Multi-modal single-cell genomics-based investigation of the pathogenesis of HLA-B27+ axial spondyloarthritis



Liying Jin

Brasenose College

University of Oxford

Supervisors: Doctor Stephen Nicholas Sansom, Professor Paul Bowness

Thesis submitted for the degree of

DPhil in Molecular and Cellular Medicine

Michaelmas Term 2025

I. Abstract

Axial spondyloarthritis (AxSpA) is a chronic, HLA-B*27-associated inflammatory disease primarily affecting the sacroiliac joints and spine, but also involving peripheral joints and extra-skeletal tissues such as the eye and gastrointestinal tract. Despite known genetic associations with HLA-B*27 and IL-17/IL-23 signaling, and emerging evidence of public $CD8^+$ T-cell clones, the cellular basis of inflammation and pathogenesis of AxSpA remains incompletely understood. Currently available biologic therapies are only effective in 60-70% of patients, yet the mechanisms underlying treatment responses are still unclear.

To first investigate the effect of TNF inhibitors on AxSpA patients, I constructed an integrated single-cell PBMC atlas to characterise longitudinal immune changes in treatment responders. This analysis revealed shifts in peripheral cellular composition, including an increase in B-cell lineage abundance following TNFi treatment, as well as broader transcriptomic changes associated with biologic therapy. To then enable unbiased discovery of the cellular mechanisms underlying HLA-B*27-positive AxSpA, I next constructed a multi-modal, cross-tissue single-cell atlas incorporating gene expression, surface protein, and TCR data across blood, synovial fluid, spine, sacroiliac joint, gut and eye. This atlas provides the first unified framework for comparing immune and structural cell populations across the diverse tissues involved in AxSpA. Compositional analysis revealed both shared and tissue-specific immune cell populations, and identified several novel AxSpA-associated populations, including $ZNF683^+$ TEMRA-like $CD8^+$ T cells and distinct conventional monocytes, suggesting potential tissue-adapted effector programs

that contribute to disease. Genetically informed evaluation further highlighted NK cells, MAIT-like T cells, inflammatory monocytes, and ZNF683^{+} CD8^{+} T cells as the populations most strongly associated with AxSpA genetic-risk signals at the transcriptomic level, and their differing patterns across tissues support the concept of a gut-joint axis in the development of AxSpA. Despite methodological limitations related to sample numbers, tissue processing and batch effects, these findings collectively point toward specific immune cell subsets and inflammatory programs that may link genetic susceptibility to tissue-level immune activation in AxSpA.

II. Declaration

I declare that the work presented in this DPhil thesis is entirely my own, except where the contributions of collaborators have been explicitly acknowledged. To the best of my knowledge, this thesis contains no material previously published or written by another person, except where duly referenced. No part of this thesis has been submitted nor is it being submitted for any degree, diploma, certificate, or other qualification at the University of Oxford or any other institution.

III. Acknowledgements

I would like to start by thanking my supervisors, Steve and Paul. I am truly grateful to Steve for his patience throughout my DPhil, for generously sharing his expertise, and for the in-depth discussions and guidance through which I developed much of my computational skills. I also want to thank Paul for his thoughtful advice, his perspective and expertise, and the support he has offered both in research and in life. I feel very fortunate to have learned from both of them.

I am also thankful to my lab members and colleagues for creating such a welcoming and friendly research environment. My thanks also go to my friends, both near and far, who have been a constant source of help, motivation, and companionship.

I am deeply grateful to the patients and donors whose samples made this research possible. I also want to acknowledge all the researchers whose work I have cited, as well as those I may not have had the opportunity to mention, all of whom continue to build and expand our understanding in this field.

Finally, and above all, I wish to express my deepest gratitude and love to my family, even from afar, whose unwavering love and support have carried me throughout, and to my dog, for the comfort she brought through many difficult days. And I thank my partner, for always being there.

IV. Contents

I. Abstract **2**

II. Declaration **4**

III. Acknowledgements **5**

IV. Contents **6**

V. List of Figures **11**

VI. List of Tables **17**

VII. List of Abbreviations **18**

Chapter 1: Introduction **20**

1.1 Clinical Overview of Axial Spondyloarthritis *20*
1.1.1 Definition and classification 20
1.1.2 Global prevalence of HLA-B*27 and AxSpA 21
1.1.3 Clinical features and diagnosis 22
1.1.4 Clinical management and therapeutic strategies 24

1.2 Mechanistic Basis of Axial Spondyloarthritis *26*
1.2.1 Genetic and molecular mechanisms 26
1.2.1.1 Genetic susceptibility and GWAS findings 26
1.2.1.2 HLA-B*27 and associated theories 27
1.2.1.3 Aminopeptidase and antigen processing 31
1.2.1.4 The IL-23/IL-17 axis 32
1.2.2 Environmental and microbial Influences 33
1.2.2.1 Gut-joint axis and intestinal dysbiosis 33
1.2.2.2 Mechanical stress and local tissue responses 36

1.3 Multi-omics Approaches in Axial Spondyloarthritis Research *37*
1.3.1 Single-cell discoveries in AxSpA 37
1.3.2 Immune repertoire insights in AxSpA 40
1.3.3 Cross-tissue atlases within and beyond AxSpA 42

1.4 Thesis Aims *44*

1.5 References *45*

Chapter 2: Materials and Methods **58**

2.1 Sample Collection and Library Construction *58*
2.1.1 Overview of datasets and samples 58
2.1.2 Project-specific sample collection and library protocols 59
2.1.2.1 Projects: P170676, P180407, P190230, P190726 59

2.1.2.2 Projects: Ox_VA_PB_SF and Berlin_PB_SF 61
2.1.2.3 Project: SIJ 62
2.1.2.4 Projects: Gut_Naples and Gut_TGU 63
2.1.2.5 Project: Spine_Enthesis 64
2.1.2.6 Project: Aqh_Eye 65
2.1.2.7 Projects: CD8_SFMC_Berlin and COMBAT 65
2.1.3 HLA-B*27 status determination 66
2.2 Computational Analysis and Resources 66
2.2.1 Data preprocessing 66
2.2.1.1 scRNA-seq data processing 66
2.2.1.2 CITE-seq data processing 67
2.2.1.3 Variant calling and joint genotyping 68
2.2.1.4 scTCR-seq preprocessing 68
2.2.2 Cell demultiplexing and sample assignment 69
2.2.3 Quality control 70
2.2.4 Data normalization and integration 71
2.2.5 Data clustering and annotation 72
2.2.5.1 Neighborhood construction 72
2.2.5.2 Clustering 73
2.2.5.3 Annotation process 73
2.2.5.4 Cluster entropy calculation 74
2.2.5.5 CD4 and CD8 cell threshold decision 74
2.2.6 Compositional and abundance analysis 75
2.2.6.1 Compositional differential abundance analysis (edgeR) 75
2.2.6.2 Bayesian compositional analysis (scCODA) 75
2.2.6.3 Neighbourhood-level differential abundance analysis (Milo) 76
2.2.6.4 Co-varying compositional pattern analysis (NMF) 76
2.2.6.5 Cross-tissue compositional comparison 77
2.2.7 Differential expression analysis 77
2.2.8 GSEA pathway analysis 78
2.2.9 scDRS analysis 78
2.2.10 Statistics 79
2.3 References 80
Chapter 3: Investigation of Axial Spondyloarthritis Patients and Healthy Individuals at Single-Cell Level 84
3.1 Introduction 84
3.2 Aim 86
3.3 Results 86
3.3.1 Experimental design 86
3.3.2 PBMC scRNA-seq data processing and analysis 88
3.3.2.1 Batch-specific preprocessing 88

3.3.2.2 Integration of 4 AxSpA PBMC batches with the COMBAT dataset 91
3.3.2.3 Advanced quality control and construction of final atlas..... 94
3.3.3 Biological naïve and TNFi treated AxSpA patients exhibit a shift in immune cell composition compared to healthy controls 97
3.3.3.1 Increased B cell frequency in biologic-naïve and TNFi-treated AxSpA patients..... 97
3.3.3.2 Identification of a $CD9^+$ monocyte subset unique to AxSpA patients..... 102
3.3.4 Transcriptional differences between AxSpA patients and healthy controls 106
3.3.5 Gene expression informed evaluation of AxSpA genetic risk across immune cell types 114
3.3.6 Validation of compositional change using bulk RNA-seq data..... 117
3.4 Discussion 118
3.5 References 122
Chapter 4: A Cross-tissue, Multi-Modal Single-Cell Atlas of HLA-B*27-Positive Axial Spondyloarthritis 126
4.1 Introduction..... 126
4.2 Aim 128
4.3 Results 128
4.3.1 Experimental design and overview of samples..... 128
4.3.2 Multi-modal atlas generation..... 132
4.3.2.1 Preprocessing and ambient RNA removal 132
4.3.2.2 Cell demultiplexing 134
4.3.2.3 Initial quality control..... 138
4.3.2.4 Optimizing the data integration strategy..... 141
4.3.2.5 Overview of initial data integration 147
4.3.2.6 Sub-regional clustering and secondary quality control..... 149
4.3.2.7 T Cell cub-regional clustering and tertiary quality control..... 159
4.3.2.8 ADT data processing..... 161
4.3.3 Cell-type annotation strategy and approach 165
4.3.3.1 Global annotation and clustering framework..... 165
4.3.3.2 Classification of $CD4^+$ and $CD8^+$ T cells..... 169
4.3.3.3 Identification of MAIT And iNKT Cells 172
4.3.4 Cell-type annotation results and initial characterisation..... 179
4.3.4.1 CD4-positive T cell compartment 179
4.3.4.2 CD8-positive T cell compartment 183
4.3.4.3 NK cell compartment..... 188
4.3.4.4 MNP cell compartment..... 191
4.3.4.5 B cell compartment 194

4.3.4.6 Other cell compartment 198
4.4 Discussion 201
4.5 References 204

Chapter 5: Cellular Composition and Cross-Tissue Immune Landscape in AxSpA 210

5.1 Introduction 210
5.2 Aim 211
5.3 Results 212
5.3.1 Analytical framework for differential abundance 212
5.3.2 EdgeR results comparing cell-type composition between AxSpA, HC, and RA groups 215
5.3.2.1 Comparison strategy in core and expanded cohorts 215
5.3.2.2 CD4-positive T Cell Compartment 216
5.3.2.3 CD8-positive T Cell Compartment 220
5.3.2.4 NK cell compartment 225
5.3.2.5 Myeloid cell compartment 227
5.3.2.6 B and plasma cell compartment 233
5.3.2.7 Technical validation of compositional changes with scCODA and Milo 237
5.3.2.8 Co-varying pattern analysis 246
5.3.2.9 Cross-tissue comparison 250
5.4 Discussion 254
5.5 References 261

Chapter 6: Transcriptional and Genetic Mechanisms Underlying AxSpA Pathogenesis 266

6.1 Introduction 266
6.2 Aim 268
6.3 Results 269
6.3.1 Genetic associations at the cell type level 269
6.3.1.1 Genetic associations in the sacroiliac joint 269
6.3.1.2 Genetic associations in the spine enthesis 271
6.3.1.3 Genetic associations in ileum 275
6.3.1.4 Cross-tissue comparison of genetic associations 277
6.3.2 Disease-associated pathway analysis 279
6.3.2.1 Targeted pathway analysis in the sacroiliac joint 280
6.3.2.2 Targeted pathway analysis in spine enthesis 282
6.3.2.3 Targeted pathway analysis in ileum 285
6.3.2.4 Targeted pathway analysis in synovial fluid 287

6.4 Discussion 289

6.5 References 295

Chapter 7: Overall Discussion and Future Directions 298

7.1 Alterations in Peripheral B Cell Lineages Following TNF Inhibition in AxSpA 298

7.2 Methods and Challenges in Building the AxSpA Atlas 299

7.3 Shared T Cell and Myeloid Alterations Support Arthritogenic Peptide Hypothesis and Gut-Joint Axis in AxSpA 301

7.4 Limitations and Future Analytical Directions 307

7.5 References 311

V. List of Figures

Figure 1.1 Multiple tissue involvement of AxSpA 22

Figure 1.2 Proposed mechanisms linking HLA-B*27 to AxSpA pathogenesis 28

Figure 3.1 Diagram for experimental design and sample collection 87

Figure 3.2 Overview of preprocessing workflow for four experimental batch 89

Figure 3.3 Initial quality control and filtering of single-cell data 91

Figure 3.4 Comparison of different HVG selection strategies for data integration 94

Figure 3.5 Secondary quality control assessment after data integration 97

Figure 3.6 Integrated PBMC atlas of AxSpA patients and healthy controls 97

Figure 3.7 Visualization of compositional shifts and CNA analysis 99

Figure 3.8 Compositional analysis of PBMCs from healthy controls, biologic-naïve AxSpA patients, and anti-TNF-treated AxSpA patients. 102

Figure 3.9 Sub-clustering analysis of MNP, B cell lineage, and T cells 104

Figure 3.10 Distribution of monocyte subsets across individual samples 106

Figure 3.11 Transcriptional differences in PBMCs between AxSpA patients and healthy controls 109

Figure 3.12 Transcriptional differences between AxSpA patients and healthy controls 109

Figure 3.13 Pathway analysis using GSEA in AxSpA patients and healthy controls 111

Figure 3.14 Pathway analysis using scDRS in AxSpA patients and healthy controls 113

Figure 3.15 Integrated analysis of GWAS-identified AxSpA risk genes with single-cell gene expression data 115

Figure 3.16 Genetic risk associations between immune cell types and AxSpA GWAS traits 116

Figure 3.17 Validation of cell compositional changes using external bulk RNA-seq datasets 118

Figure 4.1 Overview of sample collection and analytical workflow for the atlas construction 131

Figure 4.2 Cell quality metrices before and after ambient RNA removal with CellBender 133

Figure 4.3 Genetic demultiplexing outcomes with GATK and souporcell 135

Figure 4.4 Hashtag demultiplexing outcomes with GMM-demux 136

Figure 4.5 Comparison of hashtag and genetic demultiplexing results 137

Figure 4.6 Quality control metrics across projects before and after filtering 139

Figure 4.7 Sample-level quality control metrics before and after cell-level filtering 141

Figure 4.8 Comparison of HVG selection strategies in T cell compartments 143

Figure 4.9 UMAPs of the integrated atlas under different scVI model parameters 145

Figure 4.10 Overview of the initially integrated atlas and sample quality metrics 148

Figure 4.11 UMAP of four major immune cell compartments in the initial atlas 149

Figure 4.12 Quality control metrics of the T-cell subregion 151

Figure 4.13 scVI models incorporating mitochondrial covariates 152

Figure 4.14 Clustering and cluster-level quality metrics of the T/NK-cell subregion 153

Figure 4.15 Clustering and quality control of the B-cell subregion 154

Figure 4.16 Clustering and quality control of the MNP cell subregion 155

Figure 4.17 Clustering and quality control of the ‘other’ cell subregion 157

Figure 4.18 UMAPs after second-round quality control 158

Figure 4.19 Expression of key marker genes in the T/NK-cell compartment 159

Figure 4.20 Summary diagram of the iterative subregional clustering and quality control process 161

Figure 4.21 Comparison of total RNA and total protein counts across CITE-seq samples 162

Figure 4.22 Density distributions of selected ADTs before and after totalVI normalization 164

Figure 4.23 Distribution of selected ADTs across expected positive and negative cell populations 165

Figure 4.24 Evaluation of clusters in the CD4 T cell region 168

Figure 4.25 CD4 and CD8 expression patterns across T-cell clusters 171

Figure 4.26 Sub-clustering of the cycling T cell cluster to distinguish $CD4^+$ and $CD8^+$ T cells 172

Figure 4.27 Comparison of TCR-based and ADT-based MAIT-cell identification in cells with both modalities 175

Figure 4.28 Comparison of MAIT and iNKT cell identification across modalities 177

Figure 4.29 Alternative totalVI model normalization and background correction for Hu.TCR.Vα7.2 ADT counts 178

Figure 4.30 UMAP of the CD4-positive T cell region and tissue composition of identified clusters 180

Figure 4.31 Dot plot summary of the CD4-positive T cell groups showing hierarchical relationships and multimodal marker expression 181

Figure 4.32 UMAP of the CD8-positive T cell region and tissue composition of identified clusters 184

Figure 4.33 Dot plot summary of the CD8-positive T cell groups showing hierarchical relationships and multimodal marker expression 185

Figure 4.34 UMAP of the NK cell region and tissue composition of identified clusters 188

Figure 4.35 Dot plot summary of the NK cell groups showing hierarchical relationships and multimodal marker expression 189

Figure 4.36 UMAP of the MNP cell region and tissue composition of identified clusters 191

Figure 4.37 Dot plot summary of the MNP cell groups showing hierarchical relationships and multimodal marker expression 193

Figure 4.38 UMAP of the B cell region and tissue composition of identified clusters 195

Figure 4.39 Dot plot summary of the B cell groups showing hierarchical relationships and multimodal marker expression 196

Figure 4.40 UMAP of the other cell region and tissue composition of identified clusters 199

Figure 4.41 Dot plot summary of the NK cell groups showing hierarchical relationships and multimodal marker expression 199

Figure 5.1 Analysis framework for compositional analysis across cohorts 214

Figure 5.2 Differential abundance of CD4-positive non-naïve T-cell subsets across tissues and patient groups at the cell type level in the core cohort 218

Figure 5.3 Differential abundance of CD4-positive non-naïve T-cell subsets at the cell type level in the expanded cohort 220

Figure 5.4 Differential abundance of CD8-positive non-naïve T-cell subsets at the cell type level in the core cohort 221

Figure 5.5 Differential abundance of CD8-positive non-naïve T-cell subsets at the cell type level in the expanded cohort 223

Figure 5.6 Differential abundance of NK cell subsets at the cell type level in the core cohort 226

Figure 5.7 Differential abundance of NK cell subsets at the cell type level in the expanded cohort 227

Figure 5.8 Differential abundance of myeloid cell subsets at the cell type level in the core cohort 228

Figure 5.9 Differential abundance of myeloid cell subsets at the minor subset level in the core cohort 230

Figure 5.10 Differential abundance of myeloid cell subsets at the cell type level in the expanded cohort 232

Figure 5.11 Differential abundance of B and plasma cell compartments at the cell type level in the core cohort 234

Figure 5.12 Differential abundance of B and plasma cell subsets at the minor subset level in the core cohort 236

Figure 5.13 scCODA analysis of AxSpA versus healthy controls across $CD8^+$ T-cell subsets in PEB in the expanded cohort 238

Figure 5.14 Milo differential abundance analysis of the $CD8^+$ T-cell region in PBMCs between AxSpA and healthy controls in the expanded cohort 239

Figure 5.15 Milo differential abundance analysis of the CD8-positive T cell region in SFMCs between AxSpA and RA in the expanded cohort 240

Figure 5.16 scCODA analysis of AxSpA versus healthy controls across myeloid cell components in the expanded cohort 241

Figure 5.17 Milo differential abundance analysis of the myeloid cell region in PBMCs between AxSpA and healthy controls in the expanded cohort 242

Figure 5.18 scCODA analysis of AxSpA versus RA across myeloid cell components in the expanded cohort 243

Figure 5.19 Milo differential abundance analysis of the myeloid cell region between AxSpA and RA in the expanded cohort 245

Figure 5.20 Determination of optimal factorization rank for NMF analysis in synovial fluid T cell data 247

Figure 5.21 NMF identifies disease-associated T-cell programs in synovial fluid in the expanded cohort 248

Figure 5.22 Tissue-associated enrichment of $CD8^+$ T-cell compartment across AxSpA tissues 252

Figure 5.23 Tissue-associated compositional differences in the myeloid cell compartment 253

Figure 5.24 Tissue-associated compositional differences in the B cell compartment 254

Figure 6.1 scDRS-based disease association across cell types in the SIJ 270

Figure 6.2 scDRS-based disease association across cell types in PEB 272

Figure 6.3 scDRS-based disease association across cell types in EST 274

Figure 6.4 scDRS-based disease association across cell types in ileum 276

Figure 6.5 Summary of scDRS-based disease association across cell types and tissue 278

Figure 6.6 Disease-associated pathway enrichment analysis in the AxSpA sacroiliac joint samples 281

Figure 6.7 Disease-associated pathway enrichment analysis in the AxSpA PEB samples 283

Figure 6.8 Disease-associated pathway enrichment analysis in the AxSpA EST samples 284

Figure 6.9 Disease-associated pathway enrichment analysis in the AxSpA ileum samples 286

Figure 6.10 Disease-associated pathway enrichment analysis in the AxSpA SFMC samples 288

Figure 7.1 Public AxSpA-associated TRBV9^{+} T-cell clones detected across AxSpA synovial fluid and AAU ocular tissue 309

VI. List of Tables

Table 2.1 Experimental batches included in this thesis 59

Table 3.1 Clinical metadata of AxSpA patients and healthy controls analysed in Chapter 3 88

Table 4.1 Overview of projects and libraries included in Chapter 4 130

Table 4.2 Evaluation metrics for scVI models 146

Table 7.1 Summary of cell-type compositional differences observed across tissues 303

VII. List of Abbreviations

AAU	Acute Anterior Uveitis
AS	Ankylosing Spondylitis
ASAS	Assessment of Spondyloarthritis International Society
AxSpA	Axial Spondyloarthritis
BH	Benjamini-Hochberg
CD	Crohn's Disease
CITE-seq	Cellular Indexing of Transcriptomes and Epitopes by Sequencing
CTL	Cytotoxic T Lymphocytes
DEGs	Differentially-Expressed Genes
DMARDs	Disease Modifying Anti-Rheumatic Drugs
ELBO	Evidence Lower Bound
ER	Endoplasmic Reticulum
EST	Entheseal Soft Tissue
FACS	Fluorescence-Activated Cell Sorting
FDA	Food And Drug Administration
GSEA	Gene Set Enrichment Analysis
GO	Gene Ontology
GWAS	Genome-Wide Association Studies
HLA	Human Leukocyte Antigen
HSP	Heat Shock Proteins
HVG	Highly Variable Genes
IBD	Inflammatory Bowel Disease
IBP	Inflammatory Back Pain
IGAS	International Genetics of Ankylosing Spondylitis Consortium
IL	Interleukin
iNKT	Invariant Natural Killer T
IQR	Interquartile Range
JAK	Janus Kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIR	Killer Immunoglobulin-Like Receptor
LD	Linkage Disequilibrium
LILR	Leukocyte Immunoglobulin-Like Receptor
MAIT	Mucosal-Associated Invariant T
MHC	Major Histocompatibility Complex
MNPs	Mononuclear Phagocytes
MSC	Mesenchymal Stem Cells
MSigDB	Human Molecular Signatures Database
NK	Natural Killer
NMF	Non-Negative Matrix Factorization
NMI	Normalized Mutual Information
nr-AxSpA	Non-Radiographic Axial Spondyloarthritis

NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
PBMC	Peripheral Blood Mononuclear Cells
PEB	Peri-Entheseal Bone
PRS	Polygenic Risk Score
PsA	Psoriatic Arthritis
RA	Rheumatoid Arthritis
r-AxSpA	Radiographic Axial Spondyloarthritis
scRNA-seq	Single-Cell RNA Sequencing
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
SpA	Spondyloarthritis
TMM	Trimmed Mean of M-Values
TNFi	Tumour Necrosis Factor Inhibitors
UC	Ulcerative Colitis
UPR	Unfolded Protein Response

Chapter 1: Introduction

1.1 Clinical Overview of Axial Spondyloarthritis

1.1.1 Definition and classification

Axial spondyloarthritis (AxSpA) is a chronic inflammatory rheumatological disease primarily affecting the axial skeleton, including the sacroiliac joints and spine (Navarro-Compan et al., 2025; Sieper & Poddubnyy, 2017). In addition to axial involvement, AxSpA may also affect peripheral joints and extra-articular organs, such as the eyes, skin and gastrointestinal tract. AxSpA belongs to the broader family of spondyloarthritis (SpA) diseases, which share overlapping clinical, immunological, and genetic features, but differ in their predominant manifestations. Based on symptom distribution, the Assessment of SpondyloArthritis international Society (ASAS) classifies SpA into axial SpA (AxSpA) where axial symptoms dominate, and peripheral SpA where non-axial features are more prominent (Rudwaleit, Landewe, et al., 2009).

Ankylosing spondylitis (AS) is the historical and still widely used term that corresponds conceptually to what is now classified as radiographic axial spondyloarthritis (r-AxSpA), characterized by structural damage visible on radiographs (van der Heijde et al., 2024). In contrast, non-radiographic axial spondyloarthritis (nr-AxSpA) presents with similar clinical features but lacks such radiographic changes. Following the updated ASAS recommendations, this thesis uses AxSpA as the overarching term and treats AS and r-AxSpA equivalent unless otherwise specified. Therefore, references to ‘AS’ in historical literature and clinical records cited throughout this thesis are interpreted as synonymous with r-AxSpA.

1.1.2 Global prevalence of HLA-B*27 and AxSpA

The worldwide distribution of AxSpA broadly mirrors the population frequency of HLA-B*27, a major histocompatibility complex class I allele strongly associated with the disease, with both varying substantially across geographic and ethnic groups. A systematic review and meta-regression estimated the prevalence of AS to range from 0.02% in Sub-Saharan Africa to 0.35% in Northern Arctic communities, with spondyloarthritis as a whole reaching 1.61% in the latter population (Stolwijk et al., 2016). HLA-B*27 carrier frequency follows a broadly parallel pattern: it is approximately 8% in mid-European populations, 10-16% in Scandinavian populations, and 25-50% among Inuit, Yupik, and Indigenous Northern American communities, while the allele is extremely rare in African populations of unmixed ancestry (Hanson & Brown, 2017; Ziade, 2017). Across diverse populations, 60-90% of patients with AxSpA carry HLA-B*27, although this proportion is lower in some populations; fewer than 60% of African American AS patients carry the allele (Jamalyaria et al., 2017).

Among the more than 200 HLA-B*27 alleles described, HLA-B*27:05 is the most widely distributed disease-associated subtype globally, and the others include HLA-B*27:02 (in Mediterranean populations) and HLA-B*27:04 (in Chinese and other Asian populations) (Khan et al., 2023). Two subtypes are notable for lacking association with typical AS: HLA-B*27:06, which is common in Southeast Asia, and HLA-B*27:09, which is found primarily on the Italian island of Sardinia. These disease-neutral subtypes help explain otherwise puzzling epidemiological observations. For example, AS is more prevalent among Chinese Indonesians, who carry the disease-associated HLA-B*27:04

and HLA-B*27:05, despite an overall HLA-B*27 frequency of only about 5%. In contrast, native Indonesians have HLA-B*27 frequencies of up to 12%, but these are largely composed of the non-pathogenic HLA-B*27:06 (Khan et al., 2023). These population-level differences in subtype distribution partly account for the geographic variation in AxSpA prevalence.

1.1.3 Clinical features and diagnosis

AxSpA is primarily characterized by axial manifestation but also involves multiple regions and tissue types, including peripheral joints and entheses as well as extra-musculoskeletal organs (**Figure 1.1**).

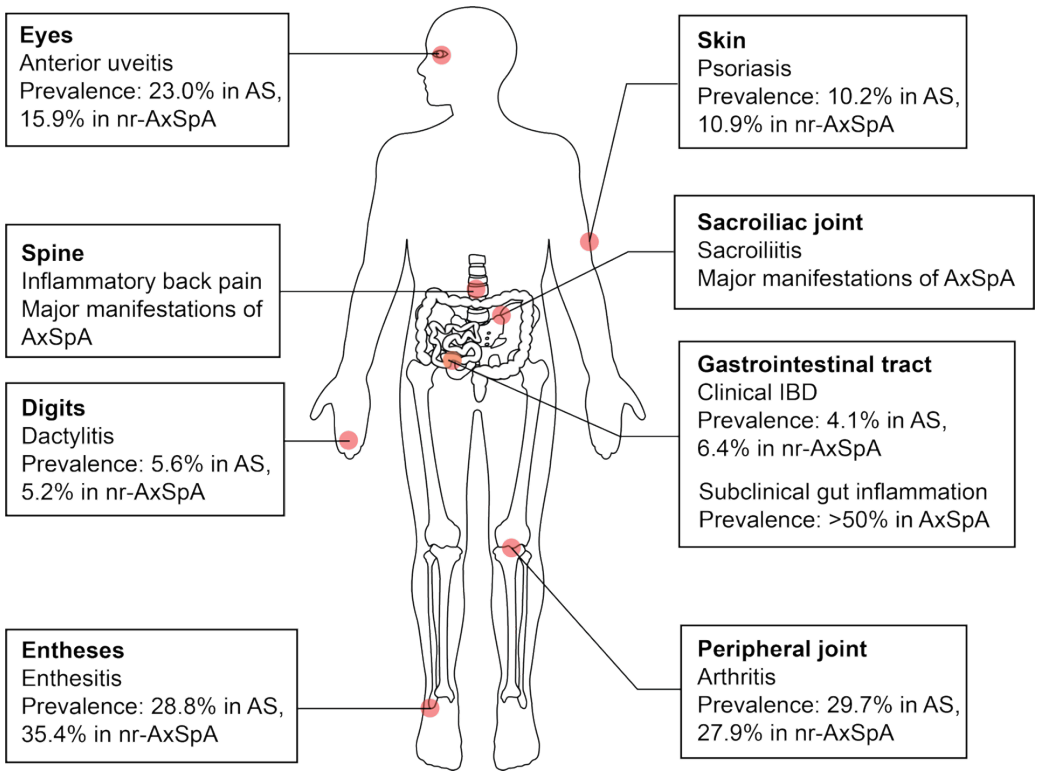


Figure 1.1 Multiple tissue involvement of AxSpA
Clinical manifestation of axial spondyloarthritis across multiple regions and tissue types, figure adapted from (Robinson et al., 2021).

Chronic back pain is the predominant clinical presentation of AxSpA and can progress to structural joint damage, including bony fusion of vertebral joints in advanced cases (Sieper & Poddubnyy, 2017). This hallmark symptom, frequently accompanied by morning stiffness, plays a central role in both clinical diagnosis and disease classification (Rudwaleit, Landewe, et al., 2009; Rudwaleit et al., 2004). Such characteristic pain patterns defined as inflammatory back pain (IBP) are closely linked to underlying inflammatory processes and are identified by established criteria (Rudwaleit et al., 2006; Sieper et al., 2009). IBP typically has a gradual onset predominantly affecting the spine and buttocks. Symptoms usually improve with physical activity or non-steroidal anti-inflammatory drugs (NSAIDs) but worsen during periods of rest (Sieper et al., 2009).

Peripheral musculoskeletal manifestations also frequently occur in AxSpA, with peripheral arthritis and enthesitis being most common (de Winter et al., 2016). Peripheral arthritis typically manifests as asymmetrical inflammation affecting one or a few joints usually in the lower limbs (Schett et al., 2017). Enthesitis, characterized by inflammation at tendon insertion sites, most often affects the Achilles tendon and plantar fascia causing localized tenderness (de Winter et al., 2016). Less commonly patients may experience dactylitis, a diffuse swelling of a finger or toe (de Winter et al., 2016). Although the prevalence and diagnostic relevance of these peripheral features vary geographically, their presence significantly aids in clinical recognition of AxSpA (Lopez-Medina et al., 2021; Rudwaleit et al., 2004).

Extra-musculoskeletal manifestations further contribute to the clinical picture of AxSpA with acute anterior uveitis being the most common, affecting

approximately 20–30% of patients (de Winter et al., 2016; Rudwaleit et al., 2004). Uveitis typically presents as acute unilateral inflammation strongly associated with HLA-B*27 positivity and frequently recurs in alternating eyes (de Winter et al., 2016; Rademacher et al., 2020). Psoriasis occurs less frequently and affects about 10% of patients (Rudwaleit et al., 2004; Stolwijk et al., 2015). Clinical inflammatory bowel disease (IBD) including Crohn's disease (CD) or ulcerative colitis (UC) affects approximately 5–10% of patients, while subclinical gut inflammation occurs in over half of cases (Leirisalo-Repo et al., 1994; Rudwaleit, Haibel, et al., 2009; Rudwaleit et al., 2004).

Peripheral and extra-musculoskeletal features are helpful indicators for axial spondyloarthritis when present, but since they occur less frequently than axial symptoms, their absence doesn't exclude the disease and these additional manifestations can develop later during the disease course.

1.1.4 Clinical management and therapeutic strategies

There is currently no cure for AxSpA and the main goals of management are to control symptoms, prevent structural damage and improve health-related quality of life (Navarro-Compan et al., 2025). Management approaches include both non-pharmacological and pharmacological strategies guided by evolving recommendations from the ASAS (Ramiro et al., 2023).

Non-pharmacological management such as physical exercise and physiotherapy is widely recognized to relieve symptoms, although there is no evidence that these interventions prevent structural spinal changes like syndesmophyte formation. Therefore, pharmacological treatment remains important with non-steroidal anti-inflammatory drugs (NSAIDs) typically used as first-line therapy.

Conventional disease-modifying antirheumatic drugs (DMARDs) are generally ineffective for axial disease, while biologic DMARDs such as tumor necrosis factor (TNF) inhibitors, first approved in 2003, and interleukin (IL)-17 inhibitors, first approved in 2016, are recommended as second-line treatments and are currently the most effective options across the AxSpA spectrum (Danve & Deodhar, 2022; Ward et al., 2019). Despite their success, only about half of patients respond adequately.

Deeper insights into AxSpA pathogenesis have guided the exploration of additional therapeutic targets beyond TNF and IL-17 pathways. Inhibition of JAK–STAT signaling has shown promising efficacy in phase III trials, with tofacitinib and upadacitinib now approved for radiographic AxSpA (Ortolan et al., 2023; Veale et al., 2019). In contrast, blockade of IL-12/23 or IL-23 have not demonstrated clinical benefit (Baeten et al., 2018; Deodhar et al., 2019), possibly because IL-17 can be produced independently of IL-23 or because IL-23 acts primarily during early disease stages, though the precise mechanism remains uncertain.

More recently, a proof-of-concept study achieved long-term remission in a patient with ankylosing spondylitis through selective depletion of TRBV9^{+} T cells, a subset that includes the public clonotype implicated in HLA-B*27-associated AxSpA (Britanova et al., 2023). The therapeutic potential of this approach was further supported by clinical use of BCD-180, a humanized anti-TRBV9^{+} monoclonal antibody tested in Russia, which was reported to demonstrate significant clinical efficacy, good tolerability, and low immunogenicity in patients with radiographic AxSpA (Nasonov et al., 2025). While the precise role of the public TRBV9^{+} clonotype in AxSpA pathogenesis

remains under investigation, these initial findings suggest that depletion of the broader TRBV9^{+} T-cell subset can yield meaningful clinical benefit and provide additional support for its potential involvement in disease mechanisms. Collectively, these developments signal a continued shift toward mechanism-informed and precision-oriented therapeutic strategies in AxSpA.

1.2 Mechanistic Basis of Axial Spondyloarthritis

1.2.1 Genetic and molecular mechanisms

1.2.1.1 Genetic susceptibility and GWAS findings

As AxSpA is a relatively recent clinical concept, genetic studies have largely been limited to AS. Among inflammatory diseases, AS exhibits one of the strongest genetic components of susceptibility, with HLA-B*27 showing the most significant association (Brown et al., 1997). Several case-control association studies using genotype arrays have significantly improved our understanding of genetic susceptibility in AS (Stadler et al., 2024). A landmark GWAS by the International Genetics of Ankylosing Spondylitis Consortium (IGAS) in 2013 explained 24.4% of AS heritability, of which HLA-B*27 accounted for 20.1%, and non-MHC loci contributed an additional 4.3% (International Genetics of Ankylosing Spondylitis et al., 2013).

Of the non-MHC contribution, several susceptibility loci map to genes involved in the IL-23 signaling pathway (e.g., *IL23R, TYK2, IL6R, IL27, IL1R2, IL1R1*), aminopeptidase activity (e.g., *ERAP1, ERAP2, LNPEP, PEPPS, TBX21*), and lymphocyte activation and differentiation (e.g., *IL7R, RUNX1, ZMIZ1*), suggesting these pathways play key roles in disease pathogenesis.

Together, these many small-effect variants contribute meaningfully to disease susceptibility. Polygenic risk scores (PRS), which aggregate thousands of GWAS-derived variants, show excellent discriminatory capacity for AS, achieving an AUC of 0.924 in European cohorts, suggesting potential complementary diagnostic value when considered alongside HLA-B*27 testing, sacroiliac MRI, and CRP (Li et al., 2021). Whether PRS adds incremental value in routine clinical practice remains to be established.

AxSpA also shares several clinical manifestations with IBD and uveitis, and these shared features are mirrored by overlapping genetic susceptibility factors observed in related disorders such as CD, UC, and psoriasis, suggesting an underlying molecular framework that differs from traditional, organ-based classifications (Hemminki et al., 2010; International Genetics of Ankylosing Spondylitis et al., 2013). This genetic overlap can also help identify novel susceptibility loci of AS; a cross-phenotype genetic analyses identified 17 novel AS-associated loci and moderate genetic overlap among these clinically related diseases (Ellinghaus et al., 2016).

Given that HLA-B*27 accounts for the large majority of the identified genetic contribution, the following section examines the mechanisms by which this allele may influence AxSpA pathogenesis.

*1.2.1.2 HLA-B*27 and associated theories*

The association between HLA-B*27 and AS was first identified in 1973 (Brewerton et al., 1973; Schlosstein et al., 1973), and remains one of the strongest known links between a common genetic variant and any human disease (Braun & Sieper, 2023). However, how HLA-B*27 contributes to disease pathogenesis is still incompletely understood. Although approximately

90% of patients carry an HLA-B*27 allele, its presence alone is not sufficient to cause disease, as less than 5% of the carriers develop AxSpA, suggesting that additional genetic, environmental and immunological factors are also required for disease onset (Haroon, 2015).

Several hypotheses have been proposed to explain how HLA-B*27 may influence disease pathogenesis, including the arthritogenic peptide, unfolded protein response (UPR), and HLA-B27 homodimer theories. The three pathogenesis theories based on HLA-B*27 are summarized in **Figure 1.2**.

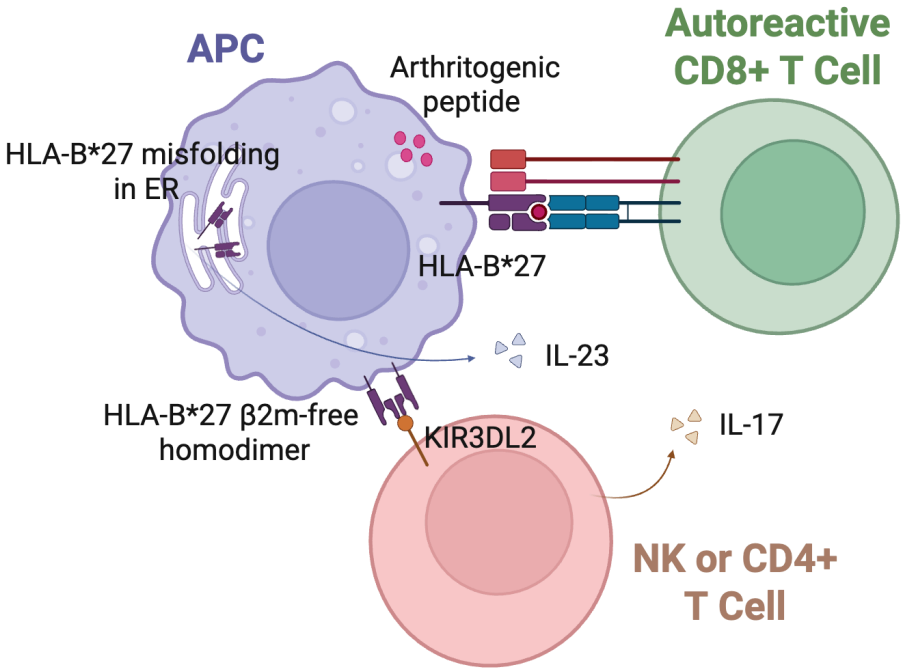


Figure 1.2 Proposed mechanisms linking HLA-B*27 to AxSpA pathogenesis
This schematic illustrates three major hypotheses: the arthritogenic peptide mechanism, in which HLA-B*27 presents arthritogenic peptides to $CD8^+$ T cells; the ER misfolding and unfolded protein response pathway, where improperly folded HLA-B*27 heavy chains induce ER stress and enhance IL-23 production; and the homodimer model, in which cell-surface HLA-B*27 heavy-chain dimers engage KIR-expressing T cells and promote IL-17 secretion. APC: antigen presenting cell. ER: endoplasmic reticulum. Figure created with biorender.com.

The first, the arthritogenic peptide hypothesis, centers on the immunological role of HLA-B*27 as a class I major histocompatibility complex (MHC) molecule that presents intracellular peptides to $CD8^+$ T cells (Benjamin & Parham, 1990). It proposes that AxSpA develops due to a cytotoxic T lymphocyte (CTL)-mediated response against joint-specific self-peptides presented by HLA-B*27 through their shared structural motif. Normally, these peptides are expressed at levels too low to trigger immune reaction, but exposure to bacterial or viral infections with structurally similar peptides could prime T cells to recognizing low-abundance self-peptides, leading to persistent inflammation and tissue damage. Although the specific arthritogenic peptides remain unidentified, indirect evidence supporting molecular mimicry process continues to accumulate. Early studies identified conserved TCR sequences in $CD8^+$ T cells which could recognize self-peptides presented by HLA-B*27 (May et al., 2002). And subsequent work showed that AS patients carrying the HLA-B*2705 subtype displayed significant T-cell responses to a self-peptide resembling an EBV-derived peptide (Fiorillo et al., 2000). More recently, multiple studies have consistently identified a shared $CD8^+$ TCRβ motif encoded by the TRBV9-TRBJ2.3 combination and characterized by CDR3β sequence of the form CASSVG(V/I/L)(Y/F)STDTQYF. This motif and several highly homologous variants have been shown to be enriched in peripheral blood and synovial fluid of HLA-B*27-positive AS patients, compared with both healthy donors and HLA-B*27-positive controls (Faham et al., 2017; Hanson et al., 2020; Komech et al., 2018). Single-cell sequencing identified clonal expansion of $CD8^+TRBV9^+$ T cells expressing the AS-enriched public TRBV9-CDR3β -TRBJ2.3 motif, characterized by CDR3β sequences ending in Y/FSTDTQ. Subsequent peptide-

screening and structural analyses showed that these TCRs can recognize both self-derived and bacterial peptides bound to HLA-B*2705, supporting a cross-reactive mechanism between microbial and self-antigens in HLA-B*27-associated disease (Yang et al., 2022).

The second theory is known as the UPR and endoplasmic reticulum (ER) stress hypothesis. Unlike the arthritogenic peptide model, this hypothesis focuses on the intrinsic biochemical and structural properties of HLA-B*27 rather than the peptides it presents. Experimental evidence suggests that certain structural features of HLA-B*27, particularly within its peptide-binding groove, may predispose a fraction of heavy chains to misfold during assembly in the ER (Mear et al., 1999). This misfolding has been proposed to contribute to ER stress and activation of the UPR, potentially leading to downstream inflammatory effects. Supporting evidence comes primarily from animal models: In transgenic rats, upregulation of HLA-B*27 in bone-marrow derived macrophages was associated with increased misfolded heavy chains and UPR activation (Turner et al., 2007), and this misfolding and the UPR response were correlated with increased production of IL-23 (DeLay et al., 2009), which in turn promotes the differentiation of Th17 cells and activation of other IL-17-producing populations. While these findings support a link between HLA-B27 misfolding and the IL-23/IL-17 axis, direct evidence in human spondyloarthritis remains limited and inconsistent.

A third hypothesis proposes that cell-surface expression of HLA-B27 heavy-chain homodimers, which lack β2m, may aberrantly interact with immune receptors on lymphocytes. Recombinant HLA-B27 homodimers have been shown to bind killer immunoglobulin-like receptors (KIRs) and leukocyte

immunoglobulin-like receptors (LILRs). Notably, $CD4^+$ T cells expressing KIR3DL2, a receptor that selectively recognizes HLA-B27 homodimers, are increased in the blood of AxSpA patients (Chan et al., 2005). These $KIR3DL2^+CD4^+$ T cells produce IL-17 upon stimulation with cells expressing HLA-B27 dimers (Bowness et al., 2011), providing a possible mechanism linking HLA-B27 dimers to IL-17-mediated inflammation in AxSpA.

Among these three hypotheses, the single-cell work presented in this thesis primarily provides cellular correlates consistent with the arthritogenic peptide hypothesis, through transcriptomic and phenotypic profiling of $CD8^+$ T cells and other antigen-responsive populations across AxSpA-affected tissues. Signatures potentially related to ER stress and the unfolded protein response are examined indirectly through targeted pathway analysis in relevant cell types. The homodimer-mediated mechanism is not directly interrogated, as it requires surface-level biochemical assays outside the scope of the present single-cell analyses.

1.2.1.3 Aminopeptidase and antigen processing

Beyond HLA-B*27 itself, genetic studies have revealed strong associations between AxSpA and genes encoding endoplasmic reticulum aminopeptidases, particularly ERAP1 and ERAP2, which play critical roles in peptide trimming and antigen processing within the MHC class I pathway. These enzymes refine precursor peptides to optimal lengths before loading onto HLA-B*27, thereby shaping the repertoire of antigens available for immune recognition (Hattori & Tsujimoto, 2013; Stratikos & Stern, 2013). The association of ERAP1 with AS occurs exclusively in HLA-B*27 positive individuals, indicating a gene-gene epistatic interaction (Evans et al., 2011). Functional

studies have shown that AS-risk *ERAP1* variants can alter trimming efficiency and substrate specificity, leading to changes in the HLA-B*27 peptidome and peptide stability (Garcia-Medel et al., 2012; Sanz-Bravo et al., 2015). These alterations may influence antigen presentation and immune tolerance, providing a potential mechanistic link between *ERAP1* polymorphisms and AxSpA pathogenesis, although the precise pathways remain to be clarified.

In contrast, the association of *ERAP2* with AS appears independent from HLA-B*27, occurring in both B*27-positive and -negative individuals (International Genetics of Ankylosing Spondylitis et al., 2013; Robinson et al., 2015). Experimental evidence suggests that ERAP2 may act either in coordination with or independently of ERAP1 to influence the HLA-B*27 peptidome, although its overall impact on HLA-B*27 folding, ER stress, and cytokine expression remains uncertain due to conflicting results (Martin-Esteban et al., 2016; Zhang et al., 2017). Collectively, these observations support a model in which aminopeptidase polymorphisms alter peptide processing and antigen presentation, thereby contributing to immune dysregulation and disease susceptibility in AxSpA. However, the precise mechanisms remain unclear, as it is still uncertain whether ERAP1 and ERAP2 contribute by disrupting the normal HLA-B*27 peptidome, generating pathogenic peptides, or affecting the stability and dimerization of HLA-B27 complexes.

1.2.1.4 The IL-23/IL-17 axis

The IL-23/IL-17 axis plays a pivotal role in the pathogenesis of chronic inflammatory diseases, and accumulating genetic, experimental, and clinical evidence supports its critical involvement in AxSpA.

Genetic studies first implicated this axis through variants in *IL23R* and downstream signaling molecules such as STAT3 and TYK2, which have been repeatedly associated with AS, linking inherited susceptibility to dysregulated IL-23 signaling (Ellinghaus et al., 2016; Wellcome Trust Case Control et al., 2007). Dysregulation of this pathway has been demonstrated in AS across both peripheral blood and inflamed tissues, with elevated IL-17 and IL-23 levels in serum and synovial fluid (Jandus et al., 2008; Mei et al., 2011; Wendling et al., 2007) and increased frequencies of Th17, KIR3DL2^{+} CD4^{+}, and IL-23R^{+} γδ T cells (Bowness et al., 2011; Jandus et al., 2008; Kenna et al., 2012; Zhang et al., 2012). Within inflamed tissues, IL-17 expression has been detected in mast cells within peripheral synovial tissue from SpA patients (Noordenbos et al., 2012), and in neutrophils and mononuclear cells in the facet joints of AS patients (Appel et al., 2011). In the gut, IL-23 expression has been localized to infiltrating monocytes and Paneth cells (Ciccia et al., 2009), while in spinal facet joints, IL-23 is overexpressed in bone marrow myeloid cells compared with osteoarthritis controls (Appel et al., 2013). Together, these findings indicate widespread activation of the IL-23/IL-17 pathway across multiple disease-relevant tissues.

1.2.2 Environmental and microbial Influences

1.2.2.1 Gut-joint axis and intestinal dysbiosis

A strong clinical and genetic link between gut inflammation and AxSpA highlights the importance of the gut-joint axis in disease pathogenesis. Clinically, subclinical gut inflammation is present in approximately *50*-60% of AxSpA patients (De Keyser et al., 1998), and 5-10% *develop diagnosed IBD (Mielants et al., 1991; Mielants et al., 1995); genetically, strong* overlap between

SpA and IBD further supports this association with shared polymorphisms in genes involved in IL-23/IL-17 signaling such as *IL-23R, IL-12B*, and *STAT3 (Ellinghaus et al., 2016; Reveille, 2012)*. Furthermore, the clinical observation of reactive arthritis following enteric infections with *specific pathogens* strengthens the gut-joint inflammatory connection *(Gill & Rosenbaum, 2020).*

Alterations in the gut microbiome, termed dysbiosis, have been consistently observed in AxSpA patients and strongly support the existence of a gut-joint axis. Compared to healthy controls, AxSpA patients show reduced microbial diversity and a distinct microbial composition characterized by shifts in specific bacterial groups (Breban et al., 2017; Gracey et al., 2019; Yin et al., 2020; Zhou et al., 2020). Genetic factors such as HLA-B*27 positivity influence the microbiome even before disease onset, indicating that dysbiosis may occur early and actively contribute to disease risk rather than simply being a result of inflammation (Asquith et al., 2017; Asquith et al., 2019). Supporting this, AxSpA patients have enriched IgA-coated microbes associated with inflammatory and metabolic changes, further suggesting dysbiosis plays an important role in AxSpA pathogenesis (Gill et al., 2022). Understanding these microbiome changes could open new possibilities for microbiome-targeted therapies in AxSpA.

Microbial dysbiosis has been associated with alterations in gut barrier function in AxSpA, including increased intestinal permeability that may facilitate greater exposure of the mucosal immune system to luminal antigens (Asquith et al., 2014). Consistent with this, AxSpA patients show molecular evidence for barrier impairment, with decreased levels of essential tight junction proteins such as claudin-4 and occludin alongside elevated zonulin, a marker indicating

enhanced gut permeability (Ciccia et al., 2017; Rizzo et al., 2018). Ileal biopsies from AxSpA patients further reveal mucosal and endothelial barrier disruption accompanied by increased bacterial products in circulation (Ciccia et al., 2017). Importantly, similar barrier dysfunction observed in healthy relatives of AxSpA patients suggests that gut leakiness may precede clinical disease onset rather than being a consequence of inflammation (Martinez-Gonzalez et al., 1994).

Microbial dysbiosis also significantly influence mucosal immunity through activation of both innate and adaptive pathways within the gut in AxSpA patients. In response to altered microbiota, Paneth cells produce higher levels of IL-23 and IL-7, promoting differentiation of ILC3 and mucosal-associated invariant T (MAIT) cells, which further contribute to local inflammation through IL-17 and IL-22 secretion (Ciccia et al., 2012; Ciccia et al., 2009; Ciccia et al., 2016). Despite elevated IL-23 expression in the gut, Th17-mediated IL-17 production is not consistently observed, potentially due to a compensatory rise in regulatory T-cells and IL-10 production that may limit Th17 responses in the gut (Ciccia et al., 2010). Dysbiosis may also be sustained by insufficient cleaning of bacteria by specific gut macrophage subsets, and may perpetuate the dysbiosis-driven gut inflammation in AS patients (Ciccia et al., 2014; Ciccia et al., 2016).

Emerging evidence suggests that mucosal immune activation in AxSpA may contribute to inflammatory responses beyond the gut. Pro-inflammatory $CX3CR1^{+}CD59^{+}TL1A^{+}IL\text{-}23^{+}$ mononuclear phagocytes, which are expanded in the inflamed ileum of AxSpA patients, are also detectable in peripheral blood, synovial fluid, and bone marrow, often expressing the gut-homing marker CCR9 (Ciccia et al., 2018). These cells can enhance ILC3 activation in vitro,

suggesting a possible mechanism by which gut-primed innate signals could be propagated systemically. Experimental work further shows that IL-23 can drive enthesitis and new bone formation in a mouse model by acting on IL-23R^{+}RORγt^{+} entheseal resident T cells (Sherlock et al., 2012), illustrating how an IL-23-rich environment may translate into site-specific pathology. In addition, AxSpA-enriched public TRBV9-CDR3β-TRBJ2.3 motif, together with its peptide-selection hits from enteric bacteria, raise the possibility that microbial antigen exposure precedes or accompanies their presence at the gut (Yang et al., 2022). Collectively, these findings support a model in which mucosal immune activation and antigen-driven lymphocyte responses may influence inflammatory processes beyond the gut, although the precise pathways linking these compartments remain to be clarified.

1.2.2.2 Mechanical stress and local tissue responses

Mechanical stress and local tissue responses are potential key contributors to AxSpA pathogenesis, particularly at entheseal sites, which experience significant mechanical strain due to their anatomical location and function. Entheses, the sites of tendon and ligament attachment to bone, are frequently inflamed in AxSpA, and patients often present with either axial or peripheral enthesitis. Evidence from animal models further support the pivotal role of mechanical stress, demonstrating that mechanical strain alone can trigger entheseal inflammation and subsequent SpA-related new bone formation, even in the absence of overt injury (Jacques et al., 2014).

At the cellular level, mechanically stressed entheses exhibit increased vascularization, macrophage infiltration, and disrupted tissue architecture (McGonagle et al., 2002). Local immune responses are further amplified by

tissue resident immune cells, including ILC3s, γδT-cells, MAIT cells, and myeloid cells such as neutrophils and monocytes, which can produce pro-inflammatory cytokines including IL-23 and IL-17 upon stimulation (Bridgewood et al., 2019; Cuthbert et al., 2019; Stavre et al., 2022; Zhang et al., 2017). In SpA mice models, IL-23 responsive cells, particularly γδT-cells, have been identified as primary sources of IL-17 production at entheseal sites, establishing a critical link between mechanical stress and type 17 immune activation (Reinhardt et al., 2016; Sherlock et al., 2012). These findings support a model in which genetic predispositions, such as *HLA-B*27* or *IL23R* polymorphisms, may amplify susceptibility to mechanical stress-induced inflammation, potentially aggravated by gut dysbiosis or barrier dysfunction.

Collectively, these insights suggest that microdamage induced by mechanical stress at the enthesis acts as a key trigger for localized immune activation in AxSpA, promoting inflammation that can evolve into chronic disease and structural joint damage. Further investigation of these mechanisms in human AxSpA tissues is essential for developing targeted therapeutic strategies.

1.3 Multi-omics Approaches in Axial Spondyloarthritis Research

1.3.1 Single-cell discoveries in AxSpA

Given the complex genetic and environmental factors that shape the immunopathogenic landscape of AxSpA, single-cell multi-omics approaches are increasingly used to dissect the cellular and molecular mechanisms underlying disease. Recent advances in high-throughput single-cell technologies including

single-cell RNA sequencing (scRNA-seq), single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq), and cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) enable the comprehensive profiling of individual immune cells at unprecedented resolution and have become powerful tools for mapping cellular networks in complex immune-mediated disorders. In AxSpA, the application of these technologies has substantially deepened our understanding of the immune landscape and revealed key cellular mechanisms driving disease progression.

In the peripheral blood context, several key immune alterations have been observed in AxSpA. Yu et al. applied scATAC-seq and reported significant changes in the proportion of CD8+ T cells, identifying differential accessibility of transcription factor (TF) motifs implicated in IL-17 and TNF signaling pathways in T cells, monocytes, and dendritic cells (Yu et al., 2020). Liu et al. characterized monocytes and NK cells using scRNA-seq and highlighted $CD14^{low}CD16^{high}$ monocytes as potential predictors of response to anti-TNF-α therapy and implicated memory-like NK cells appear to play a role in disease pathogenesis independent of TNF-α treatment (Liu et al., 2021). A decrease in NK cell cytotoxicity profiles has also been noted in another research, along with distinct RNA-binding protein (RBP) regulatory patterns, highlighting immune dysregulation in the peripheral blood of AS patients (Ren et al., 2022). Moreover, Xu et al. further integrated scRNA-seq and scATAC-seq, demonstrating abnormal expression and chromatin accessibility in the NF-κB pathway, particularly involving in $CD8^+$ T cells, suggesting a potential regulatory mechanism driving AxSpA progression. (Xu et al., 2021). CITE-seq analysis further revealed overexpression of CD52 across multiple immune cell subsets,

as well as alterations in cytotoxic gene expression (*GZMH, GZMB, NKG7*) in NK and $CD8^+$ TEM cells alongside impaired functionality of regulatory T cells indicated by decreased *CD39* expression (Alber et al., 2022).

Altered immune cell frequencies have also been identified with scRNA-seq approach: increased numbers of B cells, dendritic cells, $CD14^+$ monocytes, and $CD8^+$ T cells were observed in peripheral blood from AS patients, with several RBP genes involved in regulating immune functions, which could be implicated in the disease's progression (Ren et al., 2024). Furthermore, anti-TNFα treatment was found to impact peripheral immune composition by reducing inflammatory interactions between classical monocytes and granulocytes, through the CXCL8/2-CXCR1/2 signaling pathway (Zhao et al., 2023).

Within inflamed tissues, single-cell investigations remain limited but have begun to clarify tissue-specific immune environments in AxSpA. Synovial fluid analysis identified distinct regulatory T cell subsets, including a $CD8^+$ Treg subset and a Th17-like $RORC^+$ Treg subset that expressed IL-10 and LAG-3 (Simone et al., 2021). These synovial Tregs showed upregulation of interferon signatures and TNF receptor genes, alongside marked clonal expansion. Expanded $CD8^+$ T cell clonotypes with the HLA-B*27 specific motif were detected in synovial fluid, expressing markers suggestive of mucosal differentiation and antigen specificity (Paley et al., 2024). Moreover, a clonally expanded CTL subset in synovial fluid was found to express inhibitory receptors (*PD-1, TIGIT, LAG-3*) while retaining effector functions, indicating resistance to exhaustion as a potential mechanism driving chronic inflammation in AS (Tang et al., 2025).

Collectively, these single-cell studies have revealed distinct immune alterations at both peripheral and tissue levels in AxSpA, highlighting dysregulated immune cell subsets, altered transcriptional and epigenetic landscapes, and tissue-specific immune responses. Such insights offer critical guidance for understanding disease mechanisms and identifying therapeutic opportunities.

1.3.2 Immune repertoire insights in AxSpA

Immune-repertoire analysis encompassing T-cell receptor (TCR) and B-cell receptor (BCR) sequencing provides critical insight into clonal expansions of lymphocytes, antigen-specific responses and potential pathogenic mimicry in autoimmune and inflammatory disorders. With the methods of single-cell TCR/BCR sequencing, researchers can now integrate receptor clonotypes with transcriptomic and phenotypic states, offering a high-resolution view of adaptive immune dynamics in disease. In AxSpA, the well-established “arthritogenic peptide” hypothesis proposes that specific peptides presented by HLA-B*27 molecules could activate autoreactive T cells and lead to the inflammation, which has been described above. Consequently, much of the immune repertoire research in AxSpA focused on determining whether T cells undergo disease-specific clonal expansion in HLA-B*27 positive individuals.

Bulk TCR repertoire sequencing studies have provided supporting evidence for this hypothesis. Faham et al. conducted high-throughput profiling of peripheral blood TCR repertoires in HLA-B*27-positive and -negative AS patients, as well as HLA-B*27-positive healthy controls, and identified distinct TCRβ CDR3 motifs specifically enriched in B*27-positive AS patients (Faham et al., 2017). Subsequent studies have reinforced these findings: Komech et al.

conducted deep TCRβ sequencing of peripheral blood and synovial fluid from AS patients, identifying characteristic disease-associated TCRβ motifs (Komech et al., 2018), and Hanson et al. profiled both CD4 and CD8 TCR repertoires in AS patients and HLA-B*27 matched healthy controls, independently confirming enrichment of previously identified disease-associated CDR3 motifs, notably CASSVG(V/I/L)(Y/F)STDTQYF and CASSVGLFSTDTQYF (Hanson et al., 2020). Additional supportive studies include identifications of oligo-clonal expansions involving both $CD4^+$ and $CD8^+$ T cells sharing identical TCRs enriched in SpA joint fluid (Zheng et al., 2019), as well as dynamic changes in TCR/BCR V(D)J usage and clonal diversity across disease stages and remission (Ke et al., 2025).

At single-cell resolution, TCR sequencing studies have further clarified antigen-driven clonal expansions in AxSpA. Yang et al. identified a conserved $CD8^+$ TCRβ motif (TRBV9-CDR3β: Y/FSTDTQ-BJ2.3), consistently paired with TCRα variable region TRAV21, which was enriched in synovial fluid of HLA-B*27-positive AxSpA patients and ocular fluid from patients with acute anterior uveitis, providing evidence of shared T-cell clonotypes across HLA-B*27-associated conditions (Yang et al., 2022). Similarly, Paley et al. identified mucosally differentiated $CD8^+$ T cells recognizing bacterial peptide YeiH presented by HLA-B*27, supporting intestinal origin and antigen-specific expansion of these cells in related B*27-associated conditions (Paley et al., 2024).

Together, these bulk and single-cell immune repertoire analyses suggest that AxSpA involves antigen-driven T-cell expansions characterized by recurrent TCRβ motifs, particularly enriched in HLA-B*27-positive patients. The

identification of similar CDR3 sequences across different HLA-B*27-associated conditions further supported the possibility of systemic antigenic priming, highlighting TCR/BCR sequencing as a promising approach to explore potential biomarkers and therapeutic targets in AxSpA. Although immune repertoire analysis provides an important window into antigen-driven hypotheses in AxSpA, it is not a primary focus of this thesis; selected receptor-level observations are integrated with the cellular findings in the overall discussion, and broader repertoire-level analysis is noted as a future direction.

