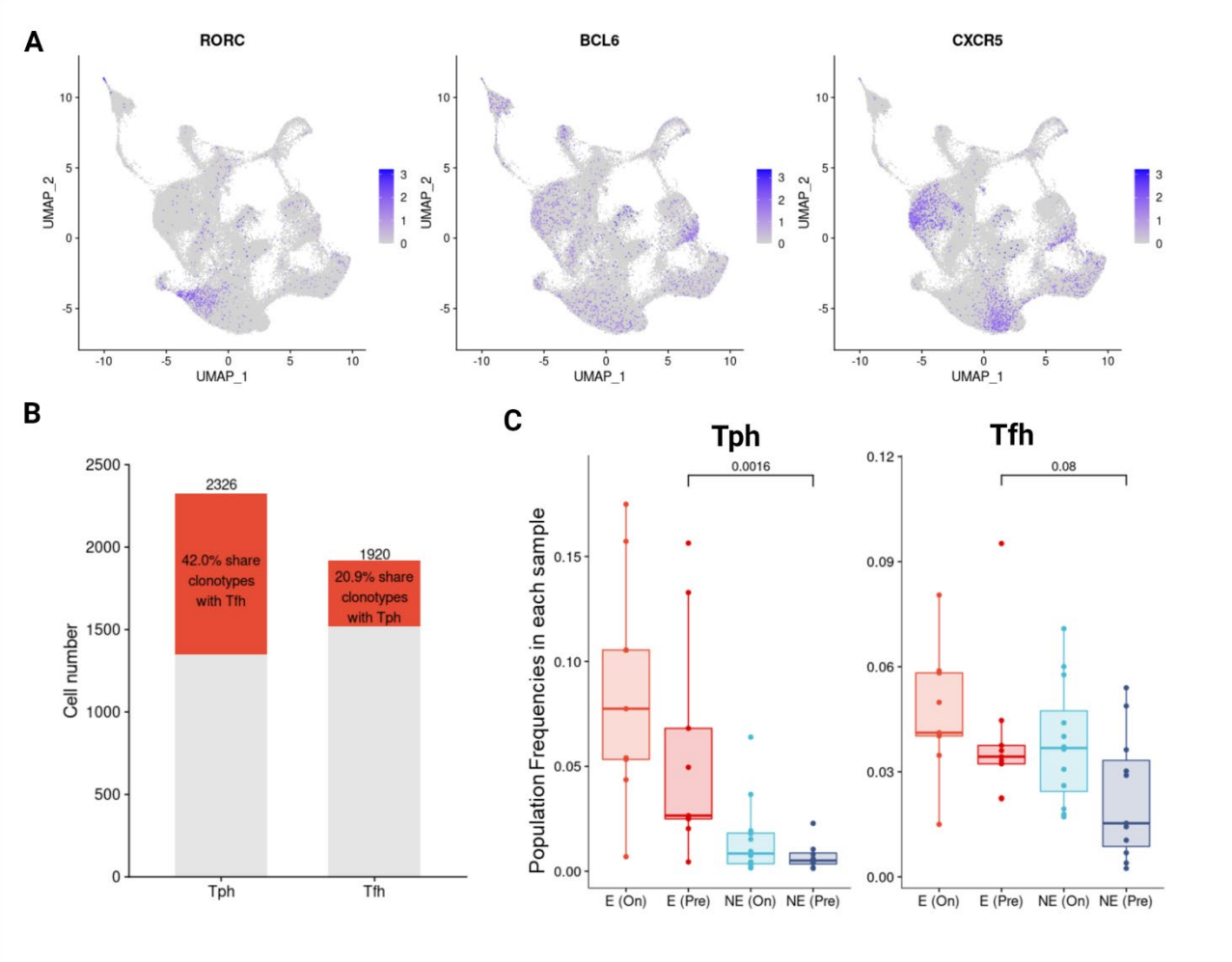


Figure 5.6 Tph cells were derived from Tph in breast cancer. (A) Expression of Th17 and Tph feature genes projected to UMAP of T cells from breast cancer tumor tissue. (B) TCR sharing between Tph and Tfh population. (C) Frequency of Tph and Tfh cells in whole T cells, comparing Es (patients with clonotype expansion) versus NEs (patients with limited or no clonotype expansion).

3. Discussion

In this Chapter, I employed a single-cell transcriptional map of T cells from AS synovial fluids and GWAS discoveries to identify pathogenetic cellular subsets. The analysis revealed a CXCR6+ Th17/Tph population exhibiting enriched expression of AS risk genes, which also demonstrated increased frequency in synovial fluids compared to blood. These cells displayed multiple pro-inflammatory phenotypes that have been previously implicated in either AS or other immune-mediated inflammatory diseases. Firstly, the expression of RORC suggests its capacity for IL-17 production, which has been established as a crucial pathogenic cytokine in AS. Secondly, its unique expression of *CXCR6* and *IFNG* linked it to the Th17 pathology model proposed by the Kuchroo group. These *CXCR6*+*IFNG*+ Th17 cells may represent a pathogenic subset derived from gut homeostatic and stem-like Th17 cells. Thirdly, these cells exhibited Tph features of being positive for *CXCL13* and *PDCD1* and negative for *CXCR5* expression. Tph has been recently implicated in several inflammatory conditions, including rheumatoid arthritis, ulcerative colitis, and systemic lupus erythematosus, as described in Chapter 1. No published studies have directly demonstrated the involvement of Tph in AS pathology, while the report of B cells infiltration at the inflamed joint of AS might imply this[197]. My study demonstrates the presence of activated Tph cells in AS joints, and these cells also exhibit a Th17 feature. Considering this together with the Kuchroo group's findings, I hypothesize that these Tph cells could derive from Th17 cells in AS. I inferred a developmental trajectory from the *CXCR3*-*TCF7*+ population to *CXCR3*+*TCF7*+ Th17 to *CXCR6*+ Th17/Tph and showed the transcription programs mediating this development. My analysis suggests that IL-1 and IL-23 might initiate this development, and strong antigen activation eventually triggers the gain of a Tph phenotype. This could imply a generalized mechanism of Tph generation in chronic diseases: Cytokines prime effector features, and then repetitive antigen engagement induces CXCL13. In this hypothesized scenario, the precursors of Tph could be diverse types of effector helper cells like Th1, Th2, Th17, and Tfh, depending on the local immune

response polarization signaling. I investigated cellular maps of T cells in Crohn's disease and breast cancer to examine this hypothesis. Substantial T cell receptor repertoire sharing between Tfh and Tph in breast cancer tumor tissues supported this model, while Tph was not able to be definitively defined in the ileum biopsies of Crohn's disease patients. My analysis also described the loss of *TCF7* along the development of Tph, which indicates a loss of stemness.

Although further validation is necessary, the pathogenic subset identified in this study through the integration of scRNA-seq mapping and genetic risk association can be explained based on its transcriptional features and cross-disease knowledge. In a separate study not described in this thesis, I found that RA GWAS risk genes exhibit enriched expression in sub-lining fibroblasts, but not lining ones, which aligns with the finding that sub-lining fibroblasts drive inflammation. Together with my results showed that the enrichment of AS risk genes in the pathogenic monocyte subset (Chapter 4), these findings support the value of the method used in my thesis. It is worth noting that the method used in this study may be oversimplified as it did not consider the varying levels of heritability contribution by different risk loci and the direction of gene expression altered by risk alleles. Alkes Price and Aviv Regev have developed frameworks that employ the local LD scoring method to partition heritability from summary statistics, which could potentially enhance prediction accuracy[190,191]. However, my work is constrained by the availability of GWAS summary statistics, and the direction of genetic risk effects has been demonstrated to be context-dependent, making it challenging to account for in this type of analysis. Overall, although a more stringent method is required, this newly developed strategy of utilizing single-cell profiling and risk gene association has demonstrated promise in aiding the identification of pathogenic populations.

The analysis of *CXCR6*+ Th17/Tph development in this study provides valuable insights into two key topics: Tph development and Th17 plasticity. The developmental route of Tph has not been extensively investigated, with most of the published work thus far coming from the Hiroyuki Yoshitomi group, who proposed that *CXCR5*-*CXCL13*+ T cells represent a distinct lineage from

Th1, Th2, Th17, Treg, and CXCR5+ Tfh cells[140,144]. Based on this theory, they explored the potential Tph differentiation condition and found the combination of TGFβ and IL-2 depletion can best give rise to Tph cells. However, my analysis suggests that Tph could represent a cell state and be derived from effector cells, such as Th17 in AS and Tfh in breast cancer. Further investigation is necessary to determine which scenario is more likely to be true, and ongoing work in my lab conducted by my colleagues is testing whether different types of purified effector cells can give rise to Tph cells. My study also provides further insight into how "good" and "stem-like" Th17 cells display plasticity and develop into a polyfunctional state in the inflamed joints of AS. Some *CXCR3*- Th17 cells, which are likely matured in the gut and recruited to the inflamed joints, harbor *TCF7* expression, indicating their capacity for self-renewal. IL-1β and IL-23 prime these cells to gain the Th1 or so-called “pathogenic” Th17 feature, and further TCR activation induces IFN-γ and CXCL13 production. Both cytokines and TCR stimulation have the potential to induce the expression of the key tissue-resident chemokine receptor CXCR6. This gain of Tph phenotype could lead to the formation of tertiary lymphoid structures (TLS) and facilitate autoantibody production, which might not be attenuated by IL-17A blockade therapy. Interpreting the results from synovial fluids is complicated, as there are other possibilities, such as the stem-like populations being leaked from blood rather than joint tissue, and only CXCR6+ cells locally contributing to the inflammation. Therefore, further investigation using patient synovium tissue would be necessary to assess the theory raised in this study.

In summary, this study identifies *CXCR6*+ Th17/Tph as a "pathogenic" cellular population in AS and infers *TCF*+ stem-like Th17 cells as their precursors. Consolidating these findings could potentially offer new therapeutic strategies for AS patients.

Chapter 6: Overall discussion and future directions

1. Overall discussion

There is a clear unmet need in the treatment of ankylosing spondylitis (AS), and this DPhil study aims to advance our understanding of AS pathology by investigating cross-disease immunopathology and integrating new techniques.

This study began by examining the shared genetic risk across AS and related diseases, with the goal of identifying therapeutic targets that explain fundamental immunopathology across conditions or provide drug repurposing opportunities. This inquiry led to the quantification of overlapping genetic risk among AS and 30 other conditions, providing evidence supporting shared genetic risks between IMIDs and opportunities of developing drug(s) for new targets and repurposing approved drug(s) to new indications. My data support the potential for repurposing Cyclosporine, which is currently used to treat inflammatory bowel disease, to AS by targeting the shared risk gene *PPIF*. The study provided genetics-supported arguments that repurposing/developing TYK2/JAK2 for AS may reduce safety concerns caused by current pan-JAK inhibitors. Moreover, using the *SH2B3* locus as an example, I showed that a single risk locus can directly affect multiple genes (multiple "causal" gene model). To consolidate the association between risk variants and gene expression, the study integrated data from multiple epigenetic profiling techniques to explore the genomic regulation in the selected two loci (*PPIF* and *SH2B3*). The study also identified transcriptional factors participating in the genomic regulation and examined risk gene expression in a single-cell transcriptional map of AS synovial fluids to infer the biological context where risk genes take effect.

Despite some progress, this exploratory journey during the 1st year of my DPhil has prompted me to review the strategy of starting from identifying interesting risk genes followed by mechanism studies of these genes for translational studies. eQTLs and chromatin looping data do provide

great insights into the cell types in which risk genes potentially exhibit their pathogenic functions. However, a major challenge beyond cell type is to identify the disease-relevant cellular context(s) for functional studies. For example, following the identification of a list of AS risk genes linked to CD4+ T cells by eQTL datasets, I found it very challenging to design assays to discover novel biology beyond the well-established Th17 polarization. A genetic screen of these risk genes using siRNA or CRISPR in a Th17 differentiation/expansion assay might indeed generate some new knowledge in this topic, but will unlikely unveil fundamental unknown immune responses involved in AS inflammation. Therefore, after the DPhil transfer, I decided to move my research focus to the discovery of novel disease-relevant cellular program(s) and use risk genes to assist this task and help the identification of potential regulators.