1.3.3 Cross-tissue atlases within and beyond AxSpA

Despite substantial progress in defining the immunological features of AxSpA, lots of insights have been derived from peripheral blood, where many disease-associated signatures are weak or absent. Evidence from inflammatory diseases increasingly highlights that key pathogenic processes often localize within affected tissues rather than in the circulation. Tissue-resident T cells provide a clear example: they display transcriptional programs that are distinct from circulating T cells (Mackay et al., 2013), many remain permanently within tissues and do not recirculate to blood (Gebhardt et al., 2013; Jiang et al., 2012; Mueller et al., 2013), and pathogenic immune memory can persist in tissue even when not markable from the blood. Consistent with this idea, scTCR sequencing in HLA-B*27-positive individuals with AS and/or AAU showed that expanded TCR clones (including those bearing the AS-associated TRBV9-Y/FSTDTQ-BJ2.3 motif) were enriched by at least 10-fold in synovial fluid and by 10- to 100-fold in ocular tissue compared with paired blood samples, indicating that major clonal expansions can be concentrated within inflamed tissues and may be under-represented or undetectable in peripheral blood

(Yang et al., 2022). Beyond AxSpA, studies in other immune-mediated diseases further illustrate the limitations of blood-based profiling. In rheumatoid arthritis, GZMK-expressing $CD8^+$ T cells are abundantly concentrated within inflamed synovial tissue rather than circulation, where they drive local complement activation (Donado et al., 2025). Similarly, a recent study of recurrent airway inflammation showed that GZMK-expressing $CD8^+$ T cells and complement-activating effector activity are strongly enriched in diseased airway tissues, while GZMK signals in the blood are minimal, demonstrating how critical immune drivers can be missed when analyses focus solely on peripheral blood (Lan et al., 2025). This highlights the importance of studying immune processes directly within affected anatomical sites.

Across diverse immune-mediated diseases, cross-tissue atlases have broadened our understanding by revealing how stromal and immune programs vary across organs and jointly contribute to pathogenesis. Large-scale reference maps, such as the GTEx single-nucleus cross-tissue atlas, demonstrated that disease-associated genes map to distinct cell types across organs, revealing tissue-specific and shared pathogenic programs (Eraslan et al., 2022). Cross-tissue fibroblast atlases showed that pathological conditions reshape fibroblast states across organs, revealing shared immune-modulating and tissue-remodeling fibroblast subsets whose distributions differ between inflammation and cancer (Gao et al., 2024), and demonstrating that chronic inflammatory diseases converge on common pathogenic fibroblast phenotypes across tissues (Korsunsky et al., 2022). Cross-tissue monocyte and macrophage atlases further show that integrating samples from multiple organs is essential for distinguishing universal activation states from tissue-restricted

adaptations (Mulder et al., 2021). Finally, integrative joint studies, such as the cross-tissue knee OA atlas, illustrate how pathological remodelling reflects coordinated changes across multiple tissues rather than isolated alterations in a single compartment (Raut et al., 2025).

Collectively, these findings indicate that cross-tissue comparison uncovers shared and unique pathogenic modules, tissue-specific differentiation trajectories, and multicellular interactions that are not detectable through single-tissue profiling alone, which further support the need for multi-tissue, single-cell approaches to uncover disease mechanisms in AxSpA. Together, this motivates a comprehensive, multimodal cross-tissue atlas of AxSpA to define shared and site-specific immune states across blood, musculoskeletal tissues, and extra-articular sites relevant to disease pathogenesis.

1.4 Thesis Aims

1.4.1 To investigate the peripheral immune landscape of AxSpA and evaluate the impact of TNF inhibitor therapy on it by analysing single-cell datasets derived from peripheral blood.

1.4.2 To construct a comprehensive, multi-modal single-cell atlas of AxSpA across multiple affected tissues including blood, synovial fluid, spine, ileum, sacroiliac joint, and eye, together with matched controls where available, providing a resource to explore cross-tissue immune patterns.

1.4.3 To identify and characterise both tissue-specific and tissue-shared immune cell populations within the multi-modal atlas, in order to determine key cellular players and their potential roles in orchestrating immune responses across disease-relevant sites.

1.4.4 To delineate disease-associated changes at the transcriptomic and proteomic levels within primary tissues from AxSpA patients compared with appropriate controls, thereby enhancing understanding of the immunopathological mechanisms underlying AxSpA.

1.5 References

Alber, S., Kumar, S., Liu, J., Huang, Z. M., Paez, D., Hong, J., Chang, H. W., Bhutani, T., Gensler, L. S., & Liao, W. (2022). Single Cell Transcriptome and Surface Epitope Analysis of Ankylosing Spondylitis Facilitates Disease Classification by Machine Learning. *Front Immunol*, *13*, 838636. https://doi.org/10.3389/fimmu.2022.838636

Appel, H., Maier, R., Bleil, J., Hempfing, A., Loddenkemper, C., Schlichting, U., Syrbe, U., & Sieper, J. (2013). In situ analysis of interleukin-23- and interleukin-12-positive cells in the spine of patients with ankylosing spondylitis. *Arthritis Rheum*, *65*(6), 1522–1529. https://doi.org/10.1002/art.37937

Appel, H., Maier, R., Wu, P., Scheer, R., Hempfing, A., Kayser, R., Thiel, A., Radbruch, A., Loddenkemper, C., & Sieper, J. (2011). Analysis of IL-17(+) cells in facet joints of patients with spondyloarthritis suggests that the innate immune pathway might be of greater relevance than the Th17-mediated adaptive immune response. *Arthritis Res Ther*, *13*(3), R95. https://doi.org/10.1186/ar3370

Asquith, M., Davin, S., Stauffer, P., Michell, C., Janowitz, C., Lin, P., Ensign-Lewis, J., Kinchen, J. M., Koop, D. R., & Rosenbaum, J. T. (2017). Intestinal Metabolites Are Profoundly Altered in the Context of HLA-B27 Expression and Functionally Modulate Disease in a Rat Model of Spondyloarthritis. *Arthritis Rheumatol*, *69*(10), 1984–1995. https://doi.org/10.1002/art.40183

Asquith, M., Elewaut, D., Lin, P., & Rosenbaum, J. T. (2014). The role of the gut and microbes in the pathogenesis of spondyloarthritis. *Best Pract Res Clin Rheumatol*, *28*(5), 687–702. https://doi.org/10.1016/j.berh.2014.10.018

Asquith, M., Sternes, P. R., Costello, M. E., Karstens, L., Diamond, S., Martin, T. M., Li, Z., Marshall, M. S., Spector, T. D., le Cao, K. A., Rosenbaum, J. T., & Brown, M. A. (2019). HLA Alleles Associated With Risk of Ankylosing Spondylitis and Rheumatoid Arthritis Influence the Gut Microbiome. *Arthritis Rheumatol*, *71*(10), 1642–1650. https://doi.org/10.1002/art.40917

Baeten, D., Ostergaard, M., Wei, J. C., Sieper, J., Jarvinen, P., Tam, L. S., Salvarani, C., Kim, T. H., Solinger, A., Datsenko, Y., Pamulapati, C., Visvanathan, S., Hall, D. B., Aslanyan, S., Scholl, P., & Padula, S. J. (2018). Risankizumab, an IL-23 inhibitor, for ankylosing spondylitis: results of a randomised, double-blind, placebo-controlled, proof-of-

concept, dose-finding phase 2 study. *Ann Rheum Dis*, *77*(9), 1295–1302. https://doi.org/10.1136/annrheumdis-2018-213328

Benjamin, R., & Parham, P. (1990). Guilt by association: HLA-B27 and ankylosing spondylitis. *Immunol Today*, *11*(4), 137–142. https://doi.org/10.1016/0167-5699(90)90051-a

Bowness, P., Ridley, A., Shaw, J., Chan, A. T., Wong-Baeza, I., Fleming, M., Cummings, F., McMichael, A., & Kollnberger, S. (2011). Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis. *J Immunol*, *186*(4), 2672–2680. https://doi.org/10.4049/jimmunol.1002653

Braun, J., & Sieper, J. (2023). Fifty years after the discovery of the association of HLA B27 with ankylosing spondylitis. *RMD Open*, *9*(3). https://doi.org/10.1136/rmdopen-2023-003102

Breban, M., Tap, J., Leboime, A., Said-Nahal, R., Langella, P., Chiocchia, G., Furet, J. P., & Sokol, H. (2017). Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. *Ann Rheum Dis*, *76*(9), 1614–1622. https://doi.org/10.1136/annrheumdis-2016-211064

Brewerton, D. A., Hart, F. D., Nicholls, A., Caffrey, M., James, D. C., & Sturrock, R. D. (1973). Ankylosing spondylitis and HL-A 27. *Lancet*, *1*(7809), 904–907. https://doi.org/10.1016/s0140-6736(73)91360-3

Bridgewood, C., Watad, A., Russell, T., Palmer, T. M., Marzo-Ortega, H., Khan, A., Millner, P. A., Dunsmuir, R., Rao, A., Loughenbury, P., Wittmann, M., Cuthbert, R. J., & McGonagle, D. G. (2019). Identification of myeloid cells in the human enthesis as the main source of local IL-23 production. *Ann Rheum Dis*, *78*(7), 929–933. https://doi.org/10.1136/annrheumdis-2018-214944

Britanova, O. V., Lupyr, K. R., Staroverov, D. B., Shagina, I. A., Aleksandrov, A. A., Ustyugov, Y. Y., Somov, D. V., Klimenko, A., Shostak, N. A., Zvyagin, I. V., Stepanov, A. V., Merzlyak, E. M., Davydov, A. N., Izraelson, M., Egorov, E. S., Bogdanova, E. A., Vladimirova, A. K., Iakovlev, P. A., Fedorenko, D. A.,…Chudakov, D. M. (2023). Targeted depletion of TRBV9(+) T cells as immunotherapy in a patient with ankylosing spondylitis. *Nat Med*, *29*(11), 2731–2736. https://doi.org/10.1038/s41591-023-02613-z

Brown, M. A., Kennedy, L. G., MacGregor, A. J., Darke, C., Duncan, E., Shatford, J. L., Taylor, A., Calin, A., & Wordsworth, P. (1997). Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. *Arthritis Rheum*, *40*(10), 1823–1828. https://doi.org/10.1002/art.1780401015

Chan, A. T., Kollnberger, S. D., Wedderburn, L. R., & Bowness, P. (2005). Expansion and enhanced survival of natural killer cells expressing the killer immunoglobulin-like receptor KIR3DL2 in spondylarthritis. *Arthritis Rheum*, *52*(11), 3586–3595. https://doi.org/10.1002/art.21395

Ciccia, F., Accardo-Palumbo, A., Alessandro, R., Rizzo, A., Principe, S., Peralta, S., Raiata, F., Giardina, A., De Leo, G., & Triolo, G. (2012). Interleukin-22 and interleukin-22-producing NKp44+ natural killer cells in subclinical gut inflammation in ankylosing spondylitis. *Arthritis Rheum*, *64*(6), 1869–1878. https://doi.org/10.1002/art.34355

Ciccia, F., Accardo-Palumbo, A., Giardina, A., Di Maggio, P., Principato, A., Bombardieri, M., Rizzo, A., Alessandro, R., Ferrante, A., Principe, S.,

Peralta, S., Conte, F., Drago, S., Craxi, A., De Leo, G., & Triolo, G. (2010). Expansion of intestinal CD4+CD25(high) Treg cells in patients with ankylosing spondylitis: a putative role for interleukin-10 in preventing intestinal Th17 response. *Arthritis Rheum*, *62*(12), 3625–3634. https://doi.org/10.1002/art.27699

Ciccia, F., Alessandro, R., Rizzo, A., Accardo-Palumbo, A., Raimondo, S., Raiata, F., Guggino, G., Giardina, A., De Leo, G., Sireci, G., & Triolo, G. (2014). Macrophage phenotype in the subclinical gut inflammation of patients with ankylosing spondylitis. *Rheumatology (Oxford)*, *53*(1), 104–113. https://doi.org/10.1093/rheumatology/ket323

Ciccia, F., Bombardieri, M., Principato, A., Giardina, A., Tripodo, C., Porcasi, R., Peralta, S., Franco, V., Giardina, E., Craxi, A., Pitzalis, C., & Triolo, G. (2009). Overexpression of interleukin-23, but not interleukin-17, as an immunologic signature of subclinical intestinal inflammation in ankylosing spondylitis. *Arthritis Rheum*, *60*(4), 955–965. https://doi.org/10.1002/art.24389

Ciccia, F., Guggino, G., Rizzo, A., Alessandro, R., Luchetti, M. M., Milling, S., Saieva, L., Cypers, H., Stampone, T., Di Benedetto, P., Gabrielli, A., Fasano, A., Elewaut, D., & Triolo, G. (2017). Dysbiosis and zonulin upregulation alter gut epithelial and vascular barriers in patients with ankylosing spondylitis. *Ann Rheum Dis*, *76*(6), 1123–1132. https://doi.org/10.1136/annrheumdis-2016-210000

Ciccia, F., Guggino, G., Zeng, M., Thomas, R., Ranganathan, V., Rahman, A., Alessandro, R., Rizzo, A., Saieva, L., Macaluso, F., Peralta, S., Di Liberto, D., Dieli, F., Cipriani, P., Giacomelli, R., Baeten, D., & Haroon, N. (2018). Proinflammatory CX3CR1+CD59+Tumor Necrosis Factor-Like Molecule 1A+Interleukin-23+ Monocytes Are Expanded in Patients With Ankylosing Spondylitis and Modulate Innate Lymphoid Cell 3 Immune Functions. *Arthritis Rheumatol*, *70*(12), 2003–2013. https://doi.org/10.1002/art.40582

Ciccia, F., Rizzo, A., & Triolo, G. (2016). Subclinical gut inflammation in ankylosing spondylitis. *Curr Opin Rheumatol*, *28*(1), 89–96. https://doi.org/10.1097/BOR.0000000000000239

Cuthbert, R. J., Watad, A., Fragkakis, E. M., Dunsmuir, R., Loughenbury, P., Khan, A., Millner, P. A., Davison, A., Marzo-Ortega, H., Newton, D., Bridgewood, C., & McGonagle, D. G. (2019). Evidence that tissue resident human enthesis gammadeltaT-cells can produce IL-17A independently of IL-23R transcript expression. *Ann Rheum Dis*, *78*(11), 1559–1565. https://doi.org/10.1136/annrheumdis-2019-215210

Danve, A., & Deodhar, A. (2022). Treatment of axial spondyloarthritis: an update. *Nat Rev Rheumatol*, *18*(4), 205–216. https://doi.org/10.1038/s41584-022-00761-z

De Keyser, F., Elewaut, D., De Vos, M., De Vlam, K., Cuvelier, C., Mielants, H., & Veys, E. M. (1998). Bowel inflammation and the spondyloarthropathies. *Rheum Dis Clin North Am*, *24*(4), 785–813, ix–x. https://doi.org/10.1016/s0889-857x(05)70042-9

de Winter, J. J., van Mens, L. J., van der Heijde, D., Landewe, R., & Baeten, D. L. (2016). Prevalence of peripheral and extra-articular disease in ankylosing spondylitis versus non-radiographic axial spondyloarthritis: a

meta-analysis. *Arthritis Res Ther*, *18*(1), 196. https://doi.org/10.1186/s13075-016-1093-z

DeLay, M. L., Turner, M. J., Klenk, E. I., Smith, J. A., Sowders, D. P., & Colbert, R. A. (2009). HLA-B27 misfolding and the unfolded protein response augment interleukin-23 production and are associated with Th17 activation in transgenic rats. *Arthritis Rheum*, *60*(9), 2633–2643. https://doi.org/10.1002/art.24763

Deodhar, A., Gensler, L. S., Sieper, J., Clark, M., Calderon, C., Wang, Y., Zhou, Y., Leu, J. H., Campbell, K., Sweet, K., Harrison, D. D., Hsia, E. C., & van der Heijde, D. (2019). Three Multicenter, Randomized, Double-Blind, Placebo-Controlled Studies Evaluating the Efficacy and Safety of Ustekinumab in Axial Spondyloarthritis. *Arthritis Rheumatol*, *71*(2), 258–270. https://doi.org/10.1002/art.40728

Donado, C. A., Theisen, E., Zhang, F., Nathan, A., Fairfield, M. L., Rupani, K. V., Jones, D., Johannes, K. P., Accelerating Medicines Partnership, R. A. S. L. E. N., Raychaudhuri, S., Dwyer, D. F., Jonsson, A. H., & Brenner, M. B. (2025). Granzyme K activates the entire complement cascade. *Nature*, *641*(8061), 211–221. https://doi.org/10.1038/s41586-025-08713-9

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., Park, Y. R., Raychaudhuri, S., Pouget, J. G., Hubenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H. J., Franke, L., Pers, T. H., Weersma, R. K., Collij, V.,…Franke, A. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*, *48*(5), 510–518. https://doi.org/10.1038/ng.3528

Eraslan, G., Drokhlyansky, E., Anand, S., Fiskin, E., Subramanian, A., Slyper, M., Wang, J., Van Wittenberghe, N., Rouhana, J. M., Waldman, J., Ashenberg, O., Lek, M., Dionne, D., Win, T. S., Cuoco, M. S., Kuksenko, O., Tsankov, A. M., Branton, P. A., Marshall, J. L.,…Regev, A. (2022). Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*, *376*(6594), eabl4290. https://doi.org/10.1126/science.abl4290

Evans, D. M., Spencer, C. C., Pointon, J. J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A., Pirinen, M., Stone, M. A., Appleton, L., Moutsianas, L., Leslie, S., Wordsworth, T., Kenna, T. J., Karaderi, T., Thomas, G. P., Ward, M. M., Weisman, M. H.,…Wellcome Trust Case Control, C. (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet*, *43*(8), 761–767. https://doi.org/10.1038/ng.873

Faham, M., Carlton, V., Moorhead, M., Zheng, J., Klinger, M., Pepin, F., Asbury, T., Vignali, M., Emerson, R. O., Robins, H. S., Ireland, J., Baechler-Gillespie, E., & Inman, R. D. (2017). Discovery of T Cell Receptor beta Motifs Specific to HLA-B27-Positive Ankylosing Spondylitis by Deep Repertoire Sequence Analysis. *Arthritis Rheumatol*, *69*(4), 774–784. https://doi.org/10.1002/art.40028

Fiorillo, M. T., Maragno, M., Butler, R., Dupuis, M. L., & Sorrentino, R. (2000). CD8(+) T-cell autoreactivity to an HLA-B27-restricted self-epitope correlates with ankylosing spondylitis. *J Clin Invest*, *106*(1), 47–53. https://doi.org/10.1172/JCI9295

Gao, Y., Li, J., Cheng, W., Diao, T., Liu, H., Bo, Y., Liu, C., Zhou, W., Chen, M., Zhang, Y., Liu, Z., Han, W., Chen, R., Peng, J., Zhu, L., Hou, W., & Zhang, Z. (2024). Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation. *Cancer Cell*, *42*(10), 1764–1783 e1710. https://doi.org/10.1016/j.ccell.2024.08.020

Garcia-Medel, N., Sanz-Bravo, A., Van Nguyen, D., Galocha, B., Gomez-Molina, P., Martin-Esteban, A., Alvarez-Navarro, C., & de Castro, J. A. (2012). Functional interaction of the ankylosing spondylitis-associated endoplasmic reticulum aminopeptidase 1 polymorphism and HLA-B27 in vivo. *Mol Cell Proteomics*, *11*(11), 1416–1429. https://doi.org/10.1074/mcp.M112.019588

Gebhardt, T., Mueller, S. N., Heath, W. R., & Carbone, F. R. (2013). Peripheral tissue surveillance and residency by memory T cells. *Trends Immunol*, *34*(1), 27–32. https://doi.org/10.1016/j.it.2012.08.008

Gill, T., & Rosenbaum, J. T. (2020). Putative Pathobionts in HLA-B27-Associated Spondyloarthropathy. *Front Immunol*, *11*, 586494. https://doi.org/10.3389/fimmu.2020.586494

Gill, T., Stauffer, P., Asquith, M., Laderas, T., Martin, T. M., Davin, S., Schleisman, M., Ramirez, C., Ogle, K., Lindquist, I., Nguyen, J., Planck, S. R., Shaut, C., Diamond, S., Rosenbaum, J. T., & Karstens, L. (2022). Axial spondyloarthritis patients have altered mucosal IgA response to oral and fecal microbiota. *Front Immunol*, *13*, 965634. https://doi.org/10.3389/fimmu.2022.965634

Gracey, E., Dumas, E., Yerushalmi, M., Qaiyum, Z., Inman, R. D., & Elewaut, D. (2019). The ties that bind: skin, gut and spondyloarthritis. *Curr Opin Rheumatol*, *31*(1), 62–69. https://doi.org/10.1097/BOR.0000000000000569

Hanson, A., & Brown, M. A. (2017). Genetics and the Causes of Ankylosing Spondylitis. *Rheum Dis Clin North Am*, *43*(3), 401–414. https://doi.org/10.1016/j.rdc.2017.04.006

Hanson, A. L., Nel, H. J., Bradbury, L., Phipps, J., Thomas, R., Le Cao, K. A., Kenna, T. J., & Brown, M. A. (2020). Altered Repertoire Diversity and Disease-Associated Clonal Expansions Revealed by T Cell Receptor Immunosequencing in Ankylosing Spondylitis Patients. *Arthritis Rheumatol*, *72*(8), 1289–1302. https://doi.org/10.1002/art.41252

Haroon, N. (2015). Does a positive HLA-B27 test increase your risk of mortality? *J Rheumatol*, *42*(4), 559–560. https://doi.org/10.3899/jrheum.150017

Hattori, A., & Tsujimoto, M. (2013). Endoplasmic reticulum aminopeptidases: biochemistry, physiology and pathology. *J Biochem*, *154*(3), 219–228. https://doi.org/10.1093/jb/mvt066

Hemminki, K., Li, X., Sundquist, K., & Sundquist, J. (2010). Familial association of inflammatory bowel diseases with other autoimmune and related diseases. *Am J Gastroenterol*, *105*(1), 139–147. https://doi.org/10.1038/ajg.2009.496

International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., Joo, K. B., Shim, S. C., Weisman, M., Ward, M., Zhou, X., Garchon, H. J., Chiocchia, G., Nossent, J.,…Brown, M. A.

(2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*, *45*(7), 730–738. https://doi.org/10.1038/ng.2667

Jacques, P., Lambrecht, S., Verheugen, E., Pauwels, E., Kollias, G., Armaka, M., Verhoye, M., Van der Linden, A., Achten, R., Lories, R. J., & Elewaut, D. (2014). Proof of concept: enthesitis and new bone formation in spondyloarthritis are driven by mechanical strain and stromal cells. *Ann Rheum Dis*, *73*(2), 437–445. https://doi.org/10.1136/annrheumdis-2013-203643

Jamalyaria, F., Ward, M. M., Assassi, S., Learch, T. J., Lee, M., Gensler, L. S., Brown, M. A., Diekman, L., Tahanan, A., Rahbar, M. H., Weisman, M. H., & Reveille, J. D. (2017). Ethnicity and disease severity in ankylosing spondylitis a cross-sectional analysis of three ethnic groups. *Clin Rheumatol*, *36*(10), 2359–2364. https://doi.org/10.1007/s10067-017-3767-6

Jandus, C., Bioley, G., Rivals, J. P., Dudler, J., Speiser, D., & Romero, P. (2008). Increased numbers of circulating polyfunctional Th17 memory cells in patients with seronegative spondylarthritides. *Arthritis Rheum*, *58*(8), 2307–2317. https://doi.org/10.1002/art.23655

Jiang, X., Clark, R. A., Liu, L., Wagers, A. J., Fuhlbrigge, R. C., & Kupper, T. S. (2012). Skin infection generates non-migratory memory CD8+ T(RM) cells providing global skin immunity. *Nature*, *483*(7388), 227–231. https://doi.org/10.1038/nature10851

Ke, D., Zhang, R., Dai, H., Su, Y., Yang, L., Guo, D., & Xu, J. (2025). Single-cell immunomics reveals the immunological hallmarks about the onset and different outcomes for ankylosing spondylitis patients. *Clin Rheumatol*, *44*(9), 3549–3559. https://doi.org/10.1007/s10067-025-07549-y

Kenna, T. J., Davidson, S. I., Duan, R., Bradbury, L. A., McFarlane, J., Smith, M., Weedon, H., Street, S., Thomas, R., Thomas, G. P., & Brown, M. A. (2012). Enrichment of circulating interleukin-17-secreting interleukin-23 receptor-positive gamma/delta T cells in patients with active ankylosing spondylitis. *Arthritis Rheum*, *64*(5), 1420–1429. https://doi.org/10.1002/art.33507

Khan, M. A., Yong, S. B., & Wei, J. C. (2023). Ankylosing spondylitis: History, epidemiology, and HLA-B27. *Int J Rheum Dis*, *26*(3), 413–414. https://doi.org/10.1111/1756-185X.14547

Komech, E. A., Pogorelyy, M. V., Egorov, E. S., Britanova, O. V., Rebrikov, D. V., Bochkova, A. G., Shmidt, E. I., Shostak, N. A., Shugay, M., Lukyanov, S., Mamedov, I. Z., Lebedev, Y. B., Chudakov, D. M., & Zvyagin, I. V. (2018). CD8+ T cells with characteristic T cell receptor beta motif are detected in blood and expanded in synovial fluid of ankylosing spondylitis patients. *Rheumatology (Oxford)*, *57*(6), 1097–1104. https://doi.org/10.1093/rheumatology/kex517

Korsunsky, I., Wei, K., Pohin, M., Kim, E. Y., Barone, F., Major, T., Taylor, E., Ravindran, R., Kemble, S., Watts, G. F. M., Jonsson, A. H., Jeong, Y., Athar, H., Windell, D., Kang, J. B., Friedrich, M., Turner, J., Nayar, S., Fisher, B. A.,…Raychaudhuri, S. (2022). Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. *Med*, *3*(7), 481–518 e414. https://doi.org/10.1016/j.medj.2022.05.002

Lan, F., Li, J., Miao, W., Sun, F., Duan, S., Song, Y., Yao, J., Wang, X., Wang, C., Liu, X., Wang, J., Zhang, L., & Qi, H. (2025). GZMK-expressing CD8(+) T cells promote recurrent airway inflammatory diseases. *Nature*, *638*(8050), 490–498. https://doi.org/10.1038/s41586-024-08395-9

Leirisalo-Repo, M., Turunen, U., Stenman, S., Helenius, P., & Seppala, K. (1994). High frequency of silent inflammatory bowel disease in spondylarthropathy. *Arthritis Rheum*, *37*(1), 23–31. https://doi.org/10.1002/art.1780370105

Li, Z., Wu, X., Leo, P. J., De Guzman, E., Akkoc, N., Breban, M., Macfarlane, G. J., Mahmoudi, M., Marzo-Ortega, H., Anderson, L. K., Wheeler, L., Chou, C. T., Harrison, A. A., Stebbings, S., Jones, G. T., Bang, S. Y., Wang, G., Jamshidi, A., Farhadi, E.,…Group, T. A. (2021). Polygenic Risk Scores have high diagnostic capacity in ankylosing spondylitis. *Ann Rheum Dis*, *80*(9), 1168–1174. https://doi.org/10.1136/annrheumdis-2020-219446

Liu, J., Tang, Y., Huang, Y., Gao, J., Jiang, S., Liu, Q., Ma, Y., Qian, X., Qian, F., Reveille, J. D., He, D., Zou, H., Jin, L., Zhu, Q., Pu, W., & Wang, J. (2021). Single-cell analysis reveals innate immunity dynamics in ankylosing spondylitis. *Clin Transl Med*, *11*(3), e369. https://doi.org/10.1002/ctm2.369

Lopez-Medina, C., Molto, A., Sieper, J., Duruoz, T., Kiltz, U., Elzorkany, B., Hajjaj-Hassouni, N., Burgos-Vargas, R., Maldonado-Cocco, J., Ziade, N., Gavali, M., Navarro-Compan, V., Luo, S. F., Monti, S., Tae-Jong, K., Kishimoto, M., Pimentel-Santos, F. M., Gu, J., Schiotis, R.,…Dougados, M. (2021). Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-PerSpA study. *RMD Open*, *7*(1). https://doi.org/10.1136/rmdopen-2020-001450

Mackay, L. K., Rahimpour, A., Ma, J. Z., Collins, N., Stock, A. T., Hafon, M. L., Vega-Ramos, J., Lauzurica, P., Mueller, S. N., Stefanovic, T., Tscharke, D. C., Heath, W. R., Inouye, M., Carbone, F. R., & Gebhardt, T. (2013). The developmental pathway for CD103(+)CD8+ tissue-resident memory T cells of skin. *Nat Immunol*, *14*(12), 1294–1301. https://doi.org/10.1038/ni.2744

Martin-Esteban, A., Guasp, P., Barnea, E., Admon, A., & Lopez de Castro, J. A. (2016). Functional Interaction of the Ankylosing Spondylitis-Associated Endoplasmic Reticulum Aminopeptidase 2 With the HLA-B*27 Peptidome in Human Cells. *Arthritis Rheumatol*, *68*(10), 2466–2475. https://doi.org/10.1002/art.39734

Martinez-Gonzalez, O., Cantero-Hinojosa, J., Paule-Sastre, P., Gomez-Magan, J. C., & Salvatierra-Rios, D. (1994). Intestinal permeability in patients with ankylosing spondylitis and their healthy relatives. *Br J Rheumatol*, *33*(7), 644–647. https://doi.org/10.1093/rheumatology/33.7.644

May, E., Dulphy, N., Frauendorf, E., Duchmann, R., Bowness, P., Lopez de Castro, J. A., Toubert, A., & Marker-Hermann, E. (2002). Conserved TCR beta chain usage in reactive arthritis; evidence for selection by a putative HLA-B27-associated autoantigen. *Tissue Antigens*, *60*(4), 299–308. https://doi.org/10.1034/j.1399-0039.2002.600404.x

McGonagle, D., Marzo-Ortega, H., O'Connor, P., Gibbon, W., Hawkey, P., Henshaw, K., & Emery, P. (2002). Histological assessment of the early

enthesitis lesion in spondyloarthropathy. *Ann Rheum Dis*, *61*(6), 534–537. https://doi.org/10.1136/ard.61.6.534

Mear, J. P., Schreiber, K. L., Munz, C., Zhu, X., Stevanovic, S., Rammensee, H. G., Rowland-Jones, S. L., & Colbert, R. A. (1999). Misfolding of HLA-B27 as a result of its B pocket suggests a novel mechanism for its role in susceptibility to spondyloarthropathies. *J Immunol*, *163*(12), 6665–6670.

Mei, Y., Pan, F., Gao, J., Ge, R., Duan, Z., Zeng, Z., Liao, F., Xia, G., Wang, S., Xu, S., Xu, J., Zhang, L., & Ye, D. (2011). Increased serum IL-17 and IL-23 in the patient with ankylosing spondylitis. *Clin Rheumatol*, *30*(2), 269–273. https://doi.org/10.1007/s10067-010-1647-4

Mielants, H., De Vos, M., Goemaere, S., Schelstraete, K., Cuvelier, C., Goethals, K., Maertens, M., Ackerman, C., & Veys, E. M. (1991). Intestinal mucosal permeability in inflammatory rheumatic diseases. II. Role of disease. *J Rheumatol*, *18*(3), 394–400.

Mielants, H., Veys, E. M., De Vos, M., Cuvelier, C., Goemaere, S., De Clercq, L., Schatteman, L., & Elewaut, D. (1995). The evolution of spondyloarthropathies in relation to gut histology. I. Clinical aspects. *J Rheumatol*, *22*(12), 2266–2272.

Mueller, S. N., Gebhardt, T., Carbone, F. R., & Heath, W. R. (2013). Memory T cell subsets, migration patterns, and tissue residence. *Annu Rev Immunol*, *31*, 137–161. https://doi.org/10.1146/annurev-immunol-032712-095954

Mulder, K., Patel, A. A., Kong, W. T., Piot, C., Halitzki, E., Dunsmore, G., Khalilnezhad, S., Irac, S. E., Dubuisson, A., Chevrier, M., Zhang, X. M., Tam, J. K. C., Lim, T. K. H., Wong, R. M. M., Pai, R., Khalil, A. I. S., Chow, P. K. H., Wu, S. Z., Al-Eryani, G.,…Ginhoux, F. (2021). Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity*, *54*(8), 1883–1900 e1885. https://doi.org/10.1016/j.immuni.2021.07.007

Nasonov, E. L., Mazurov, V. I., Lila, A. M., Dubinina, T. V., Gaidukova, I. Z., Lapshina, S. A., Klimenko, A. A., Somov, D. V., Lukyanov, S. A., Chudakov, D. M., Zvyagin, I. V., Britanova, O. V., Korolev, M. A., Abdulganieva, D. I., Krechikova, D. G., Kastanayan, A. A., Eliseeva, L. V., Samigullina, R. R., Povarova, T. V.,…Vinderskaya, G. A. (2025). The Efficacy and Safety of BCD-180, an Anti-TRBV9+ T cell Monoclonal Antibody, in Patients with Active Radiographic Axial Spondyloarthritis: 36-week Results from the Randomized, Double-Blind, Placebo-Controlled Phase 2 Clinical Study ELEFTA. *Dokl Biochem Biophys*, *522*(1), 387–403. https://doi.org/10.1134/S1607672925700140

Navarro-Compan, V., Sepriano, A., Capelusnik, D., & Baraliakos, X. (2025). Axial spondyloarthritis. *Lancet*, *405*(10473), 159–172. https://doi.org/10.1016/S0140-6736(24)02263-3

Noordenbos, T., Yeremenko, N., Gofita, I., van de Sande, M., Tak, P. P., Canete, J. D., & Baeten, D. (2012). Interleukin-17-positive mast cells contribute to synovial inflammation in spondylarthritis. *Arthritis Rheum*, *64*(1), 99–109. https://doi.org/10.1002/art.33396

Ortolan, A., Webers, C., Sepriano, A., Falzon, L., Baraliakos, X., Landewe, R. B., Ramiro, S., van der Heijde, D., & Nikiphorou, E. (2023). Efficacy and safety of non-pharmacological and non-biological interventions: a systematic literature review informing the 2022 update of the

ASAS/EULAR recommendations for the management of axial spondyloarthritis. *Ann Rheum Dis*, *82*(1), 142–152. https://doi.org/10.1136/ard-2022-223297

Paley, M. A., Yang, X., Hassman, L. M., Penkava, F., Garner, L. I., Paley, G. L., Linskey, N., Agnew, R., Arantes de Faria, P. H., Feng, A., Li, S. Y., Simone, D., Roberson, E. D., Ruzycki, P. A., Esaulova, E., Laurent, J., Feigl-Lenzen, L., Springer, L. E., Liu, C.,…Yokoyama, W. M. (2024). Mucosal signatures of pathogenic T cells in HLA-B*27+ anterior uveitis and axial spondyloarthritis. *JCI Insight*, *9*(16). https://doi.org/10.1172/jci.insight.174776

Rademacher, J., Poddubnyy, D., & Pleyer, U. (2020). Uveitis in spondyloarthritis. *Ther Adv Musculoskelet Dis*, *12*, 1759720X20951733. https://doi.org/10.1177/1759720X20951733

Ramiro, S., Nikiphorou, E., Sepriano, A., Ortolan, A., Webers, C., Baraliakos, X., Landewe, R. B. M., Van den Bosch, F. E., Boteva, B., Bremander, A., Carron, P., Ciurea, A., van Gaalen, F. A., Geher, P., Gensler, L., Hermann, J., de Hooge, M., Husakova, M., Kiltz, U.,…van der Heijde, D. (2023). ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. *Ann Rheum Dis*, *82*(1), 19–34. https://doi.org/10.1136/ard-2022-223296

Raut, R. D., Chakraborty, A. K., Neogi, T., Albro, M., Snyder, B., Schaer, T. P., Zhang, C., Grinstaff, M. W., & Bais, M. V. (2025). A multi-tissue human knee single-cell atlas identifies that osteoarthritis reduces regenerative tissue stem cells while increasing inflammatory pain macrophages. *Commun Biol*, *8*(1), 1146. https://doi.org/10.1038/s42003-025-08586-8

Reinhardt, A., Yevsa, T., Worbs, T., Lienenklaus, S., Sandrock, I., Oberdorfer, L., Korn, T., Weiss, S., Forster, R., & Prinz, I. (2016). Interleukin-23-Dependent gamma/delta T Cells Produce Interleukin-17 and Accumulate in the Enthesis, Aortic Valve, and Ciliary Body in Mice. *Arthritis Rheumatol*, *68*(10), 2476–2486. https://doi.org/10.1002/art.39732

Ren, C., Li, M., Zheng, Y., Cai, B., Du, W., Zhang, H., Wu, F., Tong, M., Lin, F., Wang, J., & Quan, R. (2022). Single-cell RNA-seq reveals altered NK cell subsets and reduced levels of cytotoxic molecules in patients with ankylosing spondylitis. *J Cell Mol Med*, *26*(4), 1071–1082. https://doi.org/10.1111/jcmm.17159

Ren, Z., Li, C., Wang, J., Sui, J., & Ma, Y. (2024). Single-cell transcriptome revealed dysregulated RNA-binding protein expression patterns and functions in human ankylosing spondylitis. *Front Med (Lausanne)*, *11*, 1369341. https://doi.org/10.3389/fmed.2024.1369341

Reveille, J. D. (2012). Genetics of spondyloarthritis--beyond the MHC. *Nat Rev Rheumatol*, *8*(5), 296–304. https://doi.org/10.1038/nrrheum.2012.41

Rizzo, A., Guggino, G., Ferrante, A., & Ciccia, F. (2018). Role of Subclinical Gut Inflammation in the Pathogenesis of Spondyloarthritis. *Front Med (Lausanne)*, *5*, 63. https://doi.org/10.3389/fmed.2018.00063

Robinson, P. C., Costello, M. E., Leo, P., Bradbury, L. A., Hollis, K., Cortes, A., Lee, S., Joo, K. B., Shim, S. C., Weisman, M., Ward, M., Zhou, X., Garchon, H. J., Chiocchia, G., Nossent, J., Lie, B. A., Forre, O., Tuomilehto, J., Laiho, K.,…Brown, M. A. (2015). ERAP2 is associated with ankylosing spondylitis in HLA-B27-positive and HLA-B27-negative

patients. *Ann Rheum Dis*, *74*(8), 1627–1629. https://doi.org/10.1136/annrheumdis-2015-207416
Robinson, P. C., van der Linden, S., Khan, M. A., & Taylor, W. J. (2021). Axial spondyloarthritis: concept, construct, classification and implications for therapy. *Nat Rev Rheumatol*, *17*(2), 109–118. https://doi.org/10.1038/s41584-020-00552-4
Rudwaleit, M., Haibel, H., Baraliakos, X., Listing, J., Marker-Hermann, E., Zeidler, H., Braun, J., & Sieper, J. (2009). The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum*, *60*(3), 717–727. https://doi.org/10.1002/art.24483
Rudwaleit, M., Landewe, R., van der Heijde, D., Listing, J., Brandt, J., Braun, J., Burgos-Vargas, R., Collantes-Estevez, E., Davis, J., Dijkmans, B., Dougados, M., Emery, P., van der Horst-Bruinsma, I. E., Inman, R., Khan, M. A., Leirisalo-Repo, M., van der Linden, S., Maksymowych, W. P., Mielants, H.,…Sieper, J. (2009). The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis*, *68*(6), 770–776. https://doi.org/10.1136/ard.2009.108217
Rudwaleit, M., Metter, A., Listing, J., Sieper, J., & Braun, J. (2006). Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. *Arthritis Rheum*, *54*(2), 569–578. https://doi.org/10.1002/art.21619
Rudwaleit, M., van der Heijde, D., Khan, M. A., Braun, J., & Sieper, J. (2004). How to diagnose axial spondyloarthritis early. *Ann Rheum Dis*, *63*(5), 535–543. https://doi.org/10.1136/ard.2003.011247
Sanz-Bravo, A., Campos, J., Mazariegos, M. S., & Lopez de Castro, J. A. (2015). Dominant role of the ERAP1 polymorphism R528K in shaping the HLA-B27 Peptidome through differential processing determined by multiple peptide residues. *Arthritis Rheumatol*, *67*(3), 692–701. https://doi.org/10.1002/art.38980
Schett, G., Lories, R. J., D'Agostino, M. A., Elewaut, D., Kirkham, B., Soriano, E. R., & McGonagle, D. (2017). Enthesitis: from pathophysiology to treatment. *Nat Rev Rheumatol*, *13*(12), 731–741. https://doi.org/10.1038/nrrheum.2017.188
Schlosstein, L., Terasaki, P. I., Bluestone, R., & Pearson, C. M. (1973). High association of an HL-A antigen, W27, with ankylosing spondylitis. *N Engl J Med*, *288*(14), 704–706. https://doi.org/10.1056/NEJM197304052881403
Sherlock, J. P., Joyce-Shaikh, B., Turner, S. P., Chao, C. C., Sathe, M., Grein, J., Gorman, D. M., Bowman, E. P., McClanahan, T. K., Yearley, J. H., Eberl, G., Buckley, C. D., Kastelein, R. A., Pierce, R. H., Laface, D. M., & Cua, D. J. (2012). IL-23 induces spondyloarthropathy by acting on ROR-gammat+ CD3+CD4-CD8- entheseal resident T cells. *Nat Med*, *18*(7), 1069–1076. https://doi.org/10.1038/nm.2817
Sieper, J., & Poddubnyy, D. (2017). Axial spondyloarthritis. *Lancet*, *390*(10089), 73–84. https://doi.org/10.1016/S0140-6736(16)31591-4
Sieper, J., van der Heijde, D., Landewe, R., Brandt, J., Burgos-Vagas, R., Collantes-Estevez, E., Dijkmans, B., Dougados, M., Khan, M. A.,

Leirisalo-Repo, M., van der Linden, S., Maksymowych, W. P., Mielants, H., Olivieri, I., & Rudwaleit, M. (2009). New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). *Ann Rheum Dis*, *68*(6), 784–788. https://doi.org/10.1136/ard.2008.101501

Simone, D., Penkava, F., Ridley, A., Sansom, S., Al-Mossawi, M. H., & Bowness, P. (2021). Single cell analysis of spondyloarthritis regulatory T cells identifies distinct synovial gene expression patterns and clonal fates. *Commun Biol*, *4*(1), 1395. https://doi.org/10.1038/s42003-021-02931-3

Stadler, M., Zhao, S. S., & Bowes, J. (2024). A review of the advances in understanding the genetic basis of spondylarthritis and emerging clinical benefit. *Best Pract Res Clin Rheumatol*, *38*(4), 101982. https://doi.org/10.1016/j.berh.2024.101982

Stavre, Z., Bridgewood, C., Zhou, Q., Maeda, Y., Huang, T. T., Karman, J., Khan, A., Giryes, S., Sharif, K., McGonagle, D., & Gravallese, E. M. (2022). A role for neutrophils in early enthesitis in spondyloarthritis. *Arthritis Res Ther*, *24*(1), 24. https://doi.org/10.1186/s13075-021-02693-7

Stolwijk, C., van Onna, M., Boonen, A., & van Tubergen, A. (2016). Global Prevalence of Spondyloarthritis: A Systematic Review and Meta-Regression Analysis. *Arthritis Care Res (Hoboken)*, *68*(9), 1320–1331. https://doi.org/10.1002/acr.22831

Stolwijk, C., van Tubergen, A., Castillo-Ortiz, J. D., & Boonen, A. (2015). Prevalence of extra-articular manifestations in patients with ankylosing spondylitis: a systematic review and meta-analysis. *Ann Rheum Dis*, *74*(1), 65–73. https://doi.org/10.1136/annrheumdis-2013-203582

Stratikos, E., & Stern, L. J. (2013). Antigenic peptide trimming by ER aminopeptidases--insights from structural studies. *Mol Immunol*, *55*(3-4), 212–219. https://doi.org/10.1016/j.molimm.2013.03.002

Tang, M., Qaiyum, Z., Lim, M., & Inman, R. D. (2025). Single cell immune profiling in ankylosing spondylitis reveals resistance of CD8(+) T cells to immune exhaustion. *iScience*, *28*(7), 112715. https://doi.org/10.1016/j.isci.2025.112715

Turner, M. J., Delay, M. L., Bai, S., Klenk, E., & Colbert, R. A. (2007). HLA-B27 up-regulation causes accumulation of misfolded heavy chains and correlates with the magnitude of the unfolded protein response in transgenic rats: Implications for the pathogenesis of spondylarthritis-like disease. *Arthritis Rheum*, *56*(1), 215–223. https://doi.org/10.1002/art.22295

van der Heijde, D., Molto, A., Ramiro, S., Braun, J., Dougados, M., van Gaalen, F. A., Gensler, L. S., Inman, R. D., Landewe, R. B. M., Marzo-Ortega, H., Navarro-Compan, V., Phoka, A., Poddubnyy, D., Protopopov, M., Reveille, J., Rudwaleit, M., Sampaio-Barros, P., Sepriano, A., Sieper, J.,…Baraliakos, X. (2024). Goodbye to the term 'ankylosing spondylitis', hello 'axial spondyloarthritis': time to embrace the ASAS-defined nomenclature. *Ann Rheum Dis*, *83*(5), 547–549. https://doi.org/10.1136/ard-2023-225185

Veale, D. J., McGonagle, D., McInnes, I. B., Krueger, J. G., Ritchlin, C. T., Elewaut, D., Kanik, K. S., Hendrikx, T., Berstein, G., Hodge, J., & Telliez,

J. B. (2019). The rationale for Janus kinase inhibitors for the treatment of spondyloarthritis. *Rheumatology (Oxford)*, *58*(2), 197–205. https://doi.org/10.1093/rheumatology/key070

Ward, M. M., Deodhar, A., Gensler, L. S., Dubreuil, M., Yu, D., Khan, M. A., Haroon, N., Borenstein, D., Wang, R., Biehl, A., Fang, M. A., Louie, G., Majithia, V., Ng, B., Bigham, R., Pianin, M., Shah, A. A., Sullivan, N., Turgunbaev, M.,…Caplan, L. (2019). 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network Recommendations for the Treatment of Ankylosing Spondylitis and Nonradiographic Axial Spondyloarthritis. *Arthritis Rheumatol*, *71*(10), 1599–1613. https://doi.org/10.1002/art.41042

Wellcome Trust Case Control, C., Australo-Anglo-American Spondylitis, C., Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D. P., McCarthy, M. I., Ouwehand, W. H., Samani, N. J., Todd, J. A., Donnelly, P., Barrett, J. C., Davison, D., Easton, D., Evans, D. M., Leung, H. T.,…Brown, M. (2007). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet*, *39*(11), 1329–1337. https://doi.org/10.1038/ng.2007.17

Wendling, D., Cedoz, J. P., Racadot, E., & Dumoulin, G. (2007). Serum IL-17, BMP-7, and bone turnover markers in patients with ankylosing spondylitis. *Joint Bone Spine*, *74*(3), 304–305. https://doi.org/10.1016/j.jbspin.2006.11.005

Xu, H., Yu, H., Liu, L., Wu, H., Zhang, C., Cai, W., Hong, X., Liu, D., Tang, D., & Dai, Y. (2021). Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Peripheral Mononuclear Cells in Patients With Ankylosing Spondylitis. *Front Immunol*, *12*, 760381. https://doi.org/10.3389/fimmu.2021.760381

Yang, X., Garner, L. I., Zvyagin, I. V., Paley, M. A., Komech, E. A., Jude, K. M., Zhao, X., Fernandes, R. A., Hassman, L. M., Paley, G. L., Savvides, C. S., Brackenridge, S., Quastel, M. N., Chudakov, D. M., Bowness, P., Yokoyama, W. M., McMichael, A. J., Gillespie, G. M., & Garcia, K. C. (2022). Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature*, *612*(7941), 771–777. https://doi.org/10.1038/s41586-022-05501-7

Yin, J., Sternes, P. R., Wang, M., Song, J., Morrison, M., Li, T., Zhou, L., Wu, X., He, F., Zhu, J., Brown, M. A., & Xu, H. (2020). Shotgun metagenomics reveals an enrichment of potentially cross-reactive bacterial epitopes in ankylosing spondylitis patients, as well as the effects of TNFi therapy upon microbiome composition. *Ann Rheum Dis*, *79*(1), 132–140. https://doi.org/10.1136/annrheumdis-2019-215763

Yu, H., Wu, H., Zheng, F., Zhu, C., Yin, L., Dai, W., Liu, D., Tang, D., Hong, X., & Dai, Y. (2020). Gene-regulatory network analysis of ankylosing spondylitis with a single-cell chromatin accessible assay. *Sci Rep*, *10*(1), 19411. https://doi.org/10.1038/s41598-020-76574-5

Zhang, L., Li, Y. G., Li, Y. H., Qi, L., Liu, X. G., Yuan, C. Z., Hu, N. W., Ma, D. X., Li, Z. F., Yang, Q., Li, W., & Li, J. M. (2012). Increased frequencies of Th22 cells as well as Th17 cells in the peripheral blood of patients with ankylosing spondylitis and rheumatoid arthritis. *PLoS One*, *7*(4), e31000. https://doi.org/10.1371/journal.pone.0031000

Zhang, Z., Ciccia, F., Zeng, F., Guggino, G., Yee, K., Abdullah, H., Silverberg, M. S., Alessandro, R., Triolo, G., & Haroon, N. (2017). Brief Report: Functional Interaction of Endoplasmic Reticulum Aminopeptidase 2 and HLA-B27 Activates the Unfolded Protein Response. *Arthritis Rheumatol*, *69*(5), 1009–1015. https://doi.org/10.1002/art.40033

Zhao, Z. B., Bian, Z. H., Lin, Z. M., Wu, S. F., Long, J., Cui, Y., Li, Y., Li, W., Sun, E. W., Lian, Z. X., & He, Y. (2023). Single-Cell Analysis of Patients with Axial Spondyloarthritis After Anti-TNFalpha Treatment: Experimental Data and Review of the Literature. *Clin Rev Allergy Immunol*, *65*(2), 136–147. https://doi.org/10.1007/s12016-023-08959-z

Zheng, M., Zhang, X., Zhou, Y., Tang, J., Han, Q., Zhang, Y., Ni, Q., Chen, G., Jia, Q., Yu, H., Liu, S., Robins, E., Jiang, N. J., Wan, Y., Li, Q. J., Chen, Z. N., & Zhu, P. (2019). TCR repertoire and CDR3 motif analyses depict the role of alphabeta T cells in Ankylosing spondylitis. *EBioMedicine*, *47*, 414–426. https://doi.org/10.1016/j.ebiom.2019.07.032

Zhou, C., Zhao, H., Xiao, X. Y., Chen, B. D., Guo, R. J., Wang, Q., Chen, H., Zhao, L. D., Zhang, C. C., Jiao, Y. H., Ju, Y. M., Yang, H. X., Fei, Y. Y., Wang, L., Shen, M., Li, H., Wang, X. H., Lu, X., Yang, B.,…Zhang, X. (2020). Metagenomic profiling of the pro-inflammatory gut microbiota in ankylosing spondylitis. *J Autoimmun*, *107*, 102360. https://doi.org/10.1016/j.jaut.2019.102360

Ziade, N. R. (2017). HLA B27 antigen in Middle Eastern and Arab countries: systematic review of the strength of association with axial spondyloarthritis and methodological gaps. *BMC Musculoskelet Disord*, *18*(1), 280. https://doi.org/10.1186/s12891-017-1639-5

Chapter 2: Materials and Methods

2.1 Sample Collection and Library Construction

All sample collection and library construction steps described in this section were conducted by collaborators involved in the broader research project. Sample collection and experimental procedures were carried out by Dr Davide Simone and Dr Frank Penkava, with the majority of single-cell library construction processed by Moustafa Attar in our laboratory using standardized protocols.

2.1.1 Overview of datasets and samples

An overview of the projects used to construct the single-cell atlas presented in this thesis is provided in **Table 2.1**. This includes information on project names, processing laboratories, collection site locations, data origin, sequencing protocol, total sample size, and corresponding usage across different chapters. Sample numbers reflect total samples generated for each project; where samples were excluded from downstream analysis, this is noted in the corresponding subsection. Disease and cohort breakdowns are provided in **Table 4.1**, and detailed per-sample metadata are provided in **Appendix A.1**: each sample is identified by a unique sample identifier (UID) linked to a patient identifier (PID), and is annotated with tissue type, HLA-B*27 status, disease condition, preservation method (fresh or frozen), demographic characteristics (age and sex), and treatment information.

Project Name	Centre	Data Origin	Sequencing Protocol	Sample Size	Chapter
P170676	Oxford (Bowness)	This project	3′ scRNA-seq	n=8	3,4,5,6
P180407	Oxford (Bowness)	This project	3′ scRNA-seq	n=4	3,4,5,6
P190230	Oxford (Bowness)	This project	5′ scRNA-seq + CITE-seq	n=2	3,4,5,6
P190726	Oxford (Bowness)	This project	5′ scRNA-seq + CITE-seq	n=6	3,4,5,6
Ox_VA_PB_SF	Oxford	This project	5′ scRNA-seq + CITE-seq + scTCR-seq	n=35	4,5,6
Berlin_PB_SF	Berlin	This project	5′ scRNA-seq + CITE-seq + scTCR-seq	n=27	4,5,6
Gut_Naples	Naples	This project	5′ scRNA-seq + CITE-seq + scTCR-seq	n=16	4,5,6
Gut_TGU	Oxford (TGU)	This project	5′ scRNA-seq + CITE-seq + scTCR-seq	n=4	4,5,6
SIJ	Oxford	This project	5′ scRNA-seq + CITE-seq + scTCR-seq + scBCR-seq	n=8	4,5,6
Spine_Enthesis	Leeds	This project	3′/5′ scRNA-seq + CITE-seq + scTCR-seq	n=15	4,5,6
Aqh_Eye	Oxford (Sharma)	This project	5′ scRNA-seq + scTCR-seq + scBCR-seq	n=8	4,5,6
CD8_SFMC_Berlin	Berlin	External	5′ scRNA-seq + scTCR-seq	n=17	4,5,6
COMBAT_healthy	Oxford	External	5′ scRNA-seq	n=10	3

Table 2.1 Experimental batches included in this thesis
This table summarizes each batch's source centre, data origin (generated within this project versus externally previously published), sequencing protocol, total sample number (n), and the chapters in which the data have been used.

2.1.2 Project-specific sample collection and library protocols

2.1.2.1 Projects: P170676, P180407, P190230, P190726

For these 4 projects, human blood samples were collected at the Bowness AS Clinic (Nuffield Orthopaedic Centre, Oxford) and processed in the Bowness Lab according to the established standard operating procedures (SOPs), including SOP_006 (“Logging incoming human tissue”) and SOP_007 (“Handling, storage and disposal of human tissue”). PBMCs were isolated and stored frozen prior to single-cell library preparation, which was carried out across 4 separate experimental batches.

Batch P170676 included PBMC samples from 4 anti-TNF-treated AxSpA patients and 4 biologic-naïve AxSpA patients, processed using the 10x Genomics Chromium Single Cell 3' Gene Expression Solution. Of those, 1 patient nominally classified as biologic-naïve had recently completed prior biologic treatment; this sample was therefore excluded from downstream analyses comparing treatment effects, due to the ambiguous treatment status.

Batch P180407 included samples from 2 AxSpA patients (Patient A: PID 598; Patient B: PID 681), each collected at both pre-treatment and post-treatment timepoints (4 samples in total). Cells were sorted by FACS into 2 separate fractions based on surface marker expression: a T/NK cell fraction, defined as $CD3^+$ or $CD56^+$, and a monocyte/dendritic cell fraction, defined as $CD3^-$, $CD56^-$, and $CD19^-$. The 4 samples were individually labelled with hashing antibodies and pooled across 2 sequencing channels: Channel 1 contained pre-treatment monocyte/DC cells and post-treatment T/NK cells from both patients; Channel 2 contained pre-treatment T/NK cells and post-treatment monocyte/DC cells from both patients. Libraries were prepared using the 10x Genomics Chromium Single Cell 3' Gene Expression Solution. After sequencing, each cell was assigned to a patient via genetic demultiplexing and to a broad lineage

(T/NK or monocyte/DC) via initial cell-type annotation. The combination of patient, lineage, and channel uniquely determined the pre- or post-treatment status of each cell, given the channel design described above. This multiplexing strategy was used as part of an earlier pilot study and was not repeated in subsequent batches.

Batch P190230 included 2 PBMC samples from a single anti-TNF-treated AxSpA patient collected at 2 different timepoints. Libraries were prepared using the 10x Genomics Chromium Single Cell 5' workflow with simultaneous surface protein measurement by CITE-seq, using a custom antibody panel conjugated in-house to oligonucleotide barcodes (Expedeon). The full panel is listed in **Appendix C.1**.

Batch P190726 included PBMC samples from 3 biologic-naïve AxSpA patients and 3 HLA-B*27-positive healthy controls, processed using the same 10x Genomics 5' Immune Profiling Solution with CITE-seq workflow as P190230. A separate custom antibody panel was used for this batch, conjugated in-house to oligonucleotide barcodes (Expedeon); the full panel is listed in **Appendix C.2**.

2.1.2.2 Projects: Ox_VA_PB_SF and Berlin_PB_SF

The Ox_VA_PB_SF project samples were distributed into 4 experimental pools: 3 PBMC pools containing 8, 7, and 8 samples respectively, and 1 SFMC pool containing 9 samples; some donors contributed paired PBMC and SFMC samples. To minimize potential batch effects, each pool included a balanced representation of disease conditions (AxSpA, RA, and healthy controls). Each PBMC pool was loaded across 4 sequencing channels and the SFMC pool across 5 channels (17 channels in total), targeting ~20,000 cells per channel.

Single-cell multimodal libraries were prepared using the 10x Genomics Chromium 5' workflow, generating 17 paired gene expression (GEX), antibody-derived tag (ADT; CITE-seq antibody-capture, TotalSeq-C Human Universal Cocktail, **Appendix C.3**), and T-cell receptor (VDJ-T) libraries, with VDJ-T libraries each targeting approximately 10,000 T cells. Bulk RNA-seq was also performed for each sample to support genetic demultiplexing of the pooled libraries.

In addition to the pooled datasets, 3 fresh SFMC samples from donors not included in the pools were processed independently using the same 10x Genomics Single Cell 5' workflow, generating matched GEX, ADT, and VDJ-T libraries with each library targeting approximately 7,500 cells.

The Berlin_PB_SF project followed a similar strategy: samples were distributed into 4 experimental pools, comprising 2 PBMC pools containing 6 and 7 samples respectively, and 2 SFMC pools each containing 7 samples, with balanced representation across AxSpA, peripheral SpA (pSpA), PsA, and monoarthritis. Each pool was loaded across 5 sequencing channels (20 channels in total), targeting ~20,000 cells per channel. Library preparation followed the same strategy as for the Ox_VA_PB_SF project, using the 10x Genomics Chromium Single Cell 5′ workflow to generate GEX, ADT (TotalSeq-C Human Universal Cocktail; **Appendix C.3**), and VDJ-T libraries, with VDJ-T libraries each targeting approximately 10,000 T cells.

2.1.2.3 Project: SIJ

Samples from the SIJ project consisted of fresh cells isolated by enzymatic dissociation from sacroiliac joint tissues of 4 patients diagnosed with AxSpA. Tissue was obtained by CT-guided percutaneous bone marrow biopsy

performed as part of a clinical procedure for intra-articular corticosteroid injection (triamcinolone acetonide, Kenalog 40 mg) for symptomatic sacroiliac pain. Prior to each procedure, a recent SIJ MRI (performed within three months) was reviewed to identify regions of active bone marrow oedema on STIR or T1 fat-saturated post-contrast sequences, and these regions were mapped to guide needle placement. Procedures were performed using a 14G Bonopty co-axial biopsy set (Apriomed) on a GE Discovery CT 590RT scanner. Following standard sterile preparation and CT localisation, bone marrow biopsy was performed in the region of mapped bone marrow oedema, after which corticosteroid was injected through the same access.

For 1 patient (PID 495), cells were further separated by FACS into $CD45^+$ and $CD45^-$ fractions, which were processed as 2 separate channels; for the remaining 3 patients, the SIJ samples were each processed on a single channel. In addition, paired PBMC samples were collected from all 4 donors and processed on individual channels. Sequencing was conducted using the 10x Genomics Chromium Single Cell 5' workflow, generating GEX, ADT, and VDJ-T libraries. VDJ-B libraries were also generated for some channels but are not analysed in this thesis. One SIJ sample was subsequently excluded from downstream analysis due to low quality, leaving 8 of 9 samples included in the atlas.

2.1.2.4 Projects: Gut_Naples and Gut_TGU

In the Gut_Naples project, $CD45^+$ ileal lamina propria cells and PBMCs were collected from 3 patients with AxSpA and 5 healthy controls by the Ciccia lab in Naples, with samples stored frozen prior to processing. All donors contributed paired ileum and PBMC samples. Samples were distributed across

3 experimental pools, each multiplexed using hashing antibodies for donor identification: Pool 1 contained the ileum and PBMC samples from 2 healthy controls (1 channel); Pool 2 contained ileum samples from 3 AxSpA patients and 3 healthy controls (2 channels); and Pool 3 contained PBMC samples from the same 6 donors as Pool 2 (2 channels). Sequencing was conducted using the 10x Genomics Chromium Single Cell 5' workflow, generating 5 paired GEX, ADT, and VDJ-T libraries across the 3 pools.