The second project is an implementation of the aforementioned idea. In this section, I aimed to test the possibility that T cell activation can be upstream of monocyte activation in a "germ-free" environment like AS joints. Through my experimentation, I was able to prove that this is indeed true and that T cell-activated monocytes can support Th17 response in PBMCs from AS patients by producing IL-1β and IL-23. Single-cell transcriptional profiling of this cellular program revealed that not all monocytes respond equally to this activation, with only a small population expressing IL-23, and these cells being transcriptionally distinct from others. Interestingly, I found that the transcriptional phenotypes of this population generated *in vitro* highly resemble a hyper-inflammatory CD14+ myeloid subset present in AS synovial fluids. These data strongly suggest that T-cell-induced monocyte activation in AS synovium represents a novel and important cellular program involved in AS pathology. Thus, the "novel disease-relevant cellular program" was identified, and I then brought in genetics to look for potential therapeutic targets that can modulate this program. I firstly asked if any subsets of T-cell-instructed monocytes are enriched for AS risk genes and found that only the IL-23+ cells, the subset phenocopies CD14+ myeloid cells in AS SF, reached the significancy. Encouraged by the genetic support, I then sought to identify a

regulator leading to the generation of IL23A+ monocytes. Among the genes shared between the monocyte-specific risk gene and the marker gene of this IL23A+ subset, *ERN1* caught my attention due to the well-known role of its protein product IRE1-α as the sensor of ER stress. ER stress was previously associated with AS pathology in the context of HLA-B27 misfolding, and I hypothesized that it could also contribute to T cell-instructed monocyte activation. My colleague, Jinyi Zhao, utilized both siRNA knockdown and chemical inhibitor to perturbate IRE1-α and demonstrated that suppressing *ERN1* in monocytes can indeed reduce T-cell-induced IL-23, IL-1β and CXCL2 production. This work provides a possible explanation for how HLA-B27 enhances immune response through ER stress. A recent study has illustrated that endoplasmic reticulum (ER) stress in intestinal epithelial cells can orchestrate a cellular program conducive to the development of Th17 cells[198]. Taken together, these findings suggest that HLA-B27-induced ER stress, if stands true, has the potential to modify the immune program through diverse mechanisms. This project nicely supports the rationale that integrating single-cell transcriptional mapping and genetic association discoveries can be an effective strategy for the identification of novel pathology.

Subsequently, I modified this method and applied it to identify CD4+ T cell subsets with enriched risk gene expression as candidate pathogenic cell types/states. By measuring the summed expression of the risk gene at a single-cell level, I discovered a polyfunctional RORC+CXCR6+IFN-γ+CXCL13+ CD4 subset that is highly enriched for AS risk genes. It appeared that two major CD4+ T cell-mediated pathogenic programs converged on this population. Based on the transcriptional similarity between this subset and others, as well as current knowledge, I inferred *TCF7*+ Th17 cells as the precursor of these cells. This is consistent with the stem-like to effector Th17 conversion model proposed by the Kuchroo group using in model neuroinflammation model (EAE). For a long time, the mechanism behind how Th17 cells maintain a homeostatic state while also developing into pathogenic states mediating chronic

inflammation has been unclear. The Kuchroo group proposed that the gut maintains a *TCF7*+ IFN-γ- stem-like and homeostatic Th17 population, which could give rise to a CXCR6+IFN-γ+ population in the spleen. These converted "angry" Th17 cells can be recruited to inflamed brain tissues and contribute to the disease. These cells are highly similar to the subset of CD4+ T cells that I found in AS SF supporting a pathogenic role of the latter population in AS. We are now developing a cellular assay to generate these cells *in vitro* for a CRISPR screen to identify potential regulator(s).

The observation that these *RORC*+*CXCR6*+ cells have features of Tph (*PDCD1*+*CXCL13*+*CXCR5*-) led me to consider the possibility that Tph could be derived from effector cells rather than naïve CD4+ T-cells. My examination of Tph in breast cancer supports this theory, demonstrating substantial overlapping of T cell receptor repertoire between the Tph and Tfh population. The formation of tertiary lymphoid structures has been increasingly recognized as a key pathological program, as this causes tissue structure damage and intensive immune cell infiltration. Recent studies in multiple inflammatory conditions and cancer highlight the role of Tph in this process. My findings implicate Tph in AS pathology and propose a possible route for its generation.

2. Future directions

Mechanistic studies of T cell-instructed monocyte activation

This study did not investigate the specific T cell signals responsible for monocyte activation. A future study could utilize a genome-wide CRISPR loss-of-function pooled screening in monocytes under T cell activation to systematically screen for molecules involved in this program. A cross-cell-type pooled screening approach, where genomic perturbation is introduced in T cells and the read-out is observed in monocytes, would also provide information on the regulators from the T

cell perspective. Complementing this work with an investigation of T cell-monocyte interactions in AS-inflamed tissue could further inform the *in vivo* effects of identified regulators.

Examining the presence and developmental route Tph/Th17 in AS joint tissue

This study is limited by the fact that the analysis was performed on synovial fluid cells, which may not necessarily be representative of the cells present in the inflamed tissue. A transcriptional map of T cells in AS joints would be necessary to confirm the presence and phenotypes of Tph and its inferred precursors at the inflamed sites. Additionally, conducting parallel single-cell TCR characterization could potentially confirm if Tph is truly derived from stem-like Th17 cells in AS joints. A cross-disease investigation of Tph development could also provide further insight. For instance, examining whether Tph in asthma derives from Th2 could offer additional clues.

Systemic screen of pathogenic cellular populations in AS through integrating genetics and scRNA-seq

This study provides support for the utility of an integrated approach utilizing genetics and single-cell RNA sequencing in both novel biological discovery and identification of potential therapeutic targets. This methodology may be extended to other cell types for which functional genomics data is available, which will provide a comprehensive pathogenic cell interacting maps. Moreover, generating cellular maps in tissue for this study could lead to more conclusive findings.

REFERENCES

1. Ashrafi, M., Ermann, J. & Weisman, M. H. Spondyloarthritis evolution: what is in your history? *Curr Opin Rheumatol* **32**, 321–329 (2020).

2. Stolwijk, C., Onna, M. van, Boonen, A. & Tubergen, A. van. Global Prevalence of Spondyloarthritis: A Systematic Review and Meta-Regression Analysis. *Arthrit Care Res* **68**, 1320–1331 (2016).

3. Crossfield, S. S. R., Marzo-Ortega, H., Kingsbury, S. R., Pujades-Rodriguez, M. & Conaghan, P. G. Changes in ankylosing spondylitis incidence, prevalence and time to diagnosis over two decades. *Rmd Open* **7**, e001888 (2021).

4. Sieper, J. & Heijde, D. van der. Review: Nonradiographic axial spondyloarthritis: New definition of an old disease? *Arthritis Rheumatism* **65**, 543–551 (2013).

5. Ramiro, S. *et al.* ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. *Ann Rheum Dis* **82**, 19–34 (2023).

6. Rademacher, J., Poddubnyy, D. & Pleyer, U. Uveitis in spondyloarthritis. *Ther Adv Musculoskelet Dis* **12**, 1759720X20951733 (2020).

7. Stolwijk, C. *et al.* The epidemiology of extra-articular manifestations in ankylosing spondylitis: a population-based matched cohort study. *Ann Rheum Dis* **74**, 1373 (2015).

8. Orlando, A., Renna, S., Perricone, G. & Cottone, M. Gastrointestinal lesions associated with spondyloarthropathies. *World J Gastroentero* **15**, 2443–2448 (2009).

9. Leirisalo-Repo, M., Turunen, U., Stenman, S., Helenius, P. & Seppälä, K. High Frequency of Silent Inflammatory Bowel Disease in Spondylarthropathy. *Arthritis Rheumatism* **37**, 23–31 (1994).

10. Cypers, H. *et al.* Elevated calprotectin levels reveal bowel inflammation in spondyloarthritis. *Ann. Rheum. Dis.* **75**, 1357 (2016).

11. Rudwaleit, M. *et al.* The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann. Rheum. Dis.* **68**, 777 (2009).

12. Linden, S. V. D., Valkenburg, H. A. & Cats, A. Evaluation of Diagnostic Criteria for Ankylosing Spondylitis. *Arthritis Rheum.* **27**, 361–368 (1984).

13. Danve, A. & Deodhar, A. Treatment of axial spondyloarthritis: an update. *Nat Rev Rheumatol* **18**, 205–216 (2022).

14. Zochling, J., Bohl-Bühler, M. H. J., Baraliakos, X., Feldtkeller, E. & Braun, J. Nonsteroidal anti-inflammatory drug use in ankylosing spondylitis—a population-based survey. *Clin Rheumatol* **25**, 794–800 (2006).

15. Deodhar, A. & Yu, D. Switching tumor necrosis factor inhibitors in the treatment of axial spondyloarthritis. *Semin Arthritis Rheu* **47**, 343–350 (2017).

16. Heijde, D. van der *et al.* Ixekizumab, an interleukin-17A antagonist in the treatment of ankylosing spondylitis or radiographic axial spondyloarthritis in patients previously untreated with biological disease-modifying anti-rheumatic drugs (COAST-V): 16 week results of a phase 3 randomised, double-blind, active-controlled and placebo-controlled trial. *Lancet* **392**, 2441–2451 (2018).

17. FDA. FDA requires warnings about increased risk of serious heart-related events, cancer, blood clots, and death for JAK inhibitors that treat certain chronic inflammatory conditions. https://www.fda.gov/drugs/fda-drug-safety-podcasts/fda-requires-warnings-about-increased-risk-serious-heart-related-events-cancer-blood-clots-and-death (2022).

18. Lukas, C. *et al.* Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* **68**, 18 (2009).

19. Blécourt, J. J. de, Polman, A., Blécourt-Meindersma, T. de, Erlee, T. J. D. & Drion, E. F. Hereditary Factors in Rheumatoid Arthritis and Ankylosing Spondylitis. *Ann Rheum Dis* **20**, 215 (1961).

20. Brown, M. A., Laval, S. H., Brophy, S. & Calin, A. Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. *Ann Rheum Dis* **59**, 883 (2000).

21. Brown, M. A. *et al.* Susceptibility to ankylosing spondylitis in twins the role of genes, HLA, and the environment. *Arthritis Rheumatism* **40**, 1823–1828 (1997).

22. Reveille, J. D., Hirsch, R., Dillon, C. F., Carroll, M. D. & Weisman, M. H. The prevalence of HLA–B27 in the US: Data from the US National Health and Nutrition Examination Survey, 2009. *Arthritis Rheumatism* **64**, 1407–1411 (2012).

23. Bowness, P. Hla-B27. *Annu Rev Immunol* **33**, 29–48 (2015).

24. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

25. Azzouz, D. *et al.* Lupus nephritis is linked to disease-activity associated expansions and immunity to a gut commensal. *Ann Rheum Dis* **78**, 947 (2019).

26. Routy, B. *et al.* Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).

27. Costello, M. *et al.* Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis. *Arthritis Rheumatol* **67**, 686–691 (2015).

28. Asquith, M. *et al.* HLA Alleles Associated With Risk of Ankylosing Spondylitis and Rheumatoid Arthritis Influence the Gut Microbiome. *Arthritis Rheumatol* **71**, 1642–1650 (2019).

29. McKiernan, S. M. *et al.* Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* **40**, 108–114 (2004).

30. Kaslow, R. A. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV–1 infection. *Nat Med* **2**, 405–411 (1996).

31. Evans, D. M. *et al.* Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* **43**, 761–767 (2011).

32. Linden, S. M. D. V., Valkenburg, H. A., Jongh, B. M. D. & Cats, A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. *Arthritis Rheumatism* **27**, 241–249 (1984).

33. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* **48**, 510–8 (2016).

34. Risch, N. & Merikangas, K. The Future of Genetic Studies of Complex Human Diseases. *Science* **273**, 1516–1517 (1996).

35. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *Am J Hum Genetics* **110**, 179–194 (2023).

36. Klein, R. J. *et al.* Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science* **308**, 385–389 (2005).

37. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

38. Loos, R. J. F. 15 years of genome-wide association studies and no signs of slowing down. *Nat Commun* **11**, 5900 (2020).

39. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).

40. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412–1425 (2018).

41. David, T., Ling, S. F. & Barton, A. Genetics of immune-mediated inflammatory diseases. *Clin Exp Immunol* **193**, 3–12 (2018).