In the Gut_TGU project, fresh ileal tissue samples were collected from 2 AxSpA patients at Oxford TGU, with paired PBMC also collected from 1 of the 2 patients. For 1 patient (PID 275), the ileal sample was sorted by FACS into $CD45^{+}$ and $CD45^{-}$ fractions, processed as 2 separate channels. For the other patient (PID 35), the ileum and PBMC samples were each processed on a single channel. Sequencing was conducted using the 10x Genomics Chromium Single Cell 5' workflow, generating GEX, ADT, and VDJ-T libraries.

2.1.2.5 Project: Spine_Enthesis

Samples from Spine_Enthesis project were provided by the McGonagle lab in Leeds. Fresh spinal enthesis tissue obtained from spinous processes and interspinous ligaments was anatomically divided into entheseal soft tissue (EST) and peri-entheseal bone (PEB) regions before enzymatic digestion. Healthy control samples, sequenced as part of the Human Cell Atlas (HCA) project, were processed fresh and sequenced using the 10x Genomics Chromium Single Cell 3' workflow, generating GEX libraries from 4 donors with paired PEB and EST tissues (8 channels in total). AxSpA samples were collected from 4 patients and sequenced using the 10x Genomics Chromium Single Cell 5' workflow with GEX and VDJ-T profiling. AxSpA samples were

stored frozen prior to processing, with the exception of 1 patient (PID 843), whose EST and PEB samples were processed fresh; ADT libraries were also generated for this patient. Data from this project have recently been published separately (Altaie et al., 2025).

2.1.2.6 Project: Aqh_Eye

Samples from the Aqh_Eye project were provided by Sharma Lab and consisted of aqueous humour-derived cells from 6 patients with acute anterior uveitis (AAU; HLA-B*27 positive) and 2 control donors (HLA-B*27 negative without AAU). Cells from these donors were processed across 6 channels using the 10x Genomics Chromium Single Cell 5' workflow, generating GEX, VDJ-T, and VDJ-B libraries with each library targeting approximately 10,000 cells. Of the 6 channels, 4 contained cells from a single donor each, and 2 contained cells pooled from 2 donors with mixed HLA-B*27 status; for the pooled channels, genetic demultiplexing was performed to separate the donors, and the 2 HLA-B*27-negative control donors were excluded from downstream analysis. The remaining 6 AAU samples were used in the atlas. VDJ-B libraries were generated for completeness but are not analysed in this thesis.

2.1.2.7 Projects: CD8_SFMC_Berlin and COMBAT

Samples from the CD8_SFMC_Berlin and COMBAT projects were sourced from previously published community datasets. Full methodological details can be found in the original publications cited below.

The CD8_SFMC_Berlin dataset (Deschler et al., 2022) comprises 5' single-cell GEX and VDJ-T data from sorted $CD8^+$ T cells isolated from SFMC of 17 patients (10 with AS, 4 with pSpA, and 3 with RA), with mixed HLA-B*27 status across the cohort. The COMBAT dataset (COMBAT Consortium, 2022)

comprises scRNA-seq data from PBMCs of patients with various diseases alongside 10 healthy volunteers.

For this thesis, I reprocessed raw sequencing data from both projects by aligning reads to the same curated reference genome and applying identical preprocessing steps used for the other samples in the atlas. For the COMBAT dataset, 10 healthy volunteer samples were retained; in Chapter 3, since P170676 contained only male donors after exclusion of the treatment-ambiguous sample, the COMBAT cohort was further restricted to the 6 male donors for sex-matched comparisons.

2.1.3 HLA-B*27 status determination

HLA-B*27 status for all donors was determined clinically at the recruiting centres using PCR-based molecular typing. For donors recruited at Oxford, typing was performed at the Oxford Transplant Centre and included subtype determination; all Oxford-typed donors carried the HLA-B*27:05 allele. The reported HLA-B*27 status is recorded in the donor metadata (**Appendix A.1**).

2.2 Computational Analysis and Resources

2.2.1 Data preprocessing

2.2.1.1 scRNA-seq data processing

Raw sequencing data were converted to FASTQ files using Cell Ranger v7.1.0 (cellranger mkfastq). For alignment, a human genome reference was built from Ensembl v110 annotations (Harrison et al., 2024) using cellranger mkref. Reads were aligned and assigned to genes with the cellranger multi pipeline, including intronic reads.

Ambient RNA contamination was assessed and corrected using CellBender v0.3.0 (Fleming et al., 2023) remove-background command with parameters: --fpr 0.01, --learning-rate 1e-4, and --epochs 150. Parameters for --expected-cells and --total-droplets-included were manually determined by inspecting UMI curves to identify the plateau corresponding to empty droplets. Although CellBender identified additional barcodes, most represented low-quality cells (<500 UMIs) and were excluded, retaining only original Cell Ranger-assigned barcodes for downstream analysis.

2.2.1.2 CITE-seq data processing

CITE-seq data preprocessing and multi-modal integration of RNA and protein expression data were performed using totalVI (Gayoso et al., 2021), implemented in the scvi-tools framework. An AnnData object containing both RNA and protein measurements was constructed from the input data. Protein expression values were extracted and filtered to remove proteins with zero counts across all cells.

Highly variable genes (HVGs) were selected using the Seurat v3 method (Stuart et al., 2019) implemented in Scanpy v1.10.3, after excluding immune receptor genes. HVG identification was performed across batches defined by UID, selecting the top 2,000 genes exhibiting the highest variability. The totalVI model was configured with the batch key set as project name, latent distribution as "normal", two decoder layers (n_layers_decoder=2), and allowed for missing proteins (override_missing_proteins=True). For MAIT-cell identification, alternative totalVI settings were explored by varying the override_missing_proteins flag and adjusting the

empirical_protein_background_prior parameter to control initialization of protein background estimates.

2.2.1.3 Variant calling and joint genotyping

Variant calling on bulk RNA-seq data was performed for the purpose of generating donor-specific genotype information used for genetic demultiplexing of pooled single-cell libraries in this thesis. The pipeline was built on the Broad Institute's GATK toolkit (v4.1.7.0) (Van der Auwera et al., 2013) combined with WDL scripts executed through Cromwell (v51). Workflow configuration was adapted from public templates hosted on GitHub and optimized for the local environment. Separate WDL pipelines were generated for paired-end and single-end sequencing data; input JSON and TSV configuration files were prepared for each sample following the pipeline template structure.

RNA-seq reads were aligned to the same Ensembl-based reference genome using the STAR aligner (Dobin et al., 2013), and resulting BAM files were processed with Picard/GATK utilities including MergeBamAlignment and MarkDuplicates to clean and merge alignments. Reads spanning introns were processed using SplitNCigarReads, and base quality scores were recalibrated prior to variant calling. GATK's HaplotypeCaller was used to produce per-sample gVCF files, followed by VariantFiltration to remove low-quality variants.

Per-sample gVCFs were combined using GATK's CombineGVCFs command and joint genotyping was performed with GenotypeGVCFs to produce group-level VCF files. Variant calls were further filtered and curated to ensure high-confidence genotypes.

2.2.1.4 scTCR-seq preprocessing

The human VDJ reference genome used for single-cell VDJ sequencing analysis was generated using Cell Ranger (v7.1.0) based on annotations retrieved from the International ImMunoGeneTics information system (IMGT) using the Cell Ranger command-line utility fetch-imgt (Giudicelli et al., 2005). The reference comprised V-gene segments (queried using IMGT/GENE-DB query version 8.1), D-gene segments (query version 7.2), J-gene segments (query version 7.2), and C-gene segments (query version 14.1) for human TCR (TRA, TRB, TRG, TRD) and BCR (IGH, IGK, IGL) loci. All VDJ sequencing data were subsequently aligned to this generated reference for downstream analyses. Mapping and processing of the TCR sequencing data were performed by Dr Frank Penkava.

2.2.2 Cell demultiplexing and sample assignment

Genetic demultiplexing for single-cell data from projects P170676, P180407, P190230, Ox_VA_PB_SF, Berlin_PB_SF and Aqh_eye was performed using Souporcell v2.5 (Heaton et al., 2020), utilizing high-confidence VCF files generated from bulk RNA-seq data as described in the variant calling section above. Souporcell was executed through the cellhub pipeline (https://github.com/sansomlab/cellhub) on an HPC cluster. Input files included the Cell Ranger generated possorted_genome_bam.bam and corresponding barcodes.tsv.gz, together with the project-specific GRCh38 reference genome. I used the genotype-aware mode (--known_genotypes) by supplying cohort-level VCFs generated from bulk RNA-seq joint genotyping for each pool along with the corresponding list of donor sample names. The number of expected genotype clusters (k=7–9) was set according to the number of donors in each pool.

For projects P190726 and Gut_Naples, hashtag-based demultiplexing was conducted using GMM-demux (Xin et al., 2020), which is also implemented in the cellhub pipeline. Souporcell and GMM-demux were employed together for two libraries within project Gut_Naples and Aqh_eye, where either genetic or hashtag data was incomplete or unavailable.

Chromosomal markers, including *XIST* and genes located on the Y chromosome (*RPS4Y1, ZFY, ZFY-AS1, LINC00278, PCDH11Y, USP9Y, DDX3Y, UTY, TMSB4Y, NLGN4Y, TTTY14, KDM5D, EIF1AY, RPS4Y2*), provided additional validation for sample assignments.

2.2.3 Quality control

Quality control (QC) filtering was performed at multiple levels to remove low-quality cells and samples prior to downstream analysis. After ambient RNA removal and genetic demultiplexing, initial cell-level filtering was applied based on common QC metrics. Cells expressing fewer than 200 genes or containing fewer than 400 unique molecular identifiers (UMIs) were removed in analyses related to Chapter 3. For Chapter 4 analyses, only the gene-count threshold (<200 genes) was applied at this initial stage without an explicit UMI cutoff, as Chapter 4 includes rare and tissue-resident cell populations with characteristically lower RNA content; lower-quality cells were instead retained at this stage to allow them to cluster together and be removed through additional rounds of QC filtering described in Chapter 4. Additionally, mitochondrial gene percentage thresholds were dataset-specific: cells exceeding 10% mitochondrial reads in Chapter 3 or 15% mitochondrial reads in Chapter 4 were excluded. Only genetically identified singlet cells were retained for subsequent analyses.

For Chapter 3 analyses involving sorted immune cell fractions (P180407), further validation was performed by integrating genetic demultiplexing results with initial cell-type assignments based on known immune marker genes. Cells identified as misassigned or misclustered were removed from their corresponding cell-type populations before downstream analyses.

Doublet identification was performed using Scrublet v0.2.3 (Wolock et al., 2019), and cells with doublet scores exceeding the mean plus three standard deviations were removed. Additional cell filtering was applied at the cluster-sample level by identifying and removing outliers. For each cluster-sample combination, cells falling outside the range defined by 1.5× the interquartile range (IQR) above the 75th percentile or below the 25th percentile were excluded. Metrics evaluated during this filtering included total UMI counts, total gene counts, mitochondrial gene percentages, ribosomal gene percentages and Scrublet doublet scores.

Finally, at the sample level, overall quality was assessed using median-based filtering criteria. Samples with consistently low quality defined as median UMI counts or gene counts below Q1-1.5×IQR, or median mitochondrial or ribosomal gene percentages above Q3+1.5×IQR, were excluded from further analysis.

2.2.4 Data normalization and integration

Raw gene read counts were normalized by sequencing depth in each cell and log-transformed.

For the atlas in Chapter 3, HVGs were selected within each project using Seurat_v3 method implemented in scanpy.pp.highly_variable_genes() function

and the final HVG list was the union of 7,500 HVGs from each project. Dimensionality reduction and batch correction were performed using the scVI model as implemented in scvi-tools, considering 10X Genomics chemistry (5' and 3') and the sample ID for each cell as the categorical covariates to correct for. The model was trained on raw counts of the final HVG list generated above.

For Chapter 4 analyses, scVI models were trained and evaluated systematically across multiple parameter configurations to assess model performance. Network architectures were tested by varying the number of hidden layers (n_layers: 1, 2, or 3) and latent space dimensions (n_latent: 10, 20, 30 or 40). Model training and evaluation were performed using scvi-tools. The marginal log-likelihood function (trainer.test_set.marginal_ll()) was used to examine model likelihood under different configurations. Two likelihood functions: negative binomial (NB) and zero-inflated negative binomial (ZINB) were compared. Batch effects were modeled using alternative batch keys (sample, sample + tissue, or sample + protocol), and mitochondrial percentage was included as a continuous covariate in some models. Model fit was assessed using the evidence lower bound (ELBO), and biological consistency was quantified by computing normalized mutual information (NMI) between scVI-derived clusters and CellTypist-based annotations.

2.2.5 Data clustering and annotation

2.2.5.1 Neighborhood construction

Neighborhood graphs were constructed using the scvelo.pp.neighbors function, implemented with the Hierarchical Navigable Small World (HNSW) algorithm. The analysis was performed on scVI-derived latent embeddings (use_rep='X_scVI') with n_neighbors set to 20 and KNN mode enabled. The

construction used Euclidean distance as the similarity metric (metric='euclidean') and included the following HNSW-specific parameters: M=20, ef=200, and ef_construction=200. A fixed random seed (random_state=0) was applied to ensure reproducibility. For visualisation, two-dimensional UMAP embeddings were generated from the scVI latent space using scanpy.tl.umap with min_dist=0.1 and random_state=0.

2.2.5.2 Clustering

Cells were clustered using the Leiden algorithm implemented in scanpy.tl.leiden, applied to the neighbourhood graph constructed as described above. For the PBMC atlas in Chapter 3, clustering was performed at a single resolution of 0.2. For the multi-tissue atlas in Chapter 4, clustering was performed at multiple resolutions within each immune compartment. The per-compartment resolutions are detailed in **Appendix C.4**.

2.2.5.3 Annotation process

Automatic cell type annotation was performed using SingleR v2.4.1 (Aran et al., 2019) and CellTypist v1.7.1 (Dominguez Conde et al., 2022). For SingleR, reference datasets including the Human Primary Cell Atlas (Mabbott et al., 2013), Blueprint (Fernandez et al., 2016), and ENCODE (Sloan et al., 2016) collections were used to assign labels based on transcriptomic similarity using Spearman correlation. CellTypist was applied with both the “Immune_All_Low” and “Immune_All_High” pretrained models to provide predictions at coarse and fine immune cell resolution levels, respectively.

Annotation was conducted in multiple stages. First, coarse annotation was used to identify major immune and non-immune cell compartments. Within each compartment, clusters were evaluated for marker gene expression, and

clusters with ambiguous or overlapping identities were manually inspected. Clusters showing shared transcriptional signatures were merged, while those displaying heterogeneity in lineage-defining markers were subdivided into finer subsets. Marker genes were assessed using scanpy.tl.rank_genes_groups (Wilcoxon method), and cell-type assignments were cross-validated between SingleR and CellTypist predictions.

To manage and document the annotation workflow, all cluster-level labels and corresponding marker genes were recorded in a structured spreadsheet and integrated into the analysis pipeline through an automated annotation script, ensuring reproducibility and version control.

2.2.5.4 Cluster entropy calculation

Cluster purity and entropy were evaluated using ROGUE v1.0 (Liu et al., 2020) implemented in R. Normalized gene expression matrices and cluster labels were exported from anndata and imported into R for analysis. To accelerate computation, approximately 10% of cells from each cluster were randomly subsampled for analysis. Calculations were performed using the default parameters (CalculateRogue function) following the developer's guidelines.

2.2.5.5 CD4 and CD8 cell threshold decision

Thresholds for classifying cells as CD4-positive or CD8-positive were determined based on gene expression distributions. For CD4 evaluation, the expression of the *CD4* gene was used, whereas CD8 evaluation was based on the averaged expression of *CD8A* and *CD8B* genes. Expression values across all cells were binned into histograms with 100 bins. Bin counts and bin centres were calculated using the numpy.histogram function, and the

signal.find_peaks() function from the scipy v1.14.0 (Virtanen et al., 2020) was used to identify peaks. The largest peak was identified, and the threshold was set as the local minima between zero expression and this first major peak. Cells with expression values above this threshold were classified as marker-positive, while those below were considered marker-negative.

2.2.6 Compositional and abundance analysis

2.2.6.1 Compositional differential abundance analysis (edgeR)

Differential abundance analysis was performed using edgeR v3.36.0 (Robinson et al., 2010). Cell-type counts were normalized using TMM method implemented in calcNormFactors function, and samples with fewer than 10 total cells were excluded. Cell types with zero counts across all samples included in the comparison were removed prior to modelling. For each compartment, a design matrix was constructed with disease condition (AxSpA vs comparator), and the following model was fitted: counts ~ condition. Dispersion parameters were estimated using estimateDisp, and differential testing was performed using quasi-likelihood negative binomial models implemented in glmQLFit and glmQLFTest. BH multiple-testing correction was applied separately within each comparison, across the cell types tested in that comparison.

2.2.6.2 Bayesian compositional analysis (scCODA)

scCODA v0.1.9 (Buttner et al., 2021) was used as an additional compositional framework to validate differential abundance results. Cell types with zero counts across all retained samples were removed, and a compositional data object was constructed with sample labels encoded as covariates. Analysis was repeated iteratively using each cell type as the

reference. For each reference, a Hamiltonian Monte Carlo model (CompositionalAnalysis with sample_hmc) was fitted, and credible effects were extracted under specific thresholds: the default scCODA criterion, and expected FDR levels of 0.1 or 0.2. Credible effects were then aggregated across all reference cycles to identify robust compositional changes.

2.2.6.3 Neighbourhood-level differential abundance analysis (Milo)

Neighbourhood-level differential abundance analysis was performed using Milo (Dann et al., 2022) with the python milopy package v0.1.1. The neighbourgraph was precomputed as described above. Dataset was then subset to the desired cells, and neighbourhoods were constructed using milo.make_nhoods and cell counts per neighbourhood were obtained using milo.count_nhoods with the sample identifier as the grouping variable. Differential abundance testing was conducted with milo.DA_nhoods, using a design formula ~ condition and restricting the analysis to the relevant set of samples. Resulting neighbourhoods were annotated using milopy.utils.annotate_nhoods based on the custom cluster labels.

2.2.6.4 Co-varying compositional pattern analysis (NMF)

Non-negative matrix factorization (NMF) analysis was performed using the NMF R package v0.28. Cell-type frequencies were calculated per sample and used as input after normalization and metadata integration. The optimal rank was determined through a rank survey assessing cophenetic correlation, dispersion, explained variance, residuals, and RSS across multiple runs (nrun = 50). The final model was computed with 100 runs, producing basis (W) and coefficient (H) matrices, which were performed t-tests with Benjamini–Hochberg (BH) adjustment for multiple comparisons.

2.2.6.5 Cross-tissue compositional comparison

Cell-type composition across tissue was assessed using Poisson regression on per-sample count data. Samples from AxSpA patients were selected, matched between metadata and count matrices. For each cell type, a Poisson generalized linear model was fitted across tissues (*count ~ target + sample*), where *target* indicated whether the sample belonged to a given tissue and *sample* accounted for donor-specific effects. The tissue-associated coefficient and *p*-value were extracted, and false discovery rate (FDR) correction was applied using the BH method. Cell types showing positive coefficients with adjusted-$p < 0.05$ were considered significantly enriched.

2.2.7 Differential expression analysis

Genes were filtered out if their counts were insufficient for reliable testing as determined by the edgeR function filterByExpr(). Differential expression was then assessed using a negative binomial generalized linear model, fitted with glmFit and tested with glmLRT, using the model: counts ~ condition. P-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) procedure, and genes with BH-adjusted $p < 0.05$ were considered significant. Differential expression analysis was performed using DESeq2 v1.34.0 (Love et al., 2014) with a likelihood ratio test, comparing the full model: counts ~ condition + projects, to the reduced model: counts ~ projects, thereby evaluating the contribution of condition while accounting for batch (project) effects.

In parallel, the limma-voom approach implemented in the limma package v3.50.3 (Ritchie et al., 2015) was applied, using the duplicateCorrelation method to account for both paired and unpaired samples. Differentially

expressed genes (DEGs) obtained from DESeq2 and limma-voom were compared to assess consistency across methods within each cell type.

2.2.8 GSEA pathway analysis

Gene set enrichment analysis (GSEA) was performed using fgsea R package v1.12.0 (Korotkevich et al., 2021) on gene-level statistics derived from differential expression testing. For each cell type, I generated ranked gene lists using likelihood ratio statistics and mapped gene symbols to Entrez IDs via biomaRt (Durinck et al., 2009) and org.Hs.eg.db. Gene sets were compiled from Gene Ontology (BP, MF, CC), Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets, Human Molecular Signatures Database (MSigDB) and BioCarta collections, with redundant gene sets collapsed using collapsePathways. Enrichment analyses were run with a minimum gene set size of 3, a maximum size of 500, and 10,000 permutations. Significant pathways were defined using BH-adjusted p-values.

2.2.9 scDRS analysis

The original GWAS summary statistics for ankylosing spondylitis were obtained from the large-scale Immunochip-based study published in 2013 (International Genetics of Ankylosing Spondylitis et al., 2013), which included 10,619 AS cases and 15,145 controls of European and East Asian ancestry, with AS cases defined according to the modified New York criteria. The summary statistics were imported and filtered to remove variants within the extended MHC region (chromosome 6: 26Mb-34Mb), which was excluded to avoid inflation driven by long-range LD in the HLA locus. Genotype data from the 1000 Genomes Project (European population) was used as the reference

panel, and only SNPs from HapMap3 were included. Genetic mapping was performed with MAGMA v1.10 (de Leeuw et al., 2015), using the European subset of the 1000 Genomes Project as the LD reference panel. SNP coordinates were annotated to genes using a 10 kb window, and gene-level association statistics were computed using the MAGMA --pval procedure. The resulting gene-level output was used as the primary AS summary statistic input for subsequent scDRS analysis.

Additional pre-processed summary statistics for rheumatoid arthritis, Crohn's disease and ulcerative colitis were obtained from the dataset release associated with the scDRS publication (Zhang et al., 2022) and were used as additional inputs to the scDRS analysis. The rheumatoid arthritis summary statistics were derived from the trans-ethnic GWAS analysis (Okada et al., 2014), which included 29,880 RA cases and 73,758 controls of European and Asian ancestry, with RA cases fulfilling the 1987 American College of Rheumatology classification criteria or diagnosed with RA by a professional rheumatologist. The Crohn's disease and ulcerative colitis summary statistics were derived from the genome-wide meta-analysis study (de Lange et al., 2017), which combined a new GWAS with previously published International IBD Genetics Consortium summary statistics, predominantly of European ancestry, with IBD cases diagnosed using established endoscopic, histopathological and radiological criteria. When computing polygenic risk scores with compute_score function, the individual identity was included as a covariate to account for inter-individual variability.

2.2.10 Statistics

Statistical analysis was performed using SciPy v1.14.0 library in Python. Continuous variables between two groups were compared using the Student's t-test where data were normally distributed, and the Mann-Whitney U test for non-normally distributed data. Comparisons involving more than two groups were analyzed using one-way ANOVA for normally distributed data. Multiple testing was corrected using Benjamini-Hochberg (BH) method and adjusted *p*-values less than 0.05 were considered statistically significant unless otherwise indicated.

2.3 References

Altaie, A., Simone, D., McDermott, N., Owston, H., Attar, M., Jin, L., Wong, C., Loughenbury, P. R., Vishal, B., McMillan, T., Buckley, C. D., Sansom, S. N., & McGonagle, D. (2025). Healthy human enthesis stromal cells mediate immunoregulation via the CD39/CD73 adenosine ectonucleotidase pathway. *Ann Rheum Dis*. https://doi.org/10.1016/j.ard.2025.09.001

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R. P., Wolters, P. J., Abate, A. R., Butte, A. J., & Bhattacharya, M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*, *20*(2), 163–172. https://doi.org/10.1038/s41590-018-0276-y

Buttner, M., Ostner, J., Muller, C. L., Theis, F. J., & Schubert, B. (2021). scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun*, *12*(1), 6876. https://doi.org/10.1038/s41467-021-27150-6

COMBAT Consortium. (2022). A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell*, *185*(5), 916–938 e958. https://doi.org/10.1016/j.cell.2022.01.012

Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D., & Marioni, J. C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol*, *40*(2), 245–253. https://doi.org/10.1038/s41587-021-01033-z

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S. G., Heap, G., Nimmo, E. R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D. C.,…Barrett, J. C. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*, *49*(2), 256–261. https://doi.org/10.1038/ng.3760

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*, *11*(4), e1004219. https://doi.org/10.1371/journal.pcbi.1004219

Deschler, K., Rademacher, J., Lacher, S. M., Huth, A., Utzt, M., Krebs, S., Blum, H., Haibel, H., Proft, F., Protopopov, M., Rodriguez, V. R., Beltran, E., Poddubnyy, D., & Dornmair, K. (2022). Antigen-specific immune reactions by expanded CD8(+) T cell clones from HLA-B*27-positive patients with spondyloarthritis. *J Autoimmun*, *133*, 102901. https://doi.org/10.1016/j.jaut.2022.102901

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dominguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., Suchanek, O., Polanski, K., King, H. W., Mamanova, L., Huang, N., Szabo, P. A., Richardson, L., Bolt, L., Fasouli, E. S., Mahbubani, K. T., Prete, M., Tuck, L.,…Teichmann, S. A. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, *376*(6594), eabl5197. https://doi.org/10.1126/science.abl5197

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*, *4*(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

Fernandez, J. M., de la Torre, V., Richardson, D., Royo, R., Puiggros, M., Moncunill, V., Fragkogianni, S., Clarke, L., Consortium, B., Flicek, P., Rico, D., Torrents, D., Carrillo de Santa Pau, E., & Valencia, A. (2016). The BLUEPRINT Data Analysis Portal. *Cell Syst*, *3*(5), 491–495 e495. https://doi.org/10.1016/j.cels.2016.10.021

Fleming, S. J., Chaffin, M. D., Arduini, A., Akkad, A. D., Banks, E., Marioni, J. C., Philippakis, A. A., Ellinor, P. T., & Babadi, M. (2023). Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods*, *20*(9), 1323–1335. https://doi.org/10.1038/s41592-023-01943-7

Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., & Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat Methods*, *18*(3), 272–282. https://doi.org/10.1038/s41592-020-01050-x

Giudicelli, V., Chaume, D., & Lefranc, M. P. (2005). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res*, *33*(Database issue), D256–261. https://doi.org/10.1093/nar/gki010

Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K.,…Yates, A. D. (2024). Ensembl 2024. *Nucleic Acids Res*, *52*(D1), D891–D899. https://doi.org/10.1093/nar/gkad1049

Heaton, H., Talman, A. M., Knights, A., Imaz, M., Gaffney, D. J., Durbin, R., Hemberg, M., & Lawniczak, M. K. N. (2020). Souporcell: robust

clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods*, *17*(6), 615–620. https://doi.org/10.1038/s41592-020-0820-1

International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., Joo, K. B., Shim, S. C., Weisman, M., Ward, M., Zhou, X., Garchon, H. J., Chiocchia, G., Nossent, J.,…Brown, M. A. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*, *45*(7), 730–738. https://doi.org/10.1038/ng.2667

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012. https://doi.org/10.1101/060012

Liu, B., Li, C., Li, Z., Wang, D., Ren, X., & Zhang, Z. (2020). An entropy-based metric for assessing the purity of single cell populations. *Nat Commun*, *11*(1), 3155. https://doi.org/10.1038/s41467-020-16904-3

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C., & Hume, D. A. (2013). An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*, *14*, 632. https://doi.org/10.1186/1471-2164-14-632

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R. R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J. C., Carroll, R. J., Eyler, A. E., Greenberg, J. D., Kremer, J. M.,…Plenge, R. M. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, *506*(7488), 376–381. https://doi.org/10.1038/nature12873

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Sloan, C. A., Chan, E. T., Davidson, J. M., Malladi, V. S., Strattan, J. S., Hitz, B. C., Gabdank, I., Narayanan, A. K., Ho, M., Lee, B. T., Rowe, L. D., Dreszer, T. R., Roe, G., Podduturi, N. R., Tanaka, F., Hong, E. L., & Cherry, J. M. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Res*, *44*(D1), D726–732. https://doi.org/10.1093/nar/gkv1160

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888–1902 e1821. https://doi.org/10.1016/j.cell.2019.05.031

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome

Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, *43*(1110), 11 10 11–11 10 33. https://doi.org/10.1002/0471250953.bi1110s43

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E.,…SciPy, C. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst*, *8*(4), 281–291 e289. https://doi.org/10.1016/j.cels.2018.11.005

Xin, H., Lian, Q., Jiang, Y., Luo, J., Wang, X., Erb, C., Xu, Z., Zhang, X., Heidrich-O'Hare, E., Yan, Q., Duerr, R. H., Chen, K., & Chen, W. (2020). GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome Biol*, *21*(1), 188. https://doi.org/10.1186/s13059-020-02084-2

Zhang, M. J., Hou, K., Dey, K. K., Sakaue, S., Jagadeesh, K. A., Weinand, K., Taychameekiatchai, A., Rao, P., Pisco, A. O., Zou, J., Wang, B., Gandal, M., Raychaudhuri, S., Pasaniuc, B., & Price, A. L. (2022). Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet*, *54*(10), 1572–1580. https://doi.org/10.1038/s41588-022-01167-z

Chapter 3: Investigation of Axial Spondyloarthritis Patients and Healthy Individuals at Single-Cell Level

3.1 Introduction

Single-cell RNA-sequencing (scRNA-seq) analyses have identified distinct immune-cell patterns in patients with axial spondyloarthritis (AxSpA), revealing dysregulated inflammatory pathways and altered peripheral immune landscapes compared with healthy individuals, as introduced in Chapter 1. These observations highlight the perturbation of inflammatory and immune-activation pathways characteristic of the disease.

Tumour-necrosis-factor inhibitors (TNFi) have proven clinically effective in controlling inflammation and improving symptoms in AxSpA (Braun et al., 2002; Callhoff et al., 2015; van der Heijde et al., 2006). However, the cellular and molecular mechanisms underlying the treatment responses are not yet fully understood. Key outstanding questions include whether TNFi therapy restores immune profiles toward those of healthy individuals or instead reshapes peripheral immune signatures into alternative transcriptional states, and which molecular features are associated with differential therapeutic responses in AxSpA.

Several bulk transcriptomic studies have investigated molecular responses to TNFi therapy and reported modulation of TNF and interferon signalling pathways, along with altered expression of AS-associated genes such as *IL6R, IL10*, and *NOTCH1*, and differential regulation of innate and adaptive immune signatures distinguishing treatment responders (Sobral et al., 2024; Wang et al., 2017). While these studies have provided valuable insights, bulk transcriptomic approaches

inherently average gene expression across heterogeneous immune cell populations, which may obscure immune cell subtype-specific changes and rare populations that may be critical for the therapeutic response.

Advances in high-dimensional immunoprofiling technologies, including mass cytometry, scRNA-seq, and single-cell TCR/BCR sequencing, now enable detailed investigation of immune-cell states at single-cell resolution. Recent single-cell transcriptomic and proteomic profiling of peripheral blood from AxSpA patients before and after anti-TNFα therapy revealed expansion of granulocytes and inflammatory monocytes at baseline, attenuation of CXCL8/2–CXCR1/2-mediated monocyte–granulocyte interactions, and an increase in functional regulatory T cells following treatment (Zhao et al., 2023). In parallel, another single-cell multi-omics analyses have demonstrated that elevated baseline type I interferon (IFN) signatures are associated with poor TNFi response, with non-responders exhibiting enhanced IFN and Th17 activity after treatment, and AIF-1 identified as a predictive biomarker of TNFi resistance (Song et al., 2025).

Although these studies have advanced our understanding of the immune landscape in AxSpA and its modulation by TNFi, comprehensive single-cell and longitudinal studies remain limited in number and scope, and the transcriptional programs governing treatment response across immune-cell subsets are still not fully defined. Building on these emerging insights, I performed scRNA-seq of PBMCs from AxSpA patients, comparing their immune-cell transcriptional profiles with those of healthy individuals and evaluating longitudinal changes in TNFi-treated responders. The analysis incorporated computational approaches such as cell-type composition modelling using scCODA (Buttner et al., 2021), cell-level polygenic disease-risk scoring with scDRS (Zhang et al., 2022), and differential gene-

expression analysis with EdgeR (Robinson et al., 2010), to delineate both shifts in immune-cell proportions and transcriptional programs at single-cell resolution.

3.2 Aim

This chapter aims to investigate two central questions:

- What are the detailed compositional and transcriptional differences in PBMC between AxSpA patients and healthy controls as revealed by scRNA-seq?
- How does TNF inhibitor treatment influence peripheral immune signatures in AxSpA patients who respond clinically to treatment?

Addressing these questions has direct translational relevance, as it provides insights into the cellular mechanisms underlying AxSpA pathogenesis and therapeutic response. Identifying specific immune-cell populations and transcriptional changes associated with effective treatment responses may inform future therapeutic strategies and help address unmet clinical needs in AxSpA.

3.3 Results

3.3.1 Experimental design

To characterize the single-cell transcriptional landscape of PBMCs in AxSpA, scRNA-seq was performed on PBMCs from n = 8 anti-TNF-treated and n=9 biologic-naïve AxSpA patient samples, as well as n = 3 HLA-B*27-positive healthy controls generated in this study. To expand the healthy reference cohort, data from an additional n = 10 healthy individuals were integrated from the University of Oxford COMBAT project (COMBAT Consortium, 2022). For the comparative analyses with AxSpA presented later in this chapter, the COMBAT cohort was further restricted to n

= 6 male donors to match the all-male AxSpA cohort and minimise sex-related confounding. Among all samples, n = 8 were profiled using cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), providing both transcriptomic and surface protein information, while the remaining samples were analysed by scRNA-seq alone. All samples were processed across 4 experimental batches, as illustrated in the experimental design overview (**Figure 3.1**). I performed all data analyses presented in this chapter, while sample collection and library preparation were carried out by members of the Sansom, Bowness, and Bowden laboratories.

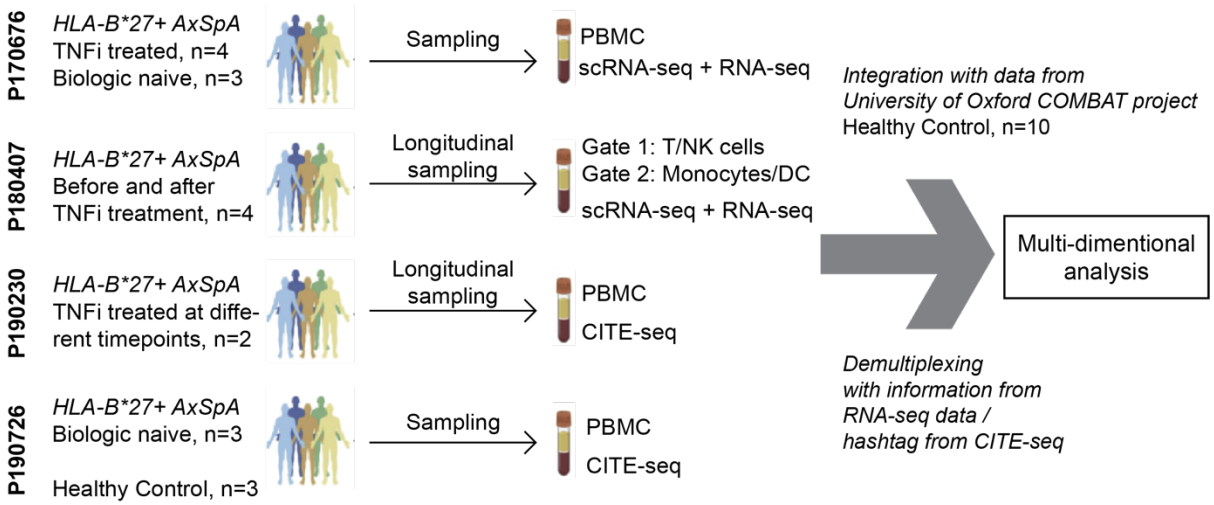


Figure 3.1 Diagram for experimental design and sample collection
PBMC samples from TNFi-treated AxSpA patients, biologic-naïve AxSpA patients, and healthy controls were collected and sequenced across 4 experimental batches (P170676, P180407, P190230 and P190726). Longitudinal samples were obtained at multiple timepoints from the same individuals in batches P180407 and P190230. Samples in project P180407 were sorted into T/NK and monocyte/dendritic cell fractions prior to sequencing. The first two batches underwent 3' scRNA-seq and bulk RNA-seq, while the latter two batches were profiled using 5' scRNA-seq with CITE-seq . n denotes the number of samples.

Clinical metadata for the samples analyzed in this chapter are summarized in **Table 3.1**. Batch-specific information, including the use of these samples across this

and the subsequent results chapters, is provided in **Table 2.1**. Comprehensive participant metadata covering all individuals studied in this thesis are provided in **Appendix** and will not be repeated in subsequent chapters.

	HC_Oxford (n=3)	HC_COMBAT (n=6)	biologic-naïve AxSpA (n=9)	TNFi-treated AxSpA (n=7)
Age (years), mean ± *SD	55.3 ± 10.1		48.3 ± 11.3	42 ± 10.1
Sex, male (%)	3 (100%)	3 (50%)	8 (88.9%)	7 (100%)
HLA-B27+ number (%)	3 (100%)	0 (0%)	8 (88.9%)	7 (100%)
BASDAI, mean ± SD	N/A	N/A	5.8 ± 1.5	3.0 ± 2.1
CRP, mean ± SD	N/A	N/A	13.7 ± 17.9	1.4 ± 1.8

Table 3.1 Clinical metadata of AxSpA patients and healthy controls analysed in Chapter 3
Summary of age, HLA-B*27 status, Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) scores, and C-reactive protein (CRP) levels for study participants included in this chapter. * *SD: Standard deviation.*

3.3.2 PBMC scRNA-seq data processing and analysis

3.3.2.1 Batch-specific preprocessing

As the 4 batches were generated using different 10x Genomics chemistries and demultiplexing strategies, each batch was preprocessed separately prior to integration. The preprocessing workflow for each batch including ambient RNA removal, demultiplexing, and initial quality-control steps is illustrated in detail in **Figure 3.2** and is also described in the **Methods**. Project P180407 underwent initial clustering and cell-type annotation for sample identity inference (**Appendix B.1**). Major lineage assignments (T/NK and Mono/DC) were used to infer sample identities: within each channel, cells were assigned to the individual whose sorted fraction they originated from.

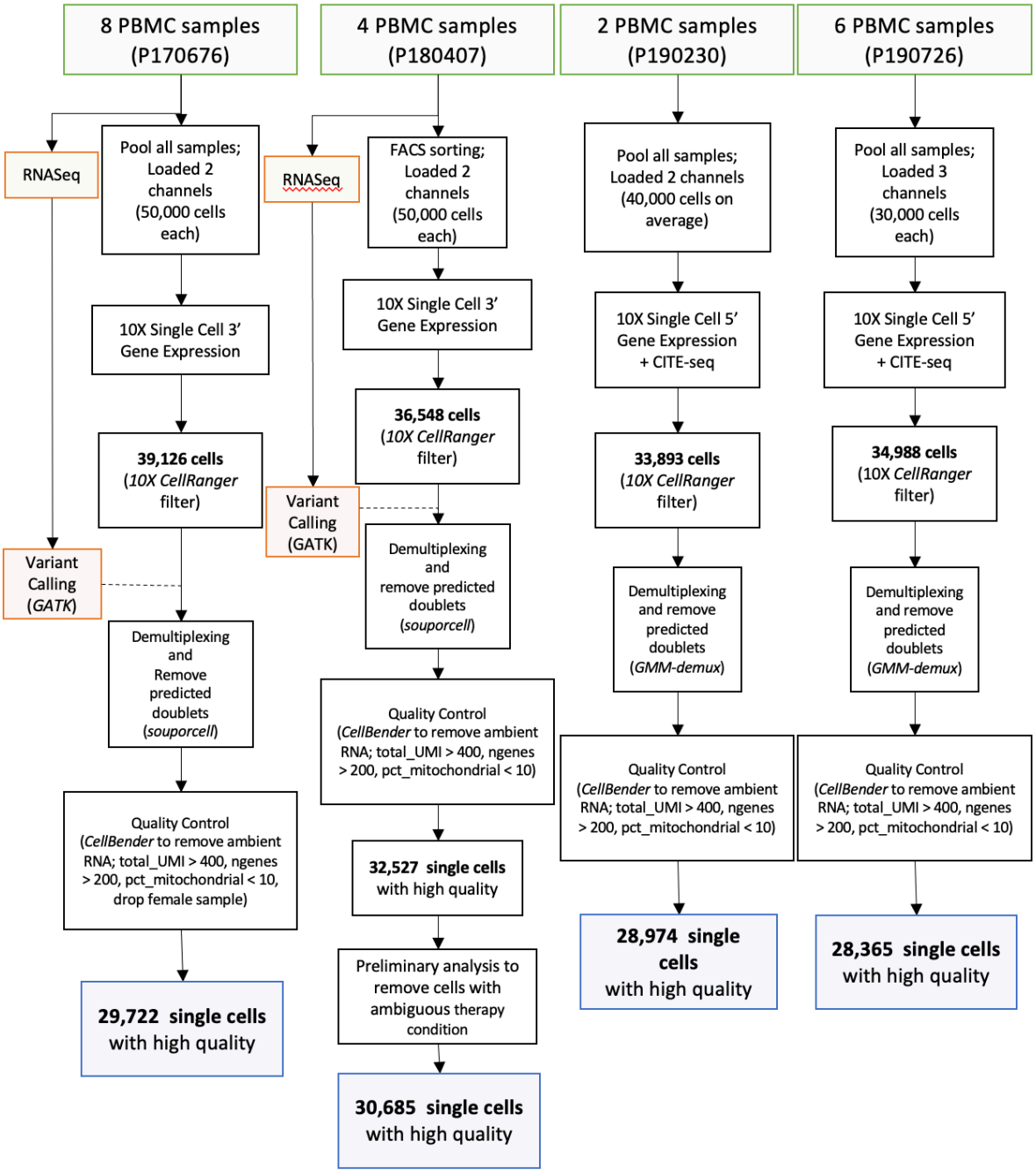


Figure 3.2 Overview of preprocessing workflow for 4 experimental batch
Schematic representation of the preprocessing steps applied to each experimental batch. Each batch underwent independent library preparation and preprocessing, followed by demultiplexing using either RNA-based genetic variation or hashtag antibodies. Batch-specific quality control was performed subsequently, with the second batch undergoing an additional preliminary analysis step to remove misclustered cells.

The distribution of total gene counts, total UMI counts, and mitochondrial gene percentages for each of the 4 batches is shown in **Figure 3.3**. To ensure removal of

low-quality cells, initial quality control was performed based on these metrics. Cells with fewer than 400 UMIs, fewer than 200 detected genes, or greater than 10% mitochondrial gene content were excluded. In addition, one sample with an ambiguous treatment condition and all predicted doublets were removed, resulting in the exclusion of 26,809 cells. After filtering, a total of 117,746 high-quality single cells were retained for downstream analyses.

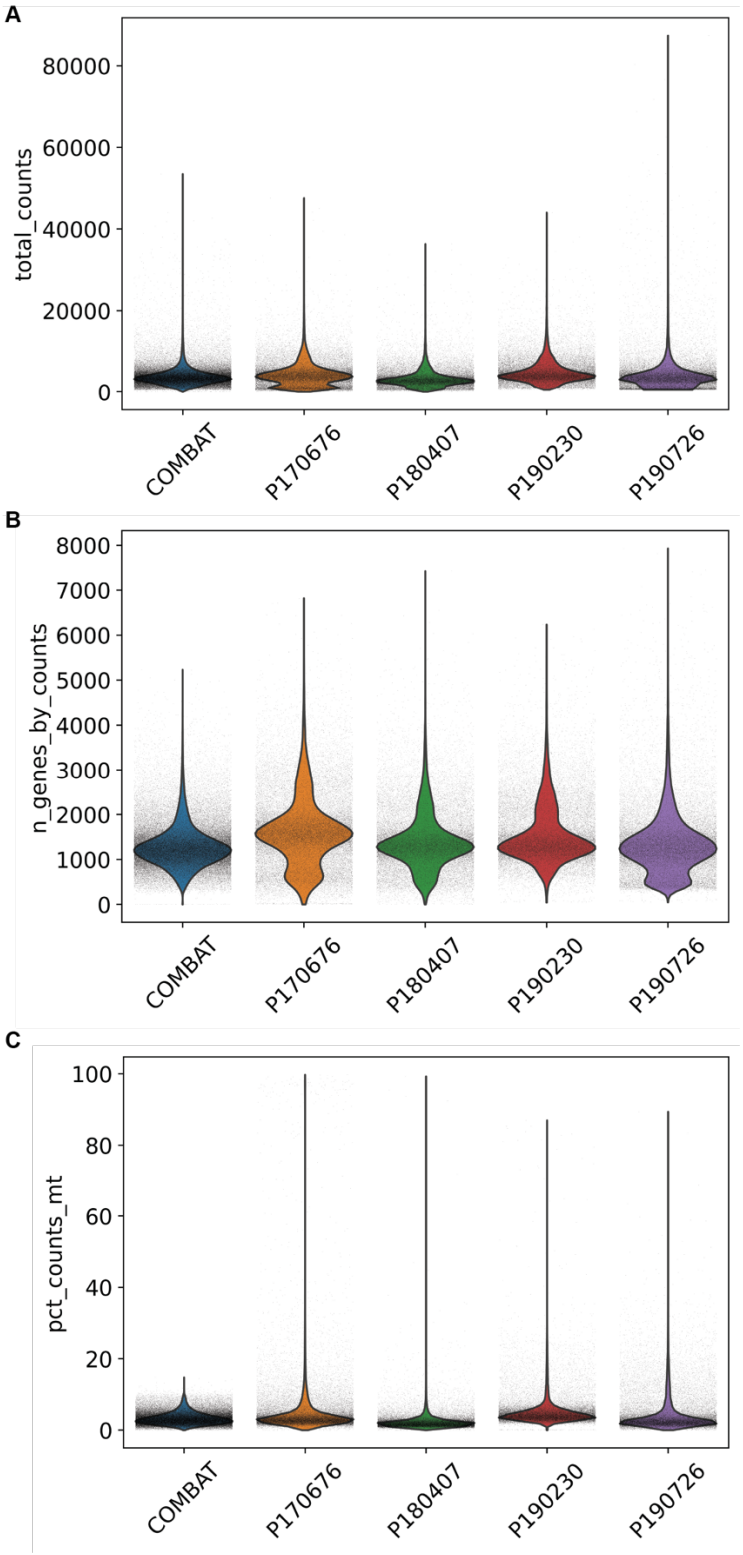


Figure 3.3 Initial quality control and filtering of single-cell data
Violin plots showing the distribution of (A) total UMI counts, (B) number of genes detected per cell, and (C) percentage of mitochondrial gene expression across each library.

3.3.2.2 Integration of 4 AxSpA PBMC batches with the COMBAT dataset

Data from the COMBAT project were integrated to include additional healthy control samples, broadening the atlas and enabling more comprehensive comparative analyses. Several integration methods, including Scanorama (Hie et al., 2024), scVI (Lopez et al., 2018), Harmony (Korsunsky et al., 2019), and BBKNN (Polanski et al., 2020), were initially evaluated. Based on visual inspection of batch and sample mixing, scVI was selected as the most effective approach. Within the scVI framework, three strategies for selecting highly variable genes (HVGs) were then assessed: (1) integration using the union of 2,000 HVGs from each batch; (2) integration using a set of 7,500 HVGs selected across all batches, and (3) integration using the union of 7,500 HVGs from each batch.

The optimal HVG strategy was determined by qualitative evaluation against expert cell-type annotations previously established in the COMBAT project. Evaluation focused on three criteria visualized in **Figure 3.4**: (1) the clarity and resolution of major and minor COMBAT cell-type annotations on the integrated UMAP; (2) effective mixing of samples across the four AxSpA batches and the COMBAT cohort; (3) mixing of cells across 3'-end and 5'-end scRNA-seq modalities. Strategy 3 (union of 7,500 HVGs per batch) was selected, as it best balanced cell-type resolution with batch and modality mixing. Strategy 1 (2,000 HVGs union per batch) showed reduced resolution of minor cell-type subsets (e.g. CD4.TREG, GDT.VD2 and CD8 subsets), while Strategy 2 (7,500 HVGs jointly selected across batches) showed less effective batch and modality mixing. This qualitative evaluation framework was subsequently extended to include quantitative metrics in the Chapter 4 integration analyses, where a wider range of tissues and modalities required more formal benchmarking.

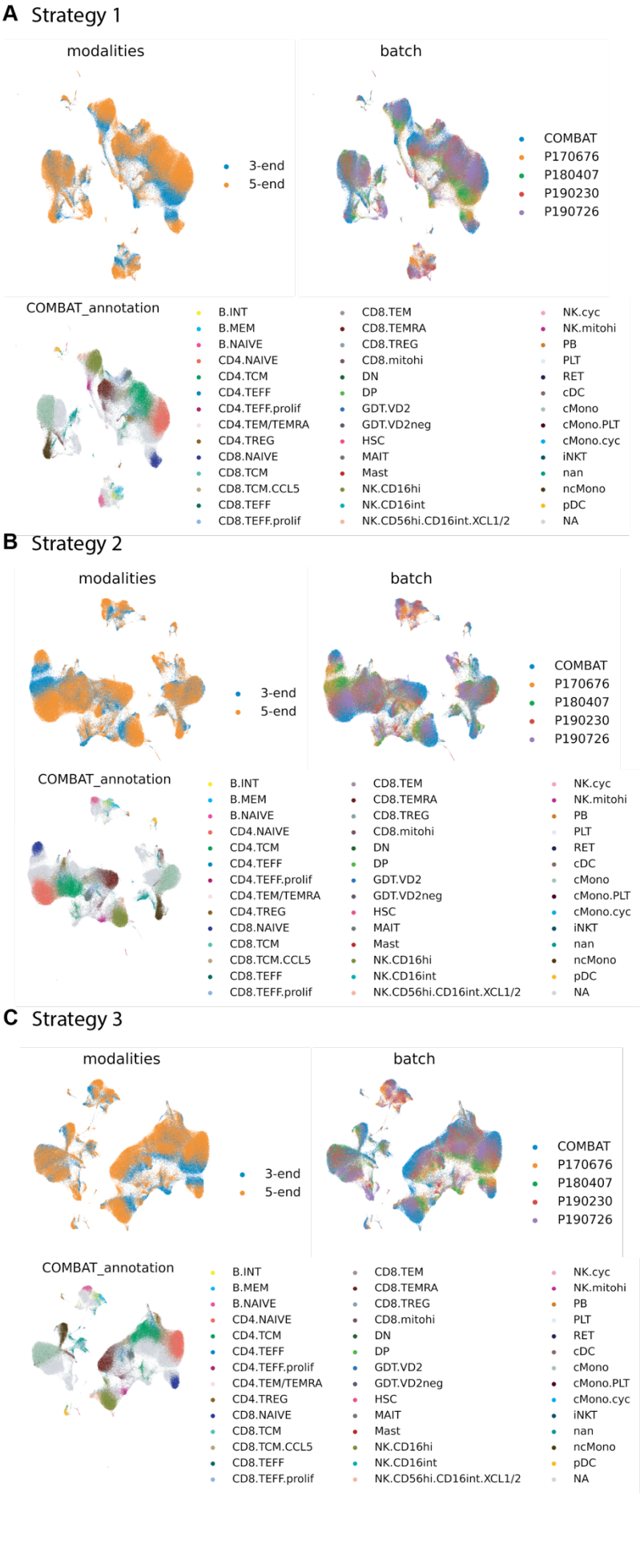

A Strategy 1
modalities
3-end
5-end
batch
COMBAT
P170676
P180407
P190230
P190726
COMBAT_annotation
B.INT
B.MEM
B.NAIVE
CD4.NAIVE
CD4.TCM
CD4.TEFF
CD4.TEFF.prolif
CD4.TEM/TEMRA
CD4.TREG
CD8.NAIVE
CD8.TCM
CD8.TCM.CCL5
CD8.TEFF
CD8.TEFF.prolif
CD8.TEM
CD8.TEMRA
CD8.TREG
CD8.mitohi
DN
DP
GDT.VD2
GDT.VD2neg
HSC
MAIT
Mast
NK.CD16hi
NK.CD16int
NK.CD56hi.CD16int.XCL1/2
NK.cyc
NK.mitohi
PB
PLT
RET
cDC
cMono
cMono.PLT
cMono.cyc
iNKT
nan
ncMono
pDC
NA
B Strategy 2
modalities
3-end
5-end
batch
COMBAT
P170676
P180407
P190230
P190726
COMBAT_annotation
B.INT
B.MEM
B.NAIVE
CD4.NAIVE
CD4.TCM
CD4.TEFF
CD4.TEFF.prolif
CD4.TEM/TEMRA
CD4.TREG
CD8.NAIVE
CD8.TCM
CD8.TCM.CCL5
CD8.TEFF
CD8.TEFF.prolif
CD8.TEM
CD8.TEMRA
CD8.TREG
CD8.mitohi
DN
DP
GDT.VD2
GDT.VD2neg
HSC
MAIT
Mast
NK.CD16hi
NK.CD16int
NK.CD56hi.CD16int.XCL1/2
NK.cyc
NK.mitohi
PB
PLT
RET
cDC
cMono
cMono.PLT
cMono.cyc
iNKT
nan
ncMono
pDC
NA
C Strategy 3
modalities
3-end
5-end
batch
COMBAT
P170676
P180407
P190230
P190726
COMBAT_annotation
B.INT
B.MEM
B.NAIVE
CD4.NAIVE
CD4.TCM
CD4.TEFF
CD4.TEFF.prolif
CD4.TEM/TEMRA
CD4.TREG
CD8.NAIVE
CD8.TCM
CD8.TCM.CCL5
CD8.TEFF
CD8.TEFF.prolif
CD8.TEM
CD8.TEMRA
CD8.TREG
CD8.mitohi
DN
DP
GDT.VD2
GDT.VD2neg
HSC
MAIT
Mast
NK.CD16hi
NK.CD16int
NK.CD56hi.CD16int.XCL1/2
NK.cyc
NK.mitohi
PB
PLT
RET
cDC
cMono
cMono.PLT
cMono.cyc
iNKT
nan
ncMono
pDC
NA

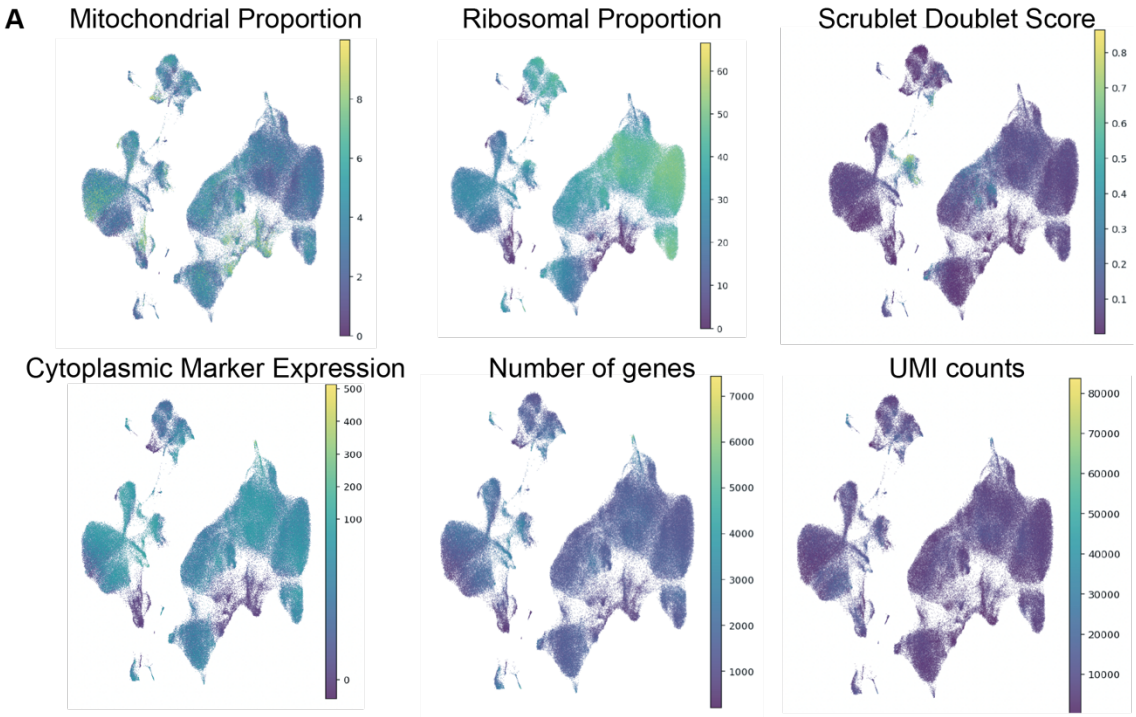
Figure 3.4 Comparison of different HVG selection strategies for data integration
(A–C) UMAP visualizations of the integrated dataset generated using three distinct HVG selection strategies: (A) Strategy 1: union of 2000 HVGs from each batch; (B) Strategy 2: combined selection of 7500 HVGs across all batches; and (C) Strategy 3: union of 7500 HVGs from each batch (final selected strategy). UMAP plots illustrate data modality, batch origin and cell subsets annotation from COMBAT project to demonstrate the integration quality for each strategy.

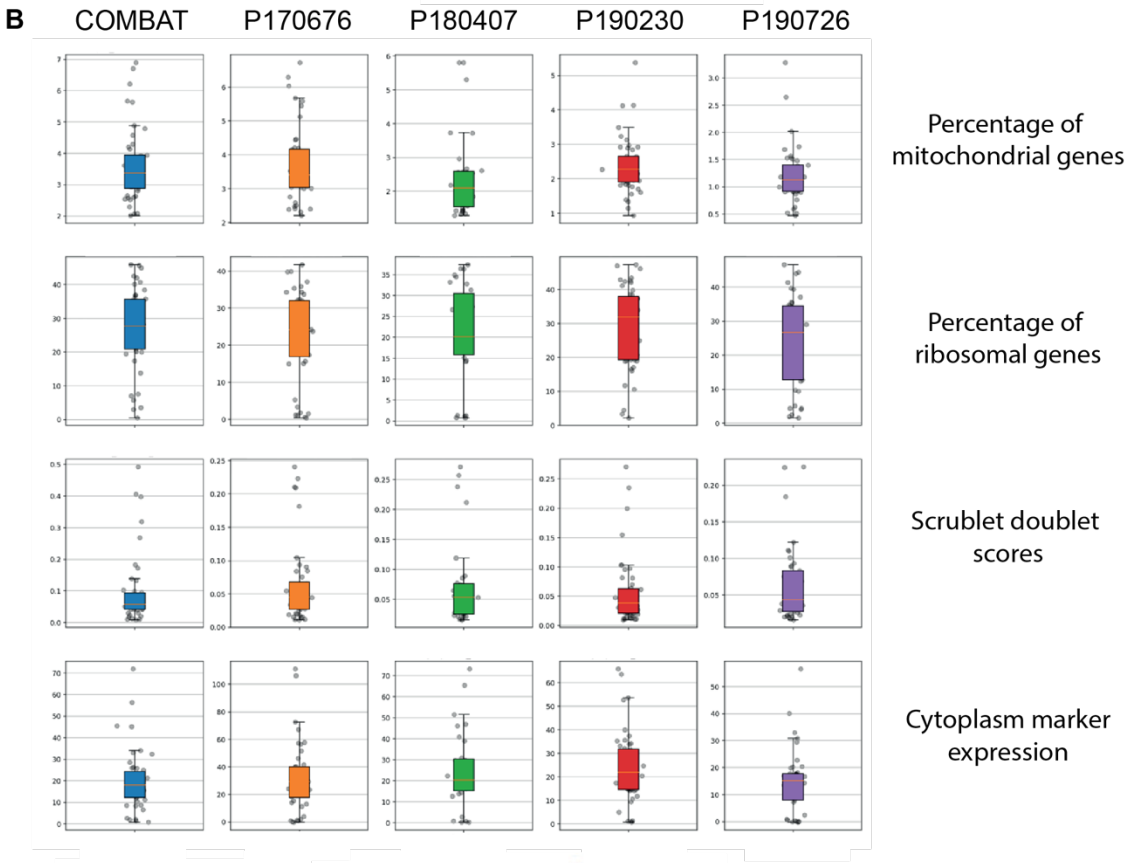
3.3.2.3 Advanced quality control and construction of final atlas

To account for technical heterogeneity arising from differences in sample preparation, sequencing chemistry and library construction across projects, quality control was performed at sample-cluster level. Doublet scores were calculated using Scrublet (Wolock et al., 2019), and cytoplasmic marker expression was defined as the average expression of *TPT1* and *TMSB4X*, two highly expressed cytoplasmic housekeeping genes. Although no prior publication has used these precise genes as a QC cutoff, they were selected because they are broadly and stably expressed according to the Human Protein Atlas, and using stable cytoplasmic markers to assess cell integrity is an established quality-control practice in single-cell RNA-seq analysis (Luecken & Theis, 2019).

For each sample-cluster group, upper and lower thresholds for quality metrices were defined based on the interquartile range (IQR). Specifically, upper and lower bounds were defined as the 75th percentile + 1.5 × IQR and 25th percentile − 1.5 × IQR, respectively, for total UMI counts and number of detected genes; for mitochondrial and ribosomal gene percentages and the doublet score, only the upper bound was applied, whereas for cytoplasmic marker genes, only the lower bound was used. Cells falling outside these thresholds were identified as outliers and removed as part of a secondary round of quality control. Applying these criteria, 6,089 predicted doublets and 16,856 low-quality cells were excluded from further analysis (**Figure 3.5**). In addition, a small number of “misclustered cells” were

removed from the P180407 project; these were cells that did not correspond to the expected cell-type fractions defined during initial sorting (i.e., cells sorted in the $CD3^+$ $CD56^+$ T/NK gate but ending up in the B-cell cluster in the P180407 dataset) and were identified based on their marker expression profiles, as described in the **Methods** section.


A
Mitochondrial Proportion
Ribosomal Proportion
Scrublet Doublet Score
Cytoplasmic Marker Expression
Number of genes
UMI counts


B
COMBAT
P170676
P180407
P190230
P190726
Percentage of mitochondrial genes
Percentage of ribosomal genes
Scrublet doublet scores
Cytoplasm marker expression

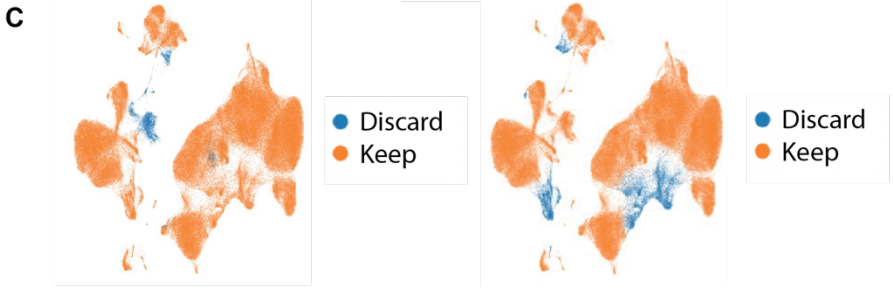

C
Discard
Keep
Discard
Keep

Figure 3.5 Secondary quality control assessment after data integration
(A) UMAP visualization of all cells after initial quality control, split by key QC metrics. (B) Boxplots displaying the distribution of 4 key quality control metrics across different projects, including mitochondrial gene percentage, ribosomal gene percentage, Scrublet doublet scores, and cytoplasmic marker gene expression. (C) UMAP plots showing cells after the removal of doublets (left) and low-quality cells (right).

Following this stringent filtering, the remaining 149,411 cells were re-integrated and re-clustered using the methods and parameters described above. This final integration and clustering step resulted in the generation of an integrated PBMC atlas as shown in **Figure 3.6**.

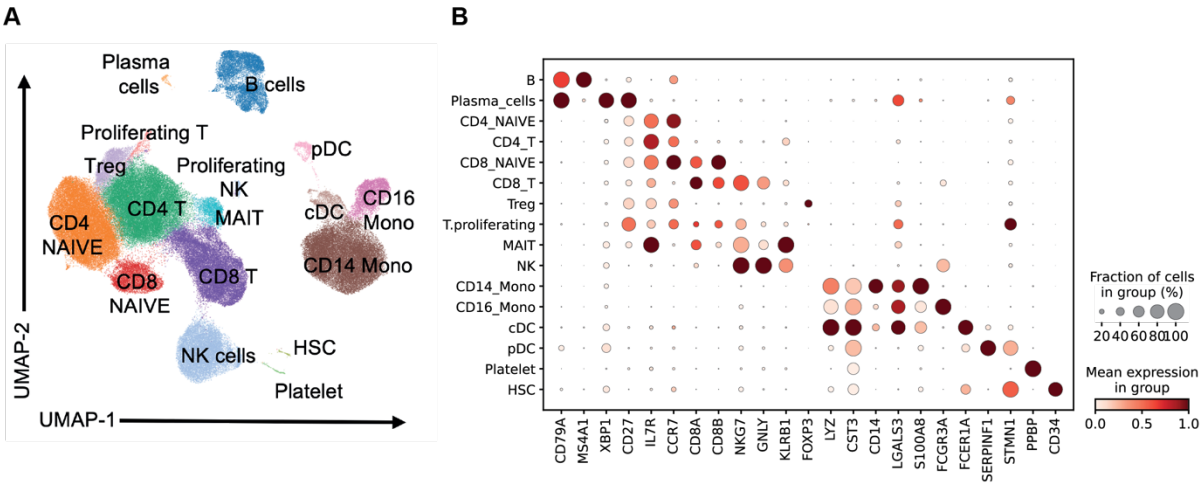


Figure 3.6 Integrated PBMC atlas of AxSpA patients and healthy controls
(A) UMAP visualization of the final integrated PBMC atlas highlighting annotated cell types. (B) Dot plot illustrating representative marker gene expression patterns across the major identified cell types.

3.3.3 Biological naïve and TNFi treated AxSpA patients exhibit a shift in immune cell composition compared to healthy controls

3.3.3.1 Increased B cell frequency in biologic-naïve and TNFi-treated AxSpA patients

Next, I explored compositional and transcriptional variation among the identified cell populations. Density plots on the integrated UMAP revealed that cells from healthy controls, biologic-naïve AxSpA patients, and anti-TNF-treated AxSpA

patients occupied distinct regions of the embedding, suggesting potential underlying transcriptional differences between groups (**Figure 3.7A**).

I observed that in the CD4 naïve T cell compartment, cells from biologic-naïve and anti-TNF-treated patients largely overlapped, but both remained distinct from healthy controls. This suggested that transcriptional differences were present in the CD4 naïve T cells in AxSpA patients even after TNF inhibitor therapy. Interestingly, within the myeloid compartment, particularly among $CD16^+$ monocytes, anti-TNF-treated cells occupied a distinct position compared to healthy and biologic-naïve samples. This observation implies that TNF inhibitor treatment may reshape the immune phenotype of $CD16^+$ monocytes, shifting them towards a potentially new transcriptional state.

To further investigate the compositional shifts visually observed in the density plots, I performed co-varying neighborhood analysis (CNA) (Reshef et al., 2022) to systematically identify statistically significant differences in cellular neighborhood abundance between AxSpA patients and healthy controls. No neighborhoods showed significant compositional changes after correcting for multiple testing (BH-adjusted $p < 0.05$) as shown in **Figure 3.7B-D**. This suggests that while subtle, localized transcriptional shifts within specific cell types (e.g. CD4 naïve T cells and $CD16^+$ monocytes) may exist, they were not sufficient to significantly alter the overall neighborhood structure. Therefore, these observed transcriptional differences likely reflect subtle, cell-type-specific changes rather than broader neighborhood-level shifts, and will be explored further in subsequent cell-type-focused analyses.

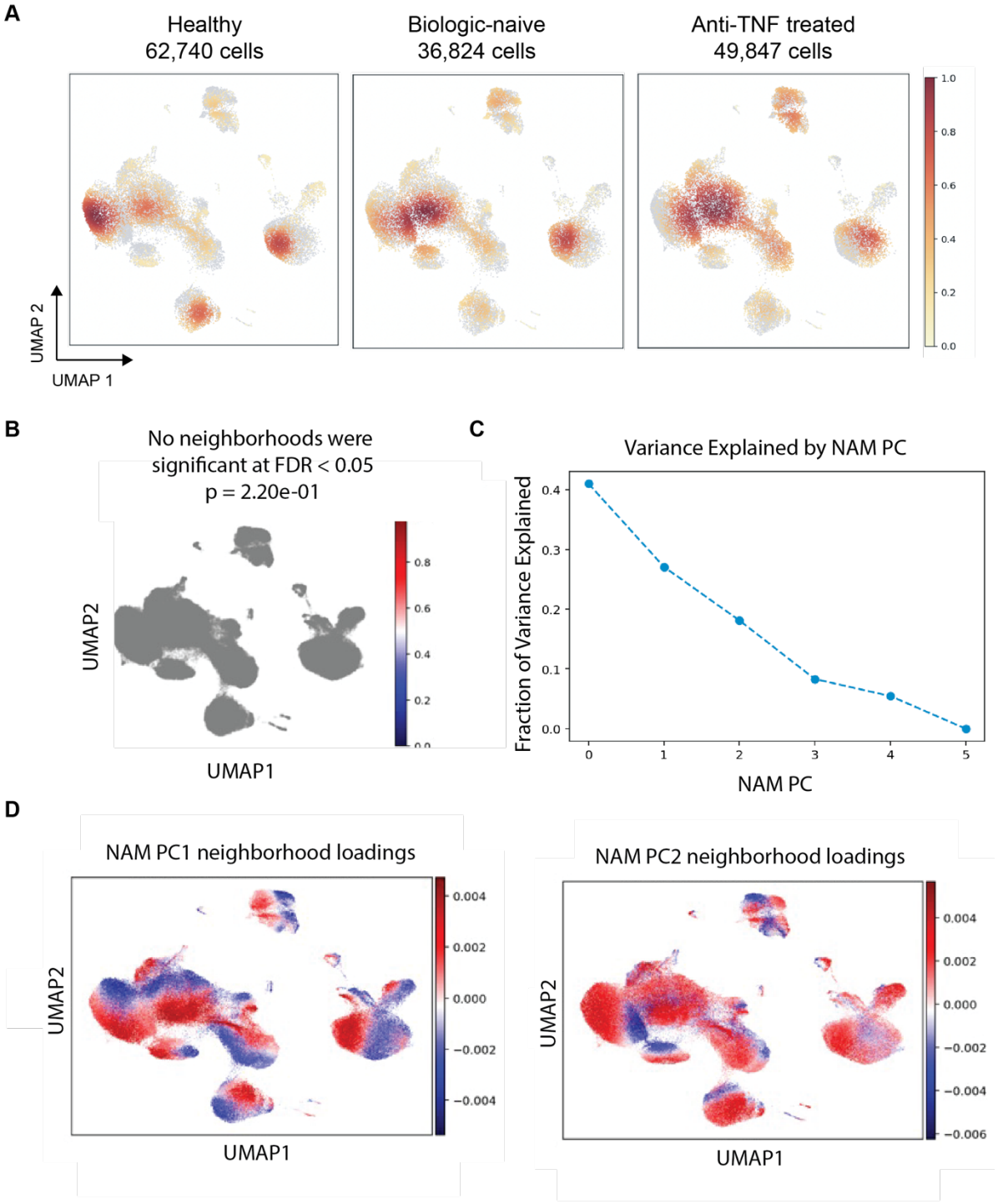
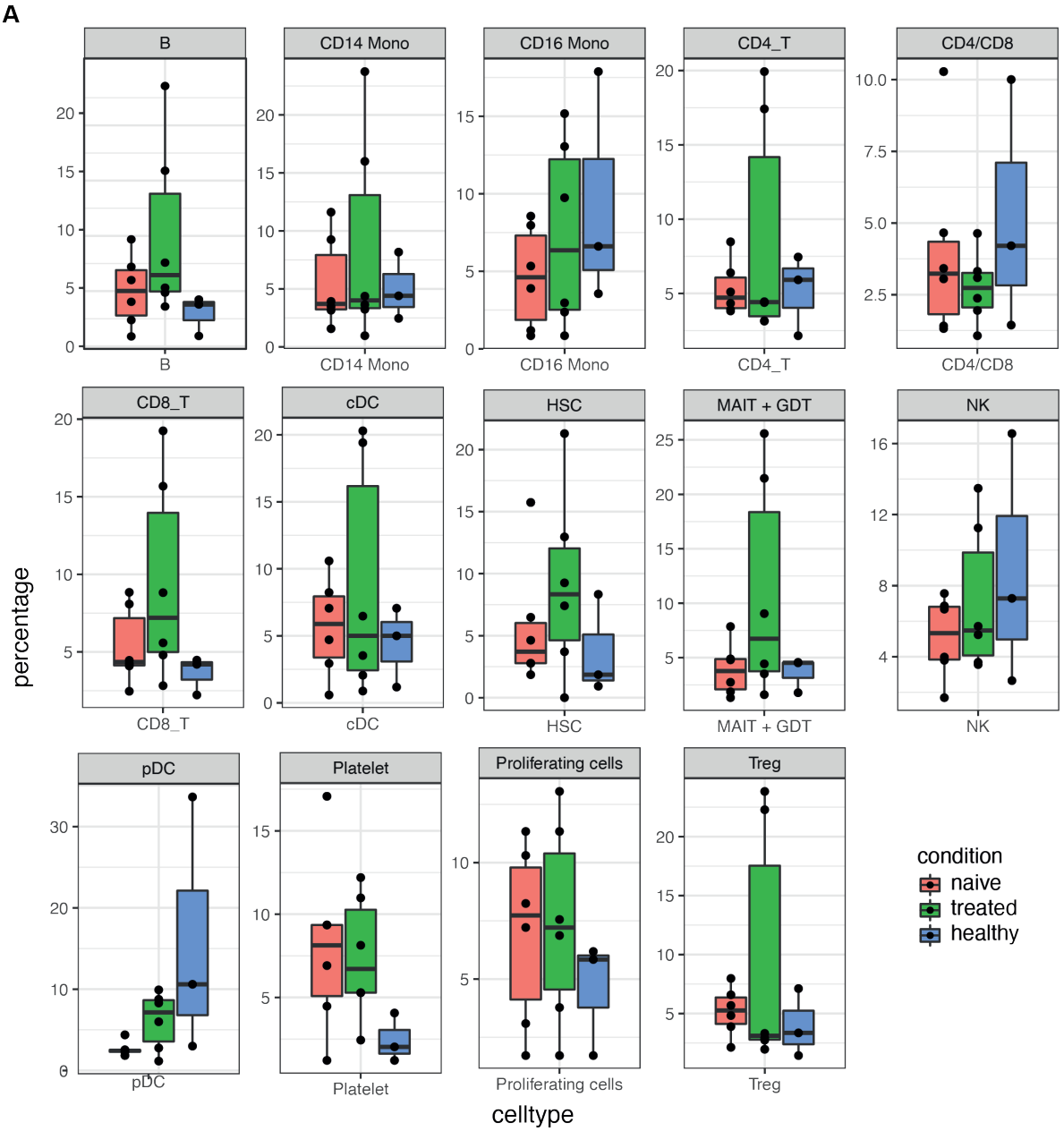


Figure 3.7 Visualization of compositional shifts and CNA analysis
(A) Density plots showing the distribution of cells from healthy controls, biologic-naïve AxSpA patients, and anti-TNF-treated AxSpA patients overlaid on the integrated UMAP embedding. Higher density regions indicate an increased abundance of cells within specific transcriptional spaces. No neighborhoods exhibited significant compositional differences between AxSpA and healthy control conditions after multiple-testing correction (adjusted p-value < 0.05). (B) Results of CNA showing the assessment of neighborhood abundance differences across conditions. (C) Variance explained by first five neighborhood abundance matrix (NUM) principal components. (D) UMAP visualization highlighting neighborhoods identified by CNA: the left panel shows NAM PC1 neighborhood loadings, and the right panel shows NUM PC2 loadings.

Next, to accurately quantify cell-type composition and minimize potential technical bias identified in the previous analysis, I excluded cells from batches that had been sorted by flow cytometry to enrich specific subsets, as well as those processed using different protocols. Boxplots illustrating the relative proportions of each cell type across conditions revealed differences in certain cell populations between AxSpA patient groups and healthy controls (**Figure 3.8A**). To assess these differences statistically, I performed t-tests comparing cell-type proportions between biologic-naïve AxSpA patients and healthy controls, and between anti-TNF-treated and biologic-naïve AxSpA patients. However, no clusters passed the false discovery rate (FDR) threshold of 0.05 after p-value adjustment, likely due to the limited sample size. Because compositional changes are inherently interdependent, I further applied the Bayesian model scCODA (Buttner et al., 2021) to robustly identify statistically credible compositional differences (**Figure 3.8B**). The scCODA analysis revealed that biologic-naïve AxSpA patients exhibited a statistically credible increase in B cell lineage compared to healthy controls, consistent with previous findings (Liu et al., 2021).

A


B
CD14 Mono
CD16 Mono
CD4_T
CD4/CD8
CD8_T
cDC
HSC
MAIT + GDT
NK
pDC
Platelet
Proliferating cells
Treg
percentage
celltype
condition
naive
treated
healthy

B

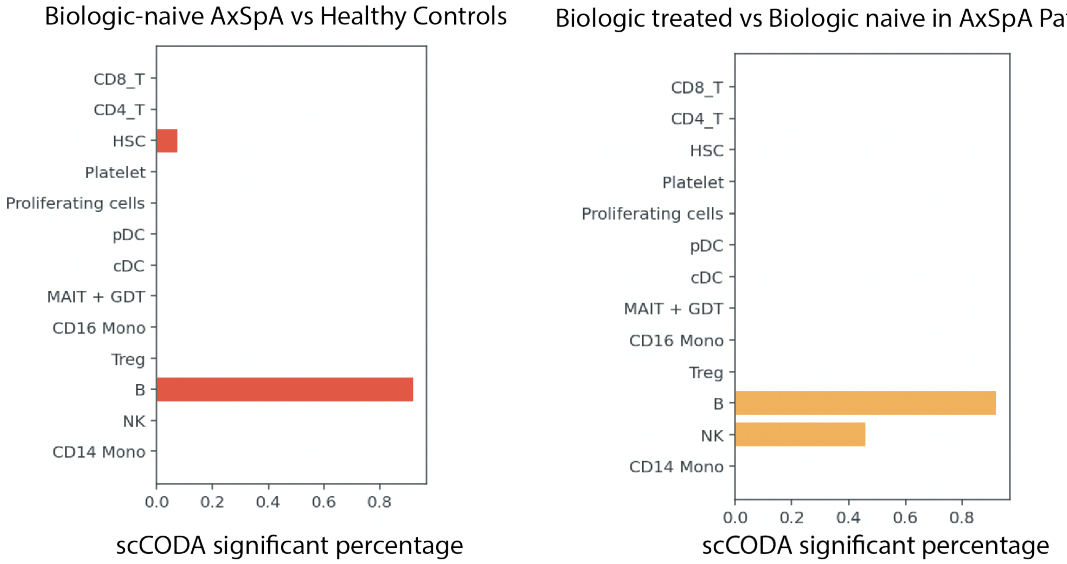

Biologic-naive AxSpA vs Healthy Controls
Biologic treated vs Biologic naive in AxSpA Patients
CD8_T
CD4_T
HSC
Platelet
Proliferating cells
pDC
cDC
MAIT + GDT
CD16 Mono
Treg
B
NK
CD14 Mono
scCODA significant percentage
scCODA significant percentage

Figure 3.8 Compositional analysis of PBMCs from healthy controls, biologic-naïve AxSpA patients, and anti-TNF-treated AxSpA patients
(A) Boxplots showing the proportions of major immune-cell populations and CD4:CD8 ratios in healthy controls, biologic-naïve AxSpA patients, and TNFi-treated AxSpA patients. (B) Bar plots summarising the scCODA results for compositional differences across conditions. The x-axis indicates the percentage of scCODA runs (n = 13, each using a different cell type as reference) in which a statistically credible difference was detected (FDR < 0.2).

3.3.3.2 Identification of a CD9⁺ monocyte subset unique to AxSpA patients

To further investigate whether specific subpopulations within major cell types contributed to the compositional differences observed between AxSpA patients and healthy controls, I performed sub-clustering analyses of mononuclear phagocytes (MNPs), T cells, and B cell compartment (including plasma cells) individually (**Figure 3.9**).

Sub-clustering of MNPs revealed distinct dendritic cell and monocyte subsets (**Figure 3.9A-B**). A plasmacytoid dendritic cell (pDC) cluster was marked by high expression of *SERPINF1*, whereas a conventional dendritic cell (cDC) cluster expressed *FCER1A* and *ITGB7*. The monocyte compartment clearly separated into CD14⁺ and CD16⁺ subsets. Within the CD14⁺ monocytes, three transcriptionally distinct subpopulations were identified, each defined by a characteristic gene-expression signature: an IL17RA⁺ subset (Mono.CD14.IL17RA) that also exhibited the highest *CD14* expression; a CCR1⁺ subset (Mono.CD14.CCR1); and a CD9⁺CD81⁺ subset (Mono.CD14.CD9). The IL17RA⁺ monocytes may reflect enhanced responsiveness to IL-17 signaling, a pathway known to play a central role in the inflammatory pathology of AxSpA. The CD9⁺ monocyte subset expressed characteristic tetraspanin genes, particularly *CD9* and *CD81*. Elevated *CD9* expression has previously been linked to anti-inflammatory macrophage phenotypes and to increased potential for cell–cell fusion, suggesting that this subset may

represent a specialized functional state relevant to chronic inflammation in AxSpA (Champion et al., 2018; Reyes et al., 2018; Tippett et al., 2013).

Sub-clustering of the B-cell lineage first separated B cells from $CD27^+$ plasma cells (**Figure 3.9C-D**). Within the naive B-cell subset, further separation yielded a memory B cell cluster distinguished by expression of *GPR183* and *NAAA*. Among conventional B cells, two additional subsets emerged: one marked by high expression of *CXCR4* and *APLP2* (B.sub1.CXCR4 cluster) and another characterised by elevated expression of *DBI*, *IFIT3*, and *TFRC* (B.sub2 cluster). The $CXCR4^+$ subset may reflect a maturation or migratory phenotype, consistent with prior observations of *CXCR4* involvement in B-cell trafficking and differentiation (Giorgiutti et al., 2024). The second subset, characterized by *IFIT3* expression, which can be induced by interferon signaling or viral infection, may represent a responsive or activated state (Fensterl & Sen, 2015; Wu et al., 2023).

Sub-clustering of T cells separated the major naïve and memory compartments (**Figure 3.9E-F**). Naïve T cells were characterised by high expression of *CCR7*, *TCF7*, and *SELL*, consistent with a resting phenotype. Within the memory compartment, $CD4^+$ memory T cells expressed *CD28*, whereas $CD8^+$ memory T cells showed elevated expression of *LAG3*, *NKG7*, and *CCL5*, indicative of cytotoxic effector-memory features. A distinct regulatory T-cell (Treg) cluster was identified by expression of *FOXP3, TIGIT,* and *IL2RA*, consistent with a suppressive phenotype. In addition, a small population of MAIT-like cells expressing *KLRB1, IL7R,* and *IL18R1* was detected, suggesting a mucosal-associated, cytokine-responsive subset.

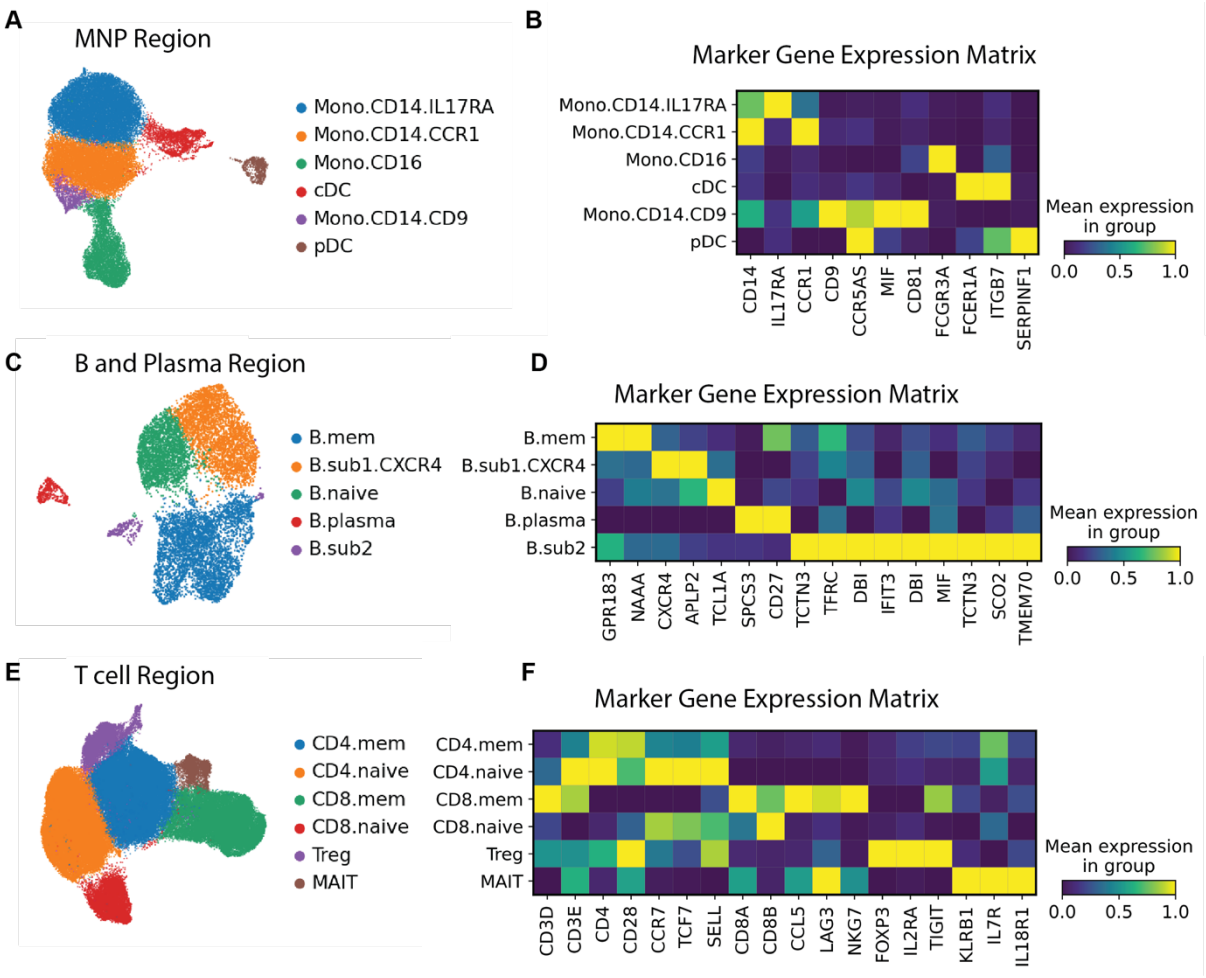


Figure 3.9 Sub-clustering analysis of MNP, B cell lineage, and T cells
(A, C, E) UMAP visualizations showing Leiden clusters identified within each major immune-cell lineage: MNPs (A), B-cell lineage (C), and T cells (E). (B, D, F) Matrix plots displaying the expression of selected marker genes used for sub-cluster annotation, corresponding to MNPs (B), B-cell lineage (D), and T cells (F). Expression values are based on scVI-normalised data.

Among the subpopulations identified within the MNP region, one monocyte subset appeared to be potentially disease associated. The distinct CD9^{+} monocyte subset, characterized by high expression of *MIF* and *CCR5AS*, was observed exclusively in 2 out of the 12 AxSpA patients and was absent from the three HLA-B*27-positive healthy controls and healthy individuals from the COMBAT dataset (**Figure 3.10**). This observation suggests it may represent a patient-specific monocyte phenotype associated with AxSpA.

MIF encodes macrophage migration inhibitory factor, a pro-inflammatory cytokine increasingly recognized as a key mediator in SpA pathogenesis. Elevated serum MIF concentrations have been reported in AS patients compared with healthy controls (Ranganathan et al., 2017), supporting a potential link between increased MIF expression and altered monocyte subset profiles in disease. In both AxSpA patients and curdlan-treated SKG mice, a murine model of SpA, neutrophils have been shown to secrete increased levels of MIF, and systemic MIF overexpression in mice has been associated with SpA-like inflammatory and skeletal manifestations (Nakamura et al., 2021). Mechanistically, MIF has been reported to amplify type 3 immune responses by enhancing IL-17 and IL-22 production and by promoting the conversion of Tregs toward a Th17-like inflammatory phenotype. This process may be further reinforced through interaction with hypoxia-inducible factor 1α (HIF1A), which cooperates with MIF to enhance IL-23 driven inflammation and new bone formation (Nakamura et al., 2024).

Together, these findings suggest that this *MIF* expressing $CD9^{+}$ monocyte subset may reflect an inflammatory cell state associated with the MIF/HIF1A regulatory axis previously linked to type 3 immunity in SpA.

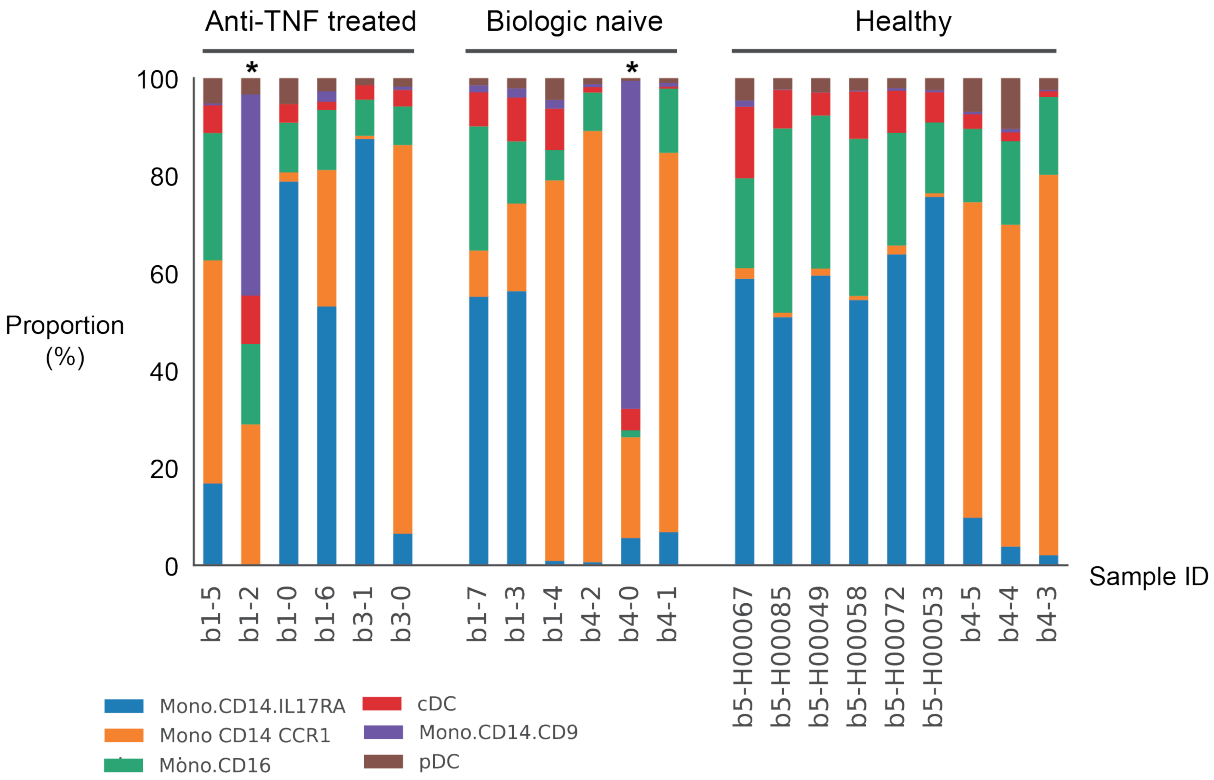


Figure 3.10 Distribution of monocyte subsets across individual samples
Bar plot showing the proportions of cells from each MNP sub-cluster in individual samples. Asterisks (*) indicate samples in which the CD9+ monocyte subset comprises more than 10% of the total myeloid cells.

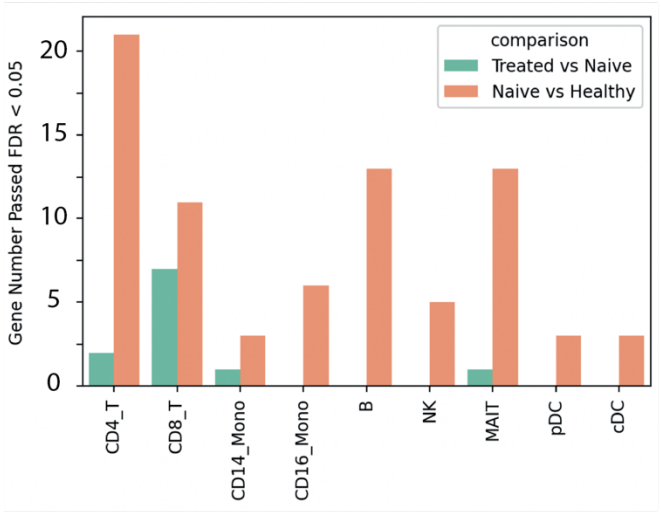
3.3.4 Transcriptional differences between AxSpA patients and healthy controls

I next sought to identify differentially expressed genes (DEGs) between anti-TNF-treated AxSpA patients, biologic-naïve AxSpA patients, and healthy controls within each cell type using pseudo-bulk analysis. DE analysis was performed using a generalized linear model implemented in edgeR, and multiple testing correction was applied using BH procedure with an FDR-adjusted p-value < 0.05 considered significant. The number of DEGs identified in each comparison and cell type is shown in **Figure 3.11A**.

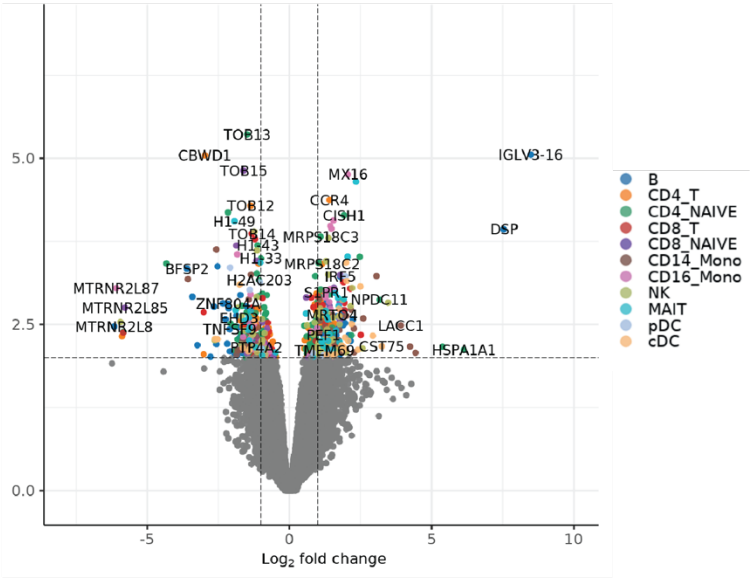
In total, 78 DEGs were identified when comparing biologic-naïve AxSpA patients with healthy controls, whereas only 11 DEGs were detected between anti-TNF-treated and biologic-naïve patients. The largest number of DEGs in the

comparison between biologic-naïve patients and healthy controls was observed within the $CD4^+$ T-cell compartment, suggesting large phenotypic shift in this population. B cells and MAIT cells were the next most transcriptionally altered populations. In contrast, the comparison between anti-TNF-treated and biologic-naïve patients revealed relatively fewer DEGs, indicating that treatment-induced transcriptional changes are modest compared with those distinguishing AxSpA patients from healthy controls. Within this comparison, $CD8^+$ T cells were the most affected immune-cell population, showing the highest number of DEGs among all cell types. The significant DEGs for all cell types are shown in **Figure 3.11B-C**, and the top three cell types with the most DEGs in each condition are shown in **Figure 3.12**.

A


comparison
Treated vs Naive
Naive vs Healthy
Gene Number Passed FDR < 0.05
0
5
10
15
20
CD4_T
CD8_T
CD14_Mono
CD16_Mono
B
NK
MAIT
pDC
cDC

B


Biologic naive AxSpA vs Healthy Control
TOB13
CBWD1
TOB15
MX16
IGLV3-16
CCR4
TOB12
CISH1
H1-49
DSP
TOB14
MRPS18C3
H1-43
H1-33
MRPS18C2
BFSP2
IRF5
H2AC203
MTRNR2L87
SIPR1
NPDC11
ZNF804A
MTRNR2L85
EHD3
MRTO4
MTRNR2L8
TNFSF9
LAGC1
PEF1
PTP4A2
TMEM69
CST75
HSPA1A1
B
CD4_T
CD4_NAIVE
CD8_T
CD8_NAIVE
CD14_Mono
CD16_Mono
NK
MAIT
pDC
cDC
5.0
2.5
0.0
-5
0
5
10
Log2 fold change

C

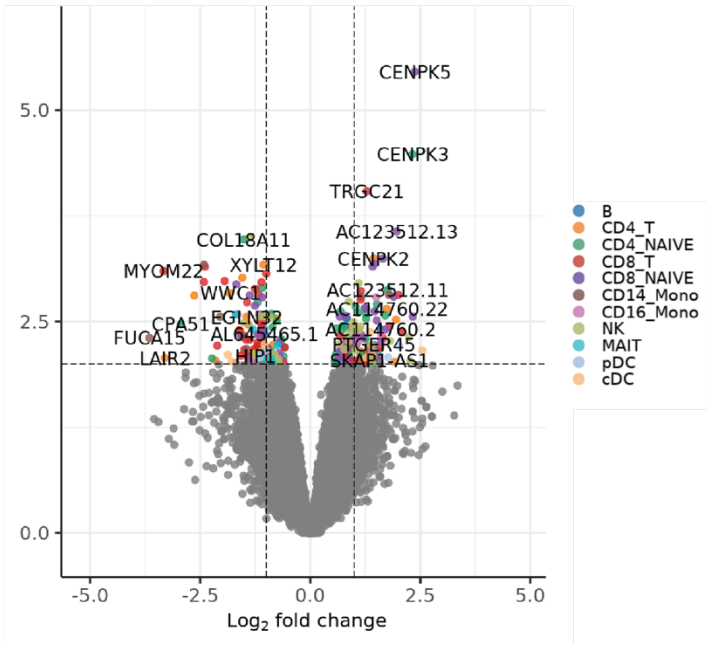

Anti-TNF treated AxSpA vs Biologic naive AxSpA
CENPK5
CENPK3
TRGC21
AC123512.13
COL18A11
CENPK2
XYLT12
MYOM22
AC123512.11
WWC1
AC114760.22
EGLN32
CPA5
AC114760.2
AL645465.1
FUCA15
PTGER45
HIP1
LAIR2
SKAP1-AS1
B
CD4_T
CD4_NAIVE
CD8_T
CD8_NAIVE
CD14_Mono
CD16_Mono
NK
MAIT
pDC
cDC
5.0
2.5
0.0
-5.0
-2.5
0.0
2.5
5.0
Log2 fold change

Figure 3.11 Transcriptional differences in PBMCs between AxSpA patients and healthy controls
(A) Bar plot summarizing the number of significantly DEGs that passed FDR threshold of 0.05, adjusted using the Benjamini–Hochberg (BH) procedure. (B) Volcano plots displaying the log fold change (logFC) and two-sided p-values from DE analyses comparing biologic-naïve AxSpA vs. healthy controls (up) and anti-TNF-treated AxSpA vs. biologic-naïve AxSpA (down). Genes are colored according to their associated cell types, with colored dots indicating genes with significant differences (BH-adjusted p< 0.01). No DEGs were shared between conditions; the DEG sets shown are condition-specific.

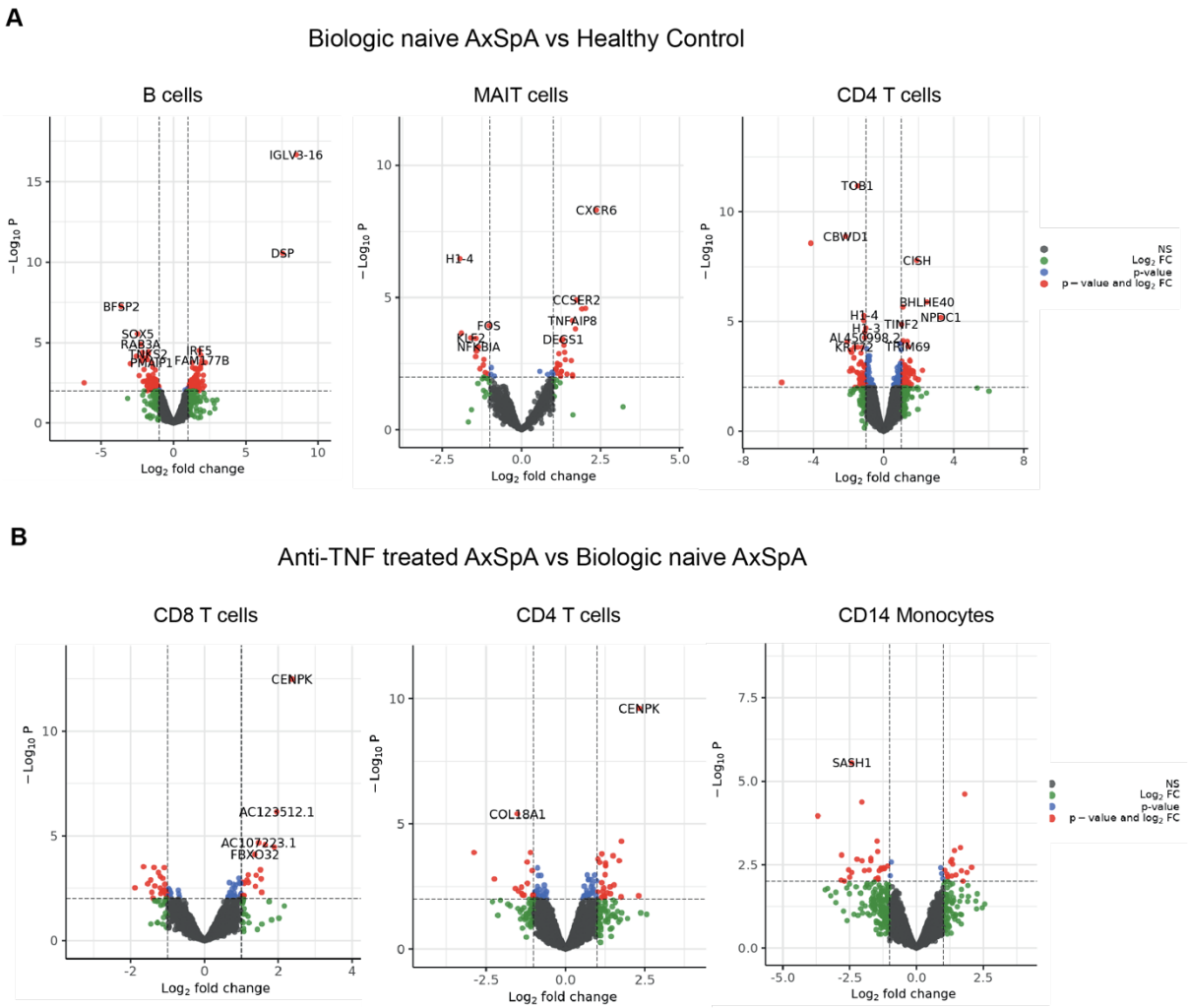


Figure 3.12 Transcriptional differences between AxSpA patients and healthy controls
(A, B) Volcano plots showing $\log_2$ fold change (colour threshold > 1 or < –1) and two-sided P-value (colour threshold < 0.001) for differential gene expression between biologic-naïve AxSpA patients and healthy controls (A), and between anti-TNF-treated and biologic-naïve AxSpA patients (B). Differential expression was tested using edgeR. Genes passing BH-adjusted p < 0.01 are labelled. The top three cell types ranked by the number of significant differentially expressed genes in each comparison are indicated.

I then performed Gene set enrichment analysis (GSEA) using fgsea package (Korotkevich et al., 2021), with gene sets obtained from Human Molecular Signatures Database (MSigDB), including Gene Ontology (GO), BioCarta gene sets and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Genes were ranked according to the edgeR differential expression results described above. GSEA revealed significant enrichment (BH-adjusted $p < 0.1$) of pathways related to cell adhesion molecules in $CD8^+$ naïve T cells from biologic-naïve AxSpA patients compared with healthy controls (**Figure 3.13A**). Following anti-TNF treatment, $CD8^+$ T cells exhibited enrichment of pathways involved in the negative regulation of cytokine production, while naïve $CD4^+$ T cells showed enrichment in pathways associated with the negative regulation of cell-cell adhesion, indicating the transcriptional shifts induced by TNF inhibitor therapy (**Figure 3.13B**).

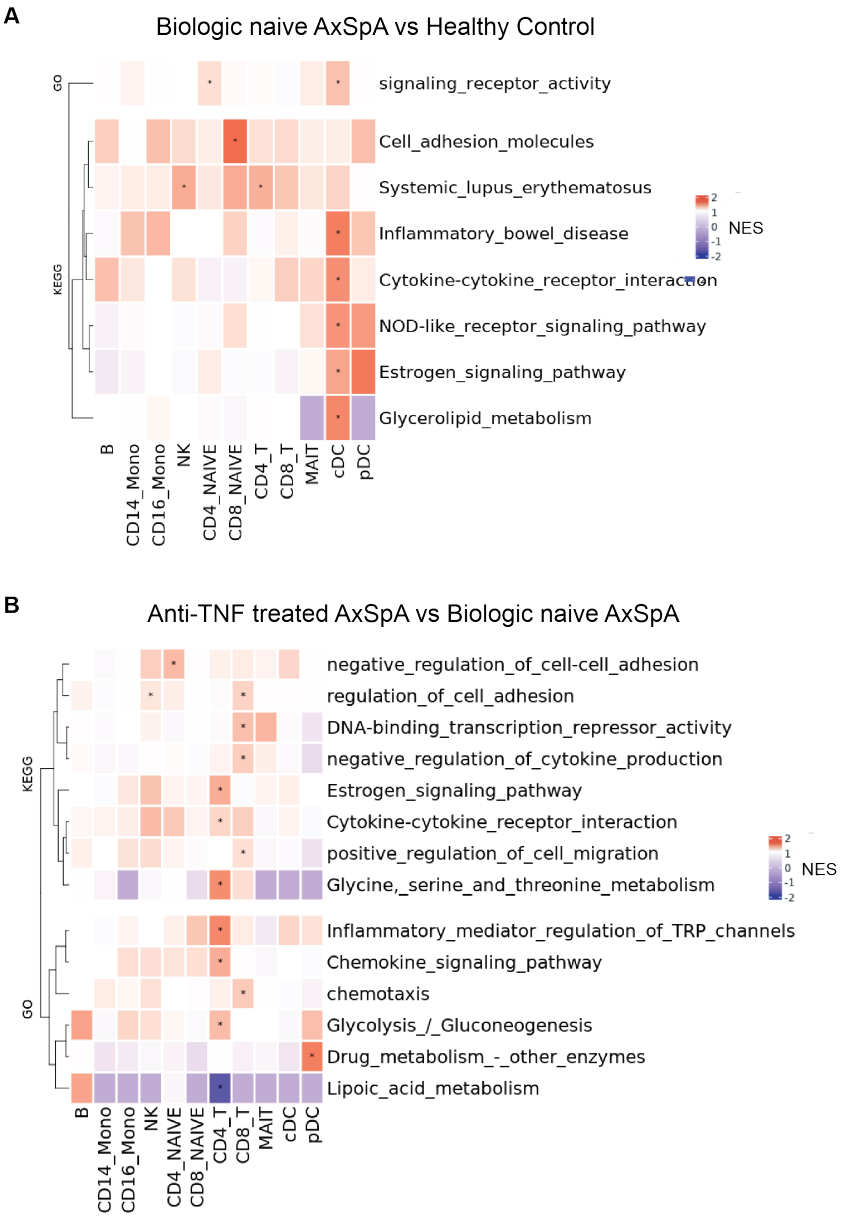


Figure 3.13 Pathway analysis using GSEA in AxSpA patients and healthy controls
Heatmaps summarising GSEA results for each cell type. Comparisons shown are (A) biologic-naïve AxSpA patients versus healthy controls and (B) anti-TNF-treated AxSpA patients versus biologic-naïve AxSpA patients. Pathways were selected based on an adjusted P-value < 0.1, with significant pathways indicated by asterisks (*).

In addition to fgsea, I also applied the single-cell Disease Relevance Score (scDRS) method to perform pathway-level analysis (Zhang et al., 2022). This approach assigns a relevance score to each cell based on the aggregate expression of genes within predefined pathways, thereby enabling the evaluation of pathway-specific enrichment relative to background cells. Given the proposed role of HLA-

B*27 misfolding in inducing endoplasmic reticulum (ER) stress and activating the unfolded protein response (UPR), I examined the involvement of key UPR signaling branches (*IRE1, PERK,* and *ATF6*), together with pathways related to antigen processing and presentation and cytokine signaling, including IL-17 and TNF production.

The scDRS pathway analysis result is shown in **Figure 3.14**. This analysis revealed elevated TNF response scores in $CD4^+$ T cells from biologic-naïve AxSpA patients compared with healthy controls. Moreover, $CD14^+$ monocytes from AxSpA patients displayed higher scDRS scores for the IRE1, ATF6, and PERK pathways and TNF production. Notably, these elevated UPR-related scores persisted in both biologic-naïve and anti-TNF-treated AxSpA datasets, suggesting a stable and potentially disease-associated transcriptional signature within the $CD14^+$ monocyte population.

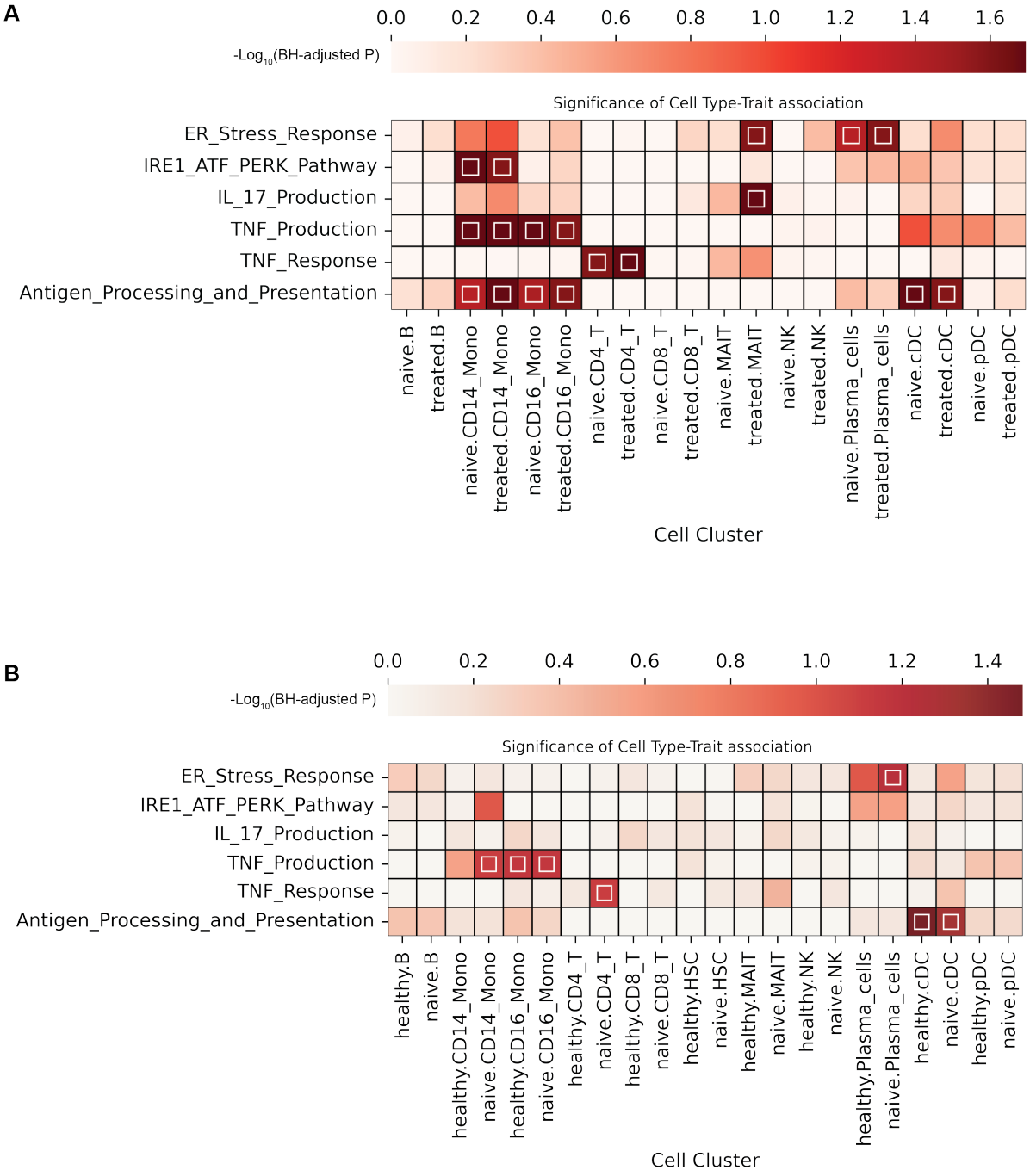


Figure 3.14 Pathway analysis using scDRS in AxSpA patients and healthy controls
Heatmaps showing cell-type-specific associations with specific pathways derived from scDRS analyses. Rows represent pathways, and columns represent cell types; heatmap color intensity indicates the -log10(BH-adjusted p-value) of association for each celltype-pathway pair. Squares mark pairs with significant heterogeneity of associations within cell types (FDR < 0.01). Comparisons shown are (A) biologic-naïve AxSpA patients versus healthy controls and (B) anti-TNF-treated AxSpA patients versus biologic-naïve AxSpA patients.

3.3.5 Gene expression informed evaluation of AxSpA genetic risk across immune cell types

I next examined how GWAS-identified AxSpA genetic risk variants are reflected in the gene expression profiles of specific immune cell populations.

From the latest GWAS investigating shared genetic susceptibility loci across ankylosing spondylitis and clinically related diseases (Ellinghaus et al., 2016), I obtained the functional annotation of AS-associated single nucleotide polymorphisms (SNPs) and examined their overlap with DEGs identified in our scRNA-seq analysis. Three coding variants had corresponding genes that were significantly differentially expressed when comparing biologic naïve AxSpA patients to healthy controls (BH-adjusted p-value < 0.05): *IRF5*, which was upregulated in B cells; *FOS*, which was downregulated in MAIT cells; and *FAM118A*, which was upregulated in $CD4^+$ T cells (**Figure 3.15A-B**). **Figure 3.15C** shows the expression of GWAS-mapped genes that passed the Open Targets Genetics locus-to-gene model, across immune cells from biologic-naïve AxSpA patients and healthy controls.

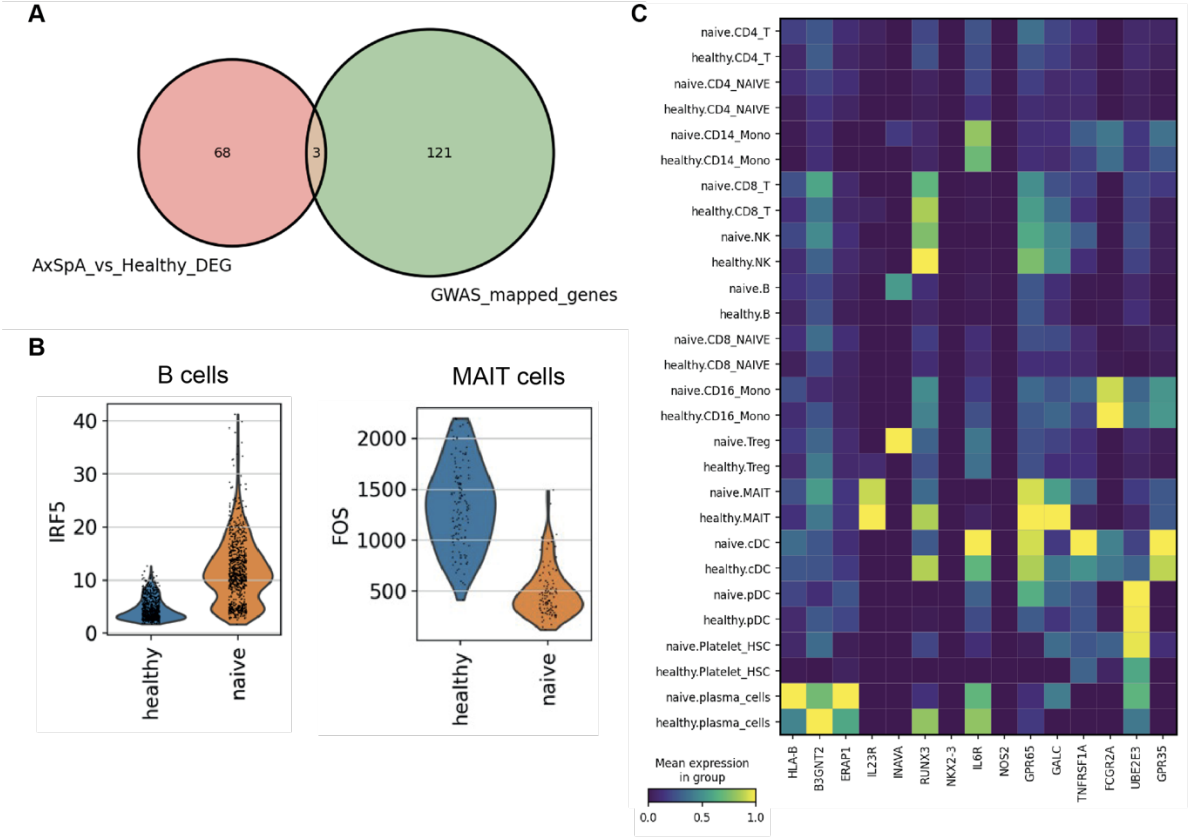


Figure 3.15 Integrated analysis of GWAS-identified AxSpA risk genes with single-cell gene expression data
(A) Venn diagram showing the overlap between DEGs identified in this study and GWAS-identified AS-associated genes. (B) Violin plots showing normalised expression of selected overlapping genes in B cells and MAIT cells. (C) Heatmap displaying the expression of prioritised GWAS-mapped genes (locus-to-gene score ⩾ 0.5 from Open Targets Genetics) and HLA-B across cell types in biologic-naïve AxSpA patients and healthy controls.

These observations raised the question of which immune cell types are most influenced by AxSpA-associated genetic risk variants, beyond individual gene-level effects. To address this, I used MAGMA to construct weighted gene lists based on the GWAS data for AS published in 2013 (International Genetics of Ankylosing Spondylitis et al., 2013), and estimated cell-type-specific associations using GWAS-derived z-scores from obtained with the scDRS algorithm. Two gene sets were evaluated: the full list of significant genes from the AS GWAS (AS_gwas_2013) and a subset excluding genes within the major histocompatibility complex (MHC) region

(AS_gwas_2013_non_MHC). Results from the non-MHC subset were considered more informative given the extensive linkage disequilibrium (LD) structure within the MHC region. In addition, I incorporated another gene set derived from 244 risk alleles identified through a cross-phenotype GWAS (AS_2016) to calculate disease correlation scores using scDRS (Ellinghaus et al., 2016). To minimise potential batch effects, the comparison between biologic-naïve AxSpA patients and healthy controls was restricted to a single experimental batch (P190726), providing a more robust evaluation of genetic risk enrichment.

At the single-cell level, scDRS scores derived from "AS_gwas_2013_non_MHC_region" were predominantly enriched in T, NK and B-cell regions on the UMAP (**Figure 3.16B**). Subsequent cell type-level analysis identified $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and plasma cells as the immune populations showing the strongest enrichment of AxSpA genetic risk signals (**Figure 3.16A**).

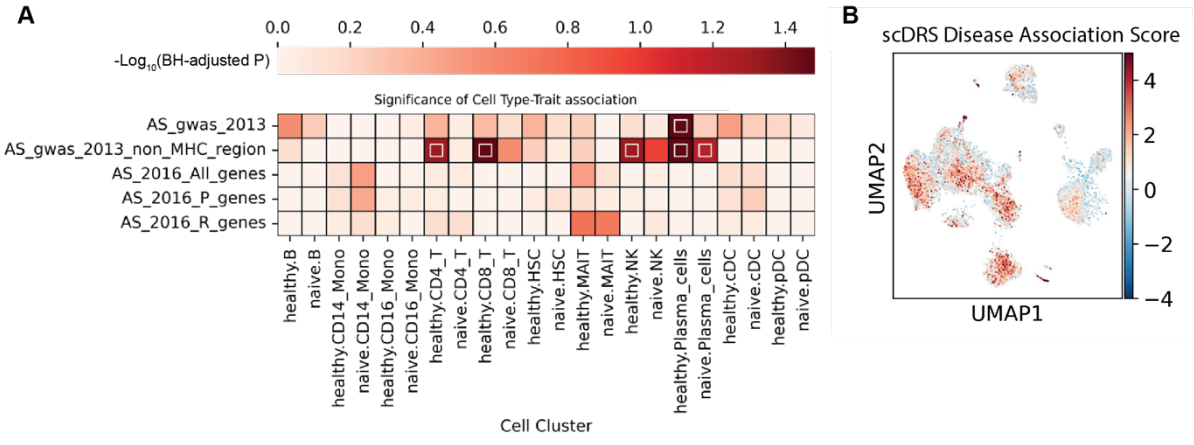


Figure 3.16 Genetic risk associations between immune cell types and AxSpA GWAS traits
(A) Heatmap showing significance of associations between immune cell types and GWAS-derived AxSpA disease risk scores calculated with scDRS, to compare biologic-naïve AxSpA patients and healthy controls. Colour intensity indicates -log10(BH-adjusted p-value); squares indicate significant heterogeneity within cell types (FDR < 0.01). "P" stands for

protective genes and “R” stands for risk genes. (B) UMAP showing cell-level AxSpA disease risk scores derived from the 2013 GWAS study for AS (AS_gwas_2013_non_MHC_region).

3.3.6 Validation of compositional change using bulk RNA-seq data

To independently validate the observed changes in cell composition, I analysed publicly available bulk RNA-seq datasets from two external studies: GSE141646, comprising PBMC samples from AxSpA patients under different treatment conditions (Wang et al., 2017), and GSE117769, containing white blood cell samples from biologic-treated AxSpA patients and healthy controls.

I first performed standard preprocessing to minimise potential batch effects arising from differences in data processing and study design. RNA-seq reads were adapter-trimmed using Trim Galore within the CGAT-core pipeline, which also generated per-sample quality control reports via FastQC. Trimmed reads were then aligned to the consistent reference genome using the STAR aligner. To further correct for batch effects between studies, the removeBatchEffect function from the limma package was applied. Following these preprocessing steps, an overview of the processed samples visualised by principal component analysis (PCA) is shown in **Figure 3.17**.