42. Fang, H., Chen, L. & Knight, J. C. From genome-wide association studies to rational drug target prioritisation in inflammatory arthritis. *Lancet Rheumatology* **2**, e50–e62 (2020).

43. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856–860 (2015).

44. Ochoa, D. *et al.* Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nat Rev Drug Discov* **21**, 551–551 (2022).

45. Bianchi, E. & Rogge, L. The IL-23/IL-17 pathway in human chronic inflammatory diseases—new insight from genetics and targeted therapies. *Genes Immun* **20**, 415–425 (2019).

46. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *New Engl J Medicine* **373**, 895–907 (2015).

47. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* **39**, 207–211 (2007).

48. Murthy, A. *et al.* A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature* **506**, 456–462 (2014).

49. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum Mol Genet* (2018) doi:10.1093/hmg/ddy314.

50. Ge, D. *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).

51. Suppiah, V. *et al.* IL28B is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. *Nat Genet* **41**, 1100–1104 (2009).

52. Tanaka, Y. *et al.* Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. *Nat Genet* **41**, 1105–1109 (2009).

53. Muir, A. J. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for IFNL3 (IL28B) Genotype and PEG Interferon-α–Based Regimens. *Clin Pharmacol Ther* **95**, 141–146 (2014).

54. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genetics* **100**, 635–649 (2017).

55. Burton, P. R. *et al.* Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* **39**, 1329–1337 (2007).

56. (TASC), A.-A.-A. S. C. *et al.* Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat Genet* **42**, 123–127 (2010).

57. (IGAS), I. G. of A. S. C. *et al.* Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* **45**, 730–738 (2013).

58. Brown, M. A. & Wordsworth, B. P. Genetics in ankylosing spondylitis – Current state of the art and translation into clinical outcomes. *Best Pract Res Clin Rheumatology* **31**, 763–776 (2017).

59. Cargill, M. *et al.* A Large-Scale Genetic Association Study Confirms IL12B and Leads to the Identification of IL23R as Psoriasis-Risk Genes. *Am J Hum Genetics* **80**, 273–290 (2007).

60. Duerr, R. H. *et al.* A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* **314**, 1461–1463 (2006).

61. Meglio, P. D. *et al.* The IL23R A/Gln381 Allele Promotes IL-23 Unresponsiveness in Human Memory T-Helper 17 Cells and Impairs Th17 Responses in Psoriasis Patients. *J Invest Dermatol* **133**, 2381–2389 (2013).

62. Bowes, J. *et al.* Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis. *Nat Commun* **6**, 6046 (2015).

63. Hanson, A. L. *et al.* Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile. *Arthritis Rheumatol* **70**, 255–265 (2018).

64. Tran, T. M. *et al.* Paradoxical Effects of Endoplasmic Reticulum Aminopeptidase 1 Deficiency on HLA–B27 and Its Role as an Epistatic Modifier in Experimental Spondyloarthritis. *Arthritis Rheumatol* **75**, 220–231 (2023).

65. Khare, S. D., Hansen, J., Luthra, H. S. & David, C. S. HLA-B27 heavy chains contribute to spontaneous inflammatory disease in B27/human beta2-microglobulin (beta2m) double transgenic mice with disrupted mouse beta2m. *J Clin Invest* **98**, 2746–2755 (1996).

66. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).

67. Lange, K. M. de *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256–261 (2017).

68. Nasrallah, R. *et al.* A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by Treg cells. *Nature* **583**, 447–452 (2020).

69. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* (2021) doi:10.1016/j.cell.2021.03.056.

70. Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet* **51**, 1494–1505 (2019).

71. Consortium, T. Gte. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

72. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).

73. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e16 (2018).

74. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).

75. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

76. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119 (2015).

77. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* **78**, 554-565.e7 (2020).

78. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell* **78**, 539-553.e8 (2020).

79. Fullwood, M. J. *et al.* An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).

80. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611–1627 (2015).

81. Wang, P. *et al.* In situ Chromatin Interaction Analysis Using Paired-End Tag Sequencing. *Curr Protoc* **1**, e174 (2021).

82. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919–922 (2016).

83. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).

84. Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* **49**, 1602–1612 (2017).

85. Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet* **53**, 110–119 (2021).

86. Annunziato, F., Romagnani, C. & Romagnani, S. The 3 major types of innate and adaptive cell-mediated effector immunity. *J Allergy Clin Immun* **135**, 626–635 (2015).

87. Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J Exp Medicine* **201**, 233–240 (2005).

88. Park, H. *et al.* A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* **6**, 1133–1141 (2005).

89. Harrington, L. E. *et al.* Interleukin 17–producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat Immunol* **6**, 1123–1132 (2005).

90. Rouvier, E., Luciani, M. F., Mattéi, M. G., Denizot, F. & Golstein, P. CTLA-8, cloned from an activated T cell, bearing AU-rich messenger RNA instability sequences, and homologous to a herpesvirus saimiri gene. *J Immunol Baltim Md 1950* **150**, 5445–56 (1993).

91. Yao, Z. *et al.* Herpesvirus Saimiri encodes a new cytokine, IL-17, which binds to a novel cytokine receptor. *Immunity* **3**, 811–821 (1995).

92. Yao, Z. *et al.* Human IL-17: a novel cytokine derived from T cells. *J Immunol Baltim Md 1950* **155**, 5483–6 (1995).

93. Fossiez, F. *et al.* T cell interleukin-17 induces stromal cells to produce proinflammatory and hematopoietic cytokines. *J Exp Medicine* **183**, 2593–2603 (1996).

94. Chabaud, M., Fossiez, F., Taupin, J. L. & Miossec, P. Enhancing effect of IL-17 on IL-1-induced IL-6 and leukemia inhibitory factor production by rheumatoid arthritis synoviocytes and its regulation by Th2 cytokines. *J Immunol Baltim Md 1950* **161**, 409–14 (1998).

95. Jovanovic, D. V. *et al.* IL-17 Stimulates the Production and Expression of Proinflammatory Cytokines, IL-β and TNF-α, by Human Macrophages. *J Immunol* **160**, 3513–3521 (1998).