Using CIBERSORTx (Newman et al., 2019) I performed cellular deconvolution of the bulk RNA-seq datasets to estimate the predicted abundance of immune cell types. The analysis showed a significant increase in the predicted proportions of B cells and plasma cells in anti-TNF-treated AxSpA samples compared with biologic-naïve patients (**Figure 3.17B**). Additionally, AxSpA patients exhibited higher predicted plasma cell proportions compared with healthy controls, whereas predicted B-cell proportions did not differ significantly between these two groups (**Figure 3.17D**). These findings were consistent with the scCODA results, confirming that

anti-TNF therapy significantly increased the proportion of B-cell lineage cells compared with biologic-naïve AxSpA patients, and suggest that the higher B-cell lineage abundance in AxSpA patients versus healthy controls may be primarily driven by plasma cells.

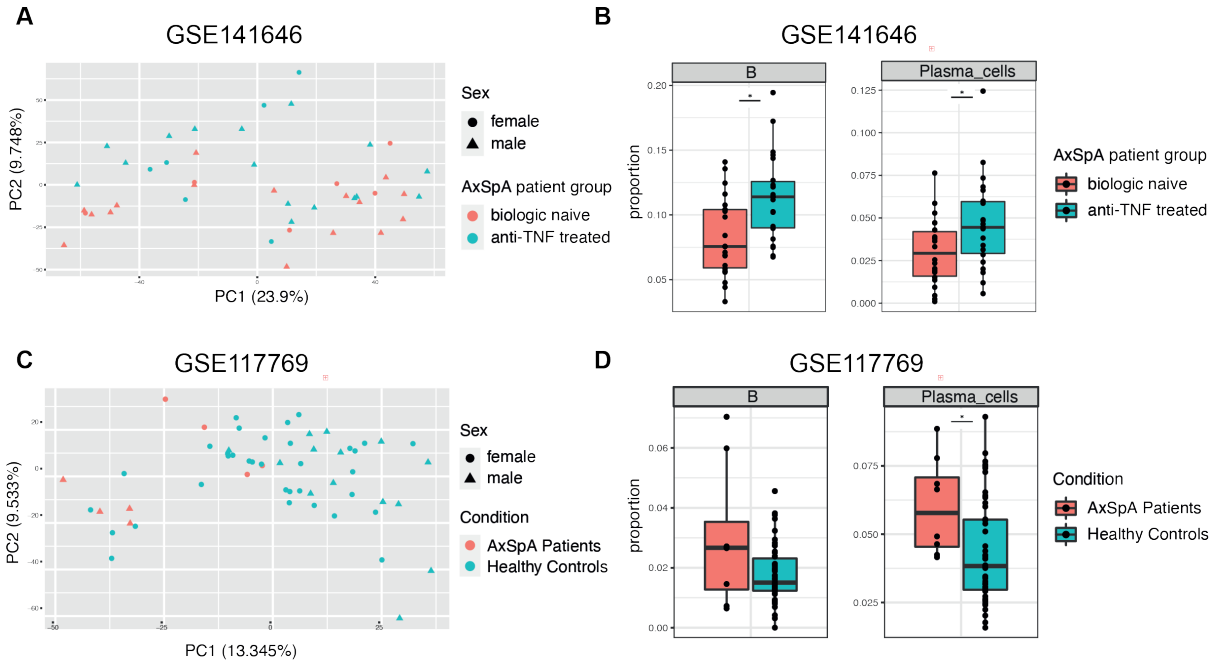


Figure 3.17 Validation of cell compositional changes using external bulk RNA-seq datasets
(A, C) PCA of bulk RNA-seq samples from AxSpA patients and healthy controls from datasets GSE141646 (A) and GSE117769 (C). (B, D) Box plots illustrating estimated percentages of immune cell populations following deconvolution of bulk RNA-seq data using CibersortX. Statistical significance was assessed by paired t-test for GSE141646 and unpaired t-test for GSE117769 ($p<0.05$).

3.4 Discussion

In this chapter, I constructed a PBMC atlas comprising biologic-naïve AxSpA patients, TNF-treated patients, and healthy control individuals using data generated in our laboratory and further integrated it with healthy control PBMC data from the large COMBAT project.

In the compositional analysis, scCODA revealed a consistent increase in the B-cell lineage following TNF inhibitor treatment, as well as higher proportions in biologic-naïve AxSpA patients compared with healthy controls. External bulk RNA-seq deconvolution supported this finding and further suggested that the lineage-level increase in AxSpA versus healthy controls was driven primarily by plasma cells, whereas TNF inhibitor therapy expanded both B cells and plasma cells. Although deconvolution-based estimates are reference-dependent and can be confounded by transcriptional similarity between B cells and plasma cells, the directional consistency with the single-cell observation across an independent cohort strengthens confidence in the finding. This pattern aligns with previous reports in AS, where multiple studies observed increased frequencies of circulating plasmablasts and plasma cells (Chen et al., 2016; Lin et al., 2009; Wilbrink, Spoorenberg, Arends, et al., 2021; Wilbrink, Spoorenberg, Verstappen, et al., 2021). The expansion of plasma cells may reflect enhanced B-cell activation and differentiation under chronic inflammatory conditions, contributing to local immune responses or antibody production, although their precise pathogenic role remains uncertain. Future work, such as B-cell receptor (BCR) repertoire sequencing, could help clarify whether the observed B-cell lineage expansion reflects antigen-driven clonal responses or a broader polyclonal increase.

Sub-clustering of the myeloid compartment identified a $CD9^{+}$ monocyte subset characterized by elevated *MIF* and *CCR5AS* expression, present in a subset of AxSpA patients but absent in healthy controls. MIF has been proposed to play a key role in AxSpA pathogenesis, with studies showing the MIF-producing neutrophils promote disease severity in a murine model of SpA (Nakamura et al., 2021), and elevated serum MIF levels have also described in AS patients (Ranganathan et al.,

2017). These findings suggest that increased *MIF* expression in this monocyte subset may be linked to disease-associated inflammatory pathways within the myeloid compartment. *CCR5AS,* a long non-coding RNA that regulates CCR5 expression (Kulkarni et al., 2019), was also highly expressed in this subset. Inflammatory monocytes with upregulated CCR5 have been reported in rheumatoid arthritis, where CCR5 signalling facilitates monocyte migration and activation within synovial tissue (Braun et al., 2002; Katschke et al., 2001). Together the $MIF^{+}CCR5AS^{+}$ $CD9^{+}$ monocytes identified here may represent a disease-associated inflammatory and migratory population that contributes to AxSpA pathogenesis through recruitment to inflamed tissues. However, as this cluster was detected in only two of twelve patients, the observation remains preliminary and requires validation in larger, independent cohorts.

The DE analysis indicated that anti-TNF treatment had the greatest transcriptomic impact on $CD8^{+}$ T cells, while GSEA revealed enrichment of pathways related to chemotaxis and the positive regulation of cell migration following therapy. These findings suggest that TNF inhibition may modulate the migratory and trafficking capacity of $CD8^{+}$ T cells, potentially reducing their recruitment to inflammatory sites. Such effects are consistent with the known role of TNF in promoting leukocyte adhesion and migration, and may contribute to the therapeutic mechanisms through which TNF inhibitors alleviate inflammation in AxSpA. Additionally, elevated trait scores for UPR signaling pathways were consistently observed in $CD14^{+}$ monocytes from both biologic-naïve and anti-TNF-treated AxSpA patients compared with healthy controls. Although evidence from human studies regarding UPR involvement in AxSpA remains mixed, these results align with reports showing increased expression of UPR-related genes, including BiP, CHOP, and

XBP1, in $CD14^+$ monocytes from AxSpA patients relative to healthy controls (Rezaiemanesh et al., 2022).

To further elucidate how genetic risk factors contribute to AxSpA pathogenesis, I employed scDRS analysis to determine which cell types showed the strongest expression enrichment for genes associated with AxSpA GWAS loci. The results highlighted CD4+ T cells, CD8+ T cells, NK cells and plasma cells as potential disease-relevant cell types. Interestingly, CD4+ and CD8+ T cells from healthy controls exhibited higher genetic disease scores compared to those from biologic-naïve AxSpA patients. The scDRS authors previously discussed similar observations, suggesting that some disease-associated genes might be highly expressed under healthy conditions, where their normal expression is essential for proper cellular function. Reduced expression of these critical genes in diseased individuals could reflect impaired functionality rather than absence of disease-related activity. My findings in peripheral blood differ from those in inflamed tissues, where different cell populations exhibit elevated genetic risk scores. This contrast highlights the importance of distinguishing between circulating and tissue-resident immune cells, which will be explored further in the following chapters.

In summary, this integrated single-cell PBMC atlas of AxSpA patients and healthy controls revealed distinct transcriptomic and compositional alterations in peripheral immune cells, with the most consistent change being an expansion of the B-cell lineage that is most pronounced under TNF inhibitor therapy and appears to be driven by plasma cells. Beyond this central finding, the analysis identified a potentially disease-associated $CD14^+CD9^+$ monocyte subset, and prioritized key genes such as IRF5 and FOS, which were both differentially expressed and genetically implicated in disease susceptibility. Moreover, genetic risk enrichment

analyses highlighted adaptive immune populations, including T, NK, and B cells, as having stronger disease-associated signatures. This initial PBMC atlas establishes a framework for the analytical pipeline used in subsequent multi-tissue integration and provides preliminary discoveries to be further validated in larger cohorts. Future analyses will explore whether these peripheral disease-associated signals are conserved or modulated within the local tissue contexts in AxSpA patients.

3.5 References

Braun, J., Brandt, J., Listing, J., Zink, A., Alten, R., Golder, W., Gromnica-Ihle, E., Kellner, H., Krause, A., Schneider, M., Sorensen, H., Zeidler, H., Thriene, W., & Sieper, J. (2002). Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. *Lancet*, *359*(9313), 1187–1193. https://doi.org/10.1016/s0140-6736(02)08215-6

Buttner, M., Ostner, J., Muller, C. L., Theis, F. J., & Schubert, B. (2021). scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun*, *12*(1), 6876. https://doi.org/10.1038/s41467-021-27150-6

Callhoff, J., Sieper, J., Weiss, A., Zink, A., & Listing, J. (2015). Efficacy of TNFalpha blockers in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis: a meta-analysis. *Ann Rheum Dis*, *74*(6), 1241–1248. https://doi.org/10.1136/annrheumdis-2014-205322

Champion, T. C., Partridge, L. J., Ong, S. M., Malleret, B., Wong, S. C., & Monk, P. N. (2018). Monocyte Subsets Have Distinct Patterns of Tetraspanin Expression and Different Capacities to Form Multinucleate Giant Cells. *Front Immunol*, *9*, 1247. https://doi.org/10.3389/fimmu.2018.01247

Chen, M., Zhang, L., Ren, Y., Zhang, K., Yang, Y., Fang, Y., Yan, X., Peng, D., Gao, C., & Li, S. (2016). Defective Function of CD24(+)CD38(+) Regulatory B Cells in Ankylosing Spondylitis. *DNA Cell Biol*, *35*(2), 88–95. https://doi.org/10.1089/dna.2015.3046

COMBAT Consortium. (2022). A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell*, *185*(5), 916–938 e958. https://doi.org/10.1016/j.cell.2022.01.012

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., Park, Y. R., Raychaudhuri, S., Pouget, J. G., Hubenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H. J., Franke, L., Pers, T. H., Weersma, R. K., Collij, V.,…Franke, A. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*, *48*(5), 510–518. https://doi.org/10.1038/ng.3528

Fensterl, V., & Sen, G. C. (2015). Interferon-induced Ifit proteins: their role in viral pathogenesis. *J Virol*, *89*(5), 2462–2468. https://doi.org/10.1128/JVI.02744-14

Giorgiutti, S., Rottura, J., Korganow, A. S., & Gies, V. (2024). CXCR4: from B-cell development to B cell-mediated diseases. *Life Sci Alliance*, *7*(6). https://doi.org/10.26508/lsa.202302465

Hie, B. L., Kim, S., Rando, T. A., Bryson, B., & Berger, B. (2024). Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nat Protoc*, *19*(8), 2283–2297. https://doi.org/10.1038/s41596-024-00991-3

International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., Joo, K. B., Shim, S. C., Weisman, M., Ward, M., Zhou, X., Garchon, H. J., Chiocchia, G., Nossent, J.,…Brown, M. A. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*, *45*(7), 730–738. https://doi.org/10.1038/ng.2667

Katschke, K. J., Jr., Rottman, J. B., Ruth, J. H., Qin, S., Wu, L., LaRosa, G., Ponath, P., Park, C. C., Pope, R. M., & Koch, A. E. (2001). Differential expression of chemokine receptors on peripheral blood, synovial fluid, and synovial tissue monocytes/macrophages in rheumatoid arthritis. *Arthritis Rheum*, *44*(5), 1022–1032. https://doi.org/10.1002/1529-0131(200105)44:5<1022::AID-ANR181>3.0.CO;2-N

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012. https://doi.org/10.1101/060012

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. R., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*, *16*(12), 1289–1296. https://doi.org/10.1038/s41592-019-0619-0

Kulkarni, S., Lied, A., Kulkarni, V., Rucevic, M., Martin, M. P., Walker-Sperling, V., Anderson, S. K., Ewy, R., Singh, S., Nguyen, H., McLaren, P. J., Viard, M., Naranbhai, V., Zou, C., Lin, Z., Gatanaga, H., Oka, S., Takiguchi, M., Thio, C. L.,…Carrington, M. (2019). CCR5AS lncRNA variation differentially regulates CCR5, influencing HIV disease outcome. *Nat Immunol*, *20*(7), 824–834. https://doi.org/10.1038/s41590-019-0406-1

Lin, Q., Gu, J. R., Li, T. W., Zhang, F. C., Lin, Z. M., Liao, Z. T., Wei, Q. J., Cao, S. Y., & Li, L. (2009). Value of the peripheral blood B-cells subsets in patients with ankylosing spondylitis. *Chin Med J (Engl)*, *122*(15), 1784–1789.

Liu, D., Liu, B., Lin, C., & Gu, J. (2021). Imbalance of Peripheral Lymphocyte Subsets in Patients With Ankylosing Spondylitis: A Meta-Analysis. *Front Immunol*, *12*, 696973. https://doi.org/10.3389/fimmu.2021.696973

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat Methods*, *15*(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2

Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*, *15*(6), e8746. https://doi.org/10.15252/msb.20188746

Nakamura, A., Jo, S., Nakamura, S., Aparnathi, M. K., Boroojeni, S. F., Korshko, M., Park, Y. S., Gupta, H., Vijayan, S., Rockel, J. S., Kapoor, M., Jurisica, I., Kim, T. H., & Haroon, N. (2024). HIF-1alpha and MIF enhance neutrophil-driven type 3 immunity and chondrogenesis in a murine spondyloarthritis model. *Cell Mol Immunol*, *21*(7), 770–786. https://doi.org/10.1038/s41423-024-01183-5

Nakamura, A., Zeng, F., Nakamura, S., Reid, K. T., Gracey, E., Lim, M., Leng, L., Jo, S., Park, Y. S., Kusuda, M., Machhar, R., Boroojeni, S. F., Wu, B., Rossomacha, E., Kim, T. H., Ciccia, F., Rockel, J. S., Kapoor, M., Inman, R. D.,…Haroon, N. (2021). Macrophage migration inhibitory factor drives

pathology in a mouse model of spondyloarthritis and is associated with human disease. *Sci Transl Med*, *13*(616), eabg1210. https://doi.org/10.1126/scitranslmed.abg1210
Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M., & Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*, *37*(7), 773–782. https://doi.org/10.1038/s41587-019-0114-2
Polanski, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A., & Park, J. E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics*, *36*(3), 964–965. https://doi.org/10.1093/bioinformatics/btz625
Ranganathan, V., Ciccia, F., Zeng, F., Sari, I., Guggino, G., Muralitharan, J., Gracey, E., & Haroon, N. (2017). Macrophage Migration Inhibitory Factor Induces Inflammation and Predicts Spinal Progression in Ankylosing Spondylitis. *Arthritis Rheumatol*, *69*(9), 1796–1806. https://doi.org/10.1002/art.40175
Reshef, Y. A., Rumker, L., Kang, J. B., Nathan, A., Korsunsky, I., Asgari, S., Murray, M. B., Moody, D. B., & Raychaudhuri, S. (2022). Co-varying neighborhood analysis identifies cell populations associated with phenotypes of interest from single-cell transcriptomics. *Nat Biotechnol*, *40*(3), 355–363. https://doi.org/10.1038/s41587-021-01066-4
Reyes, R., Cardenes, B., Machado-Pineda, Y., & Cabanas, C. (2018). Tetraspanin CD9: A Key Regulator of Cell Adhesion in the Immune System. *Front Immunol*, *9*, 863. https://doi.org/10.3389/fimmu.2018.00863
Rezaiemanesh, A., Mahmoudi, M., Amirzargar, A. A., Vojdanian, M., Babaie, F., Mahdavi, J., Rajabinejad, M., Jamshidi, A. R., & Nicknam, M. H. (2022). Upregulation of Unfolded Protein Response and ER Stress-Related IL-23 Production in M1 Macrophages from Ankylosing Spondylitis Patients. *Inflammation*, *45*(2), 665–676. https://doi.org/10.1007/s10753-021-01575-z
Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616
Sobral, D., Fernandes, A. F., Bernardes, M., Pinto, P., Santos, H., Lagoas-Gomes, J., Tavares-Costa, J., Silva, J. A. P., Dias, J. M., Bernardo, A., Gaillard, J. C., Armengaud, J., Benes, V., Domingues, L., Maia, S., Branco, J. C., Coelho, A. V., & Pimentel-Santos, F. M. (2024). Molecular Profiling of Axial Spondyloarthritis Patients Reveals an Association between Innate and Adaptive Cell Populations and Therapeutic Response to Tumor Necrosis Factor Inhibitors. *Biomolecules*, *14*(3). https://doi.org/10.3390/biom14030382
Song, W., Lee, E. E., Park, S., Choi, B., Kim, M. G., Ban, S. Y., Choi, S. R., Kim, J. Y., Kim, S. U., Kim, J. I., Shin, E. C., Jung, I., Lee, J. S., Lee, E. Y., & Single-Cell Atlas of Immune Diseases, C. (2025). Type 1 interferon signature and allograft inflammatory factor-1 contribute to refractoriness to TNF inhibition in ankylosing spondylitis. *Nat Commun*, *16*(1), 5531. https://doi.org/10.1038/s41467-025-60445-6
Tippett, E., Cameron, P. U., Marsh, M., & Crowe, S. M. (2013). Characterization of tetraspanins CD9, CD53, CD63, and CD81 in monocytes and macrophages in HIV-1 infection. *J Leukoc Biol*, *93*(6), 913–920. https://doi.org/10.1189/jlb.0812391
van der Heijde, D., Kivitz, A., Schiff, M. H., Sieper, J., Dijkmans, B. A., Braun, J., Dougados, M., Reveille, J. D., Wong, R. L., Kupper, H., Davis, J. C., Jr., &

Group, A. S. (2006). Efficacy and safety of adalimumab in patients with ankylosing spondylitis: results of a multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum*, *54*(7), 2136–2146. https://doi.org/10.1002/art.21913

Wang, X. B., Ellis, J. J., Pennisi, D. J., Song, X., Batra, J., Hollis, K., Bradbury, L. A., Li, Z., Kenna, T. J., & Brown, M. A. (2017). Transcriptome analysis of ankylosing spondylitis patients before and after TNF-alpha inhibitor therapy reveals the pathways affected. *Genes Immun*, *18*(3), 184–190. https://doi.org/10.1038/gene.2017.19

Wilbrink, R., Spoorenberg, A., Arends, S., van der Geest, K. S. M., Brouwer, E., Bootsma, H., Kroese, F. G. M., & Verstappen, G. M. (2021). CD27(-)CD38(low)CD21(low) B-Cells Are Increased in Axial Spondyloarthritis. *Front Immunol*, *12*, 686273. https://doi.org/10.3389/fimmu.2021.686273

Wilbrink, R., Spoorenberg, A., Verstappen, G., & Kroese, F. G. M. (2021). B Cell Involvement in the Pathogenesis of Ankylosing Spondylitis. *Int J Mol Sci*, *22*(24). https://doi.org/10.3390/ijms222413325

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst*, *8*(4), 281–291 e289. https://doi.org/10.1016/j.cels.2018.11.005

Wu, Y. Y., Xing, J., Li, X. F., Yang, Y. L., Shao, H., & Li, J. (2023). Roles of interferon induced protein with tetratricopeptide repeats (IFIT) family in autoimmune disease. *Autoimmun Rev*, *22*(11), 103453. https://doi.org/10.1016/j.autrev.2023.103453

Zhang, M. J., Hou, K., Dey, K. K., Sakaue, S., Jagadeesh, K. A., Weinand, K., Taychameekiatchai, A., Rao, P., Pisco, A. O., Zou, J., Wang, B., Gandal, M., Raychaudhuri, S., Pasaniuc, B., & Price, A. L. (2022). Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet*, *54*(10), 1572–1580. https://doi.org/10.1038/s41588-022-01167-z

Zhao, Z. B., Bian, Z. H., Lin, Z. M., Wu, S. F., Long, J., Cui, Y., Li, Y., Li, W., Sun, E. W., Lian, Z. X., & He, Y. (2023). Single-Cell Analysis of Patients with Axial Spondyloarthritis After Anti-TNFalpha Treatment: Experimental Data and Review of the Literature. *Clin Rev Allergy Immunol*, *65*(2), 136–147. https://doi.org/10.1007/s12016-023-08959-z

Chapter 4: A Cross-tissue, Multi-Modal Single-Cell Atlas of HLA-B*27-Positive Axial Spondyloarthritis

4.1 Introduction

AxSpA is characterized by inflammation affecting multiple tissues, including axial and peripheral musculoskeletal structures, and also extra-articular organs such as the gut and eyes (Navarro-Compan et al., 2025; Sieper & Poddubnyy, 2017). In Chapter 1, I outlined the clinical features, summarized current knowledge of the underlying pathogenesis investigated in various studies, and discussed potential mechanisms implicated in AxSpA. Chapter 3 presented a detailed single-cell atlas of PBMCs from AxSpA patients before and after TNF inhibitor treatment, but this analysis was primarily focused on circulating immune populations. As a result, the insights obtained thus far remain largely confined to the peripheral blood compartment, and the broader immune landscape across affected tissues has yet to be characterized.

Inflammation in AxSpA is thought to involve complex interactions between local tissues, such as the gut, enthesis, and synovium, and circulating immune components. Studies focusing on multiple affected sites have provided important insights into how these tissues may influence one another; for example, work in reactive arthritis, a condition within the SpA family, has highlighted the potential cause-and-effect relationships in which enthesitis and synovitis develop following distant genitourinary or gastrointestinal infection with *Campylobacter, Salmonella, Yersinia* or *Shigella* (Ajene et al., 2013). However, in AxSpA, our understanding of these inter-tissue relationships remains limited, in part because no integrated framework has yet examined multiple AxSpA-relevant tissues simultaneously.

The value of such cross-tissue analyses has been demonstrated in other immune-mediated inflammatory disease. For instance, a single-cell study constructed a stromal atlas across synovium, intestine, lung, and salivary gland and identified shared pathogenic fibroblast phenotypes across four chronic inflammatory diseases (Korsunsky et al., 2022). This work highlighted both common and tissue-specific activation states that would not have been apparent from single-tissue studies alone. Without a comparable integrated framework in AxSpA, it remains difficult to determine how systemic and local immune environments interact and how they collectively contribute to disease mechanisms.

To address this gap, I extended the initial atlas and analytical approaches developed in Chapter 3 by broadening the scope to construct a cross-tissue, multi-modal atlas across AxSpA and relevant control conditions. In this chapter, I combined single-cell transcriptomic (scRNA-seq), proteomic (CITE-seq), and T-cell receptor sequencing (scTCR-seq) data generated from peripheral blood, multiple musculoskeletal tissues including sacroiliac joints, spinal entheses (comprising peri-entheseal bone and entheseal soft tissue), and synovial fluid, as well as extra-articular sites such as the gut and eye. These datasets are integrated with matched samples from healthy controls and comparator groups with other relevant inflammatory conditions, with rheumatoid arthritis (RA) the primary comparator, to generate a unified framework for annotation and cross-tissue comparison, forming the foundation for downstream analyses in the next chapter.

This integrated multi-modal, multi-tissue atlas developed in this chapter aims to comprehensively identify and annotate immune populations and their distinct cellular status across different conditions. By establishing well-resolved cellular clusters, the atlas provides a robust framework for subsequent analyses of

transcriptomic shifts, pathway-level differences and potential genetic contribution to disease. Together, this detailed characterization of cell types will provide a foundation for deeper exploration into the mechanisms underlying AxSpA pathogenesis, which will be described in the next chapter.

4.2 Aim

This chapter aims to address several key questions. First, it examines how scRNA-seq, CITE-seq and scTCR-seq data can be jointly analyzed to leverage the complementary information provided by each modality, enabling the construction of a robust multi-modal atlas from various tissues and supporting cross-tissue comparisons. Second, it investigates strategies for integrating large-scale single-cell gene expression data and performing clustering to identify distinct immune cell populations across diverse tissue and disease contexts. Finally, it focuses on the accurate annotation of the immune populations, characterizing each cluster by its marker genes and examining its distribution across tissue and disease conditions.

By constructing such an integrated multi-modal single-cell atlas, this chapter establishes a foundational resource for comprehensive characterization of the immune landscape across AxSpA-affected tissues. The identification and annotation of cellular subsets within the atlas provide a robust framework for the subsequent analyses in the next chapter, which will investigate transcriptomic, pathway-level and genetic contributions to AxSpA, thereby advancing our understanding of disease pathogenesis.

4.3 Results

4.3.1 Experimental design and overview of samples

The samples analyzed in this chapter were collected across multiple projects and patient cohorts, as summarized in **Table 4.1**, which provides the number of samples per project and per multiplexing pool. In addition to the previously described projects (P170676, P180407, P190230 and P190726) used in Chapter 3 for the PBMC atlas, this chapter includes newly generated samples from additional AxSpA and comparator cohorts.

Sample collection was carried out by collaborators as part of the broader research project. Processing of the majority of samples for single-cell analysis, including all newly generated samples, was carried out in our laboratory by research associate Moustafa Attar following the standard protocols for the 10x Genomics Chromium single-cell platform. Detailed library-level processing metadata for each project are described in the **Methods** section.

Project Name	Modality	Demultiplexing Method	Tissue	Cohort and sample number
P170676	GEX	Genetic	PB	n=8 AxSpA
P180407	GEX	Genetic	PB	n=2 AxSpA
	GEX	Genetic	PB	n=2 AxSpA
P190230	GEX, ADT	Hashtag	PB	n=2 AxSpA
P190726	GEX, ADT	Hashtag	PB	n=3 AxSpA, n=3 HC
Ox_VA_PB_SF	GEX, ADT, VDJ-T	Genetic	PB	n=4 AxSpA, n=2 RA, n=2 HC
	GEX, ADT, VDJ-T	Genetic	PB	n=3 AxSpA, n=2 RA, n=2 HC
	GEX, ADT, VDJ-T	Genetic	PB	n=4 AxSpA, n=2 RA, n=2 HC
	GEX, ADT, VDJ-T	Genetic	SF	n=4 AxSpA, n=5 RA
	GEX, ADT, VDJ-T	Single donor per library	SF	n=3 AxSpA
Berlin_PB_SF	GEX, ADT, VDJ-T	Genetic	PB	n=3 AxSpA, n=1 pSpA, n=2 PsA
	GEX, ADT, VDJ-T	Genetic	PB	n=3 AxSpA, n=2 pSpA, n=1 PsA, n=1 Ma

	GEX, ADT, VDJ-T	Genetic	SF	n=3 AxSpA, n=2 pSpA, n=1 PsA, n=1 Ma
	GEX, ADT, VDJ-T	Genetic	SF	n=4 AxSpA, n=1 pSpA, n=2 PsA
Gut_Naples	GEX, ADT, VDJ-T	Hashtag	ILM, PB	n=2 HC (ILM), n=2 HC (PB)
	GEX, ADT, VDJ-T	Hashtag	ILM	n=3 AxSpA, n=3 HC
	GEX, ADT, VDJ-T	Hashtag	PB	n=3 AxSpA, n=3 HC
Gut_TGU	GEX, ADT, VDJ-T	Single donor per library	ILM	n=3 AxSpA
	GEX, ADT, VDJ-T	Single donor per library	PB	n=1 AxSpA
SIJ	GEX, ADT, VDJ-T, VDJ-B	Single donor per library	SIJ, PB	n=5 AxSpA (SIJ), n=4 AxSpA (PB)
Spine_Enthesis	GEX, ADT, VDJ-T	Single donor per library	PEB	n=5 AxSpA, n=4 HC
	GEX, ADT, VDJ-T	Single donor per library	EST	n=2 AxSpA, n=4 HC
Aqh_Eye	GEX, VDJ-T, VDJ-B	Genetic	Eye	n=2 Aa, n=2 Ni
	GEX, VDJ-T, VDJ-B	Single donor per library	Eye	n=4 Aa
CD8_SFMC_Berlin	GEX, VDJ-T	Single donor per library	SF	n=10 AxSpA, n=3 RA, n=4 pSpA

Table 4.1 Overview of projects and libraries included in Chapter 4
Overview of all datasets used in Chapter 4, listing the project name, modality, demultiplexing method, tissue of origin, and cohort with sample numbers. Tissue abbreviations: PB = peripheral blood; SF = synovial fluid; SIJ = sacroiliac joint; ILM = ileum; PEB = peri-entheseal bone; EST = entheseal soft tissue. Disease abbreviations: AxSpA = axial spondyloarthritis; Aa = acute anterior uveitis; RA = rheumatoid arthritis; pSpA = peripheral spondyloarthritis; PsA = psoriatic arthritis; Ma = monoarthritis; HC = healthy control; Ni = no information. Modality abbreviations: GEX = gene expression library; ADT = antibody-derived tag library (CITE-seq); VDJ-T = single-cell T-cell receptor library; VDJ-B = single-cell B-cell receptor library.

Libraries were excluded from downstream analysis if they exhibited extremely low numbers of recovered cells or detected genes, or if they contained abnormal barcode structures. An overview of the sample numbers retained after this initial quality-control step across anatomical sites, along with the analytical workflow, is shown in **Figure 4.1**.

Across peripheral blood and synovial fluid, samples were obtained from individuals with AxSpA, RA and other inflammatory conditions including pSpA, PsA and monoarthritis. Combined with previously published scRNA-seq of sorted memory $CD8^+$ T cells from SFMC (Deschler et al., 2022), a total of n = 68 PBMC and n = 26 SFMC samples were included. For musculoskeletal tissues, the dataset comprised n = 4 sacroiliac joint samples (all from AxSpA), n = 9 spinal peri-entheseal bone samples, and n = 6 spinal entheseal soft tissue samples from AxSpA patients and healthy controls. Additional extra-articular samples included n = 11 ileum samples (AxSpA and healthy controls) and n = 5 aqueous humour samples from patients with acute anterior uveitis (AAU). Newly collected samples underwent CITE-seq and VDJ sequencing for T and B cells, although B-cell VDJ data were not analyzed further in this chapter. My work presented here began with processing the raw sequencing output and constructing an integrated multi-modal atlas across all tissues.

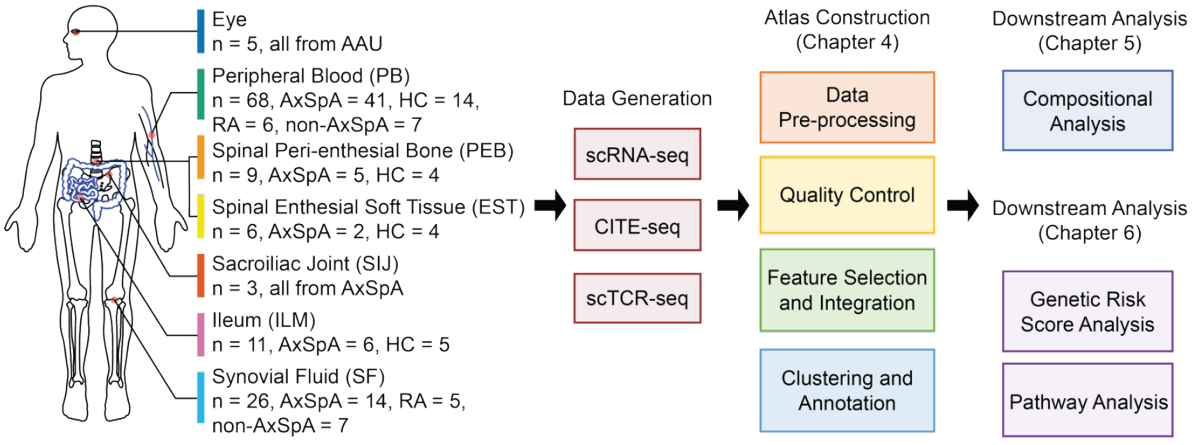


Figure 4.1 Overview of sample collection and analytical workflow for the atlas construction
The left panel shows illustration of anatomical sites and number of samples collected from each tissue site included in this chapter. Tissue types represented include peripheral blood (PB), synovial fluid (SF), sacroiliac joints (SIJ), spinal peri-entheseal bone (PEB), entheseal soft tissue (EST), ileum (ILM) and aqueous humour (Eye). Sample numbers represent the total from AxSpA patients, healthy controls (HC), RA patients, and non-AxSpA inflammatory

conditions include pSpA, PsA, and monoarthritis. The right panels summarize the overall analytical workflow for the generated datasets and highlights key steps in atlas construction and downstream analysis presented in this thesis.

4.3.2 Multi-modal atlas generation

4.3.2.1 Preprocessing and ambient RNA removal

To ensure accurate and up-to-date genomic annotations, I generated a reference genome using the latest Ensembl release 110 (Harrison et al., 2024) and aligned the raw sequencing data to this updated reference. Mapping was performed using the Cell Ranger pipeline, which assigned reads to genes using an Ensembl v110-based reference and included intronic reads in the quantification. This approach provided more precise gene assignment and reliable quantification at the single-cell level for downstream analyses.

Ambient RNA is a recognized source of background noise in droplet-based single-cell data, and correcting it enhances the reliability of downstream analyses (Janssen et al., 2023). Following the alignment, I further reduced ambient RNA contamination using CellBender (Fleming et al., 2023) to improve overall data quality. **Figure 4.2** demonstrates the cell quality matrices before and after the ambient RNA removal step.

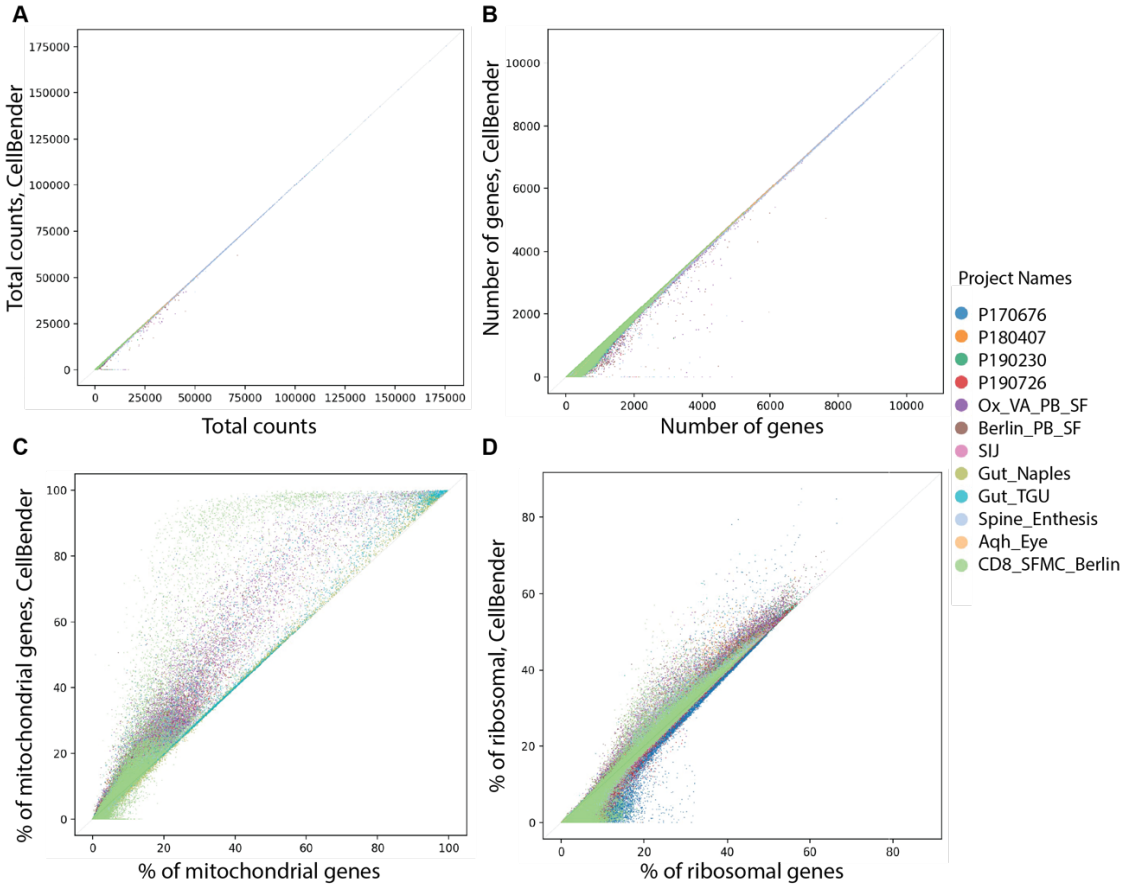


Figure 4.2 Cell quality metrices before and after ambient RNA removal with CellBender
Scatter plots comparing cell quality metrics before and after CellBender processing. Each point represents an individual cell, coloured by project. The x-axis shows values derived from raw Cell Ranger output, and the y-axis shows values following ambient RNA correction with CellBender. Quality control metrics displayed include: (A) total counts, (B) number of genes detected, (C) percentage of mitochondrial genes, and (D) percentage of ribosomal genes.

Following CellBender processing, some cells showed a reduction in the number of detected genes, reflecting the removal of background transcripts that had inflated gene counts in the raw data. For cells with high mitochondrial gene content, the percentage of mitochondrial reads often increased after ambient RNA removal (**Figure 4.2C**). This pattern indicates that cells with high mitochondrial reads were largely low-quality or damaged, with a substantial proportion of their transcriptome attributable to ambient RNA rather than true cellular expression. Once this

background signal was removed, the relative contribution of mitochondrial transcripts became more apparent. This behaviour is consistent with observations from droplet-based scRNA-seq studies, where elevated mitochondrial gene proportions are a well-established indicator of damaged, dying, or low-quality cells (Ilicic et al., 2016).

4.3.2.2 Cell demultiplexing

In several projects, samples from different individuals were pooled together and processed through the same library construction workflow to minimize batch effects. Single cells were subsequently assigned to the sample identity using appropriate demultiplexing methods. As summarized in **Table 4.1**, different multiplexing strategies were applied across projects, including genetic demultiplexing and hashtag-based approaches.

For samples with available bulk RNA-seq data, I performed genetic demultiplexing by first calling variants using the GATK pipeline (Van der Auwera et al., 2013) to derive sample-specific genotype information, and then assigning single cells to their corresponding individuals using the souporcell method (Heaton et al., 2020). The proportion of cells classified by genetic demultiplexing in the P170676, P180407, Ox_VA_PB_SF and Berlin_PB_SF projects is shown in **Figure 4.3**. Souporcell assigned each cell as a singlet, doublet, or unassigned, and only singlet cells were retained for downstream analysis.

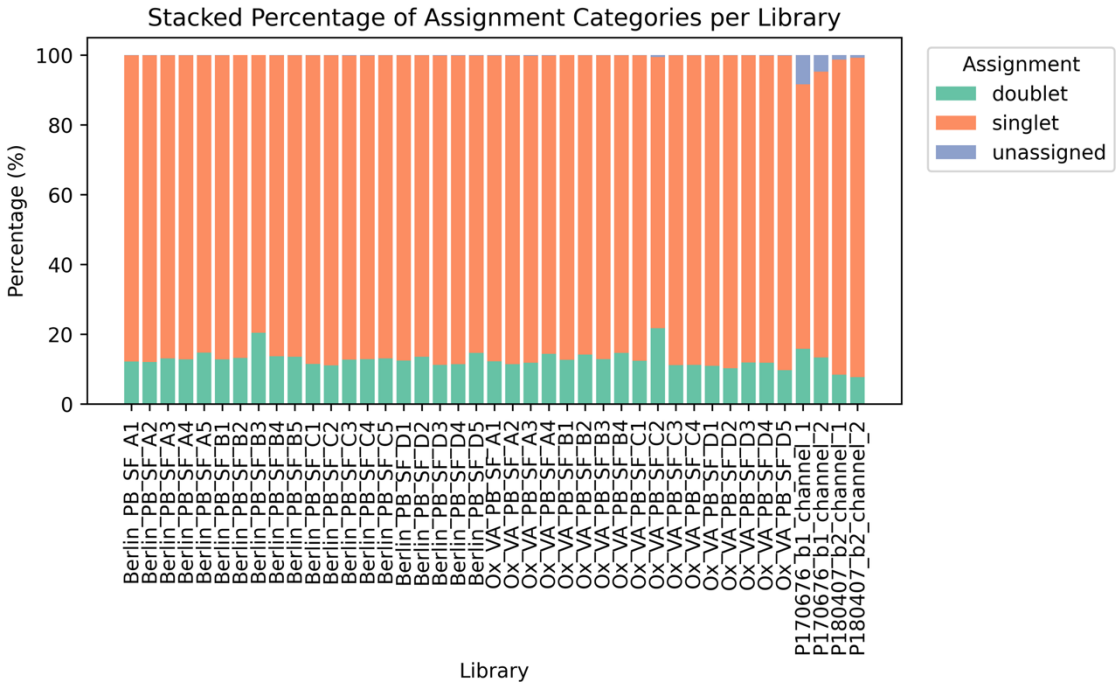


Figure 4.3 Genetic demultiplexing outcomes with GATK and souporcell
Bar plots showing the proportions of singlet, doublet, and unassigned cells generated by the souporcell genetic demultiplexing workflow for each individual library within the Berlin_PB_SF, Ox_VA_PB_SF, P170676 and P180407 projects.

For samples with CITE-seq hashtag data, cell assignment was performed using GMM-Demux (Xin et al., 2020). The proportion of cells assigned by this method for each library is shown in **Figure 4.4**. GMM-Demux classifies cells as singlets, multi-sample multiplets (MSMs), or negatives, and only singlet cells were retained for downstream analysis.

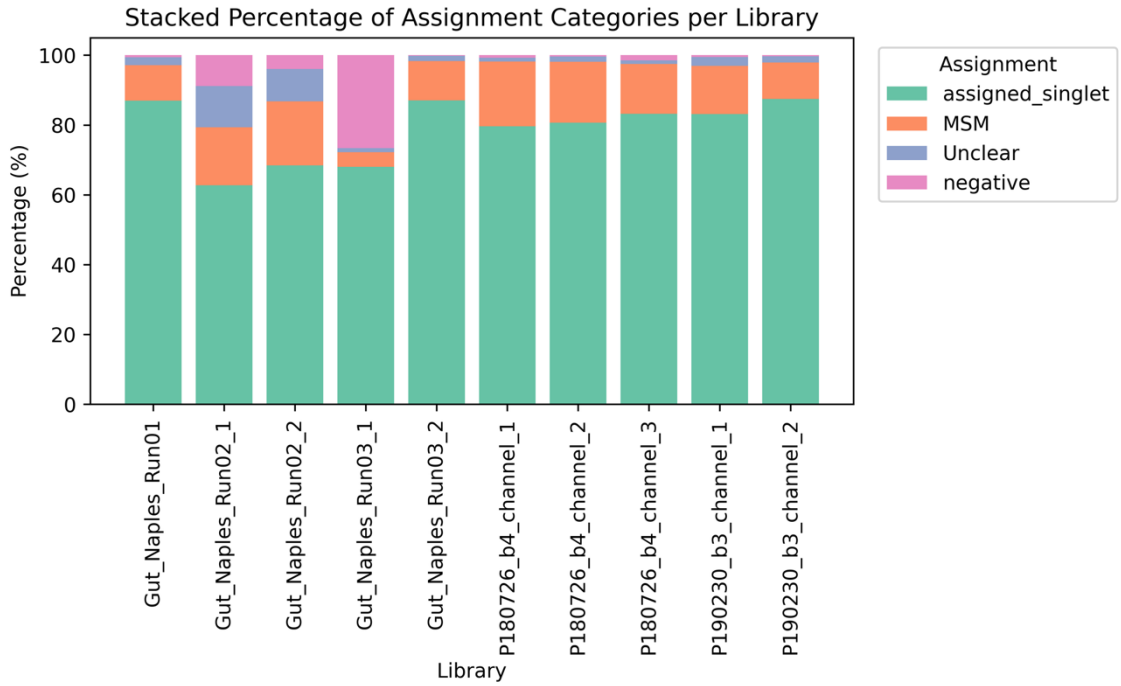


Figure 4.4 Hashtag demultiplexing outcomes with GMM-demux
Bar plots showing the proportions of singlet, MSM, negative and unclear cells generated by the GMM-demux for each individual library within the Gut_Naples, P180407, P190230 projects.

As shown in **Figure 4.4**, the library Gut_Naples_Run03_1 contained a relatively large proportion of cells with negative hashtag labels, which was due to a failure of hashtag labelling for one hashtag in a pool of three individuals. Because this library also lacked bulk genotype information, I additionally performed genetic demultiplexing using souporcell without prior genotype references, and subsequently assigned cells to this individual by integrating the negative hashtag pool from the GMM-Demux results with the singlet clusters inferred by souporcell. As illustrated in **Figure 4.5**, the results from both methods showed clear correspondence: Hashtag_1 aligned with Souporcell cluster 2, Hashtag_2 aligned with Souporcell cluster 1, and the negative hashtag population corresponded to Souporcell cluster 0.

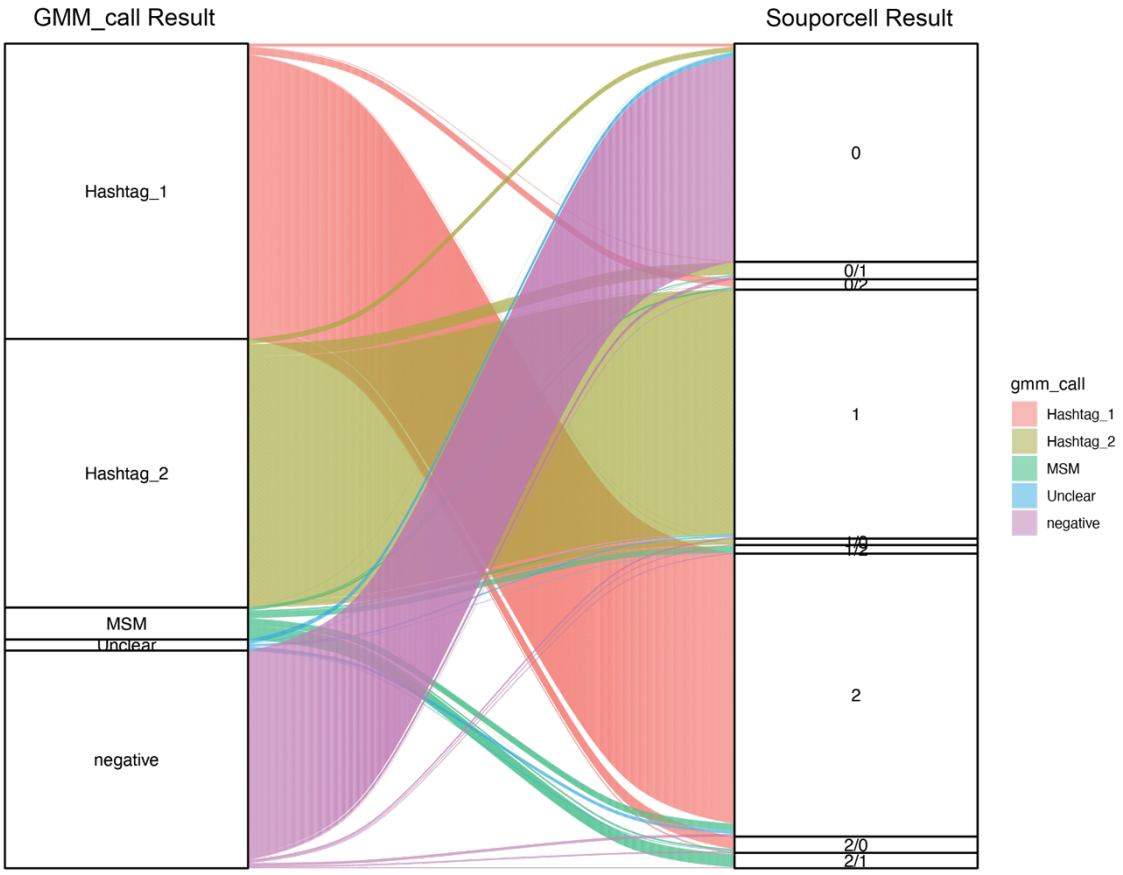


Figure 4.5 Comparison of hashtag and genetic demultiplexing results
Sankey plot illustrating the concordance between hashtag-based (GMM_Demux, left) and genetic-based (Souporcell, right) demultiplexing methods for library Gut_Naples_Run03_1. Categories in hashtag demultiplexing include singlet assignments ('Hashtag_1', 'Hashtag_2'), MSM, unclear, and negative (non-assigned) cells. Souporcell genetic demultiplexing categories show single numbers indicating singlet assignments, and combined numbers indicating identified doublets. Flows between categories are colored according to the hashtag assignments.

This example illustrates the consistency between genetic-based and hashtag-based demultiplexing approaches. In general, when a complete set of hashtags or genetic data was available, I used that primary method for sample assignment. For the Gut_Naples_Run03_1 library where one hashtag was missing, combining both approaches provided a reliable solution for resolving cell identities. Because each batch included multiple disease conditions, any cells that could not be confidently assigned to an individual after demultiplexing process were excluded from downstream analysis.

Together, I acquired 1,105,766 ambient RNA-removed and confidently demultiplexed cells, which were then subjected to downstream quality control.

4.3.2.3 Initial quality control

To retain high-quality cells, I applied the following QC criteria: cells with >15% mitochondrial gene content or fewer than 200 detected genes were removed. Predicted doublets were excluded, defined as cells with a Scrublet score greater than the mean plus three standard deviations within each project. The cells before and after this initial quality control are displayed at the project level in **Figure 4.6**. Following this first round of quality control, a total of 1,056,524 cells remained and were carried forward for the initial integration analysis.

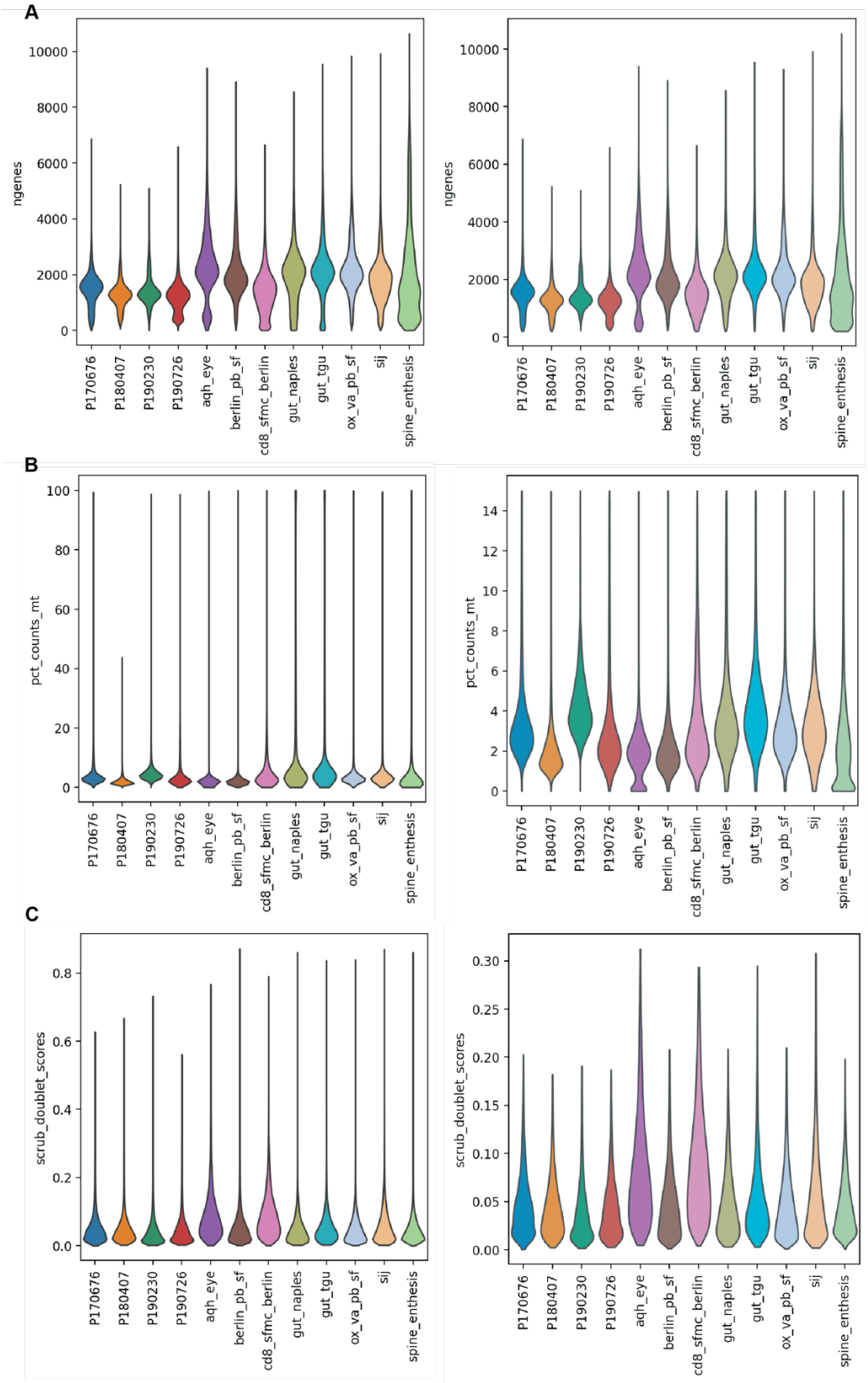


Figure 4.6 Quality control metrics across projects before and after filtering
(A) Number of detected genes, (B) percentage of mitochondrial gene expression, and (C) Scrublet doublet scores, shown for each project. For each panel, the left plot displays values before quality-control filtering and the right plot shows values after filtering.

I next inspected quality metrics at the sample level before and after the initial cell-level filtering (**Figure 4.7**). Prior to filtering, quality profiles varied across tissues, and samples with elevated mitochondrial proportions were mainly drawn from the Gut_Naples and Gut_TGU projects, corresponding to the lieum tissue. This pattern is consistent with reports that gut-derived scRNA-seq samples often contain cells with higher mitochondrial gene proportions (Thomas et al., 2024). After removing high-mitochondrial and low-gene-count cells at the cell level, no remaining samples were identified as outliers regarding the percentage of mitochondrial, indicating that these low-quality cells had been successfully excluded.

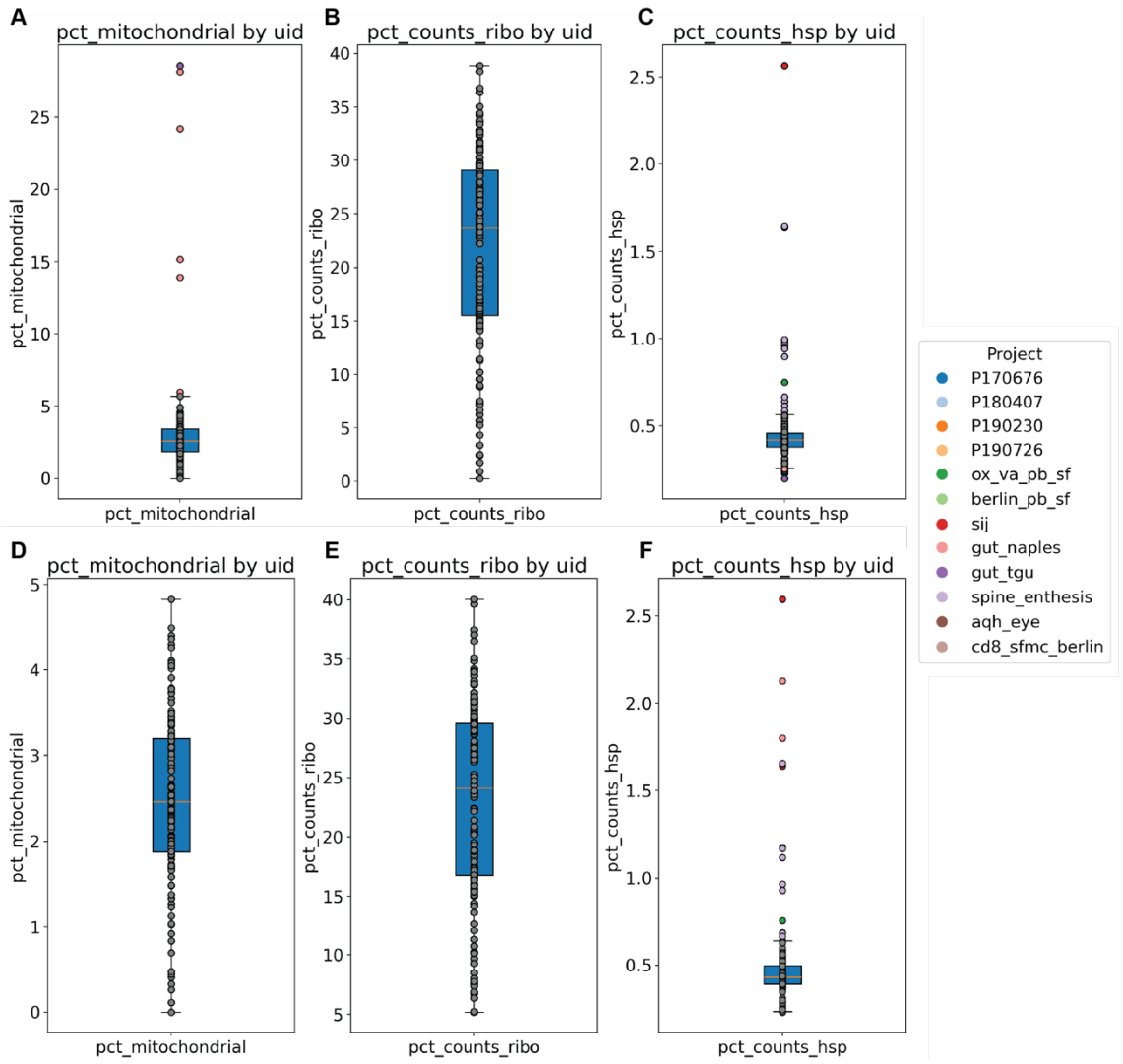


Figure 4.7 Sample-level quality control metrics before and after cell-level filtering
Box plots showing the distribution of sample-level median quality control metrics across projects before and after cell filtering. Panels (A–C) display the median percentage of mitochondrial genes, ribosomal genes, and heat-shock protein genes per sample prior to filtering. Panels (D–F) show the corresponding metrics per sample after filtering. Dots are colored by different project.

I also inspected an additional criterion based on the percentage of heat-shock protein (HSP) gene expression. Samples from the sacroiliac joint and spinal enthesis generally showed higher HSP percentages, as shown in **Figure 4.7C** and **Figure 4.7F**. Elevated HSP expression has been reported to be associated with dissociation-induced cellular stress (van den Brink et al., 2017), and in this context may partly reflect mechanical forces applied during the dissociation of bone-associated tissues. At this stage, I did not filter cells based on HSP expression; instead, I retained this signal for use during integration and applied this criterion in a later round of filtering, as described below.

4.3.2.4 Optimizing the data integration strategy

Since the samples were collected across multiple projects and cohorts, selecting an appropriate integration strategy was a critical step in atlas construction. As an overview, I compared several integration methods, followed by different strategies for selecting highly variable genes (HVG). Based on these comparisons, I selected scVI as the integration framework, as it demonstrated better batch-effect removal performance while preserving biologically meaningful structure. I then assessed different models with varying hyperparameters and tested alternative ways of incorporating continuous and categorical covariates. Using custom performance criteria, I determined the optimal parameters for the initial integration. The steps and evaluations underlying these decisions are described in the following analysis.

First, I focused on selecting an appropriate integration method. I compared scVI (Lopez et al., 2018), Harmony (Korsunsky et al., 2019), Scanorama (Hie et al., 2024), and BBKNN (Polanski et al., 2020), applying each method with comparable parameter settings and treating individuals as batches. The goal of this initial integration step was to achieve clear separation of major cell compartments for subsequent sub-region analyses. Among the methods tested, scVI achieved the most consistent and distinct separation of major cell compartments, including T cells, B cells and mononuclear phagocytes (MNPs). Benchmarking studies have also shown that scVI performs well in both biological conservation and batch-effect correction across diverse single-cell datasets (Luecken et al., 2022). Based on its performance in my dataset and its demonstrated strengths, scVI was selected as the integration method for downstream analyses.

After selecting scVI, I evaluated its performance across different stages, beginning with HVG selection. Using the default scVI parameters, I compared different strategies: (i) using all genes, (ii) selecting the top 1000 HVGs per sample and taking the union, (iii) selecting the top 2000 HVGs per sample and taking the union, (iv) selecting the top 3000 HVGs across the atlas, and (v) selecting the top 2000 HVGs per disease-tissue condition and taking the union. The analysis showed that using either very small or very large HVG sets reduced the ability to distinguish cell states across tissues, whereas a combinatorial union strategy better preserved tissue-specific transcriptional variation. Based on these observations, I selected strategy (v) which preserves intrinsic within-project structure while minimizing cross-project variation likely driven by batch effects.

Figure 4.8 illustrates the comparison in the T-cell compartment as an example, where differences between strategies are most apparent. Under strategy

(v) (**Figure 4.8D**), scVI most clearly separated the expected major T-cell groups including naïve CD8 T cells, CD4 T cells, non-naïve/effector CD8 T cells, and NK-cell subsets, providing a strong basis for subsequent subregion analyses of T-cell populations.

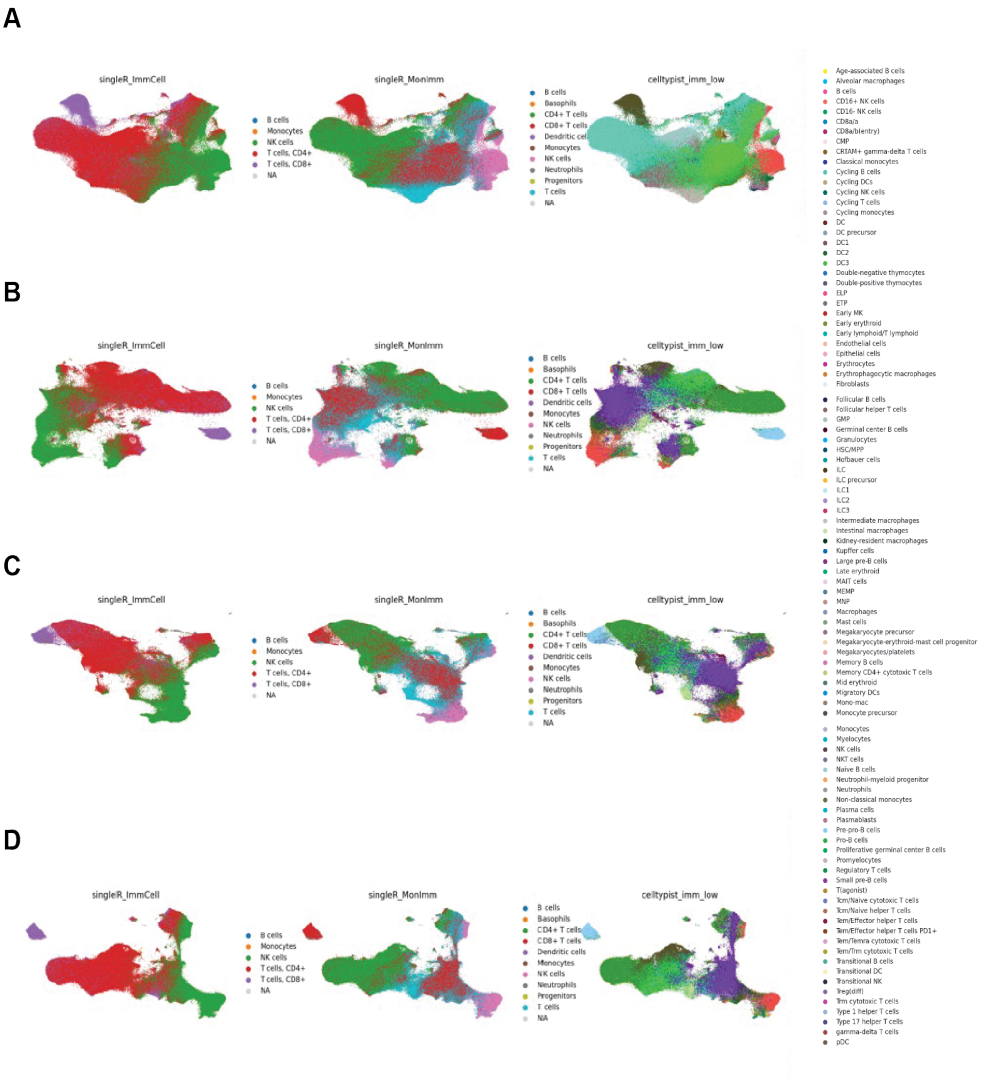


Figure 4.8 Comparison of HVG selection strategies in T cell compartments
UMAPs of the integrated atlas using 4 HVG strategies: (A) top 1000 HVGs per sample (union), (B) top 2000 HVGs per disease–tissue condition (union), (C) top 3000 HVGs across atlas and (D) top 2000 HVGs per disease–tissue condition (union), Rows correspond to strategies; columns show automatic annotation results by SingleR (left, middle) and CellTypist (right).

With the HVG strategy established, I next explored scVI model construction parameters, including the number of hidden layers used in the encoder and decoder neural networks, the dimensionality of the latent space, likelihood models, batch keys, and continuous covariates. **Figure 4.9** presents an example of UMAP comparisons across models with different numbers of network layers and latent space dimensions.

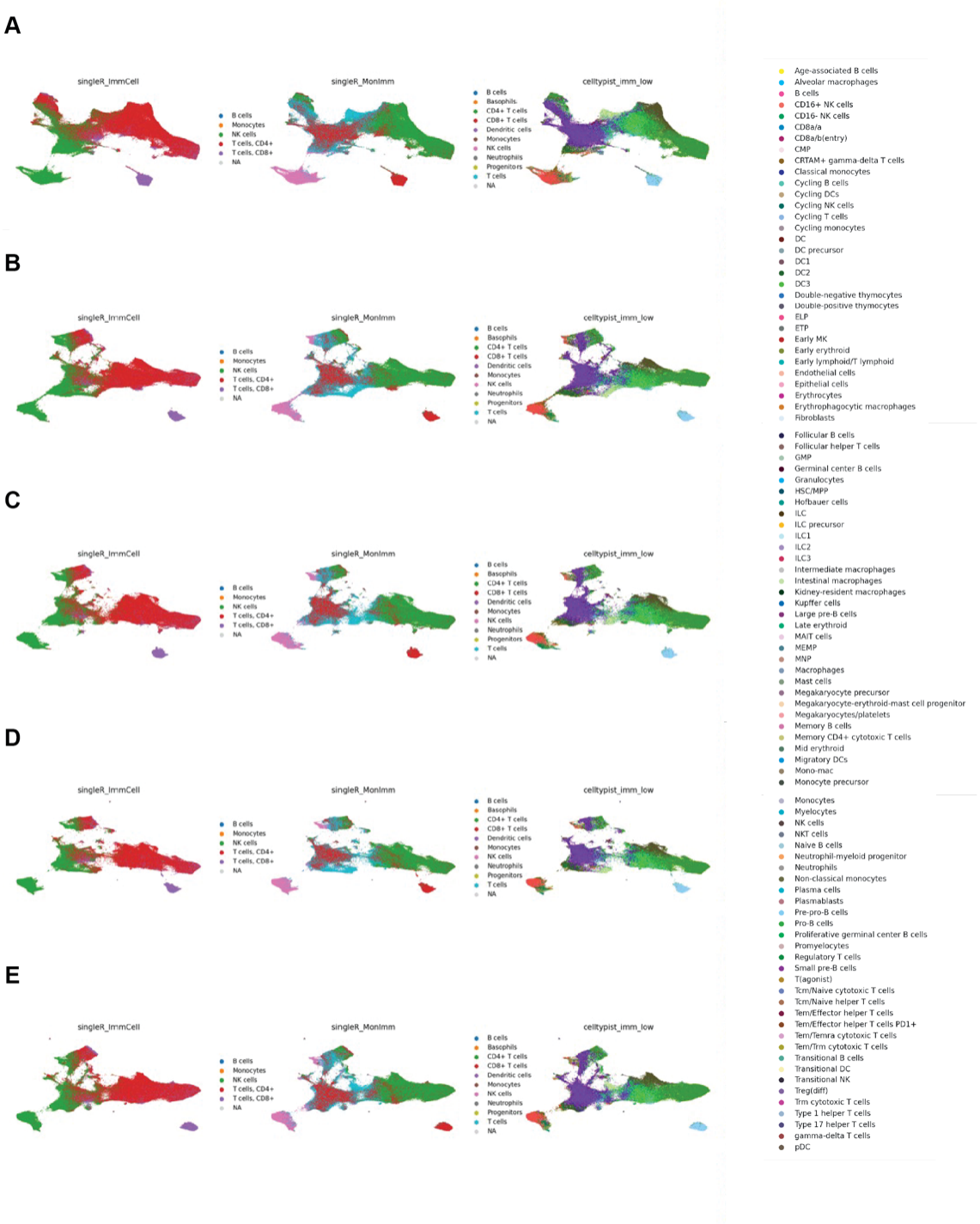


Figure 4.9 UMAPs of the integrated atlas under different scVI model parameters
(A–E) UMAP embeddings of the overall atlas generated with varying scVI model configurations. (A) n_layer=10, n_latent=2; (B) n_layer=20, n_latent=2; (C) n_layer=30, n_latent=2; (D) n_layer=40, n_latent=2; (E) n_layer=40, n_latent=3. n_layer refers to the number of layers, and n_latent indicates the size of the latent space.

To evaluate model performance, I compared evidence lower bound (ELBO) values (Kingma & Welling, 2014) and normalized mutual information (NMI) scores

(Strehl & Ghosh, 2003) between clusters and annotation labels, and used CellTypist (Dominguez Conde et al., 2022) for automated cell-type annotation (**Table 4.2**). Overall integration and sub-regional integration were performed iteratively, with decisions guided by balancing ELBO as a measure of model performance and NMI as a measure of biological consistency. Because the goals of the integration differed between stages, the evaluation criteria were weighted accordingly. For the initial integration, where the objective was to separate major immune compartments, I focused on NMI metric computed using low resolution CellTypist annotations (imm_low), which reflects broad cell-type distinctions. For subregional integration, where the aim was to resolve finer-grained populations, I prioritized NMI metric using higher resolution CellTypist annotations (imm_high) to capture more subtle cell-state differences.

Model	ELBO	NMI (imm_low)	NMI (imm_high)
n_layer: 10 n_latent: 2	-2.39×10^3	0.501	0.224
n_layer: 20 n_latent: 2	-2.41×10^3	0.543	0.227
n_layer: 30 n_latent: 2	-2.43×10^3	0.542	0.234
n_layer: 40 n_latent: 2	-2.45×10^3	0.534	0.227
n_layer: 40 n_latent: 3	-2.42×10^3	0.527	0.233

Table 4.2 Evaluation metrics for scVI models
Evaluation of scVI models trained with different numbers of network layers and latent dimensionalities. Metrics include the ELBO and NMI values computed using low- and high-resolution CellTypist annotations. All metrics were calculated using the full set of cells included in the integration.

The UMAP visualizations and evaluation metrics indicated that increased network depth or larger latent dimensionality did not consistently improve model performance and integration results. Instead, the optimal configuration depended on the specific analytical objective in each stage of the workflow. In summary, parameter choices were determined based on their behaviour in the context of each integration stage, guided by ELBO and NMI evaluations rather than a single universally optimal configuration.

4.3.2.5 Overview of initial data integration

After the initial integration, cells from different projects and different individuals were combined into a single atlas (**Figure 4.10**). As shown in **Figure 4.10B-C**, cells from projects derived from the same tissue type were well aligned in the integrated space. In contrast, tissue-specific projects such as those from the spinal enthesis (PEB and EST) formed distinct clusters, likely reflecting the presence of tissue-specific cell populations.

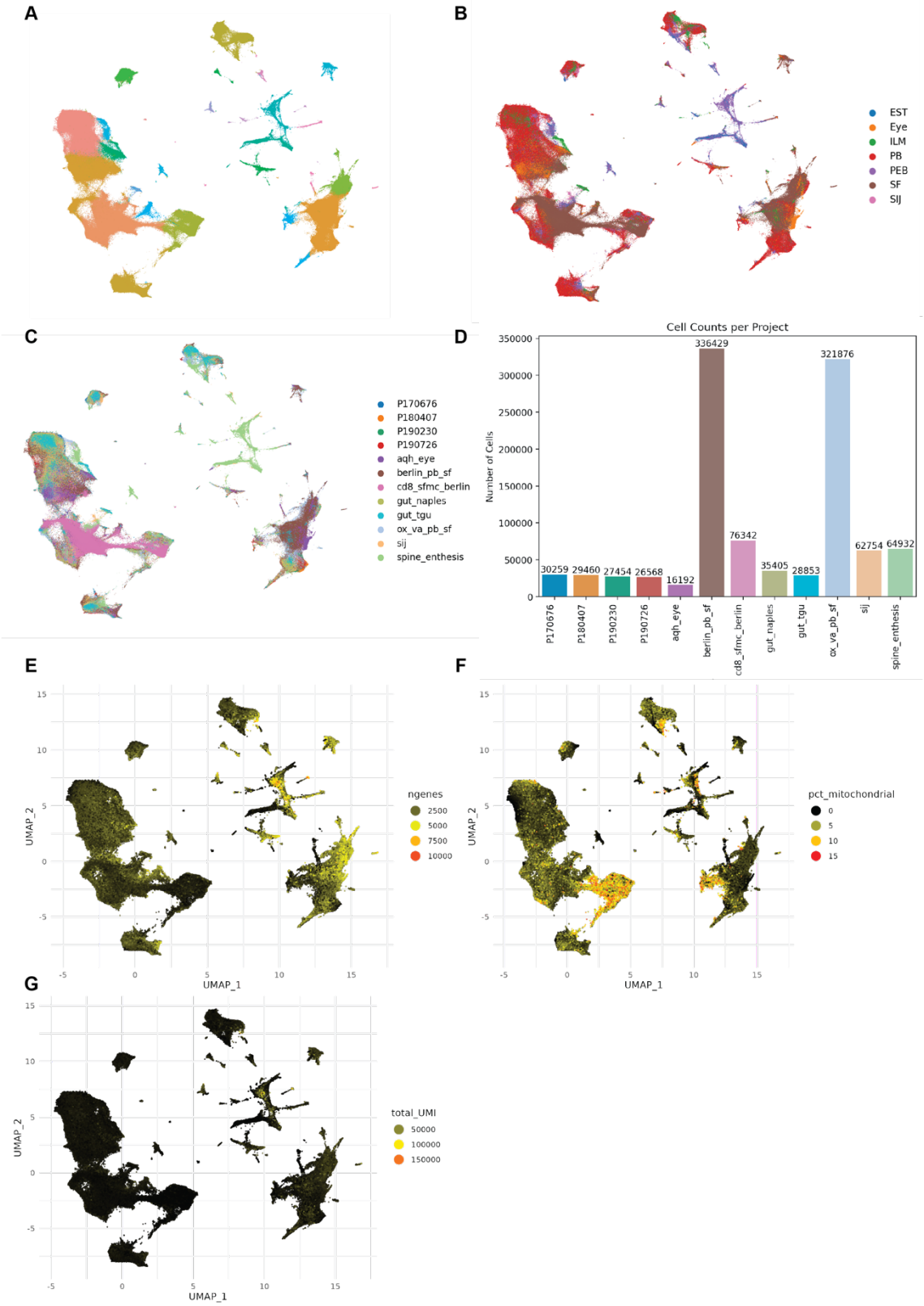


Figure 4.10 Overview of the initially integrated atlas and sample quality metrics
(A) UMAP of the integrated initial atlas at clustering resolution 0.4. (B) UMAP coloured by tissue of origin. (C) UMAP coloured by project. (D) Bar plot showing the number of cells contributed by each project. (E–G) UMAPs displaying cell-level quality metrics: (E) number of detected genes, (F) percentage of mitochondrial gene expression, and (G) total UMI counts.

The atlas revealed broad regions corresponding to major immune compartments as shown in **Figure 4.11**, but also highlighted clusters enriched for indicators of low-quality cells, such as elevated mitochondrial gene expression (**Figure 4.10F**). These observations indicated that further quality-control steps would be required following subsequent analyses to remove low-quality clusters.

Based on broad immune cell categories, the UMAP revealed 4 major regions: a T/NK cell region, a B/plasma cell region, a mononuclear phagocyte (MNP) region, and a mixed other cell region (e.g. erythrocytes and fibroblasts) (**Figure 4.11**).

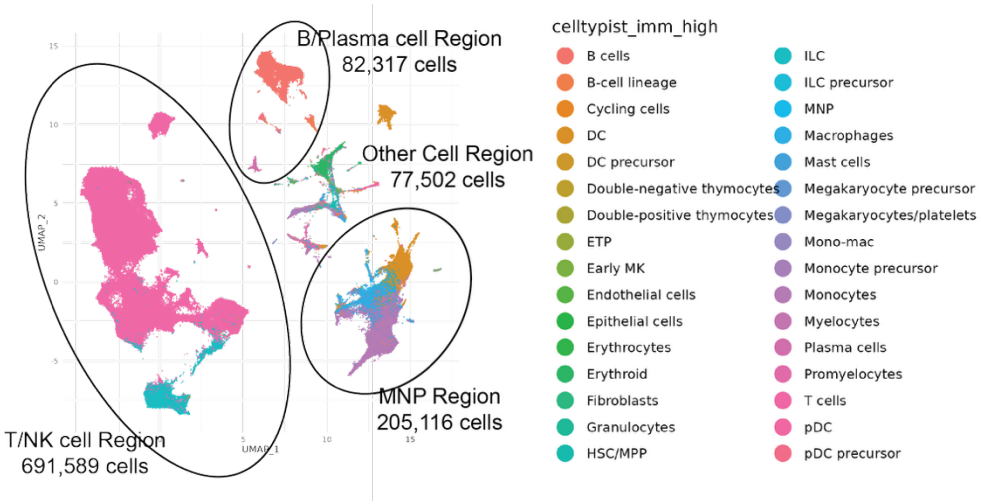


Figure 4.11 UMAP of 4 major immune cell compartments in the initial atlas
UMAP of the integrated atlas highlighting 4 major compartments: T/NK cells, B/plasma cells, MNPs, and other cells, colored by Celltypist annotations.

4.3.2.6 Sub-regional clustering and secondary quality control

To capture within-region heterogeneity and enable more detailed exploration of each compartment, I subsetted each region, re-selected variable genes, and performed re-integration and clustering for subsequent secondary round of quality control.

Because elevated HSP expression was observed in cells from the sacroiliac joint (as shown in **Figure 4.7**), I examined this signal at the single-cell level as an additional quality-control criterion. Elevated HSP expression has been reported as an indicator of cellular stress during tissue processing (van den Brink et al., 2017), and to assess it systematically I curated a canonical HSP gene list (small HSPs, HSP40/DNAJ, HSP60, HSP70, HSP90 and large HSPs; excluding TRiC/CCT chaperonins) based on HSP family reviews (Hu et al., 2022).

The updated UMAPs for the T/NK-cell region with quality metrics, including mitochondrial percentage, mean cytoplasmic gene expression, HSP gene percentage, and Scrublet doublet scores, are shown in **Figure 4.12**. The UMAP showed distinct regions enriched for mitochondrial and HSP expression, suggesting the presence of low-quality cells or stress-induced signatures likely caused by sample handling. I therefore proceeded to explore strategies for removing these unwanted cell populations.

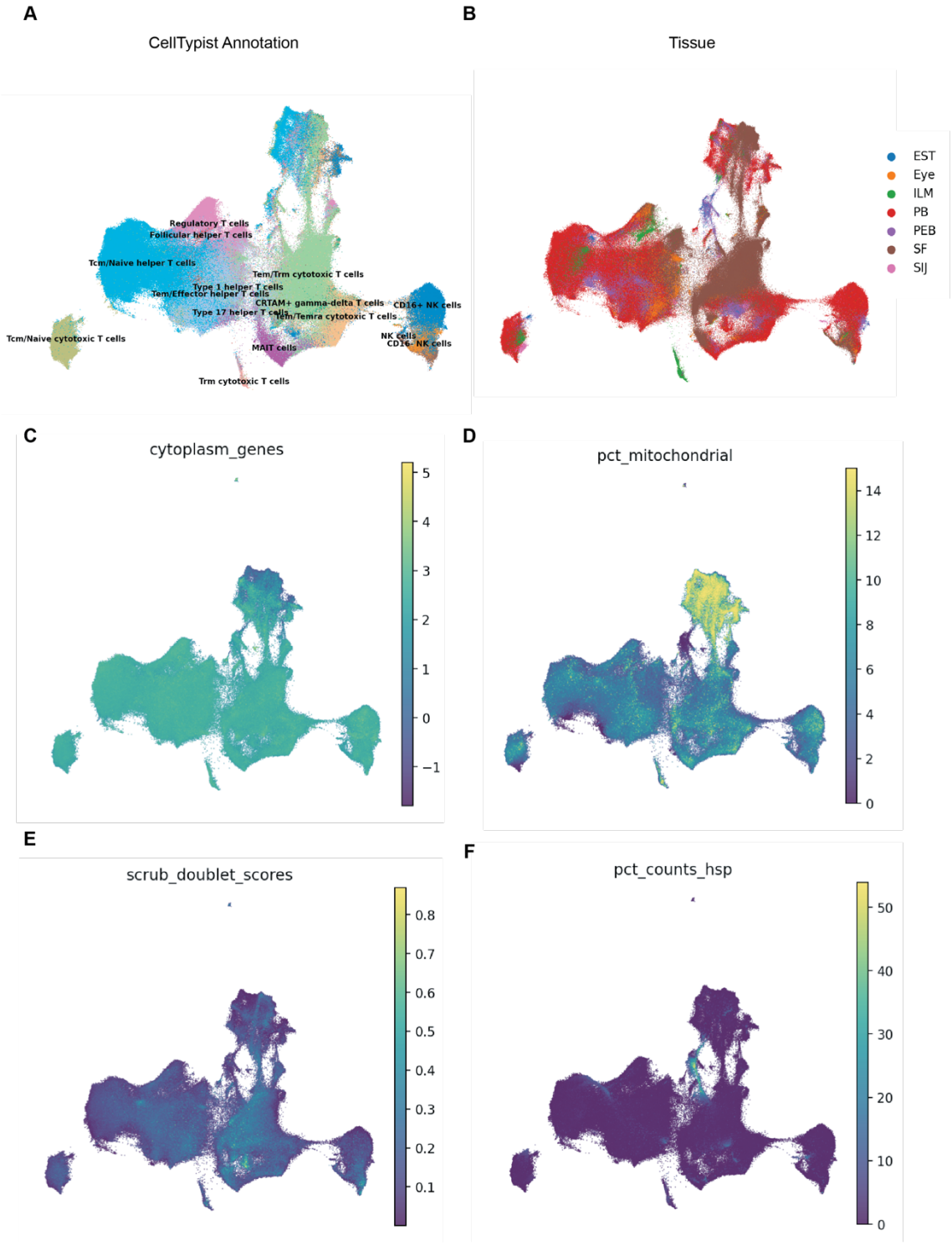


Figure 4.12 Quality control metrics of the T-cell subregion
(A) UMAP of the T-cell subregion colored by the top 15 enriched cell types predicted by CellTypist. (B) UMAP of T cell subregion colored by tissue origin. (C–E) UMAPs showing quality control metrics: (C) mean expression of cytoplasmic genes, (D) percentage of mitochondrial gene counts, (E) Scrublet doublet scores, and (F) percentage of HSP gene counts.

To address the mitochondrial signals, I tested two approaches: applying more stringent filtering thresholds and incorporating mitochondrial gene percentages as continuous covariates in the scVI model. However, these approaches did not fully mitigate the mitochondrial effects, as regions with elevated mitochondrial expression remained distinct after integration (**Figure 4.13**).

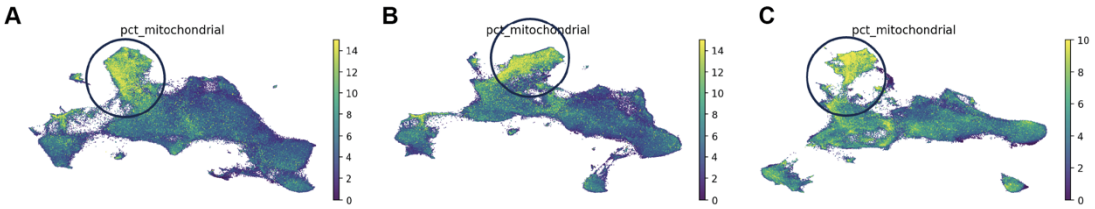


Figure 4.13 scVI models incorporating mitochondrial covariates
(A–C) UMAPs of the T-cell subregion from scVI models including mitochondrial percentage as a continuous covariate. Models are shown after filtering cells with >15% (A), >20% (B), or >25% (C) mitochondrial content.

In the model without continuous covariates (**Figure 4.12D**), the low-quality clusters appeared more clearly separated, which facilitated their identification. I therefore reverted to this model for cluster-level quality control and removed the isolated low-quality clusters accordingly. Boxplots of quality metrics for each Leiden cluster are shown in **Figure 4.14**.

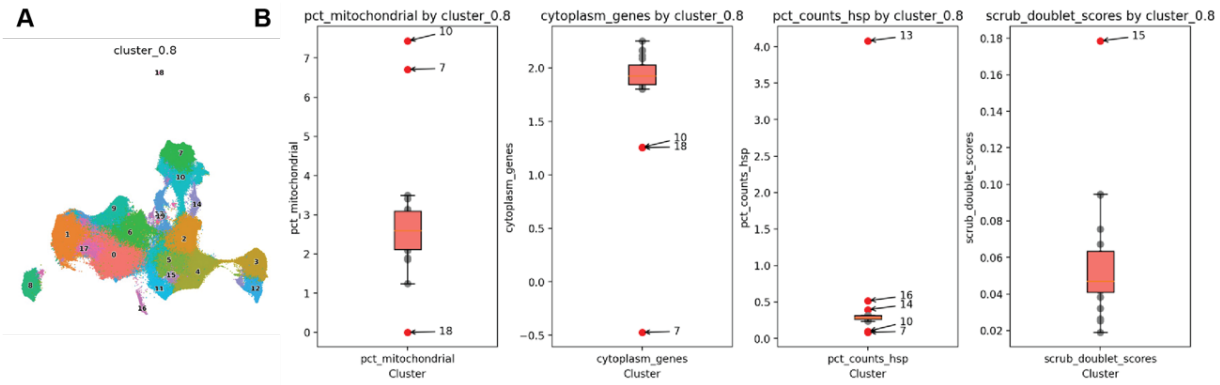


Figure 4.14 Clustering and cluster-level quality metrics of the T/NK-cell subregion
(A) UMAP of the T/NK-cell subregion at clustering resolution 0.8. (B) Boxplots showing cluster-level distributions of key quality control metrics: percentage of mitochondrial genes, mean cytoplasmic gene expression, percentage of HSP genes, and Scrublet doublet scores; red dots indicate outlier clusters.

Based on these evaluations, I excluded clusters that met the predefined outlier criteria (upper bound = 75th percentile + 1.5 × IQR; lower bound = 25th percentile – 1.5 × IQR). Specifically, clusters were removed if they showed both aberrant mitochondrial percentage (upper or lower outlier) and low cytoplasmic gene expression (lower outlier), a median HSP percentage above 1.5%, or Scrublet scores exceeding the upper outlier threshold. Following this approach, clusters 7, 10, and 18 were excluded as low-quality clusters, cluster 13 as an HSP-high cluster, and cluster 15 as a doublet cluster.

Following the same methods, I performed cluster-based quality control on the B-cell compartment (**Figure 4.15**). Clusters 5 and 12 were identified as outliers based on elevated Scrublet doublet scores. Examination of marker genes confirmed that cluster 5 expressed both *CD79A*, *MS4A1* together with *CD3E*, indicating a likely B/T-cell doublet cluster; Cluster 12 expressed *CD79A* and *MS4A1* alongside *CD14*, *LYZ*, and *CST3* and was considered as a B/monocyte doublet cluster (**Figure 4.15D**). Both clusters were therefore excluded. In addition, cluster 8 was removed as a low-quality cluster due to high mitochondrial content and low cytoplasmic gene

expression, and cluster 10 was excluded as an HSP-enriched cluster, with a median HSP percentage above 1.5%, consistent with the criteria previously.

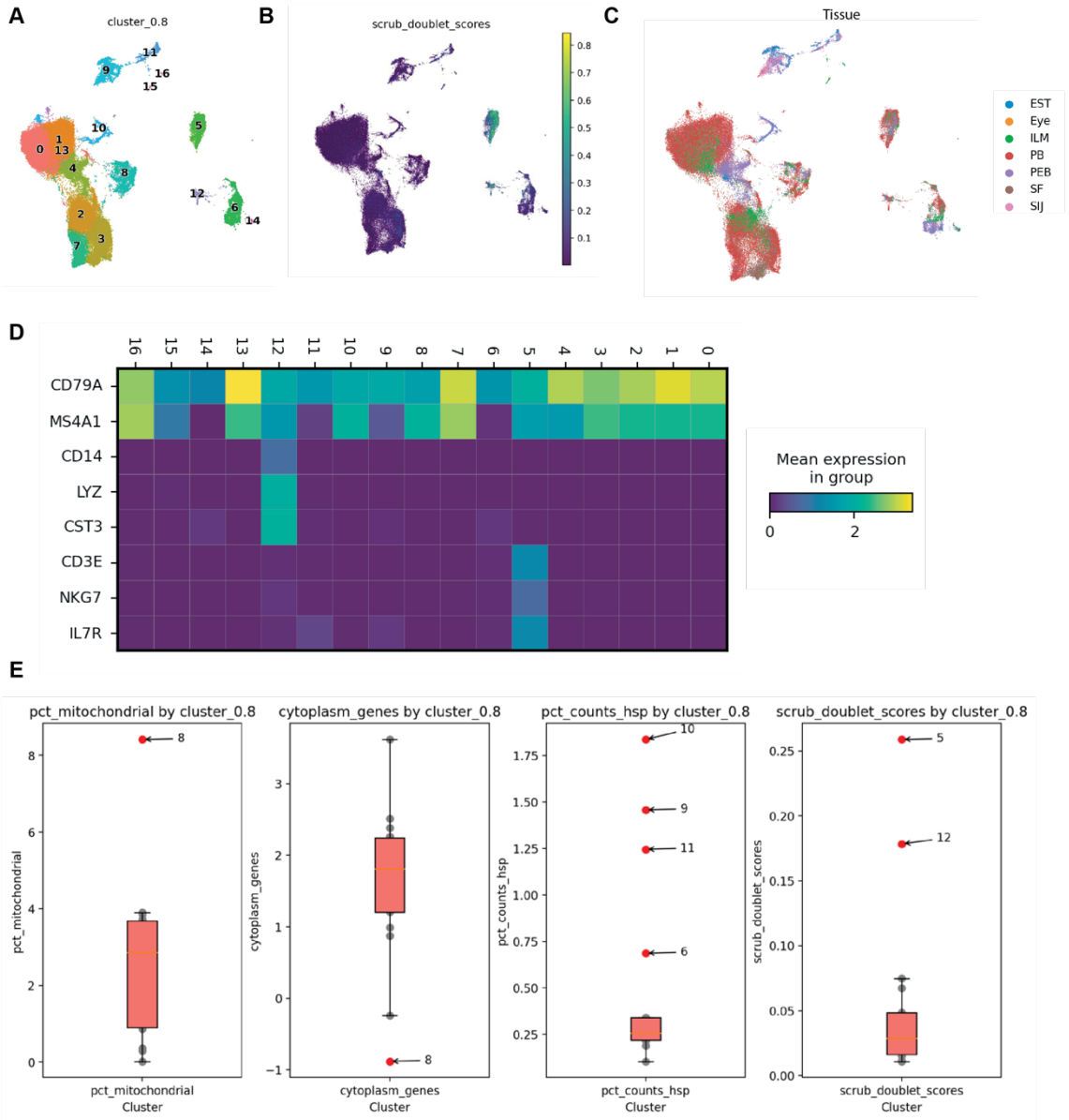


Figure 4.15 Clustering and quality control of the B-cell subregion
(A) UMAP of the B-cell subregion after clustering in resolution 0.8. (B) UMAP colored by Scrublet doublet scores. (C) UMAP colored by tissue origin. (D) Dot plot showing expression of canonical immune cell marker genes across clusters. (E) Boxplots of cluster-level quality control metrics, including mitochondrial percentage, cytoplasmic gene expression, HSP gene percentage, and Scrublet doublet scores; red dots indicate outlier clusters.

Figure 4.16 shows the cluster-based quality control of the mononuclear phagocytes (MNP) compartment. No doublet clusters were excluded from this compartment, likely because doublets involving monocytes and major cell types (e.g., T/mono or B/mono) were already identified in other regions. I excluded clusters 4 and 8 as low-quality clusters, as both showed consistent outlier profiles in mitochondrial and cytoplasmic gene proportions. Cluster 7 was removed as an HSP-enriched cluster, with a median HSP percentage above 1.5%, applying the same threshold used in other regions.

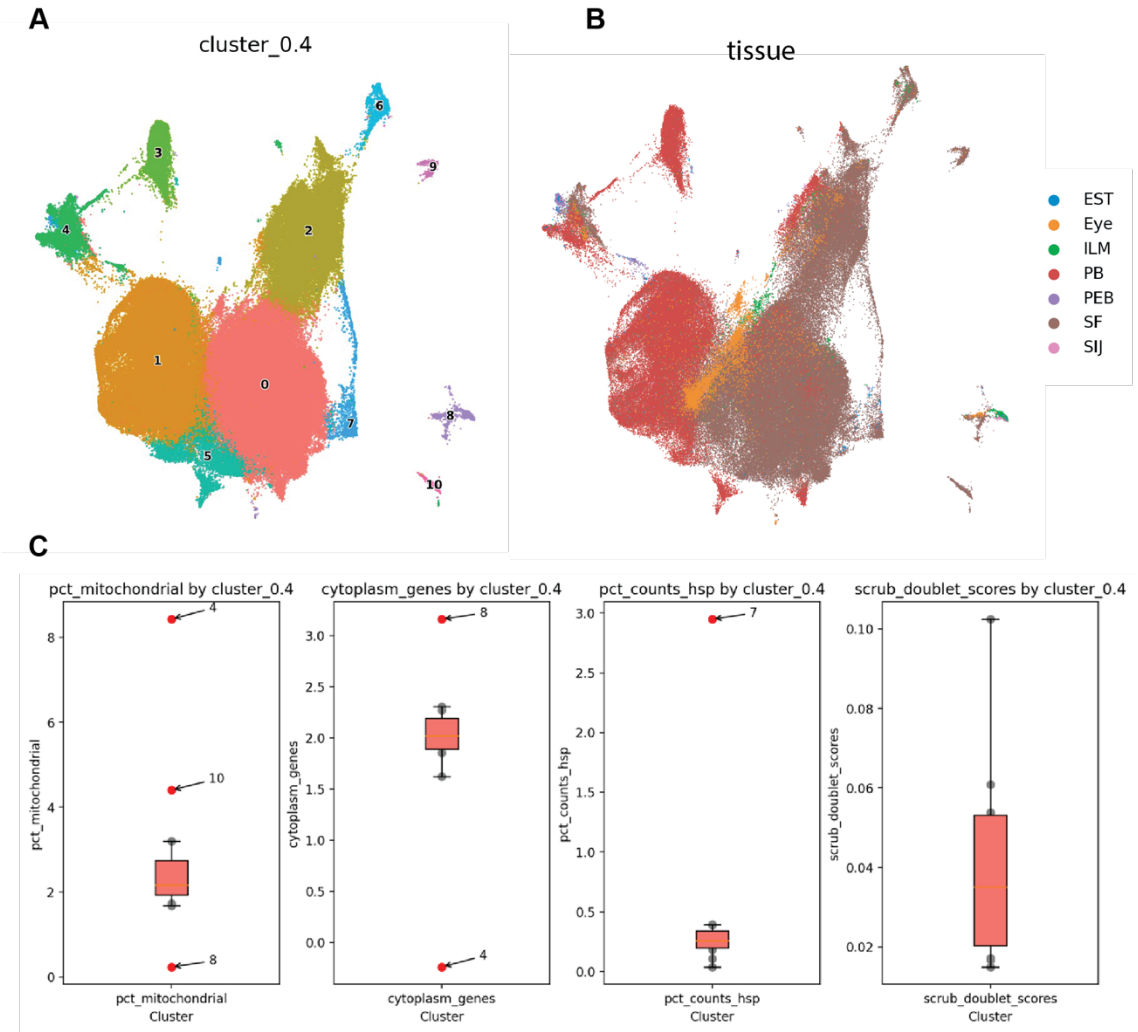


Figure 4.16 Clustering and quality control of the MNP cell subregion
(A) UMAP of the MNP-cell subregion after clustering in resolution 0.4. (B) UMAP of the MNP-cell subregion coloured by tissue. (C) Boxplots of cluster-level quality control metrics,

including mitochondrial percentage, cytoplasmic gene expression, HSP gene percentage, and Scrublet doublet scores; red dots indicate outlier clusters.

All remaining cells were grouped into the “other” cell compartment, which was primarily composed of erythrocytes and fibroblasts, along with small immune cell subsets that could not be merged into the major cell regions (**Figure 4.17**). Three doublet clusters were identified (cluster 1,9,16), and expression of canonical immune marker genes confirmed these as T-cell and classical/non-classical monocyte/DC doublets. In addition, cluster 17 was excluded as a low-quality cluster: although it was not an outlier, it showed the lowest cytoplasmic gene proportion among the remaining clusters and overall lower quality. Cluster 22 was excluded as an HSP-enriched cluster according to the same criteria applied previously.

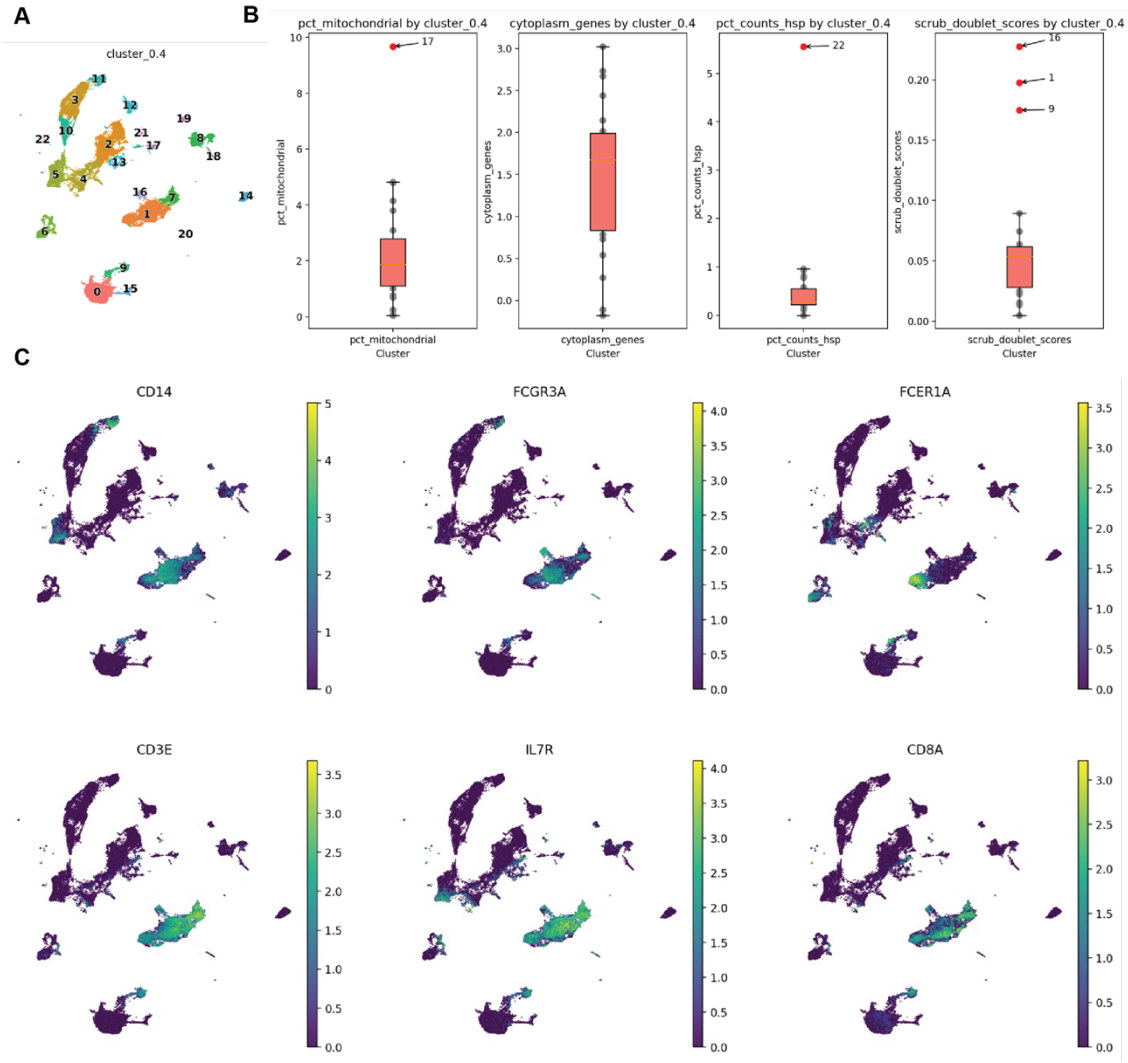


Figure 4.17 Clustering and quality control of the 'other' cell subregion
(A) UMAP of the remaining 'other' cell subregion after clustering in resolution 0.4. (B) Boxplots of cluster-level quality control metrics, including mitochondrial percentage, cytoplasmic gene expression, HSP gene percentage, and Scrublet doublet scores; red dots indicate outlier clusters. (C) UMAPs showing the distribution of canonical immune cell marker genes within this subregion.

After the second round of cluster-level quality control, I repeated the analysis pipeline for each subregion from HVG selection to Leiden clustering. In this QC round, HSP genes were not included as covariates so that HSP-high clusters could be identified and removed. In the final scVI model for annotation, HSP genes were then incorporated as continuous covariates to improve integration performance. Overall, this iterative process resulted in improved delineation of biologically

meaningful clusters, as features from contaminating cell types previously present in doublet or low-quality clusters were no longer captured among the highly variable genes.

Together, the 4 major compartments are shown in **Figure 4.16**; three of these compartments are presented in their final form, whereas the T/NK-cell compartment required additional subsetting and is described in the next section.

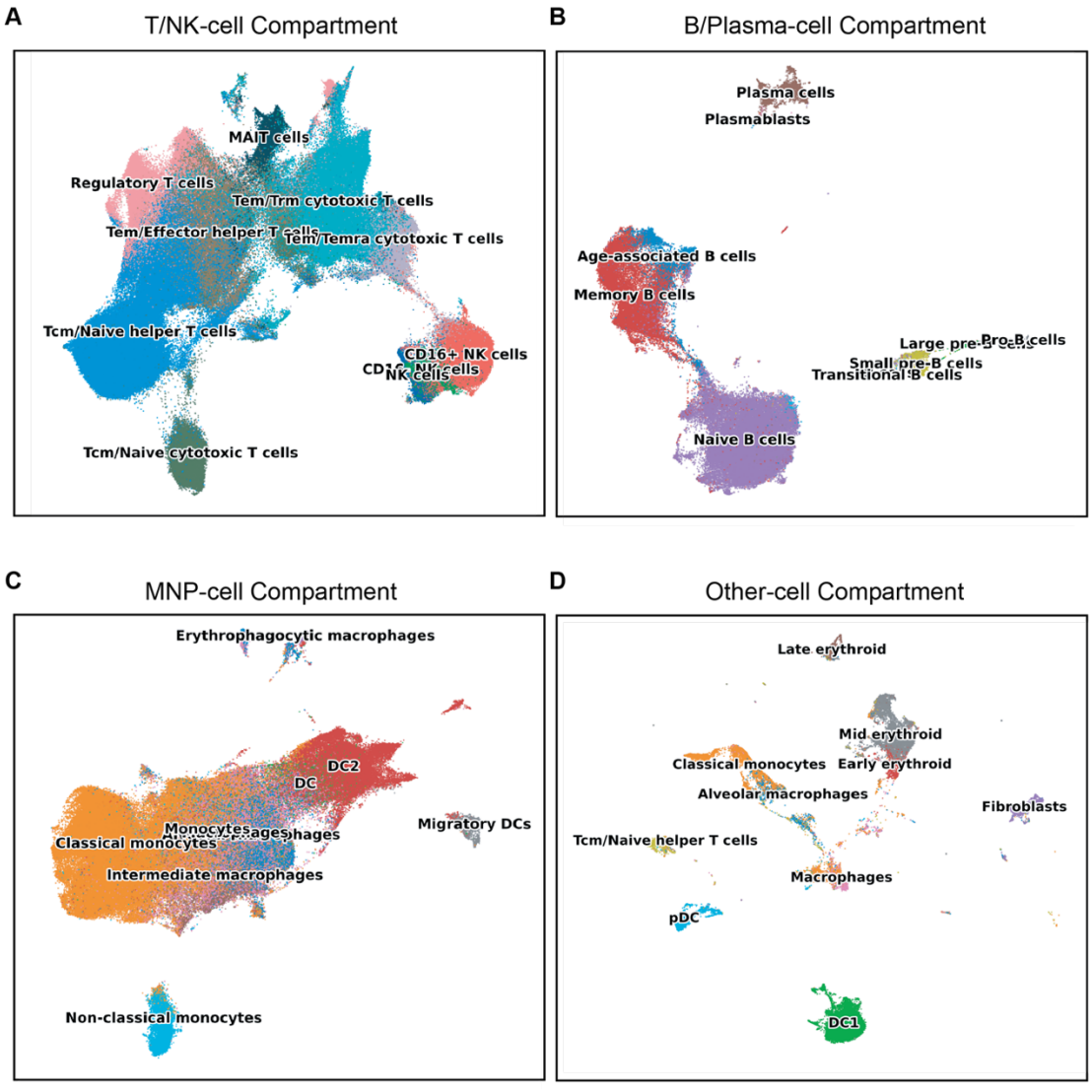


Figure 4.18 UMAPs after second-round quality control
UMAP of the (A) T/NK-cell compartment, (B) B-cell compartment, (C) MNP compartment, and (D) 'other' cell compartment after the second round of quality control, labeled by the top 10 enriched CellTypist annotation. Some labels (e.g., alveolar macrophage) reflect transcriptional similarity to the CellTypist reference rather than true tissue origin.

4.3.2.7 T Cell cub-regional clustering and tertiary quality control

Because the T/NK-cell compartment represents a large and heterogeneous population, higher-resolution subregional analysis was required. The UMAP representation in **Figure 4.19** shows that this compartment includes distinct major lymphocyte lineages (T cells marked by *CD3E*, NK cells marked by *NCAM1*), clear CD4/CD8 divisions (*CD4, CD8A, CD8B*), and a diversity of functional states, ranging from naïve (*CCR7*) to memory (*IL7R*) and regulatory (*FOXP3*) populations. These patterns highlight the biological diversity contained within this compartment and indicate the need for further refinement through subregional analysis.

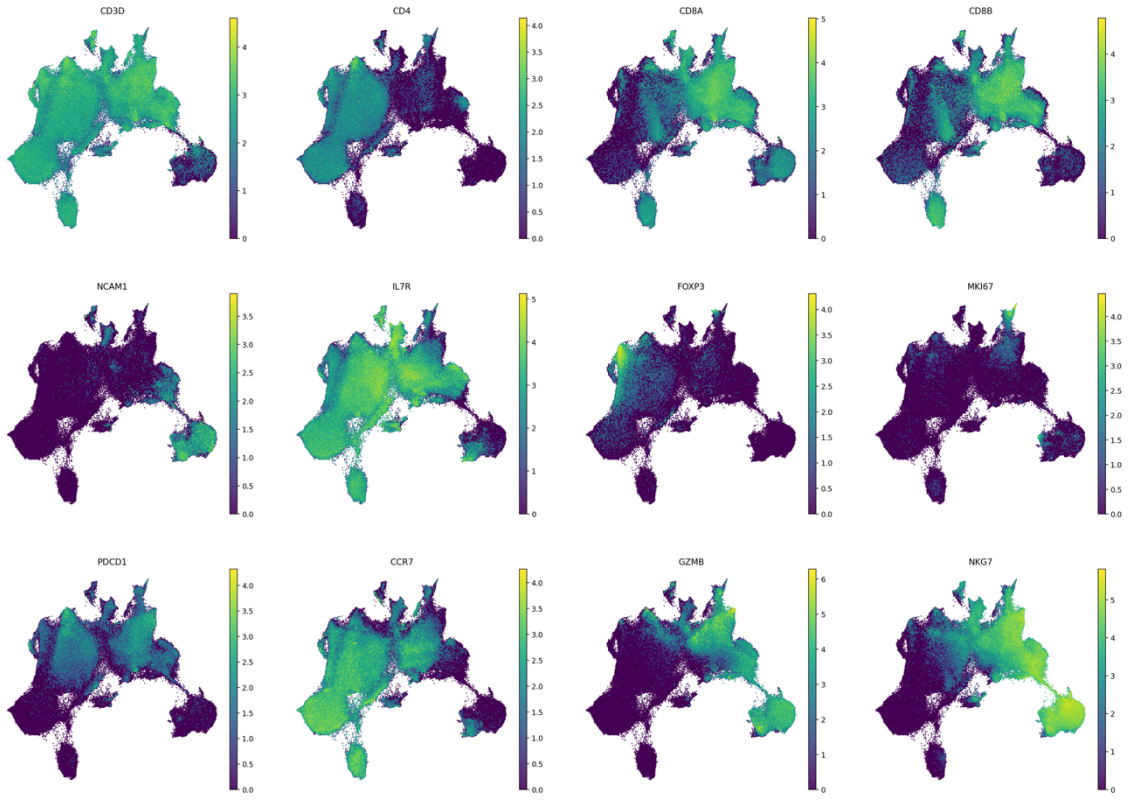


Figure 4.19 Expression of key marker genes in the T/NK-cell compartment
UMAPs of the T/NK-cell compartment showing expression of representative marker genes for major lineages and functional states.

To better characterize these populations, I subdivided the T/NK-cell region into 4 compartments: CD8 naïve, CD4, remaining CD8, and NK cells based on Leiden clustering results. For each compartment, I re-selected HVGs, performed additional rounds of integration and clustering, and carried out a third round of quality control based on the updated results.

The summary diagram illustrating the generation of the seven final cell regions is shown in **Figure 4.20**. The final round of quality control was performed retrospectively, informed by observations from downstream analyses. During pseudo-bulk analysis, I detected immunoglobulin gene expression within T-cell clusters, indicating the likely presence of T/plasma cell doublets. I therefore returned at this stage to remove cells in the final T-cell compartments that expressed immunoglobulin genes, with the exception of *IGHV1-3* and *IGLV2-14*, which have been observed in T cells in publicly available online resources and are discussed later.

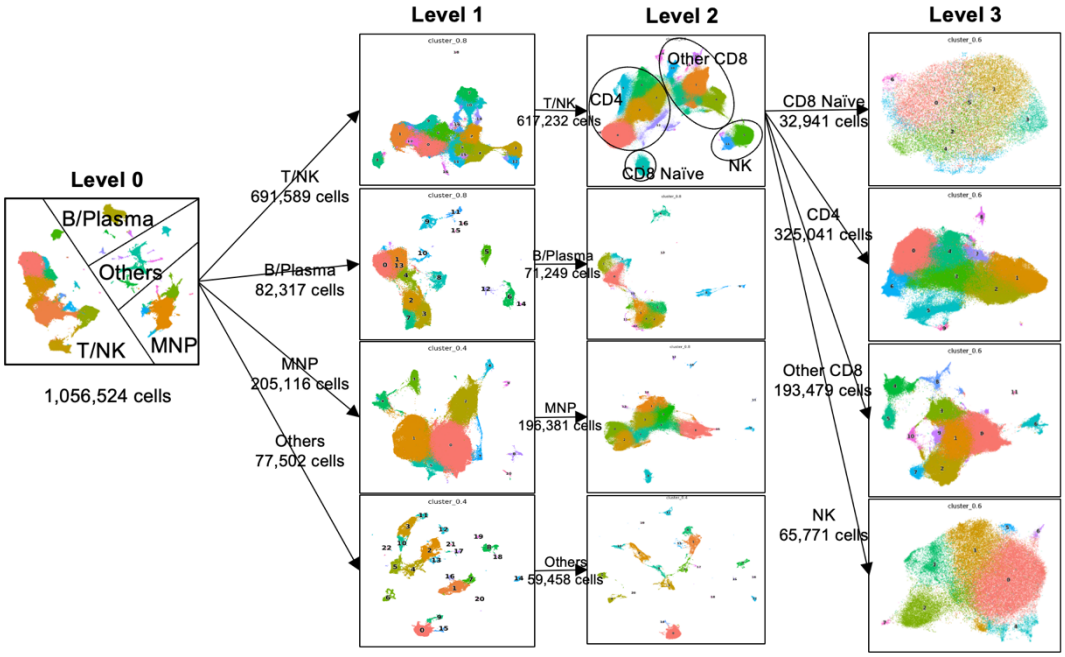


Figure 4.20 Summary diagram of the iterative subregional clustering and quality control process
UMAP overview of the atlas showing progressive subsetting from the initial integration (Level 0) into major compartments, followed by iterative re-integration and clustering at Levels 1–3. Cell numbers are indicated for each compartment.

4.3.2.8 ADT data processing

For this cross-tissue atlas, ADT data were incorporated to complement transcriptomic information by providing protein-level evidence for cell type annotation and to explore whether gene-level differences were also reflected at the protein level. The libraries processed with CITE-seq are listed in **Table 4.1**. For samples from these libraries, the corresponding RNA and ADT counts per sample are shown in **Figure 4.21**. Although all samples in the CITE-seq libraries generated RNA counts, three samples did not yield detectable ADT signals, which reflects variation in antibody staining during CITE-seq process.

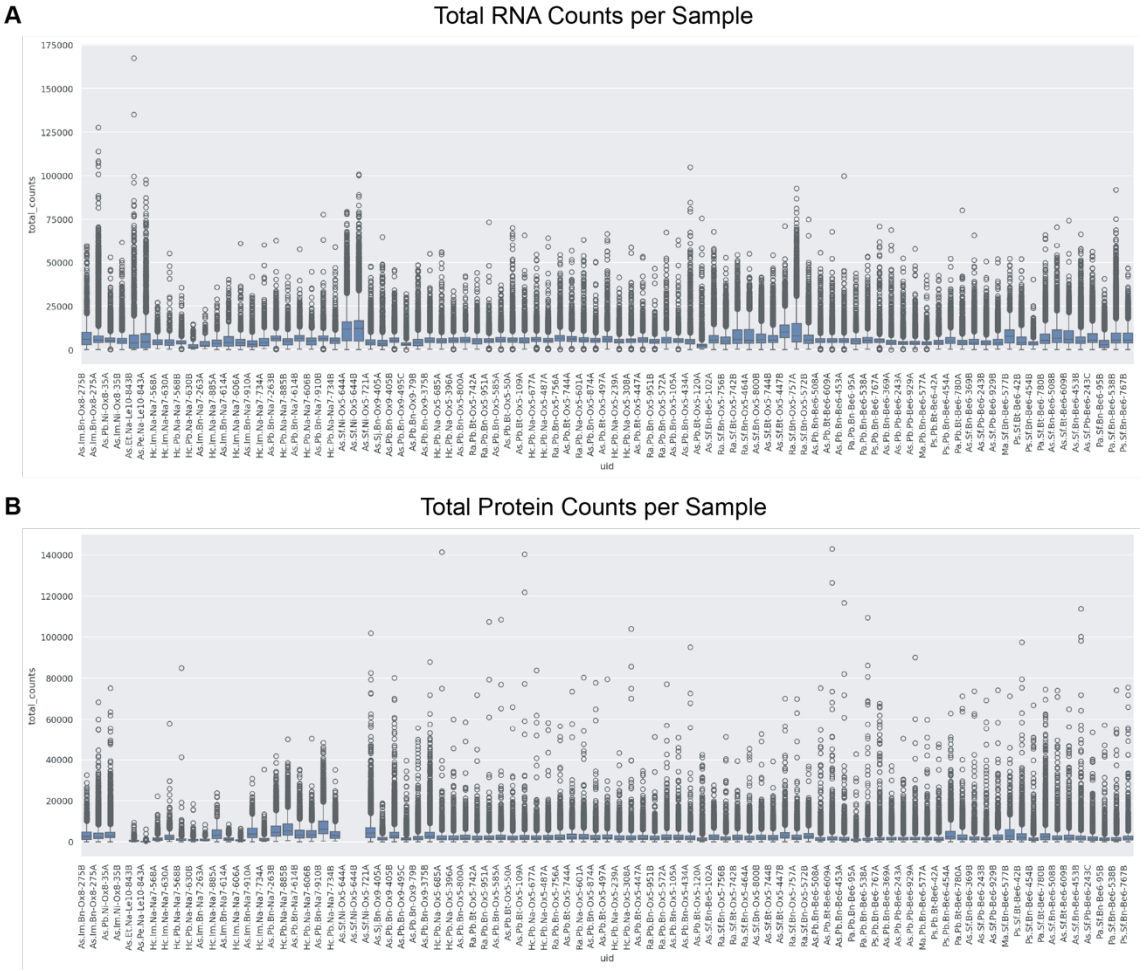


Figure 4.21 Comparison of total RNA and total protein counts across CITE-seq samples
(A) Total RNA counts per sample, showing variation in sequencing depth across donors. (B) Total protein counts per sample derived from CITE-seq data. Only samples processed with CITE-seq are included, and samples are labeled by unique identifiers (UIDs).

Among these libraries, ADT data were generated using three different antibody panels. Two pilot projects (P190230 and P190726) were processed with custom lab-designed panels (**Appendices C.1-C.2**), while all remaining projects were processed with a common panel (TotalSeq-C Human Universal Cocktail; **Appendix C.3**).

Quality control for ADT data was performed at the single-cell level. Cells with total ADT counts exceeding 80,000 or with fewer than 7 detected proteins were

excluded. After applying these filters, 767,863 cells remained and were considered to have sufficiently high-quality ADT measurements for downstream analyses.

Because ADT measurements can be affected by variation in antibody concentration, staining efficiency, and background signal, integration across libraries was required to ensure that protein expression was comparable across samples. Several approaches exist for normalising or denoising ADT data, including DSB, which is widely used to correct background contamination at the per-sample level. However, DSB does not perform batch integration or jointly model relationships between RNA and protein modalities. I therefore used totalVI (Gayoso et al., 2021), which jointly models RNA and protein data to generate a shared latent representation of both modalities and harmonize datasets across batches.

Since the first two panels were pilot experiments that used antibody sets distinct from the main panel, accordingly totalVI models were trained separately for these datasets, and downstream analyses were focused on the common antibody panel. To maintain a consistent representation of all cells during annotation, UMAP visualization and Leiden clustering were performed using the scVI model trained on gene expression data alone, with ADT information incorporated only for visualization, validation and cluster annotation.

Processed ADT signals for selected antibodies before and after totalVI-based normalization are shown in **Figure 4.22**. After normalization, the distributions transition from heterogeneous and often non-comparable raw count profiles to clearer negative/positive bimodal patterns or distinct negative, intermediate, and high expression peaks.

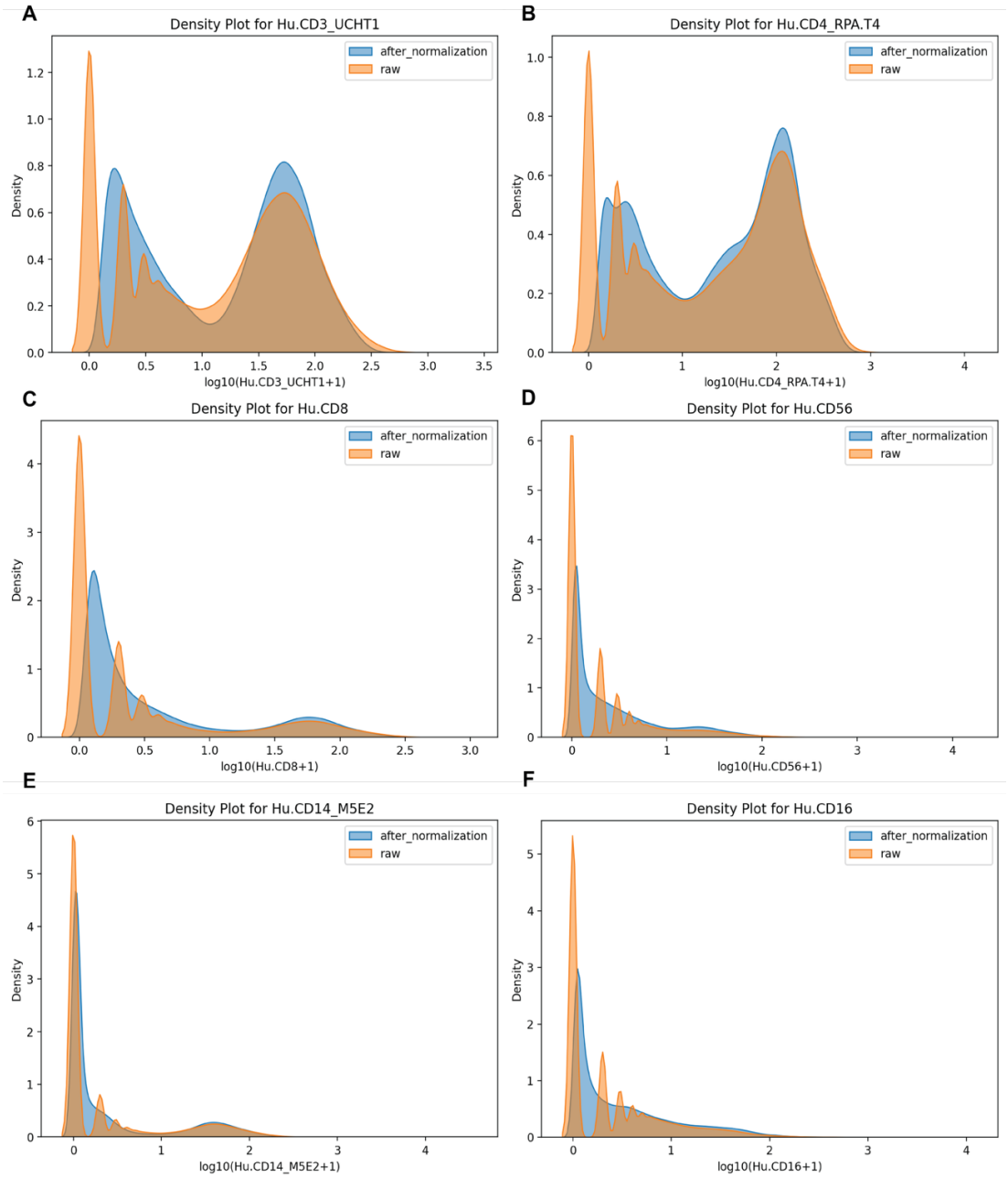


Figure 4.22 Density distributions of selected ADTs before and after totalVI normalization
Density plots showing the expression distributions of six selected ADTs. The x-axis represents log-transformed antibody signal before and after totalVI normalization, and the y-axis indicates cell density. Panels display: (A) CD3, (B) CD4, (C) CD8, (D) CD56, (E) CD14, and (F) CD16.