96. Teunissen, M. B. M., Bos, J. D., Koomen, C. W., Malefyt, R. de W. & Wierenga, E. A. Interleukin-17 and Interferon-γ Synergize in the Enhancement of Proinflammatory Cytokine Production by Human Keratinocytes. *J Invest Dermatol* **111**, 645–649 (1998).

97. Chabaud, M. *et al.* Human interleukin-17: A T cell–derived proinflammatory cytokine produced by the rheumatoid synovium. *Arthritis Rheumatism* **42**, 963–970 (1999).

98. Kotake, S. *et al.* IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. *J Clin Invest* **103**, 1345–1352 (1999).

99. Dudler, J., Renggli-Zulliger, N., Busso, N., Lotz, M. & So, A. Effect of interleukin 17 on proteoglycan degradation in murine knee joints. *Ann Rheum Dis* **59**, 529 (2000).

100. Chabaud, M., Lubberts, E., Joosten, L., Berg, W. van den & Miossec, P. IL-17 derived from juxta-articular bone and synovium contributes to joint degradation in rheumatoid arthritis. *Arthritis Res Ther* **3**, 168 (2001).

101. Bush, K. A., Farmer, K. M., Walker, J. S. & Kirkham, B. W. Reduction of joint inflammation and bone erosion in rat adjuvant arthritis by treatment with interleukin-17 receptor IgG1 Fc fusion protein. *Arthritis Rheumatism* **46**, 802–805 (2002).

102. Nakae, S., Nambu, A., Sudo, K. & Iwakura, Y. Suppression of Immune Induction of Collagen-Induced Arthritis in IL-17-Deficient Mice. *J Immunol* **171**, 6173–6177 (2003).

103. Kurasawa, K. *et al.* Increased interleukin-17 production in patients with systemic sclerosis. *Arthritis Rheumatism* **43**, 2455–2463 (2000).

104. Molet, S. *et al.* IL-17 is increased in asthmatic airways and induces human bronchial fibroblasts to produce cytokines. *J Allergy Clin Immun* **108**, 430–438 (2001).

105. Fujino, S. *et al.* Increased expression of interleukin 17 in inflammatory bowel disease. *Gut* **52**, 65 (2003).

106. Conde, C. D. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Sci New York N Y* **376**, eabl5197–eabl5197 (2022).

107. Aarvak, T., Chabaud, M., Miossec, P. & Natvig, J. B. IL-17 is produced by some proinflammatory Th1/Th0 cells but not by Th2 cells. *J Immunol Baltim Md 1950* **162**, 1246–51 (1999).

108. Albanesi, C. *et al.* Interleukin-17 is Produced by Both Th1 and Th2 Lymphocytes, and Modulates Interferon-γ- and Interleukin-4-Induced Activation of Human Keratinocytes. *J Invest Dermatol* **115**, 81–87 (2000).

109. Aggarwal, S. & Gurney, A. L. IL-17: prototype member of an emerging cytokine family. *J Leukocyte Biol* **71**, 1–8 (2002).

110. Oppmann, B. *et al.* Novel p19 Protein Engages IL-12p40 to Form a Cytokine, IL-23, with Biological Activities Similar as Well as Distinct from IL-12. *Immunity* **13**, 715–725 (2000).

111. Vermeire, K. *et al.* Accelerated collagen-induced arthritis in IFN-gamma receptor-deficient mice. *J Immunol* **158**, 5507–5513 (1997).

112. Guedez, Y. B. *et al.* Genetic ablation of interferon-gamma up-regulates interleukin-1beta expression and enables the elicitation of collagen-induced arthritis in a nonsusceptible mouse strain. *Arthritis Rheum* **44**, 2413–24 (2001).

113. Billiau, A. *et al.* Enhancement of experimental allergic encephalomyelitis in mice by antibodies against IFN-gamma. *J Immunol Baltim Md 1950* **140**, 1506–10 (1988).

114. Willenborg, D. O., Fordham, S., Bernard, C. C., Cowden, W. B. & Ramshaw, I. A. IFN-gamma plays a critical down-regulatory role in the induction and effector phase of myelin oligodendrocyte glycoprotein-induced autoimmune encephalomyelitis. *J Immunol Baltim Md 1950* **157**, 3223–7 (1996).

115. Debray-Sachs, M. *et al.* Prevention of diabetes in NOD mice treated with antibody to murine IFNγ. *J Autoimmun* **4**, 237–248 (1991).

116. Jones, L. S. *et al.* IFN-gamma-deficient mice develop experimental autoimmune uveitis in the context of a deviant effector response. *J Immunol Baltim Md 1950* **158**, 5997–6005 (1997).

117. Cua, D. J. *et al.* Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature* **421**, 744–748 (2003).

118. Ivanov, I. I. *et al.* The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. *Cell* **126**, 1121–1133 (2006).

119. Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J Exp Medicine* **204**, 1849–1861 (2007).

120. Nistala, K. *et al.* Th17 plasticity in human autoimmune arthritis is driven by the inflammatory environment. *Proc National Acad Sci* **107**, 14751–14756 (2010).

121. Zielinski, C. E. *et al.* Pathogen-induced human TH17 cells produce IFN-γ or IL-10 and are regulated by IL-1β. *Nature* **484**, 514–518 (2012).

122. Ghoreschi, K. *et al.* Generation of pathogenic TH17 cells in the absence of TGF-β signalling. *Nature* **467**, 967–971 (2010).

123. Schnell, A. *et al.* Stem-like intestinal Th17 cells give rise to pathogenic effector T cells during autoimmunity. *Cell* **184**, 6281-6298.e23 (2021).

124. Stritesky, G. L., Yeh, N. & Kaplan, M. H. IL-23 Promotes Maintenance but Not Commitment to the Th17 Lineage. *J Immunol* **181**, 5948–5955 (2008).

125. Romero-Sanchez, C. *et al.* Association between Th-17 cytokine profile and clinical features in patients with spondyloarthritis. *Clin Exp Rheumatol* **29**, 828–34 (2011).

126. Chen, W.-S. *et al.* Association of serum interleukin-17 and interleukin-23 levels with disease activity in Chinese patients with ankylosing spondylitis. *J Chin Med Assoc* **75**, 303–308 (2012).

127. Yi, K. *et al.* Single cell transcriptome and surface protein expression analysis identify OX40+GITR+ pathogenic T helper 17 in ankylosing spondylitis. *Arthritis Rheumatol* (2023) doi:10.1002/art.42476.