To further assess the effectiveness of totalVI normalization, I examined the distribution of selected ADTs across cell populations defined by the downstream annotation, focusing on groups that are biologically expected to represent positive

and negative expression states. As shown in **Figure 4.23**, ADT markers such as CD4 and CD8 display clear separation between $CD4^+$ and $CD8^+$ T cells, CD19 distinguishes B cells from non–B cells, and CD56 differentiates NK cells from other lymphocytes. These patterns demonstrate that totalVI-normalized protein signals align well with known immunophenotypic relationships, supporting the reliability of the integrated ADT data for downstream interpretation.

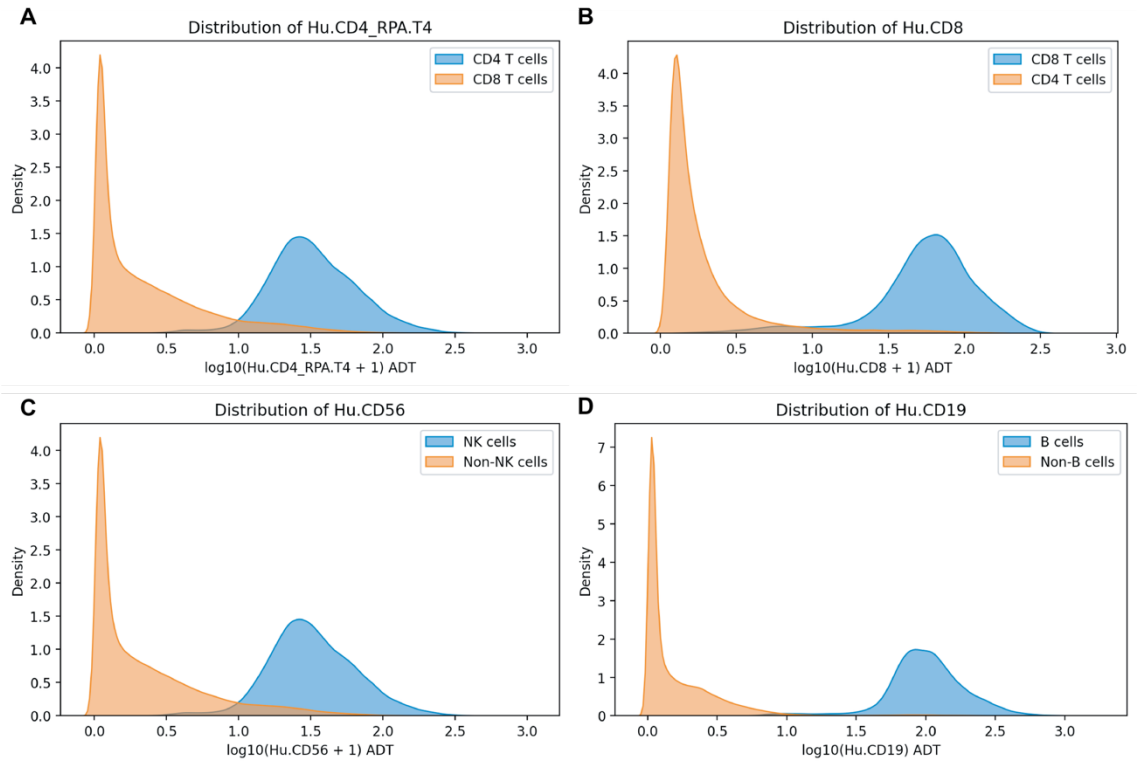


Figure 4.23 Distribution of selected ADTs across expected positive and negative cell populations
Density plots showing totalVI-normalised ADT expression for key lineage markers across biologically defined cell populations. Panels illustrate separation of: (A) CD4 expression in $CD4^+$ vs. $CD8^+$ T cells, (B) CD8 expression in $CD8^+$ vs. $CD4^+$ T cells, (C) CD56 expression in NK cells vs. non–NK cells, and (D) CD19 expression in B cells vs. non–B cells.

4.3.3 Cell-type annotation strategy and approach

4.3.3.1 Global annotation and clustering framework

In this section, I describe the strategy and approaches used to define the cell-type identities across the clusters. The annotation process combined automatic tools, manual inspection and iterative resolution tuning to ensure that clusters represented transcriptomically distinct and biologically meaningful cell populations.

During the initial stage of atlas construction, I used automatic annotation tools to provide a reference for assessing clustering results and tuning integration parameters. Preliminary annotations, generated with singleR (Aran et al., 2019) and CellTypist (Dominguez Conde et al., 2022), were manually cross-checked with canonical marker genes and guided early resolution tuning. For musculoskeletal tissues such as the spine enthesis and sacroiliac joint, I additionally evaluated annotations generated by Azimuth (Hao et al., 2021) using reference datasets from published bone marrow studies (Granja et al., 2019; Oetjen et al., 2018) and the Human Cell Atlas immune cell reference (Li et al.), to incorporate external expert annotations and assess potential bone marrow-derived populations. However, the transferred labels did not provide additional clarity for these tissues and were therefore not used for final annotation.

Building on the automatic annotations, I carried out detailed manual inspection of each cluster to ensure accurate and biologically meaningful labelling. This process involved identifying cluster-specific marker genes, assessing their specificity and sensitivity using CellxGene (Megill et al., 2021), and comparing these markers with reported signatures from published studies. Annotation decisions were further reviewed in consultation with immunologists in the research team. Although automated predictions served as a useful starting point, the final annotation labels were determined through this marker-based and literature-informed manual evaluation.

The determination of optimal clustering resolution was closely integrated with the annotation process. Rather than applying a single resolution across the atlas, I adopted a region-specific and iterative approach: a base resolution was selected for each compartment, and clusters were inspected for marker gene expression and functional distinctness. Depending on their biological characteristics, clusters were either subdivided or merged. To complement these qualitative assessments, I used ROGUE (Liu et al., 2020) to quantify cluster homogeneity and guide resolution refinement.

An example of the limitation of relying on a single Leiden clustering resolution is shown in **Figure 4.24**. In the $CD4^+$ T-cell region, a resolution of 0.7 resolved several substructures; however, clusters varied in internal homogeneity, which could be partly reflected by their ROGUE scores. At this resolution, cluster 6 displayed a low ROGUE value, indicating high entropy and potential heterogeneity. Examination of marker expression revealed that this cluster contained a mixture of $CD4^+$ and $CD8^+$ T cells, alongside a transcriptionally distinct $CXCL13^+$ subset consistent with helper T cells that are known to be functionally divergent. These observations indicate that further subdivision of this region is necessary to achieve appropriate biological resolution. However, the next Leiden resolution that successfully isolates the $CXCL13^+$ subset is 1.8, which yields more than 30 clusters overall and is therefore too granular for direct use as the final clustering.

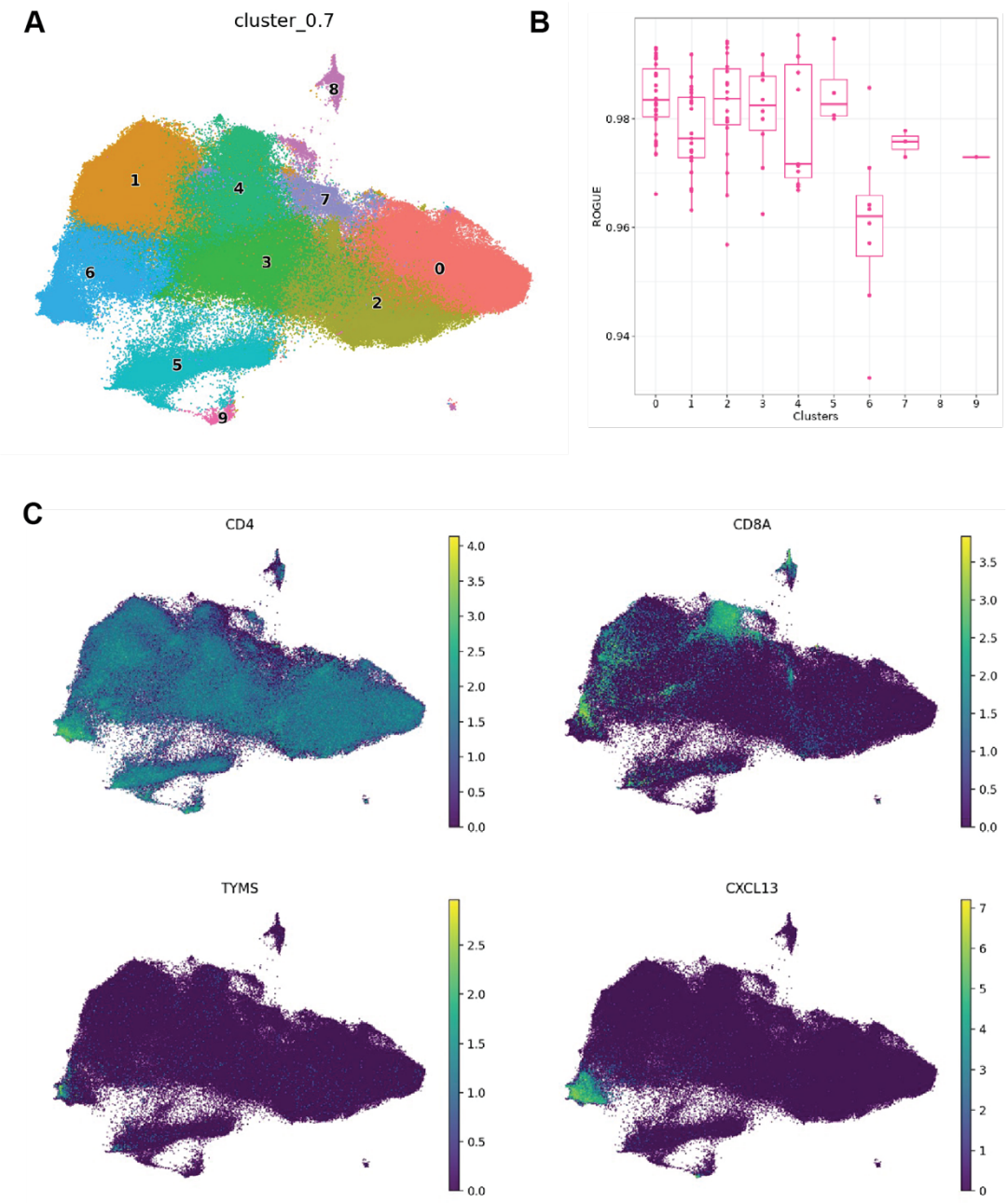


Figure 4.24 Evaluation of clusters in the CD4 T cell region
(A) Clustering of CD4 T cell region at resolution 0.7. (B) ROGUE scores for each cluster. (C) UMAP visualization showing expression of representative marker genes within the CD4 T cell region.

This example highlights the value of integrating information across multiple clustering resolutions rather than relying on a single global setting. The final clustering strategy, which incorporates intra-cluster heterogeneity, lineage structure (for example, CD4 and CD8 T cells), and transcriptomic similarity, is detailed in

Appendix C.4. Overall, cell-type definition was achieved through an integrative approach that combined computational predictions with marker-based manual review and iterative adjustment of clustering resolution. This iterative process ensured that the final atlas captured biologically meaningful populations while avoiding both over- and under-clustering. The currently described level of annotation represents the most granular level used for downstream analyses, although clusters were grouped into broader categories where appropriate for comparative studies. Additional annotation approaches, including the delineation of $CD4^+$ and $CD8^+$ T-cell subsets and the identification of MAIT and iNKT populations, are described in the following sections.

4.3.3.2 Classification of $CD4^+$ and $CD8^+$ T cells

To refine T-cell annotation at the functional lineage level, I classified each T-cell cluster according to their CD4 and CD8 expression profiles. Thresholds distinguishing positive and negative expression of these genes were determined by identifying local minima within the bimodal distributions of CD4 and CD8 expression levels. I then examined both the absolute proportion of CD4 or CD8 expressing cells and the CD4/CD8 cell number ratio within each cluster to confirm that the classification reflected true biological identity rather than technical dropout. This step helped to avoid misclassifying low-quality clusters with undetected transcripts as negative populations.

Based on these evaluations, clusters were assigned the prefixes CD4 or CD8 when they predominantly expressed the corresponding marker and showed clear separation from the opposite lineage (**Figure 4.25**).The prefix SP was used for clusters with intermediate CD8/CD4 expression ratios, in which neither marker alone defined the population. Manual inspection of expression patterns in UMAP confirmed

that cells expressing CD4 and those expressing CD8 formed distinct regions within the cluster, rather than representing double-positive T cells.

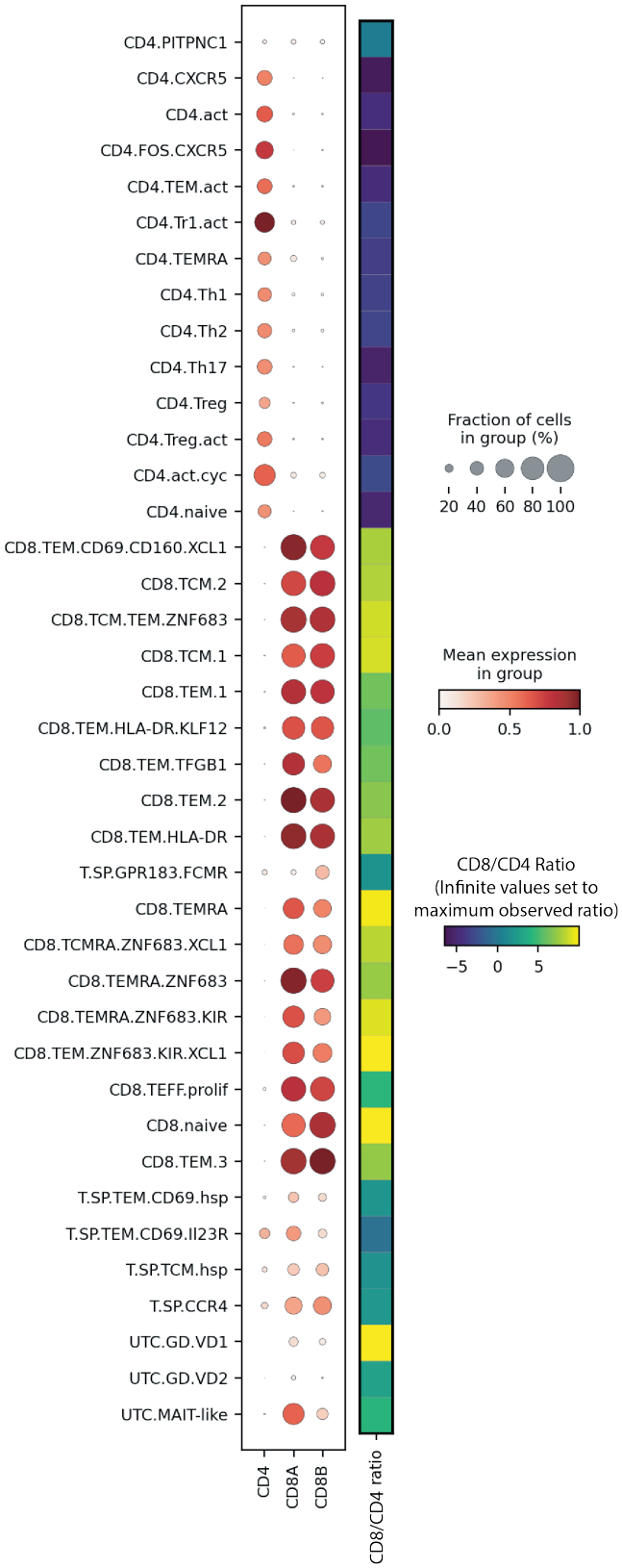


Figure 4.25 CD4 and CD8 expression patterns across T-cell clusters
Dot plot showing the expression of CD4, CD8A, and CD8B across T-cell clusters. The calculated CD8/CD4 cell ratio is shown on the right as a reference for cluster-level annotation. Infinite ratio values were capped at the maximum observed ratio for visualization.

Clusters labelled with the SP prefix were further examined to assess whether additional separation into $CD4^+$ and $CD8^+$ lineages was necessary. In the cycling T-cell cluster, both $CD4^+$ and $CD8^+$ subsets are known to coexist as distinct proliferating lineages, so I performed sub-clustering to separate them (**Figure 4.26**). In contrast, clusters such as T.SP.CCR4 remained homogeneous even at higher resolutions which indicated a unified phenotype, so this was therefore retained as single population and kept the SP prefix.

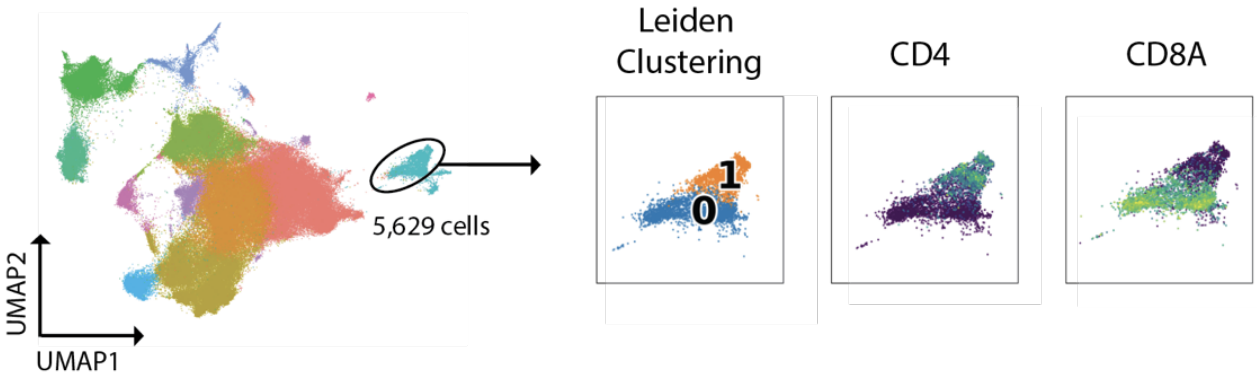


Figure 4.26 Sub-clustering of the cycling T cell cluster to distinguish $CD4^+$ and $CD8^+$ T cells
The cycling T cell cluster was extracted from the CD4 T cell region and re-clustered using the Leiden algorithm. Expression of CD4 and CD8A delineates $CD4^+$ and $CD8^+$ T cell subsets within this proliferative population. Cluster 1: CD4.act.cyc; Cluster 2: CD8.TEFF.prolif.

Overall, this classification approach established a consistent and biologically meaningful framework for distinguishing T-cell subsets based on CD4 and CD8 expression. The resulting categories - CD4, CD8, and SP clusters - provided a standardized and interpretable structure for T-cell annotation across tissues and supported subsequent analyses.

4.3.3.3 Identification of MAIT And iNKT Cells

MAIT and iNKT cells are semi-invariant T-cell subsets characterized by restricted TCR usage and distinct surface marker profiles. MAIT cells typically express an invariant TCR α-chain (TRAV1-2–TRAJ33/12/20) paired with limited β-chain diversity, recognizing microbial metabolites presented by MR1(Porcelli et al., 1993; Tilloy et al., 1999; Treiner et al., 2003). In contrast, iNKT cells express a canonical TCR α-chain (TRAV10–TRAJ18 in humans) that recognizes lipid antigens presented by CD1d (Borg et al., 2007; Kawano et al., 1997). Because their transcriptional signatures often overlap with those of conventional T cells, accurate identification requires combining information from multiple modalities. To identify MAIT and iNKT cells, I evaluated TCR chain information and ADT data, as these modalities capture TCR features and surface protein patterns, which are less distinguishable at the gene expression level. The TCR data, including preprocessing and chain annotation, were generated and processed by Dr Frank Penkava.

To evaluate the correspondence of MAIT-cell identification across modalities, I compared MAIT assignments derived from TCR and ADT information among the 369,420 cells for which both data types were available. I first examined the distribution of the Hu.TCR.VA7.2 ADT signal before and after normalisation (**Figure 4.27A**). Two approaches were then explored for defining MAIT cells using ADT modality: (i) classifying cells with detectable Hu.TCR.VA7.2 counts in the raw ADT matrix as positive, and (ii) using the normalised distribution to identify the negative peak and applying a manually selected threshold of normalised ADT value > 9 (**Figure 4.27B**).

If we treat TCR-identified MAIT cells as the reference definition for this comparison, then using the raw ADT approach, 5,007 of the 5,072 reference MAIT cells were recovered (sensitivity 0.99), and 5,007 of the 118,682 ADT-positive cells

corresponded to MAIT cells (specificity 0.69) (**Figure 4.27C**). The distribution of raw ADT counts for all cells and for TCR-defined MAIT cells is shown in **Figure 4.27D**. Using the normalised threshold-based approach, 3,945 of the 5,072 reference MAIT cells were recovered (sensitivity 0.78), and 3,945 of the 8,170 threshold-positive cells corresponded to MAIT cells (specificity 0.99) (**Figure 4.27E**). The distribution of normalised ADT values for all cells and for TCR-defined MAIT cells is shown in **Figure 4.27F**. In this section, “all cells” refers exclusively to cells with both ADT and TCR information.

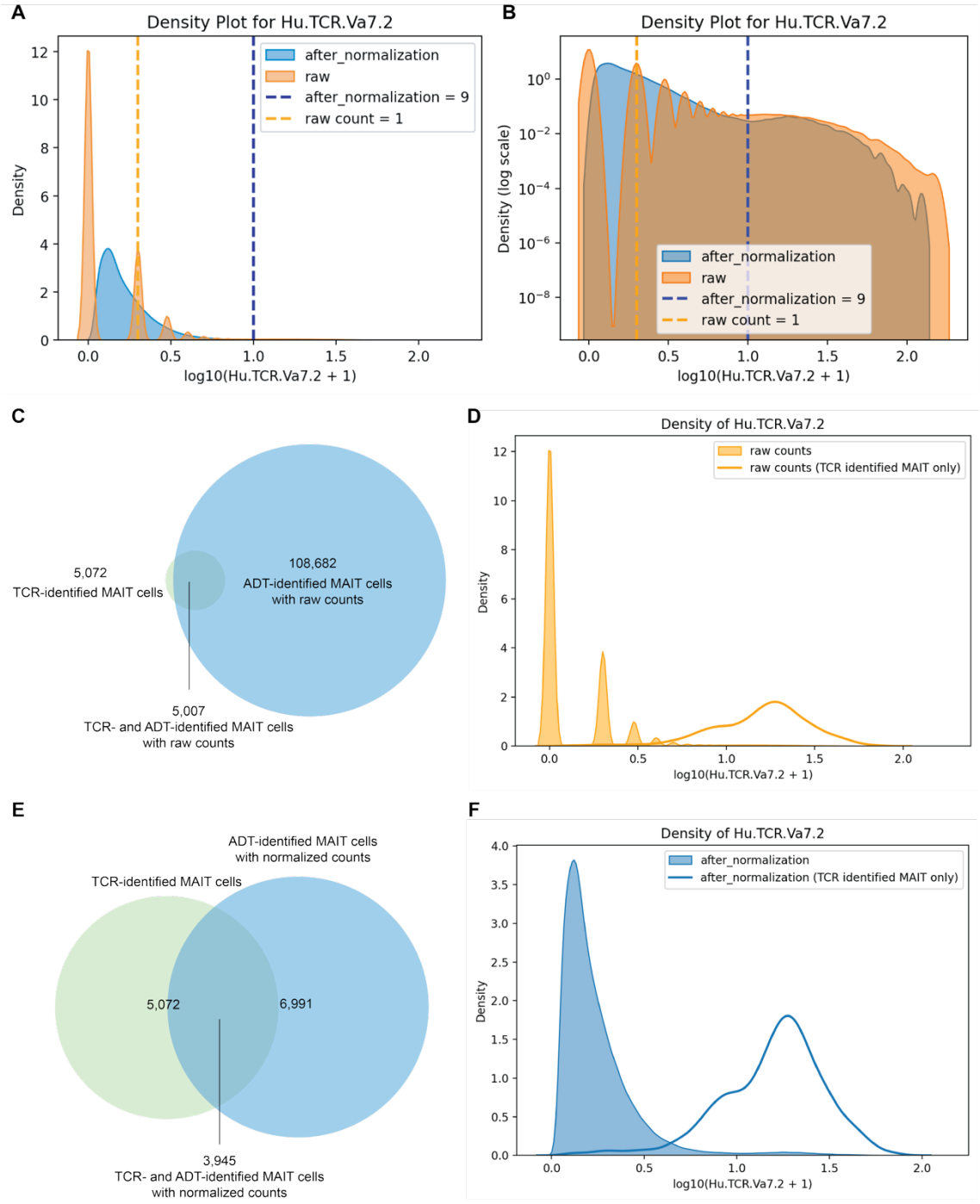


Figure 4.27 Comparison of TCR-based and ADT-based MAIT-cell identification in cells with both modalities

(A) Distribution of Hu.TCR.VA7.2 ADT signal before and after totalVI normalisation, with orange and blue lines indicating the raw ADT and normalised ADT thresholds, respectively. (B) The same Hu.TCR.VA7.2 distributions as in (A), shown on a log-density scale to visualize separation between negative and positive signal. (C) Overlap between TCR-identified MAIT cells and cells positive for Hu.TCR.VA7.2 in the raw ADT matrix. (D) Raw ADT counts for all cells and for TCR-defined MAIT cells. (E) Overlap between TCR-identified MAIT cells and cells exceeding the ADT value > 9 threshold in the normalized ADT matrix. (F) Normalized ADT values for all cells and for TCR-defined MAIT cells.

Taken together, these results show that ADT-based identification alone cannot achieve both high sensitivity and high specificity for MAIT cells: the raw ADT approach recovers most MAIT cells but includes many non-MAIT cells, whereas the threshold-based approach is more specific but omits a substantial fraction of true MAIT cells.

To evaluate whether the lower specificity reflected issues in totalVI normalisation, I compared ADT distributions across three TCR-defined groups: MAIT, iNKT, and non-MAIT αβT cells. In the raw ADT matrix, Hu.TCR.Vα7.2 expression showed broadly similar levels in MAIT and non-MAIT αβT cells (**Figure 4.28A**), supporting the need for normalisation. After totalVI normalisation, MAIT cells showed a clearer separation from non-MAIT cells (**Figure 4.28B**); however, an unexpected increase in Vα7.2 signal was also observed in iNKT cells and non-MAIT αβT populations, suggesting possible cross-modality imputation. This effect appeared restricted to Vα7.2, as other ADTs such as CD14 and CD16 remained stable and did not show similar artefacts (**Figure 4.28C-D**).

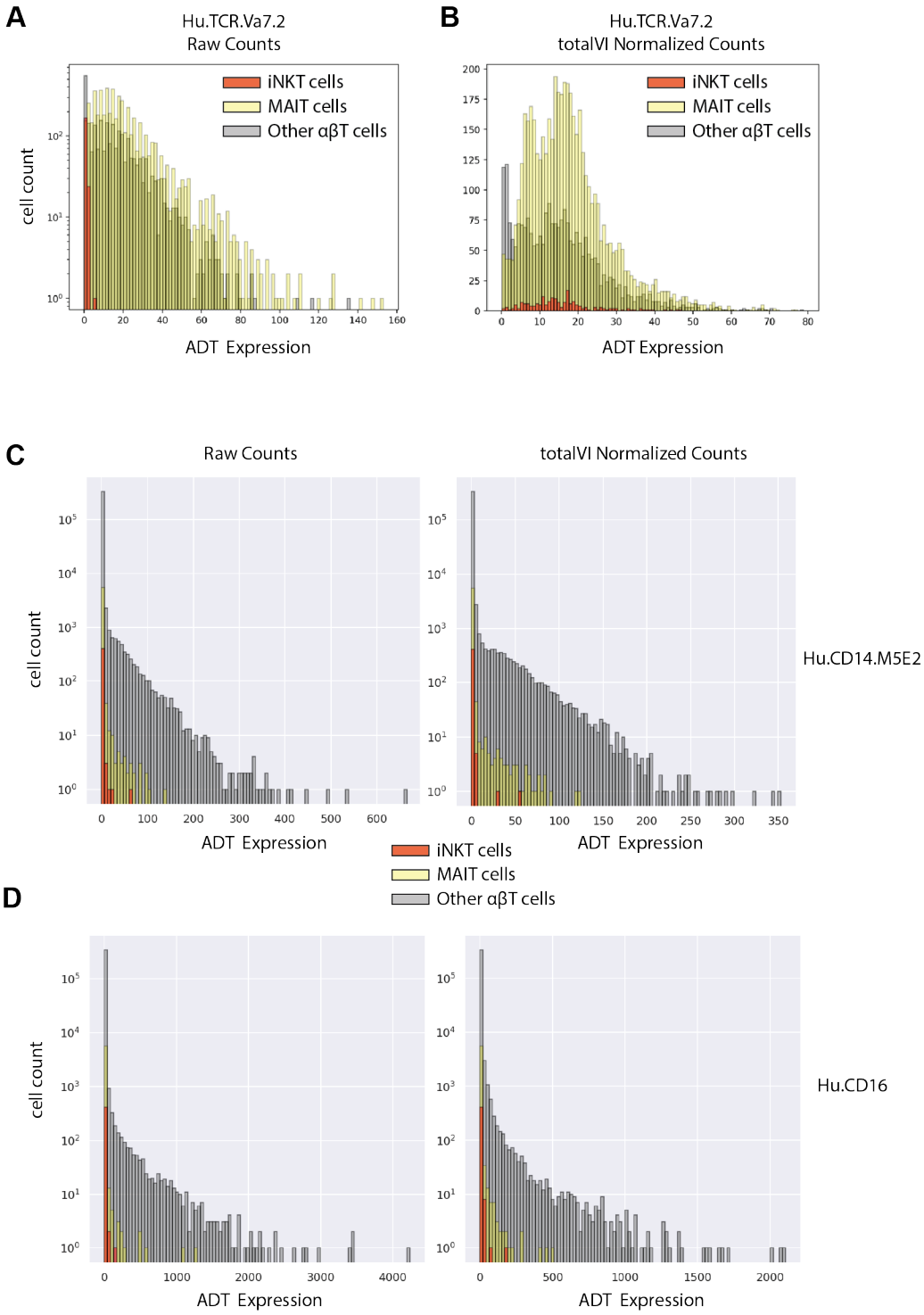


Figure 4.28 Comparison of MAIT and iNKT cell identification across modalities
(A-B) Distribution of raw and TotalVI-normalised Hu.TCR.Vα7.2 ADT counts across TCR-defined iNKT, MAIT, and non-MAIT αβ T cells. (C-D) Comparison of control markers Hu.CD14.M5E2 and Hu.CD16 before and after TotalVI normalization.

I next explored whether this type of imputation could be avoided by model tuning. Several modifications, including adjustments to missing-protein masking and

background removal, were tested but did not meaningfully reduce the issue. This limitation is likely due to the intrinsic similarity of their transcriptomic profiles; when GEX and ADT are jointly modeled and ADT values are normalized in a GEX-informed latent space, such artefacts cannot be fully avoided in our dataset (**Figure 4.29**).

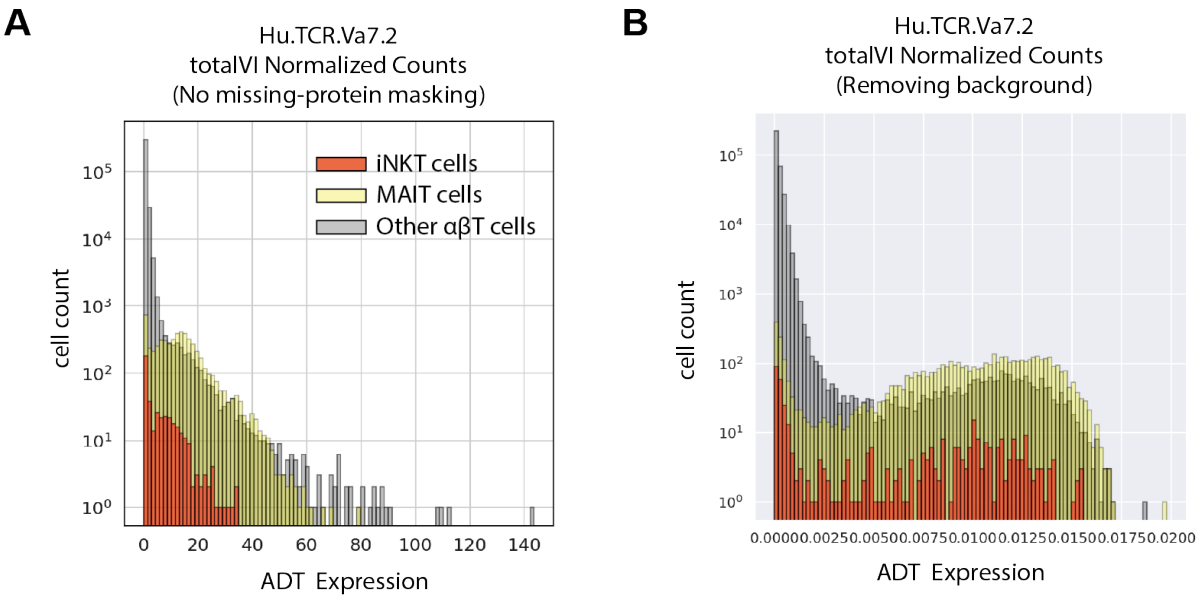


Figure 4.29 Alternative totalVI model normalization and background correction for Hu.TCR.Vα7.2 ADT counts
Models without missing-protein masking (left) and with background removal (right) are shown. Neither approach effectively reduced the artefactual VA7.2 signal, as MAIT, non-MAIT and iNKT cells were not clearly separated.

In summary, comparison of TCR-based and ADT-based MAIT-cell identification showed that neither raw nor totalVI-normalised ADT signals could achieve both high sensitivity and high specificity, and this limitation is likely driven by the intrinsic transcriptomic similarity between MAIT and non-MAIT αβT cells in this dataset. Although the ADT modality covers more cells than the TCR modality, TCR-based annotation was prioritized wherever available, as TCR features provide much more definitive evidence for MAIT and iNKT identity. Therefore cells that colocalized

with MAIT clusters in the GEX-based UMAP but lacked TCR information were designated as MAIT-like rather than MAIT cells, regardless of their ADT signal.

4.3.4 Cell-type annotation results and initial characterisation

In this section, I present the most granular level of cell annotation for the current atlas. As described above, clustering was performed independently within each region of the integrated atlas, where each region corresponds to a distinct area on the UMAP representing transcriptionally similar cells. For each compartment, I first show the UMAP of that region to illustrate how clusters were defined and positioned within the UMAP space. To complement this, I present dot plots of cell type groups, which are manually curated collections of related cell types across regions, to facilitate clearer biological interpretation and comparison. In the next chapter, these fine-grained clusters will be further combined to generate broader major and minor categories to support higher-level comparative analyses. The dot plots shown in this section were generated by my supervisor, Dr Stephen Sansom.

4.3.4.1 CD4-positive T cell compartment

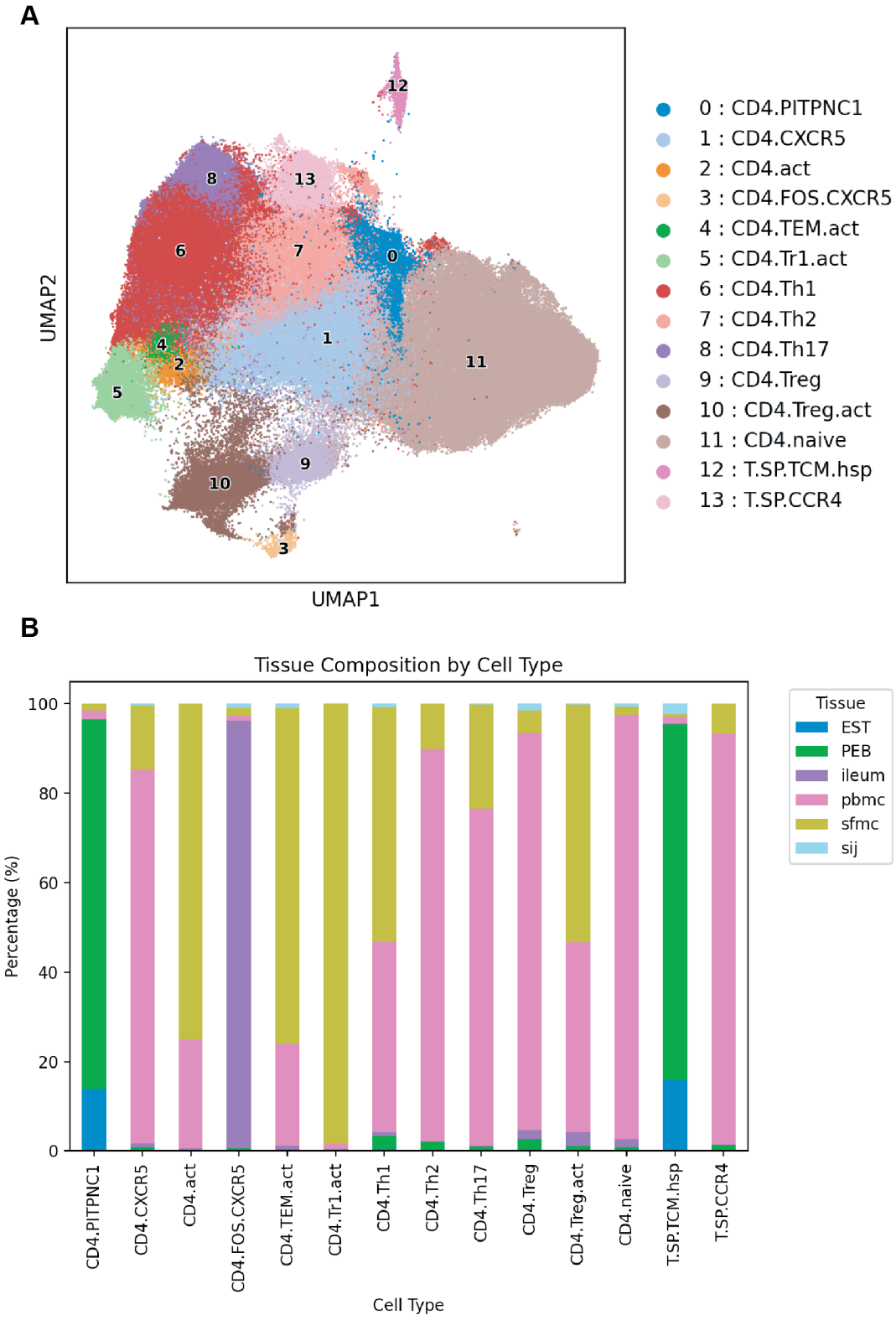


Figure 4.30 UMAP of the CD4-positive T cell region and tissue composition of identified clusters
(A) UMAP of the CD4-positive T cell region showing the final clustering and annotation. (B) Bar plot showing the tissue composition of each cluster, illustrating the relative contribution of cells from different tissues.

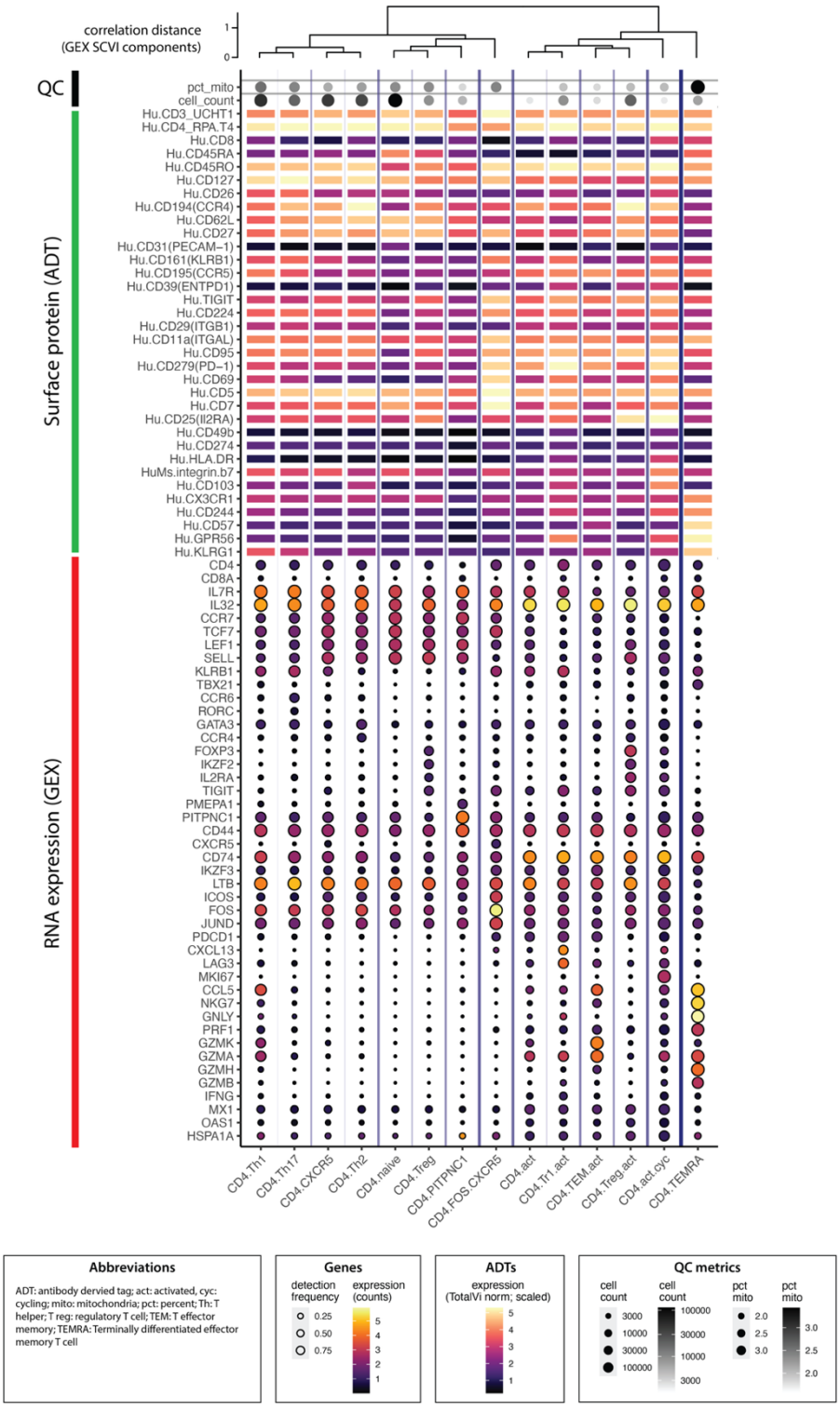


Figure 4.31 Dot plot summary of the CD4-positive T cell groups showing hierarchical relationships and multimodal marker expression
The dot plot displays quality control features, surface protein expression of key markers, and RNA expression of canonical genes, providing an integrated overview of each cluster's characteristics and marker profiles. Correlation distance hierarchy based on scVI components is shown above the heatmap.

Within the CD4-positive T-cell compartment, various transcriptionally and phenotypically distinct subsets were identified based on gene expression and ADT profiles (**Figure 4.30, Figure 4.31**). Naïve $CD4^+$ T cells (CD4.naive) expressed *CCR7* and *LEF1* and showed surface expression of CD45RA, consistent with a resting phenotype. Notably, *TCF7* was expressed in only a subset of this cluster, suggesting potential heterogeneity or transitional states within the naïve population. Effector subsets included Th1, Th2, and Th17 populations. The Th1 cluster *expressed CCL5, GZMK, GZMA*, and *TBX21 (T-bet)* together with ADT marker CD195 (CCR5), while Th17 cells were characterized by *KLRB1, CCR6,* and *RORC* expression along with a surface marker CD26. A Th2 subset was defined by *GATA3* and *CCR4* expression. A $CXCR5^+$ cell cluster (CD4.FOS.CXCR5) expressing *ICOS, JUND* and *TIGIT* likely represented circulating T follicular helper or related transitional populations.

Two regulatory T-cell clusters were distinguished: an activated Treg subset (CD4.Treg.act) expressing *FOXP3, TIGIT, IL2RA* and *CD74* with surface CD25 and CD39, consistent with an activated Treg phenotype; a less activated Treg subset (CD4.Treg) displayed high *FOXP3* and *IKZF2* expression together with stronger surface CD45RA signal, consistent with thymic-derived or resting naïve Treg precursors.

Several activated or effector-memory subsets were also observed. The CD4.TEMRA cluster expressed surface CD45RA and CD57, consistent with terminal or late-differentiated effector cells. The CD4.TEM.act population showed *GZMK, GZMA, CCL5* and *CD74* expression, indicating activated cytotoxic effector-memory characteristics. A distinct population expressing *PDCD1, LAG3, TIGIT,* and *CXCL13*, with corresponding ADT expression of CD39 (ENTPD1), TIGIT, HLA-DR, and

CD49b, was identified as Tr1-like cells (CD4.Tr1.act); this LAG3 and CD49b combination is consistent with IL-10-producing Tr1-like populations reported in both mouse and humans (Gagliani et al., 2013).

Finally, a cluster defined by *PITPNC1, PMEPA1*, and *CD44* expression (CD4.PITPNC1) likely represents an activated, adhesion-biased CD4 subset enriched for genes involved in cell adhesion and migration, and was predominantly derived from PEB and EST samples. Cycling $CD4^+$ T cells (CD4.act.cyc) were distinguished by elevated CD74 expression together with strong proliferation *and* cell cycle signatures. Together, these populations span a spectrum of $CD4^+$ T-cell states — from naive through effector, regulatory, to activated *and* cycling — thereby illustrating the transcriptional and functional diversity within this compartment.

4.3.4.2 CD8-positive T cell compartment

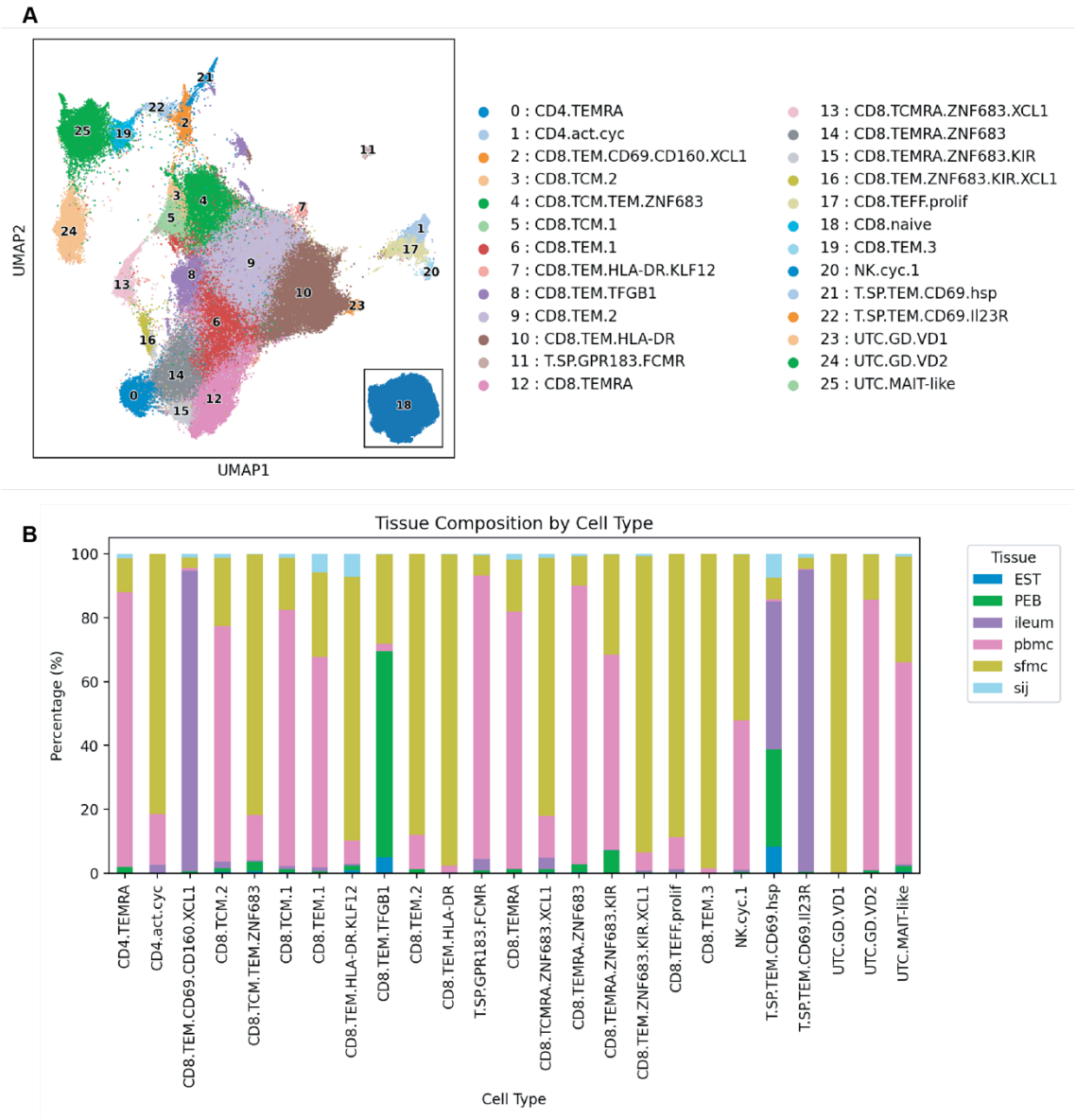


Figure 4.32 UMAP of the CD8-positive T cell region and tissue composition of identified clusters
(A) UMAP of the CD8-positive T cell region showing the final clustering and annotation. (B) Bar plot showing the tissue composition of each cluster, illustrating the relative contribution of cells from different tissues.

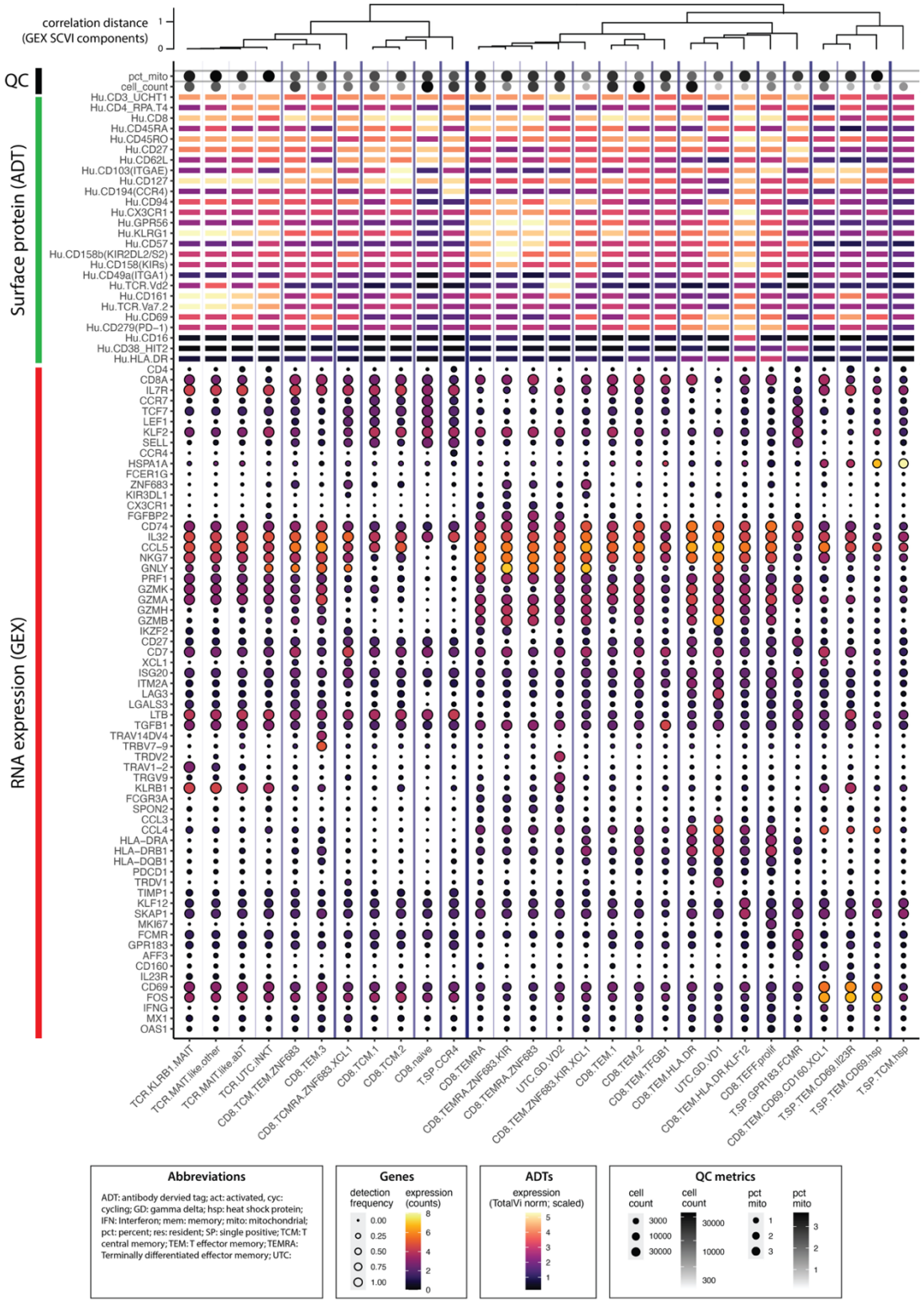


Figure 4.33 Dot plot summary of the CD8-positive T cell groups showing hierarchical relationships and multimodal marker expression
The dot plot displays quality control features, surface protein expression of key markers, and RNA expression of canonical genes, providing an integrated overview of each cluster's characteristics and marker profiles. Correlation distance hierarchy based on scVI components is shown above the heatmap.

Within the $CD8^+$ T-cell compartment, a spectrum of subsets encompassing naïve, memory, effector, and tissue-associated populations were identified (**Figure 4.29, Figure 4.30**). The naïve $CD8^+$ T cell (CD8.naive) population had already formed a distinct cluster in the overall T/NK subregion analysis, and expressed *CCR7, SELL, LEF1* and surface CD27, consistent with a resting phenotype. Two central memory–like populations (CD8.TCM.1, CD8.TCM.2) exhibited *KLF2, SELL* and *CCR7,* consistent with a central memory state. Another cluster, CD8.TCM.TEM.ZNF683, expressed *ZNF683* and *IL7R* with relatively modest *CCR7* and cytotoxic gene expression, representing an intermediate or a mixture of TCM and TEM phenotype cells.

Within the effector-memory compartment, several transcriptionally distinct subsets were identified. CD8.TEM.1 showed high *KLF2* and low *GZMB* and *GNLY* expression. CD8.TEM.2 expressed higher *HLA-DRA, HLA-DRB1, GZMB* and *GZMH*. Both CD8.TEM.HLA-DR and CD8.TEM.HLA-DR.KLF12 were predominantly derived from synovial fluid and showed high *ITM2A* and class II HLA expression (*HLA-DRA, HLA-DRB1, HLA-DQB1*); the latter additionally exhibited elevated *KLF12* and *SKAP1* expression with relatively higher proportion of cells derived from sacroiliac joint samples. A small population enriched in AxSpA synovial fluid (CD8.TEM.ZNF683.KIR.XCL1) expressed *ZNF683*, *XCL1, CD74*, and *LGALS3* and showed higher KIR expression compared other TEM cells, suggesting a tissue-resident, and activated effector phenotype associated with local inflammation.

A distinct cluster, CD8.TEM.3, largely derived from synovial fluid, showed a unique transcriptional profile while retaining effector-memory features. This cluster was marked by *TRAV14DV4 (TRAV1-2)* and *TRBV7-9* expression and exhibited higher *GNLY* levels, indicating part of the cells in this cluster had semi-invariant TCR

usage and a MAIT-like effector-memory phenotype considering the transcriptional features.

$CD8^+$ TEM clusters generally showed higher *GZMK* expression relative to other CD8 T subsets. Notably, recent work has shown that GZMK has functions beyond cytotoxicity, including the capacity to activate the complement cascade in inflamed tissues (Donado et al., 2025). Given the widespread presence of GZMK-expressing $CD8^+$ T cells in chronic inflammatory environments, the enrichment of *GZMK* in these effector memory T cell populations may reflect an additional pro-inflammatory role within AxSpA tissues, potentially contributing to local complement activation and amplifying tissue inflammation.

Terminally differentiated effector cells (CD8.TEMRA and related clusters) expressed cytotoxic effector molecules *GZMA, GZMB, GZMH*, *GNLY* and surface proteins CD45RA, GPR56, KLRG1 and CD57. A KIR-positive subset (CD8.TEMRA.ZNF683.KIR) with high CD158b (KIR2DL2/S2) ADT signal and *ZNF683* expression represented highly differentiated TEMRA cells with strong cytotoxic potential. Additional small clusters, including CD8.TEM.CD69.CD160.XCL1 and CD8.TEM.CD69.IL23R, both expressed *CD69* and were primarily derived from ileum and PEB samples, indicating tissue-specific adaptation.

A population mostly from ileum and PEB with high *HSPA1A* expression (T.SP.TCM.hsp) represented heat shock–associated memory-like cells, showing similar cytotoxic profiles to other TCM populations. Another subset, T.SP.CCR4, expressed *IL7R* and *LTB* but lacked *CCR7*, indicating a non-naïve, $CCR4^+$ resting-memory phenotype. Two γδ T-cell populations were also identified, expressing *TRDV2/TRGV9* (UTC.GD.VD2) and *TRDV1/LAG3* (UTC.GD.VD1), corresponding to $V\delta2^+$ and $V\delta1^+$ γδ T cells, respectively.

Together, these T cell subsets in the CD8 T cell region span the spectrum from naïve and central memory to terminally differentiated effector and tissue-enriched cells, capturing the transcriptional and functional diversity of cytotoxic T-cell states across tissues.

4.3.4.3 NK cell compartment

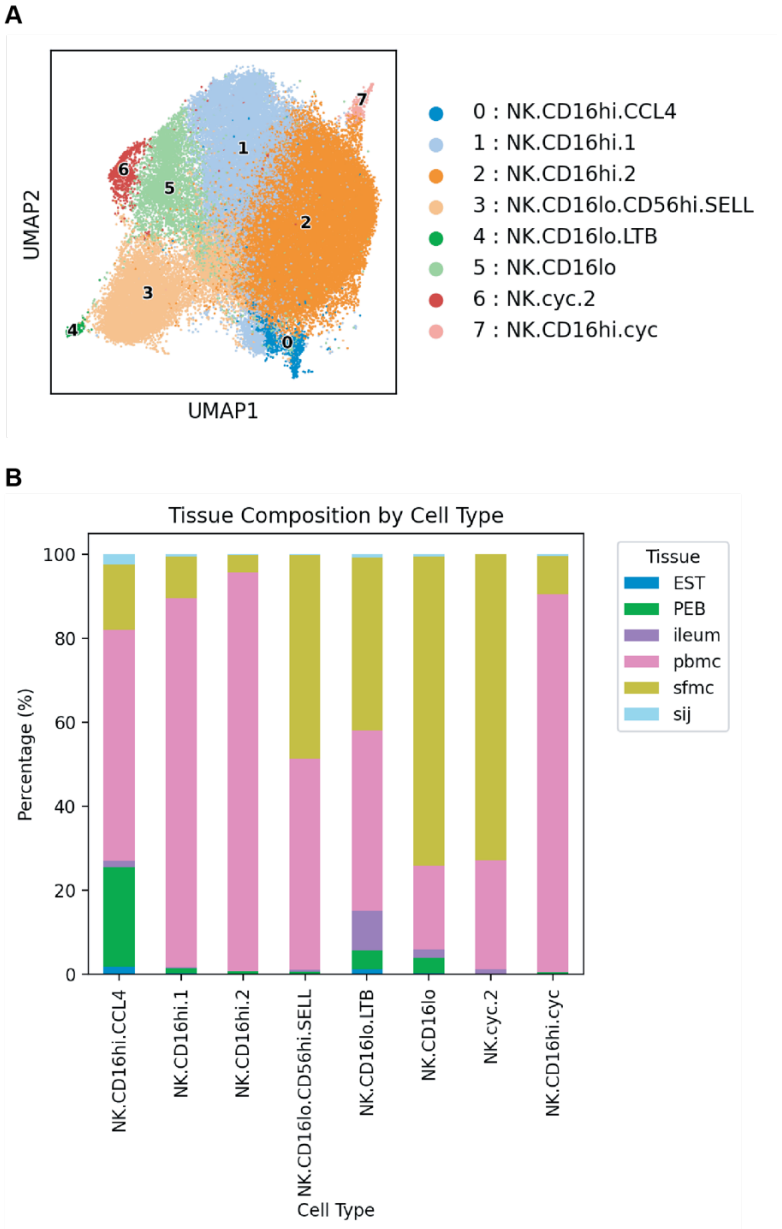


Figure 4.34 UMAP of the NK cell region and tissue composition of identified clusters (A) UMAP of the NK cell region showing the final clustering and annotation. (B) Bar plot showing the tissue composition of each cluster, illustrating the relative contribution of cells from different tissues.