128. Umemura, M. *et al.* Involvement of IL-17 in Fas ligand-induced inflammation. *Int Immunol* **16**, 1099–1108 (2004).

129. Venken, K. *et al.* RORγt inhibition selectively targets IL-17 producing iNKT and γδ-T cells enriched in Spondyloarthritis patients. *Nat Commun* **10**, 9 (2019).

130. Gracey, E. *et al.* IL-7 primes IL-17 in mucosal-associated invariant T (MAIT) cells, which contribute to the Th17-axis in ankylosing spondylitis. *Ann Rheum Dis* **75**, 2124 (2016).

131. WANG, C., LIAO, Q., HU, Y. & ZHONG, D. T lymphocyte subset imbalances in patients contribute to ankylosing spondylitis. *Exp Ther Med* **9**, 250–256 (2015).

132. Taams, L. S., Steel, K. J. A., Srenathan, U., Burns, L. A. & Kirkham, B. W. IL-17 in the immunopathogenesis of spondyloarthritis. *Nat Rev Rheumatol* **14**, 453–466 (2018).

133. Mauro, D. *et al.* Ankylosing spondylitis: an autoimmune or autoinflammatory disease? *Nat Rev Rheumatol* **17**, 387–404 (2021).

134. Weber, U. *et al.* Frequency and Anatomic Distribution of Magnetic Resonance Imaging Features in the Sacroiliac Joints of Young Athletes. *Arthritis Rheumatol* **70**, 736–745 (2018).

135. Renson, T. *et al.* High prevalence of spondyloarthritis-like MRI lesions in postpartum women: a prospective analysis in relation to maternal, child and birth characteristics. *Ann Rheum Dis* **79**, 929–934 (2020).

136. Rao, D. A. *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* **542**, 110–114 (2017).

137. Yoshitomi, H. & Ueno, H. Shared and distinct roles of T peripheral helper and T follicular helper cells in human diseases. *Cell Mol Immunol* **18**, 523–527 (2021).

138. Marks, K. E. & Rao, D. A. T peripheral helper cells in autoimmune diseases*. *Immunol Rev* **307**, 191–202 (2022).

139. Ukita, M. *et al.* CXCL13-producing CD4+ T cells accumulate in early phase of tertiary lymphoid structures in ovarian cancer. *Jci Insight* **7**, (2022).

140. Yoshitomi, H. *et al.* Human Sox4 facilitates the development of CXCL13-producing helper T cells in inflammatory environments. *Nat Commun* **9**, 3762 (2018).

141. Long, Y. *et al.* Increased circulating PD-1hiCXCR5- peripheral helper T cells are associated with disease severity of active ulcerative colitis patients. *Immunol Lett* **233**, 2–10 (2021).

142. Bocharnikov, A. V. *et al.* PD-1hi CXCR5- T peripheral helper cells promote B cells responses in lupus via MAF and IL-21. *Jci Insight* **4**, (2019).

143. Dupré, A. *et al.* Association between T follicular helper cells and T peripheral helper cells with B-cell biomarkers and disease activity in primary Sjögren syndrome. *Rmd Open* **7**, e001442 (2021).

144. Kobayashi, S. *et al.* TGF-β induces the differentiation of human CXCL13-producing CD4+ T cells. *Eur. J. Immunol.* **46**, 360–371 (2016).

145. Johnston, R. J. *et al.* Bcl6 and Blimp-1 Are Reciprocal and Antagonistic Regulators of T Follicular Helper Cell Differentiation. *Science* **325**, 1006–1010 (2009).

146. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

147. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom.* **19**, 477 (2018).

148. Berge, K. V. den *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).

149. Eriksson, J. K., Jacobsson, L., Bengtsson, K. & Askling, J. Is ankylosing spondylitis a risk factor for cardiovascular disease, and how do these risks compare with those in rheumatoid arthritis? *Ann Rheum Dis* **76**, 364 (2017).

150. Bettcher, B. M., Tansey, M. G., Dorothée, G. & Heneka, M. T. Peripheral and central immune system crosstalk in Alzheimer disease — a research prospectus. *Nat Rev Neurol* **17**, 689–701 (2021).

151. Tan, E.-K. *et al.* Parkinson disease and the immune system — associations, mechanisms and therapeutics. *Nat Rev Neurol* **16**, 303–318 (2020).

152. Baker, K. F. & Isaacs, J. D. Novel therapies for immune-mediated inflammatory diseases: What can we learn from their use in rheumatoid arthritis, spondyloarthritis, systemic lupus erythematosus, psoriasis, Crohn's disease and ulcerative colitis? *Ann Rheum Dis* **77**, 175–187 (2018).

153. Krämer, J. & Wiendl, H. What Have Failed, Interrupted, and Withdrawn Antibody Therapies in Multiple Sclerosis Taught Us? *Neurotherapeutics* **19**, 785–807 (2022).

154. Colbert, R. A. Classification of juvenile spondyloarthritis: enthesitis-related arthritis and beyond. *Nat. Rev. Rheumatol.* **6**, 477–485 (2010).

155. Phimister, E. G., Parks, B. W. & Lusis, A. J. Macrophage Accumulation in Atherosclerosis. *New Engl J Medicine* **369**, 2352–2353 (2013).

156. Fang, H. *et al.* A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat Genet* **51**, 1082–1091 (2019).

157. Hansson, M. J. *et al.* Cyclophilin D-Sensitive Mitochondrial Permeability Transition in Adult Human Brain and Liver Mitochondria. *J. Neurotrauma* **28**, 143–153 (2011).

158. Matsuda, S. & Koyasu, S. Mechanisms of action of cyclosporine. *Immunopharmacology* **47**, 119–125 (2000).

159. Géher, P. & Gömör, B. Repeated cyclosporine therapy of peripheral arthritis associated with ankylosing spondylitis. *Medical Sci Monit Int Medical J Exp Clin Res* **7**, 105–7 (2001).

160. Colombo, D. *et al.* Efficacy of cyclosporine A as monotherapy in patients with psoriatic arthritis: a subgroup analysis of the SYNERGY Study. *Italian J Dermatology Venereol* **152**, 297–301 (2017).

161. Tsai, A., Alves, M. R. & Crocker, J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife* **8**, e45325 (2019).

162. Palstra, R.-J. & Grosveld, F. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Frontiers Genetics* **3**, 195 (2012).

163. Sun, S.-C. The non-canonical NF-κB pathway in immunity and inflammation. *Nat Rev Immunol* **17**, 545–558 (2017).

164. Feng, Y. & Pauklin, S. Revisiting 3D chromatin architecture in cancer development and progression. *Nucleic Acids Res* **48**, gkaa747- (2020).

165. Zhernakova, A. *et al.* Evolutionary and Functional Analysis of Celiac Risk Loci Reveals SH2B3 as a Protective Factor against Bacterial Infection. *Am J Hum Genetics* **86**, 970–977 (2010).

166. Balcerek, J. *et al.* Lnk/Sh2b3 deficiency restores hematopoietic stem cell function and genome integrity in Fancd2 deficient Fanconi anemia. *Nat Commun* **9**, 3915 (2018).

167. Wang, W. *et al.* LNK/SH2B3 Loss of Function Promotes Atherosclerosis and Thrombosis. *Circ Res* **119**, e91–e103 (2016).

168. Allenspach, E. J. *et al.* The Autoimmune Risk R262W Variant of the Adaptor SH2B3 Improves Survival in Sepsis. *J Immunol* **207**, 2710–2719 (2021).

169. Lévy, R. *et al.* Inherited human c-Rel deficiency disrupts myeloid and lymphoid immunity to multiple infectious agents. *J. Clin. Investig.* **131**, e150143 (2021).

170. Conrad, K., Wu, P., Sieper, J. & Syrbe, U. In vivo pre-activation of monocytes in patients with axial spondyloarthritis. *Arthritis Res Ther* **17**, 179 (2015).

171. Shi, H. *et al.* GM-CSF Primes Proinflammatory Monocyte Responses in Ankylosing Spondylitis. *Front Immunol* **11**, 1520 (2020).

172. Zeng, L., Lindstrom, M. J. & Smith, J. A. Ankylosing spondylitis macrophage production of higher levels of interleukin-23 in response to lipopolysaccharide without induction of a significant unfolded protein response. *Arthritis Rheumatism* **63**, 3807–3817 (2011).

173. Mei, Y. *et al.* Increased serum IL-17 and IL-23 in the patient with ankylosing spondylitis. *Clin Rheumatol* **30**, 269–273 (2011).

174. Schröder, M. & Kaufman, R. J. THE MAMMALIAN UNFOLDED PROTEIN RESPONSE. *Annu Rev Biochem* **74**, 739–789 (2005).

175. Hetz, C. & Papa, F. R. The Unfolded Protein Response and Cell Fate Control. *Mol Cell* **69**, 169–181 (2018).

176. Janssens, S., Pulendran, B. & Lambrecht, B. N. Emerging functions of the unfolded protein response in immunity. *Nat Immunol* **15**, 910–919 (2014).

177. Tavernier, S. J. *et al.* Regulated IRE1-dependent mRNA decay sets the threshold for dendritic cell survival. *Nat Cell Biol* **19**, 698–710 (2017).

178. Martinon, F., Chen, X., Lee, A.-H. & Glimcher, L. H. TLR activation of the transcription factor XBP1 regulates innate immune responses in macrophages. *Nat Immunol* **11**, 411–418 (2010).

179. Ciccia, F. *et al.* Dysbiosis and zonulin upregulation alter gut epithelial and vascular barriers in patients with ankylosing spondylitis. *Ann Rheum Dis* **76**, 1123 (2017).

180. Guggino, G. *et al.* Inflammasome Activation in Ankylosing Spondylitis Is Associated With Gut Dysbiosis. *Arthritis Rheumatol* **73**, 1189–1199 (2021).

181. Ciccia, F. *et al.* Evidence that autophagy, but not the unfolded protein response, regulates the expression of IL-23 in the gut of patients with ankylosing spondylitis and subclinical gut inflammation. *Ann Rheum Dis* **73**, 1566 (2014).

182. Qaiyum, Z., Gracey, E., Yao, Y. & Inman, R. D. Integrin and transcriptomic profiles identify a distinctive synovial CD8+ T cell subpopulation in spondyloarthritis. *Ann Rheum Dis* **78**, 1566 (2019).

183. Guggino, G., Rizzo, A., Mauro, D., Macaluso, F. & Ciccia, F. Gut-derived CD8+ tissue-resident memory T cells are expanded in the peripheral blood and synovia of SpA patients. *Ann Rheum Dis* **80**, e174–e174 (2021).

184. Yang, X. *et al.* Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature* **612**, 771–777 (2022).

185. King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *Plos Genet* **15**, e1008489 (2019).

186. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).

187. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet* **52**, 482–493 (2020).

188. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* **52**, 1158–1168 (2020).

189. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).

190. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet* **54**, 1572–1580 (2022).

191. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet* **54**, 1479–1492 (2022).

192. Sallusto, F., Zielinski, C. E. & Lanzavecchia, A. Human Th17 subsets. *Eur J Immunol* **42**, 2215–2220 (2012).

193. Argyriou, A. *et al.* Single cell sequencing identifies clonally expanded synovial CD4+ TPH cells expressing GPR56 in rheumatoid arthritis. *Nat Commun* **13**, 4046 (2022).

194. Gao, Y., Zhao, H., Wang, P., Wang, J. & Zou, L. The roles of SOCS3 and STAT3 in bacterial infection and inflammatory diseases. *Scand J Immunol* **88**, e12727 (2018).

195. Escobar, G., Mangani, D. & Anderson, A. C. T cell factor 1: A master regulator of the T cell response in disease. *Sci Immunol* **5**, (2020).

196. Bassez, A. *et al.* A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat Med* **27**, 820–832 (2021).

197. Peng, J., Gong, Y., Zhang, Y., Wang, D. & Xiao, Z. Immunohistological analysis of active sacroiliitis in patients with axial spondyloarthritis. *Medicine* **96**, e6605 (2017).

198. Duan, J. *et al.* Endoplasmic reticulum stress in the intestinal epithelium initiates purine metabolite synthesis and promotes Th17 cell differentiation in the gut. *Immunity* **56**, 1115-1131.e9 (2023).