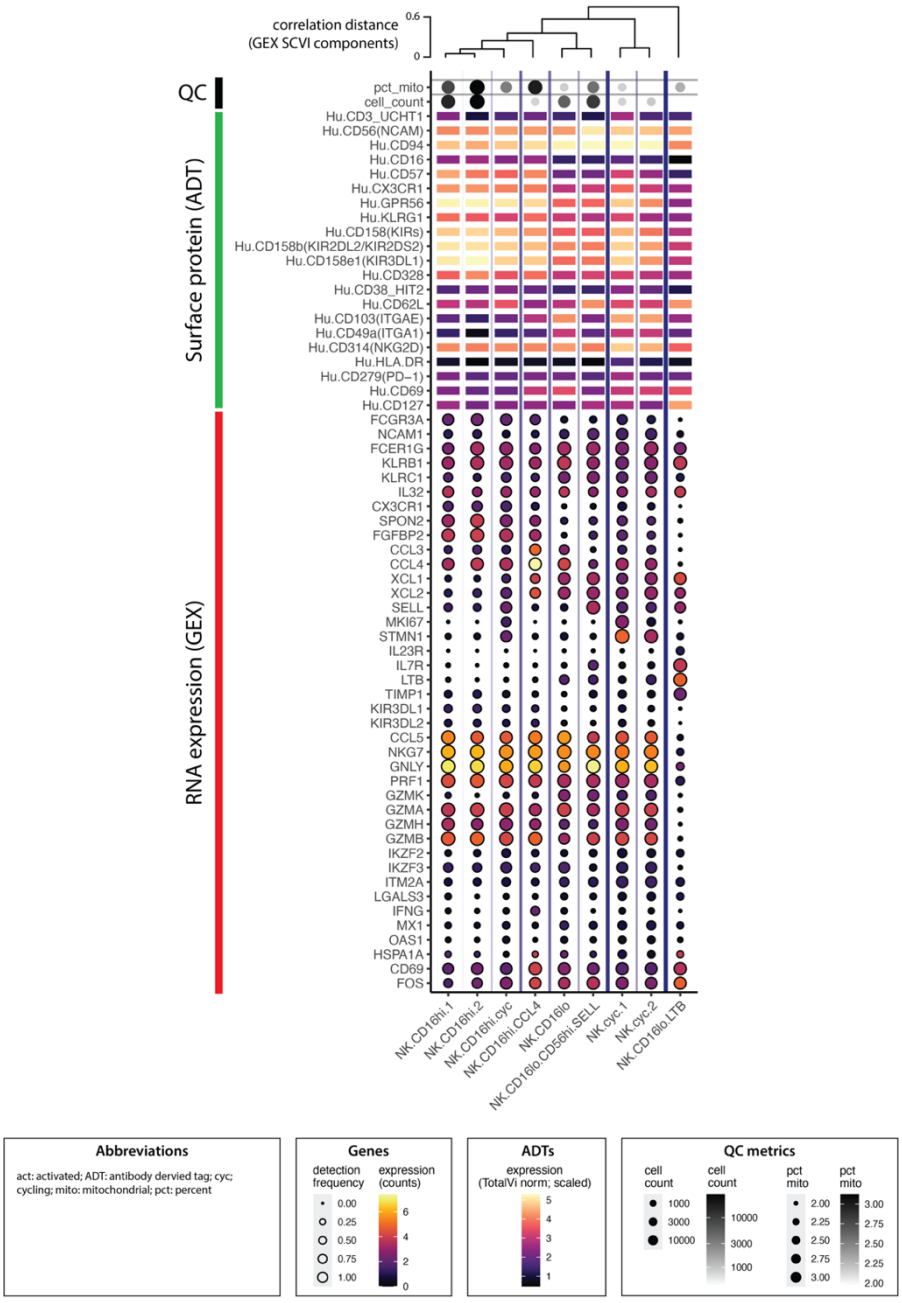


Figure 4.35 Dot plot summary of the NK cell groups showing hierarchical relationships and multimodal marker expression
The dot plot displays quality control features, surface protein expression of key markers, and RNA expression of canonical genes, providing an integrated overview of each cluster's characteristics and marker profiles. Correlation distance hierarchy based on scVI components is shown above the heatmap.

Within the NK-cell compartment, distinct CD16^{hi} and CD16^{lo} subsets were identified, supported by both *FCGR3A* gene expression and ADT-derived CD16

protein levels, reflecting canonical cytotoxic and tissue-associated NK states (**Figure 4.34, Figure 4.35**). CD16^{hi} populations showed higher *FGFBP2* expression and strong ADT signals for KIR receptors, including CD158 (KIRs), CD158b (KIR2DL2/KIR2DS2), and CD158e1 (KIR3DL1).

A distinct population NK.CD16hi.CCL4 which was characterized by high *CCL3, CCL4, XCL1/2,* was the most enriched NK cell type in PEB and in sacroiliac joint. NK cells are known producers of CCL3 and CCL4 in the activated condition (Walzer et al., 2005), and studies in mice show that CCL3 can influence both NK-cell trafficking within the bone marrow (Bernardini et al., 2008) and stromal differentiation in bone-associated niches (Yu et al., 2023). These observations suggest that this chemokine-producing NK population may represent a locally activated, bone-marrow-conditioned phenotype that is potentially associated with AxSpA-affected tissue environments.

The NK.CD16lo.CD56hi.SELL cluster expressed *SELL, XCL1*, and *IL7R,* together with surface CD56, CD94, and CD62L, and showed high *GZMK* but low *GZMH* expression, consistent with a CD56^{bright}, cytokine-producing NK subset with lymphoid-homing and immune cross-talk potential rather than strong cytotoxic activity.

Three cycling NK clusters were identified. NK.cyc.1, which clustered together with CD4^{+} and CD8^{+} T-cell cycling populations in the T-cell region, showed higher *STMN1, MKI67, TUBB* and *PCNA* expression, suggesting strong proliferative signals that may have overridden cell type specific features, leading to its initial grouping with T cells. The other two clusters, NK.cyc.2 and NK.CD16hi.cyc, displayed relatively lower overall cycling activity. When comparing these two, NK.cyc.2

expressed higher *STMN1*, whereas NK.CD16hi.cyc showed higher *MKI67* and *FCGR3A,* indicating differences in both NK-cell phenotype and proliferative states.

Altogether, these clusters define a continuum from resting cytotoxic CD16^{hi} NK cells to adaptive-like and activated CD56^{bright} subsets, exhibiting increasing cytokine and homing specialization.

4.3.4.4 MNP cell compartment

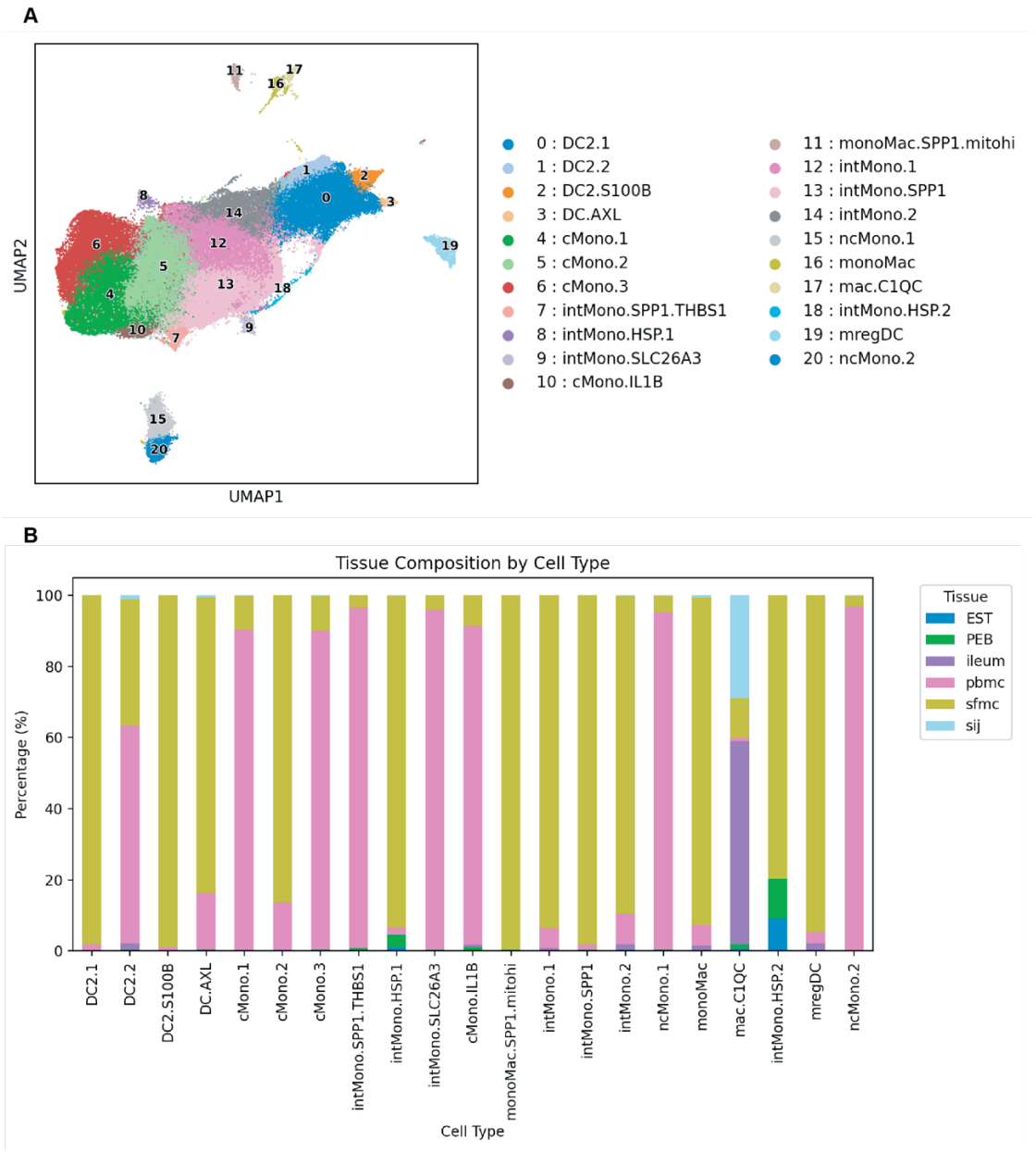


Figure 4.36 UMAP of the MNP cell region and tissue composition of identified clusters

(A) UMAP of the MNP cell region showing the final clustering and annotation. (B) Bar plot showing the tissue composition of each cluster, illustrating the relative contribution of cells from different tissues.

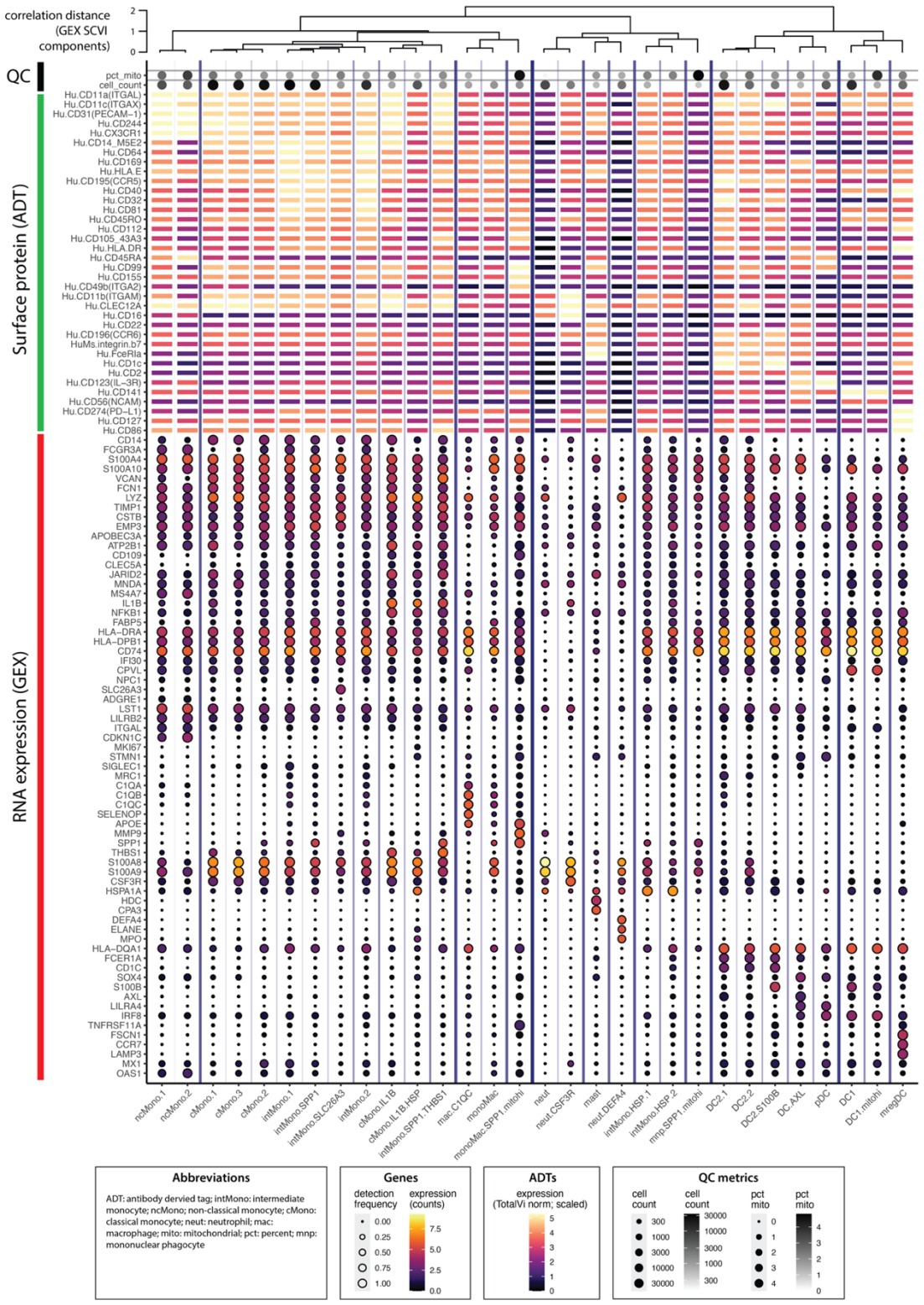


Figure 4.37 Dot plot summary of the MNP cell groups showing hierarchical relationships and multimodal marker expression
The dot plot displays quality control features, surface protein expression of key markers, and RNA expression of canonical genes, providing an integrated overview of each cluster's characteristics and marker profiles. Correlation distance hierarchy based on scVI components is shown above the heatmap.

Within the MNP compartment, a range of monocyte, macrophage, and dendritic cell subsets was identified, reflecting both circulating and tissue-resident states (**Figure 4.36, Figure 4.37**). Classical (cMono), intermediate (intMono), and non-classical (ncMono) monocytes were defined by CD14/CD16 signatures at both gene and protein levels. cMono in our atlas were divided into five clusters, with cMono.2 predominantly derived from synovial tissue. Two clusters expressed IL1B and NFKB1, consistent with a conventional inflammatory monocyte phenotype, likely reflecting activation of the IL1B transcriptional pathway via NF-κB. intMono also comprised five clusters, including an $SPP1^+$ subset enriched in synovial fluid and a $THBS1^+$ cluster primarily found in peripheral blood. ncMono expressed *LST1, CDKN1C*, and *MS4A7*, with surface markers CD11a, CD244, and CX3CR1, consistent with a patrolling phenotype.

Several macrophage-like populations were observed. mac.C1QC expressed APOE, SELENOP, and *C1Q* family genes as marker genes. $C1Q^+$ $APOE^+$ macrophages have been reported in various healthy *and* disease contexts as tissue-resident macrophages *(Horowitz et al., 2024; Revel et al., 2022)*, and in our dataset this subset was predominantly derived from sacroiliac joint and ileum samples. Another macrophage-associated marker, *SPP1,* defined the monoMac.SPP1.mitohi cluster, which expressed *SPP1, MMP9*, and *APOE* and was enriched in synovial tissue, suggesting a distinct tissue-specific macrophage phenotype.

Several dendritic cell populations were identified, consistent with the growing appreciation of DC heterogeneity revealed by recent single-cell studies (Villar & Segura, 2020). cDC1 expressed *IRF8* and HLA-DPB1, with surface CD141 detected by ADT, consistent with classical type 1 dendritic cells. cDC2 expressed *CD1C* and *FCER1A*, with corresponding CD1c surface expression. The DC2.S100B subset expressed *S100B* and *ITGB7*, representing a tissue-associated DC2 population. AXL^+ DCs expressed *AXL* and *SIGLEC6*, consistent with the recently described “AS” DC subset in the single-cell studies (Villani et al., 2017). Plasmacytoid DCs (pDCs) expressed *LILRA4* and *IRF7*, while mature regulatory DCs (mregDC) were characterized by *LAMP3, CCR7, FSCN1* gene expression and surface CD274, CD86, CD127, and HLA-DR, indicating an activated, migratory, and antigen-presenting phenotype.

Together, these results reveal MNP heterogeneity spanning classical, intermediate, and non-classical monocytes, tissue macrophages, and diverse DC populations, reflecting both circulating immune dynamics and tissue-specific adaptations in the integrated multi-tissue atlas.

4.3.4.5 B cell compartment

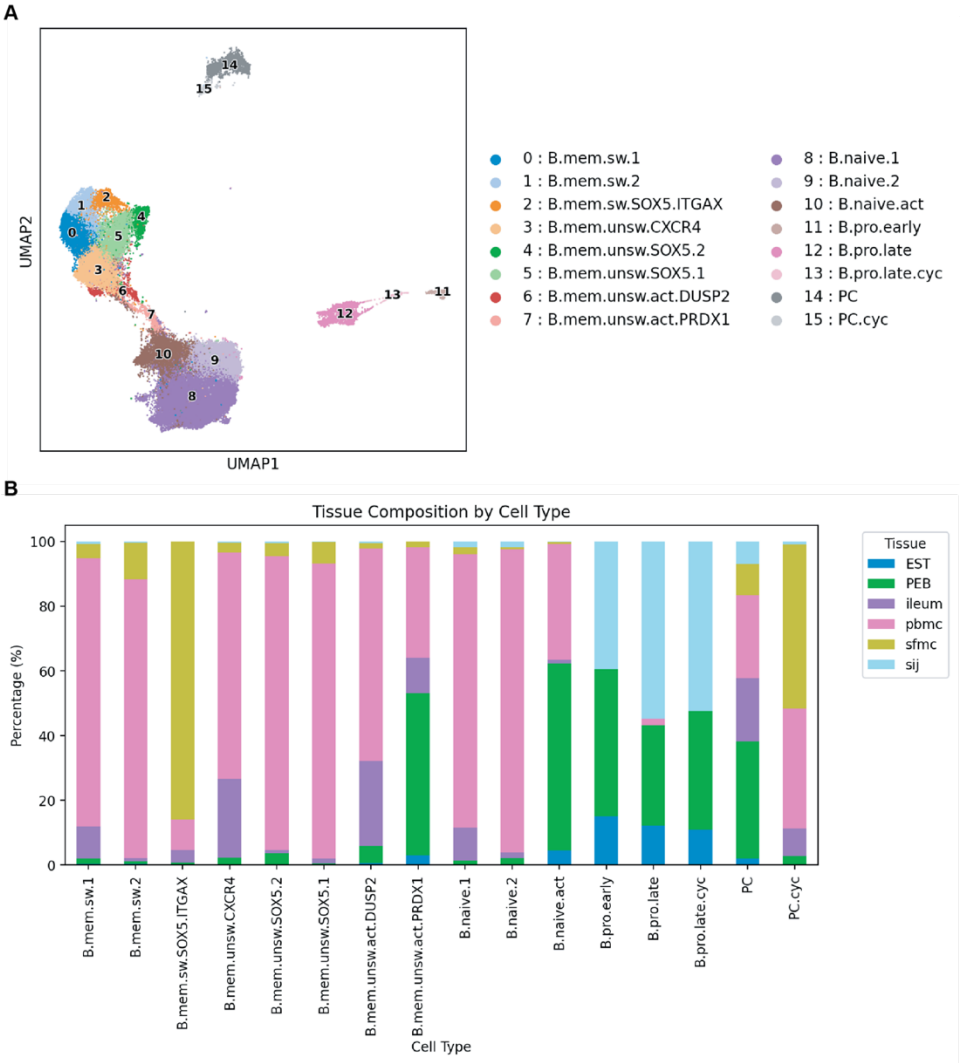


Figure 4.38 UMAP of the B cell region and tissue composition of identified clusters
(A) UMAP of the B cell region showing the final clustering and annotation. (B) Bar plot showing the tissue composition of each cluster, illustrating the relative contribution of cells from different tissues.

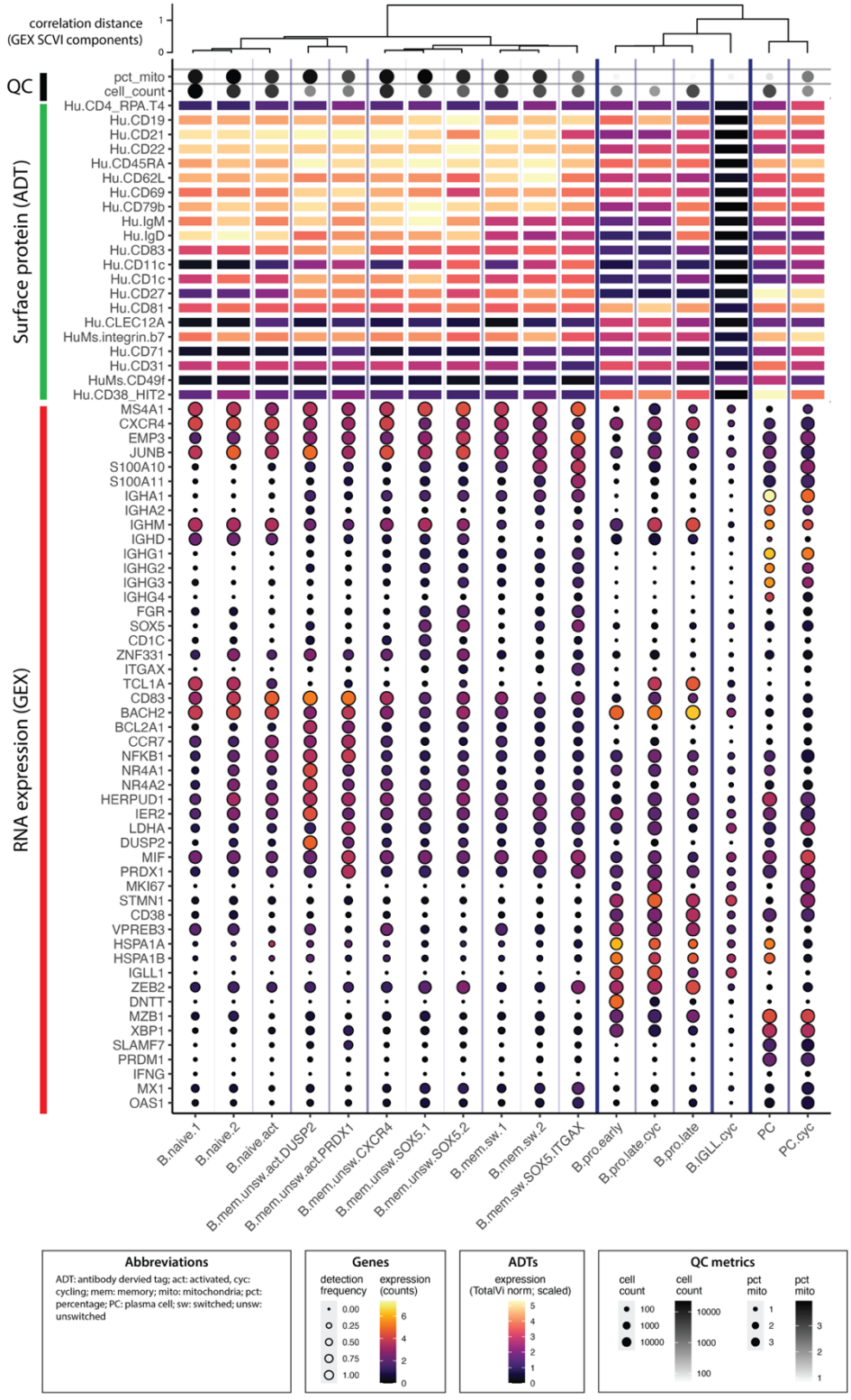


Figure 4.39 Dot plot summary of the B cell groups showing hierarchical relationships and multimodal marker expression
The dot plot displays quality control features, surface protein expression of key markers, and RNA expression of canonical genes, providing an integrated overview of each cluster's characteristics and marker profiles. Correlation distance hierarchy based on scVI components is shown above the heatmap.

The B-cell compartment displayed transcriptional and phenotypic diversity spanning naïve, unswitched memory, switched memory, and plasma states (**Figure 4.38, Figure 4.39**).

Two naïve clusters (B.naive.1, B.naive.2) expressed *TCL1A* and additional genes observed in early transitional cells. Progenitor-like clusters (B.pro.early, B.pro.late) were only observed in spine and sacroiliac joint samples and expressed *DNTT* (in B.pro.early), *CD38,* and *BACH2*. Given that terminal deoxynucleotidyl transferase (TdT) coded by *DNTT* is a canonical marker of immature B-lineage cells in bone marrow (Li et al., 1993), this pattern could reflect either local marrow-proximal progenitors or tissue sampling of marrow elements.

Memory B-cell populations were distinguished by isotype switching and activation state. Unswitched memory clusters (B.mem.unsw) expressed *IGHM* and *IGHD* with little or no expression of *IGHG, IGHA,* or *IGHE*, whereas switched memory clusters (B.mem.sw) showed the opposite pattern. An activated unswitched subset (B.mem.unsw.act.PRDX1) expressed *IGHM/IGHD* and low levels of *IGHA*, together with *PRDX1, LDHA, CCR7*, and *CD83*, *which could indicate* an early activation or transitional switching state. An activated switched-memory subset (B.mem.sw.SOX5.ITGAX) expressed *ITGAX* (CD11c) and *EMP3*, consistent with previously described $CD11c^{+}$ age-associated B cells (ABCs) frequently reported in autoimmune and inflammatory conditions (Cancro, 2020; Rubtsov et al., 2011; Xie et al., 2025). Plasma clusters expressed *XBP1* and *MZB1* with surface CD38, and rare cycling plasma/progenitor populations (PC.cyc, B.pro.late.cyc, B.IGLL.cyc) expressed *MKI67* and *STMN1.*

Overall, the B-cell compartment spans early developmental to terminally differentiated states, encompassing activated and memory subsets with evidence of distinct, tissue- and context-specific phenotypes.

4.3.4.6 Other cell compartment

All remaining cells not included in the previous different compartment are grouped together as “other” cell compartment and were analyzed together.

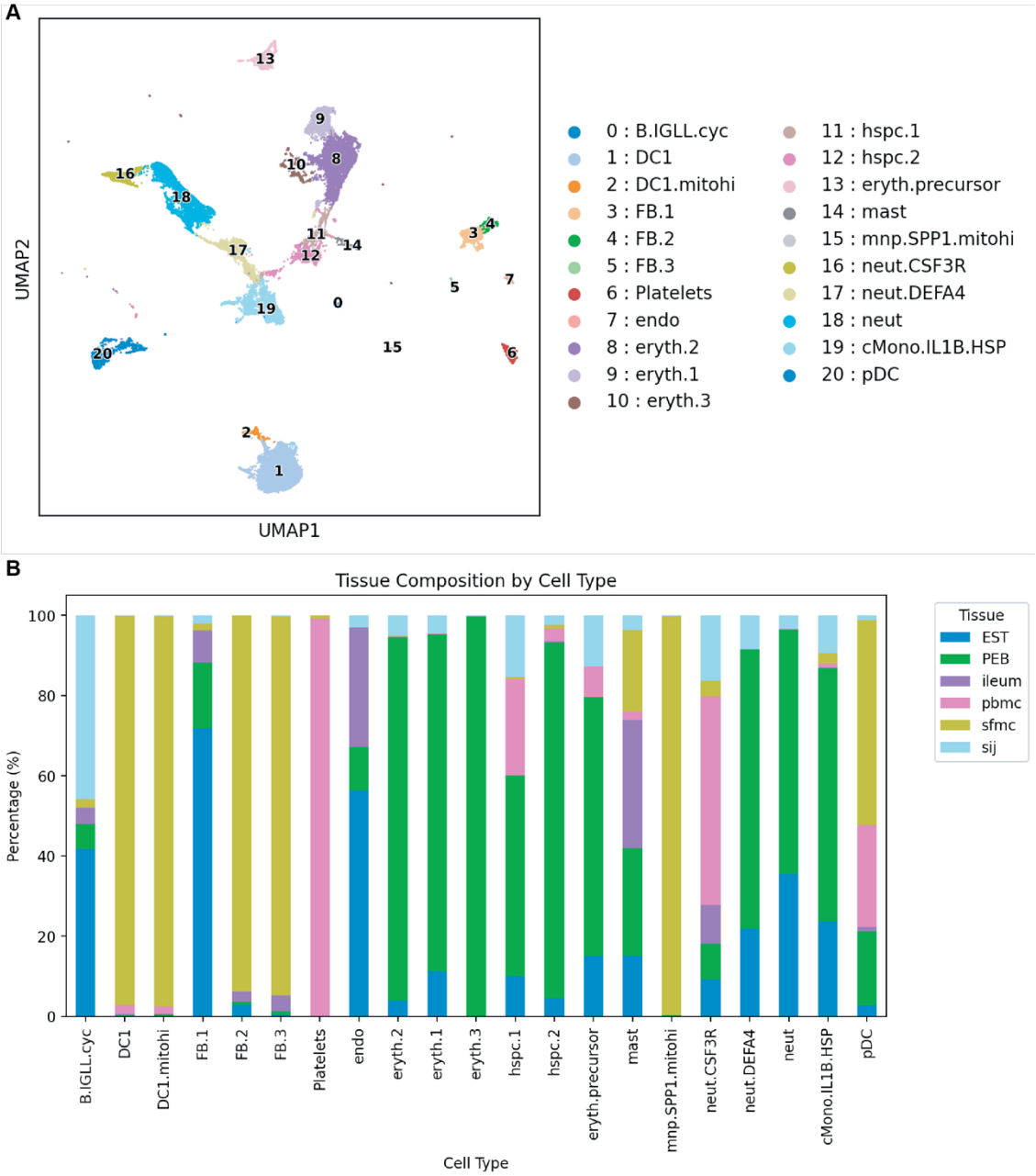


Figure 4.40 UMAP of the other cell region and tissue composition of identified clusters
(A) UMAP of the other cell region showing the final clustering and annotation. (B) Bar plot showing the tissue composition of each cluster, illustrating the relative contribution of cells from different tissues.

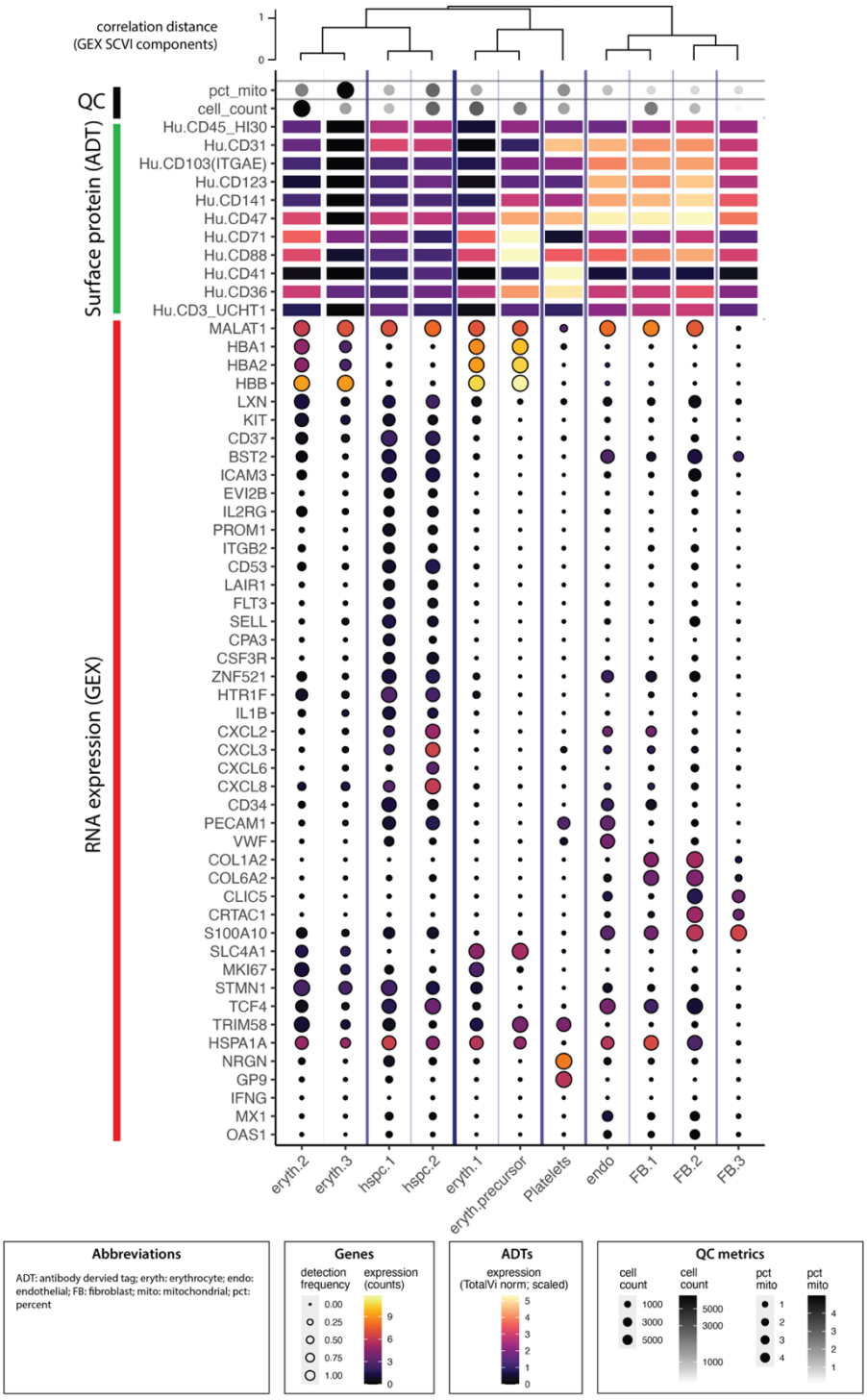


Figure 4.41 Dot plot summary of the NK cell groups showing hierarchical relationships and multimodal marker expression

The dot plot displays quality control features, surface protein expression of key markers, and RNA expression of canonical genes, providing an integrated overview of each cluster's characteristics and marker profiles. Correlation distance hierarchy based on scVI components is shown above the heatmap.

The "other" cell compartment contained a range of non-immune and progenitor-derived populations, including erythroid, stromal, and endothelial lineages (**Figure 4.40, Figure 4.41**). Erythroid populations were identified at multiple developmental stages. Hematopoietic stem and progenitor cells (hspc.1, hspc.2) expressed *CD34*, *TCF4*, and *PECAM1*, consistent with hematopoietic progenitor programs, whereas intermediate and mature erythroid clusters (eryth.precursor, eryth.1, eryth.2, eryth.3) expressed *HBB* and *SLC4A1* with surface CD71, reflecting progressive maturation and hemoglobin synthesis.

Three neutrophil clusters were identified including a general neutrophil cluster (neut) expressing *PGLYRP1,* a degranulating subset (neut.DEFA4) expressing granule-associated genes such as *MPO* and *ELANE*, and a mature/activated cluster (neut.CSF3R) characterized by high *CSF3R* expression and surface markers CD16, CD11b, and CLEC12A, representing mature, activated neutrophil states involved in antimicrobial responses and inflammation.

Stromal and structural cell types were also detected. Fibroblast-like clusters (FB.1, FB.2, FB.3) expressed *COL1A2*, *COL6A2*, and *CRTAC1*, with varying expression intensity consistent with tissue heterogeneity. Endothelial cells expressed *VWF*, consistent with vascular endothelium, while mast cells expressed *HDC* and *CPA3*. A platelet cluster expressed *NRGN* and *GP9*, consistent with a megakaryocyte-derived origin.

Together, these cell populations represent a diverse set of structural and progenitor lineages captured within the integrated atlas. Although not the primary

focus of immune annotation, they provide important context for interpreting the cellular composition of tissue microenvironments across the dataset, particularly as different tissues contribute distinct proportions of stromal and structural cells.

4.4 Discussion

This chapter describes the construction and annotation of a multi-modal, multi-tissue single-cell atlas, integrating gene expression, surface protein, and TCR sequencing data. It represents the first attempt in AxSpA to integrate single-cell data across tissues including peripheral blood, synovial fluid, spine, sacroiliac joint, gut and eye, covering both immune and structural cell populations. The aim was to generate a unified framework that allows comparison across tissues and disease conditions while maintaining sufficient resolution to capture cell-type specific variation.

One of the major challenges of this work was avoiding the introduction of technical bias during data integration. Each tissue and experiment inevitably introduced batch effects arising from sequencing, processing, or sample handling, particularly noting that the work involved collaboration across multiple sites. The key question was how to remove these technical effects while preserving meaningful biological differences. This required a clear definition of the aim at each integration step: whether to prioritize clear separation of major cell lineages or to resolve subtle, biologically meaningful subtypes. These objectives guided the choice of integration models and parameter settings throughout the analysis.

Determining the optimal clustering resolution was also an iterative process that depended on both technical evaluation and biological interpretation. In this work, I defined clusters based on marker stability and functional coherence, but I recognize

that these boundaries may evolve as understanding improves. For example, the T.SP.CCR4 cluster was retained as a single, relatively homogeneous population showing mixed CD4/CD8 features. However, later analyses (shown in the next chapter) indicated that this cluster may have disease relevance in AxSpA. In parallel, literature (Martini et al., 2024; Yang et al., 2004) has described multiple $CCR4^+$ T-cell subsets in the AxSpA context, suggesting that further exploration of this SP cluster may be needed. This illustrates that cluster resolution is not a fixed outcome but a working model—one that should be revisited when new biological or computational evidence points to additional granularity.

In reflecting on the CD4/CD8 classification process, several limitations became apparent. CD4 and CD8 gene expression were not always reliably captured due to the sparse nature of single-cell RNA-seq, complicating lineage assignment. Using a minimal threshold (e.g. at least one mapped count) resulted in a large number of double-positive cells, likely reflecting technical artefacts. Conversely, assigning CD4 or CD8 identity at the cluster level occasionally merged phenotypically similar but lineage-distinct cells; for instance, cycling T cells whose proliferative signature overrode lineage markers. I also attempted to use CITE-seq data as a reference, but normalization differences between antibody panels and RNA-dependent imputation after totalVI made it difficult to achieve consistent classification across libraries. These observations highlight the technical and conceptual challenges of defining $CD4^+$ and $CD8^+$ lineages in single-cell datasets, emphasizing the need for caution when interpreting mixed or functionally plastic populations.

The identification of MAIT and iNKT cells further demonstrated the limitations of single-modal classification. MAIT-like populations could be identified using GEX-

defined clusters, ADT markers (such as Hu.TCR.Vα7.2), or TCR-based annotations, but these methods did not yield fully consistent results. This inconsistency likely reflects both the limited discriminatory power of transcriptomic features; since MAIT cells and MAIT-like αβ T cells share similar expression profiles, and the imperfect specificity of antibody-based detection, which may not have reliably distinguished Hu.TCR.Vα7.2-expressing cells from related subsets. Moreover, totalVI normalization may have introduced unwanted imputation, as protein expression was inferred partly from gene expression patterns. In cells with highly similar transcriptomes, this process may have blurred true biological distinctions. These findings suggest that joint multi-modal integration should be applied cautiously when dealing with subtle lineage differences that rely on invariant TCR characteristics.

An unexpected observation was the presence of plasma cell-associated gene expression within several T-cell clusters, first noted during pseudo-bulk analyses. Retrospective quality control revealed expression of immunoglobulin-related genes across T cells, suggesting possible low-level contamination or biological signal leakage. Initially, I considered that this pattern might reflect ambient RNA from highly expressed plasma cell transcripts or incomplete removal of doublets during preprocessing. However, after removing ambient RNA and excluding all doublet-like clusters, residual expression of *IGHV1-3* and *IGLV2-14* persisted in T-cell populations. This is unlikely to be caused by ambient RNA, as these two genes are not among the most abundantly expressed immunoglobulins in plasma cells, nor by doublets, since B cells would not be expected to form doublets only with specific T cells but not in any other cell types. Interestingly, similar observations were noted in the Human Protein Atlas (HPA): these genes are highly specific to B and plasma cells, though the minor expression is reported in T/NK cells from HPA internal

dataset, Monaco dataset (Monaco et al., 2019) and Schmiedel dataset (Schmiedel et al., 2018). This raises the possibility of plasma-associated stress under specific conditions, or alternatively, a technical artefact arising from transcript mapping if T-cell genes share sequence similarity with those two immunoglobulin loci.

Overall, this chapter establishes a solid foundation for the downstream analyses presented in the next chapter. Through iterative quality control, parameter optimization, and careful integration of multi-modal data, I constructed a comprehensive and interpretable atlas that captures both the shared and tissue-specific aspects of immune and stromal cell heterogeneity. The iterative annotation and cross-validation processes ensure that this atlas not only reflects current biological understanding but also remains flexible to accommodate future refinement as new insights emerge.

4.5 References

Ajene, A. N., Fischer Walker, C. L., & Black, R. E. (2013). Enteric pathogens and reactive arthritis: a systematic review of Campylobacter, salmonella and Shigella-associated reactive arthritis. *J Health Popul Nutr*, *31*(3), 299–307. https://doi.org/10.3329/jhpn.v31i3.16515

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R. P., Wolters, P. J., Abate, A. R., Butte, A. J., & Bhattacharya, M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*, *20*(2), 163–172. https://doi.org/10.1038/s41590-018-0276-y

Bernardini, G., Sciume, G., Bosisio, D., Morrone, S., Sozzani, S., & Santoni, A. (2008). CCL3 and CXCL12 regulate trafficking of mouse bone marrow NK cell subsets. *Blood*, *111*(7), 3626–3634. https://doi.org/10.1182/blood-2007-08-106203

Borg, N. A., Wun, K. S., Kjer-Nielsen, L., Wilce, M. C., Pellicci, D. G., Koh, R., Besra, G. S., Bharadwaj, M., Godfrey, D. I., McCluskey, J., & Rossjohn, J. (2007). CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature*, *448*(7149), 44–49. https://doi.org/10.1038/nature05907

Cancro, M. P. (2020). Age-Associated B Cells. *Annu Rev Immunol*, *38*, 315–340. https://doi.org/10.1146/annurev-immunol-092419-031130

Deschler, K., Rademacher, J., Lacher, S. M., Huth, A., Utzt, M., Krebs, S., Blum, H., Haibel, H., Proft, F., Protopopov, M., Rodriguez, V. R., Beltran, E., Poddubnyy, D., & Dornmair, K. (2022). Antigen-specific immune reactions by expanded

CD8(+) T cell clones from HLA-B*27-positive patients with spondyloarthritis. *J Autoimmun*, *133*, 102901. https://doi.org/10.1016/j.jaut.2022.102901
Dominguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., Suchanek, O., Polanski, K., King, H. W., Mamanova, L., Huang, N., Szabo, P. A., Richardson, L., Bolt, L., Fasouli, E. S., Mahbubani, K. T., Prete, M., Tuck, L.,…Teichmann, S. A. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, *376*(6594), eabl5197. https://doi.org/10.1126/science.abl5197
Donado, C. A., Theisen, E., Zhang, F., Nathan, A., Fairfield, M. L., Rupani, K. V., Jones, D., Johannes, K. P., Accelerating Medicines Partnership, R. A. S. L. E. N., Raychaudhuri, S., Dwyer, D. F., Jonsson, A. H., & Brenner, M. B. (2025). Granzyme K activates the entire complement cascade. *Nature*, *641*(8061), 211–221. https://doi.org/10.1038/s41586-025-08713-9
Fleming, S. J., Chaffin, M. D., Arduini, A., Akkad, A. D., Banks, E., Marioni, J. C., Philippakis, A. A., Ellinor, P. T., & Babadi, M. (2023). Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods*, *20*(9), 1323–1335. https://doi.org/10.1038/s41592-023-01943-7
Gagliani, N., Magnani, C. F., Huber, S., Gianolini, M. E., Pala, M., Licona-Limon, P., Guo, B., Herbert, D. R., Bulfone, A., Trentini, F., Di Serio, C., Bacchetta, R., Andreani, M., Brockmann, L., Gregori, S., Flavell, R. A., & Roncarolo, M. G. (2013). Coexpression of CD49b and LAG-3 identifies human and mouse T regulatory type 1 cells. *Nat Med*, *19*(6), 739–746. https://doi.org/10.1038/nm.3179
Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., & Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat Methods*, *18*(3), 272–282. https://doi.org/10.1038/s41592-020-01050-x
Granja, J. M., Klemm, S., McGinnis, L. M., Kathiria, A. S., Mezger, A., Corces, M. R., Parks, B., Gars, E., Liedtke, M., Zheng, G. X. Y., Chang, H. Y., Majeti, R., & Greenleaf, W. J. (2019). Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol*, *37*(12), 1458–1465. https://doi.org/10.1038/s41587-019-0332-7
Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B.,…Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587 e3529. https://doi.org/10.1016/j.cell.2021.04.048
Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K.,…Yates, A. D. (2024). Ensembl 2024. *Nucleic Acids Res*, *52*(D1), D891–D899. https://doi.org/10.1093/nar/gkad1049
Heaton, H., Talman, A. M., Knights, A., Imaz, M., Gaffney, D. J., Durbin, R., Hemberg, M., & Lawniczak, M. K. N. (2020). Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods*, *17*(6), 615–620. https://doi.org/10.1038/s41592-020-0820-1
Hie, B. L., Kim, S., Rando, T. A., Bryson, B., & Berger, B. (2024). Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nat Protoc*, *19*(8), 2283–2297. https://doi.org/10.1038/s41596-024-00991-3

Horowitz, A., Yu, H., Pandey, S., Mishra, B., & Sahoo, D. (2024). C1QA is an invariant biomarker for tissue macrophages. *bioRxiv*, 2024.2001.2026.577475. https://doi.org/10.1101/2024.01.26.577475

Hu, C., Yang, J., Qi, Z., Wu, H., Wang, B., Zou, F., Mei, H., Liu, J., Wang, W., & Liu, Q. (2022). Heat shock proteins: Biological functions, pathological roles, and therapeutic opportunities. *MedComm (2020)*, *3*(3), e161. https://doi.org/10.1002/mco2.161

Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., & Teichmann, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol*, *17*, 29. https://doi.org/10.1186/s13059-016-0888-1

Janssen, P., Kliesmete, Z., Vieth, B., Adiconis, X., Simmons, S., Marshall, J., McCabe, C., Heyn, H., Levin, J. Z., Enard, W., & Hellmann, I. (2023). The effect of background noise and its removal on the analysis of single-cell expression data. *Genome Biol*, *24*(1), 140. https://doi.org/10.1186/s13059-023-02978-x

Kawano, T., Cui, J., Koezuka, Y., Toura, I., Kaneko, Y., Motoki, K., Ueno, H., Nakagawa, R., Sato, H., Kondo, E., Koseki, H., & Taniguchi, M. (1997). CD1d-restricted and TCR-mediated activation of valpha14 NKT cells by glycosylceramides. *Science*, *278*(5343), 1626–1629. https://doi.org/10.1126/science.278.5343.1626

Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes* International Conference on Learning Representations (ICLR), https://arxiv.org/abs/1312.6114

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. R., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*, *16*(12), 1289–1296. https://doi.org/10.1038/s41592-019-0619-0

Korsunsky, I., Wei, K., Pohin, M., Kim, E. Y., Barone, F., Major, T., Taylor, E., Ravindran, R., Kemble, S., Watts, G. F. M., Jonsson, A. H., Jeong, Y., Athar, H., Windell, D., Kang, J. B., Friedrich, M., Turner, J., Nayar, S., Fisher, B. A.,…Raychaudhuri, S. (2022). Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. *Med*, *3*(7), 481–518 e414. https://doi.org/10.1016/j.medj.2022.05.002

Li, B., Kowalczyk, M. S., Slyper, M., Jellert, G., Tabaka, M., Ashenberg, O., Waldman, J., Dionne, D., Abigail, K., Hui, M., Yang, Y., Rozenblatt-Rosen, O., & Regev, A. *A single cell immune cell atlas of human hematopoietic system* (

Li, Y. S., Hayakawa, K., & Hardy, R. R. (1993). The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. *J Exp Med*, *178*(3), 951–960. https://doi.org/10.1084/jem.178.3.951

Liu, B., Li, C., Li, Z., Wang, D., Ren, X., & Zhang, Z. (2020). An entropy-based metric for assessing the purity of single cell populations. *Nat Commun*, *11*(1), 3155. https://doi.org/10.1038/s41467-020-16904-3

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat Methods*, *15*(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2

Luecken, M. D., Buttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colome-Tatche, M., & Theis, F. J.

(2022). Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods*, *19*(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8

Martini, V., Silvestri, Y., Ciurea, A., Moller, B., Danelon, G., Flamigni, F., Jarrossay, D., Kwee, I., Foglierini, M., Rinaldi, A., Cecchinato, V., & Uguccioni, M. (2024). Patients with ankylosing spondylitis present a distinct CD8 T cell subset with osteogenic and cytotoxic potential. *RMD Open*, *10*(1). https://doi.org/10.1136/rmdopen-2023-003926

Megill, C., Martin, B., Weaver, C., Bell, S., Prins, L., Badajoz, S., McCandless, B., Pisco, A. O., Kinsella, M., Griffin, F., Kiggins, J., Haliburton, G., Mani, A., Weiden, M., Dunitz, M., Lombardo, M., Huang, T., Smith, T., Chambers, S.,…Carr, A. (2021). cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. *bioRxiv*, 2021.2004.2005.438318. https://doi.org/10.1101/2021.04.05.438318

Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y. Y., Carre, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., Zippelius, A., Pedro de Magalhaes, J., & Larbi, A. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep*, *26*(6), 1627–1640 e1627. https://doi.org/10.1016/j.celrep.2019.01.041

Navarro-Compan, V., Sepriano, A., Capelusnik, D., & Baraliakos, X. (2025). Axial spondyloarthritis. *Lancet*, *405*(10473), 159–172. https://doi.org/10.1016/S0140-6736(24)02263-3

Oetjen, K. A., Lindblad, K. E., Goswami, M., Gui, G., Dagur, P. K., Lai, C., Dillon, L. W., McCoy, J. P., & Hourigan, C. S. (2018). Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight*, *3*(23). https://doi.org/10.1172/jci.insight.124928

Polanski, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A., & Park, J. E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics*, *36*(3), 964–965. https://doi.org/10.1093/bioinformatics/btz625

Porcelli, S., Yockey, C. E., Brenner, M. B., & Balk, S. P. (1993). Analysis of T cell antigen receptor (TCR) expression by human peripheral blood CD4-8- alpha/beta T cells demonstrates preferential use of several V beta genes and an invariant TCR alpha chain. *J Exp Med*, *178*(1), 1–16. https://doi.org/10.1084/jem.178.1.1

Revel, M., Sautes-Fridman, C., Fridman, W. H., & Roumenina, L. T. (2022). C1q+ macrophages: passengers or drivers of cancer progression. *Trends Cancer*, *8*(7), 517–526. https://doi.org/10.1016/j.trecan.2022.02.006

Rubtsov, A. V., Rubtsova, K., Fischer, A., Meehan, R. T., Gillis, J. Z., Kappler, J. W., & Marrack, P. (2011). Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c(+) B-cell population is important for the development of autoimmunity. *Blood*, *118*(5), 1305–1315. https://doi.org/10.1182/blood-2011-01-331462

Schmiedel, B. J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A. G., White, B. M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J. A., McVicker, G., Seumois, G., Rao, A., Kronenberg, M., Peters, B., & Vijayanand, P. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*, *175*(6), 1701–1715 e1716. https://doi.org/10.1016/j.cell.2018.10.022

Sieper, J., & Poddubnyy, D. (2017). Axial spondyloarthritis. *Lancet*, *390*(10089), 73–84. https://doi.org/10.1016/S0140-6736(16)31591-4

Strehl, A., & Ghosh, J. (2003). Cluster ensembles --- a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, *3*(null), 583–617. https://doi.org/10.1162/153244303321897735

Thomas, T., Friedrich, M., Rich-Griffin, C., Pohin, M., Agarwal, D., Pakpoor, J., Lee, C., Tandon, R., Rendek, A., Aschenbrenner, D., Jainarayanan, A., Voda, A., Siu, J. H. Y., Sanches-Peres, R., Nee, E., Sathananthan, D., Kotliar, D., Todd, P., Kiourlappou, M.,…Buckley, C. D. (2024). A longitudinal single-cell atlas of anti-tumour necrosis factor treatment in inflammatory bowel disease. *Nat Immunol*, *25*(11), 2152–2165. https://doi.org/10.1038/s41590-024-01994-8

Tilloy, F., Treiner, E., Park, S. H., Garcia, C., Lemonnier, F., de la Salle, H., Bendelac, A., Bonneville, M., & Lantz, O. (1999). An invariant T cell receptor alpha chain defines a novel TAP-independent major histocompatibility complex class Ib-restricted alpha/beta T cell subpopulation in mammals. *J Exp Med*, *189*(12), 1907–1921. https://doi.org/10.1084/jem.189.12.1907

Treiner, E., Duban, L., Bahram, S., Radosavljevic, M., Wanner, V., Tilloy, F., Affaticati, P., Gilfillan, S., & Lantz, O. (2003). Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature*, *422*(6928), 164–169. https://doi.org/10.1038/nature01433

van den Brink, S. C., Sage, F., Vertesy, A., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., Robin, C., & van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods*, *14*(10), 935–936. https://doi.org/10.1038/nmeth.4437

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, *43*(1110), 11 10 11–11 10 33. https://doi.org/10.1002/0471250953.bi1110s43

Villani, A. C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P. L.,…Hacohen, N. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, *356*(6335). https://doi.org/10.1126/science.aah4573

Villar, J., & Segura, E. (2020). Decoding the Heterogeneity of Human Dendritic Cell Subsets. *Trends Immunol*, *41*(12), 1062–1071. https://doi.org/10.1016/j.it.2020.10.002

Walzer, T., Dalod, M., Robbins, S. H., Zitvogel, L., & Vivier, E. (2005). Natural-killer cells and dendritic cells: "l'union fait la force". *Blood*, *106*(7), 2252–2258. https://doi.org/10.1182/blood-2005-03-1154

Xie, G., Chen, X., Gao, Y., Yang, M., Zhou, S., Lu, L., Wu, H., & Lu, Q. (2025). Age-Associated B Cells in Autoimmune Diseases: Pathogenesis and Clinical Implications. *Clin Rev Allergy Immunol*, *68*(1), 18. https://doi.org/10.1007/s12016-025-09021-w

Xin, H., Lian, Q., Jiang, Y., Luo, J., Wang, X., Erb, C., Xu, Z., Zhang, X., Heidrich-O'Hare, E., Yan, Q., Duerr, R. H., Chen, K., & Chen, W. (2020). GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome Biol*, *21*(1), 188. https://doi.org/10.1186/s13059-020-02084-2

Yang, P. T., Kasai, H., Zhao, L. J., Xiao, W. G., Tanabe, F., & Ito, M. (2004). Increased CCR4 expression on circulating CD4(+) T cells in ankylosing

spondylitis, rheumatoid arthritis and systemic lupus erythematosus. *Clin Exp Immunol*, *138*(2), 342–347. https://doi.org/10.1111/j.1365-2249.2004.02617.x

Yu, D., Zhang, S., Ma, C., Huang, S., Xu, L., Liang, J., Li, H., Fan, Q., Liu, G., & Zhai, Z. (2023). CCL3 in the bone marrow microenvironment causes bone loss and bone marrow adiposity in aged mice. *JCI Insight*, *8*(1). https://doi.org/10.1172/jci.insight.159107

Chapter 5: Cellular Composition and Cross-Tissue Immune Landscape in AxSpA

5.1 Introduction

The involvement of multiple tissues is a hallmark of AxSpA, yet the precise mechanisms underlying this multi-tissue pathology remain incompletely understood. Several hypotheses have been proposed to explain how inflammation initiates and spreads across multiple anatomical sites in AxSpA. One prominent theory, derived from the arthritogenic peptide hypothesis, emphasizes the gut-joint axis, proposing that pathogenic T cells might initially be primed in the gut in response to microbial-derived peptides (Gracey et al., 2020; Qaiyum et al., 2021). This concept is supported by several lines of evidence: the frequent co-occurrence of IBD and higher proportion of sub-clinical gut inflammation observed in SpA (Rudwaleit & Baeten, 2006; Van Praet et al., 2013); microbial dysbiosis characterized by reduced bacterial diversity and altered bacterial compositions in SpA patients (Costello et al., 2015); increased intestinal permeability in SpA compared to RA (Hecquet et al., 2021); the expression of gut-associated integrins on synovial T cells (Guggino et al., 2021); and shared genetic susceptibility loci between SpA and IBD (Ellinghaus et al., 2016). However, despite this supportive evidence, the precise causal relationship remains unclear, and the presence of distinct genetic backgrounds also suggests the involvement of unique, disease-specific pathways.

Further supporting the microbiome-driven mechanism, a recent study identified shared TCR clonotypes between synovial fluid and ocular tissues in HLA-B*27-related diseases, indicating potential TCR cross-reactivity driven by

shared peptide-MHC binding motifs between self and microbial antigens (Yang et al., 2022). In addition to the microbiome-centric explanation, other potential mechanisms have been proposed. These include the systemic immune hypersensitivity hypothesis (Thiran & Vereecke, 2025), suggesting that gut inflammation may arise independently of microbial triggers; and the gut barrier dysfunction and dysbiosis hypothesis, proposing that disrupted intestinal tight junctions permit microbial components to translocate into systemic circulation, subsequently triggering immune responses at joint sites (Ciccia et al., 2017). These alternative hypotheses highlight the complexity and potentially multifactorial nature of tissue involvement in AxSpA.

In the previous chapter, I constructed a comprehensive single-cell atlas of AxSpA encompassing multiple disease-relevant tissues. This multi-modal atlas with consistent and detailed cell-type annotations serves as a robust framework for investigating immunological mechanisms underlying AxSpA. In this chapter, I leverage this resource to characterize differences in immune cell composition across tissues and between disease conditions. Employing both proportion-based and neighborhood graph-based analytical approaches, I systematically compared immune cell compositions across AxSpA patients, healthy controls, and RA patients. By performing comprehensive statistical tests, I sought to distinguish AxSpA-specific immune signatures from general inflammatory responses, thus advancing our understanding of the distinct cellular alterations associated with this disease.

5.2 Aim

This chapter aims to characterize disease-associated immune cell populations in AxSpA by comparing cellular composition across AxSpA patients, healthy controls, and individuals with other inflammatory diseases such as RA. Through cross-tissue comparisons, I aim to identify immune cell phenotypes that are shared across inflamed sites, potentially revealing common pathological signatures, as well as uncovering tissue-specific immune patterns that reflect AxSpA-related local immune responses. By identifying immune subsets specifically enriched in AxSpA and distinguishing these from general inflammatory responses, this chapter seeks to enhance our understanding of AxSpA pathogenesis and highlight potential therapeutic targets.

5.3 Results

This section presents the results of compositional analyses comparing immune cell composition across AxSpA, RA, and healthy controls. The analyses combine multiple complementary methods, including edgeR, scCODA, Milo, non-negative matrix factorisation (NMF), and cross-tissue Poisson regression. These methods were applied to both a balanced core cohort and an expanded cohort spanning peripheral blood, synovial fluid, peri-entheseal bone, entheseal soft tissue, ileum, and the sacroiliac joint. Results are organised by immune compartment, with compositional analysis supported by technical validation across complementary methods, so that the reproducibility of candidate AxSpA-associated signatures can be evaluated across methods and across tissues.

5.3.1 Analytical framework for differential abundance

In this chapter I performed cross-condition comparisons to identify disease-specific features and cross-tissue comparisons to highlight tissue-

specific features. The analyses here are built upon the atlas previously established in Chapter 4. Although significant effort was invested to reduce batch effects, certain analyses relying on raw count data may still retain residual batch-related variability. Therefore, careful consideration was taken in designing comparison matrices to minimize potential confounding and ensure robustness of the analytical conclusions presented in this thesis.

Since some batches are fully confounded with a single condition, I present results using two distinct sample groupings: (1) Core cohort: includes only batches where all comparison groups are represented, resulting in a fully crossed and statistically balanced design. Batch effects are explicitly modeled as covariates, minimizing confounding; (2) Expanded cohort: incorporates additional batches containing only a subset of comparison groups, resulting in a partially crossed and statistically imbalanced design. Although batch effects remain explicitly modeled as covariates, this design prioritizes maximizing statistical power, potentially at the expense of increased residual confounding. The specific samples included in each analytical cohort are detailed in **Appendix D.1.**

The analytical framework for compositional analysis of samples from different cohorts is illustrated in **Figure 5.1**. For both cohorts, I initially assessed the proportions of each cell type within their respective lineages and applied Student's *t*-tests to compare these proportions between conditions. However, this approach does not adequately account for the interdependencies among proportions or variations in cell counts and sequencing depths across samples. To address these limitations, I subsequently applied edgeR (Robinson et al., 2010) using trimmed mean of M-values (TMM) normalization on the cell-type

count matrix for each immune compartment, thereby accommodating variation in library size and composition. For the extended cohort analysis, I employed scCODA (Buttner et al., 2021), a Bayesian Dirichlet-multinomial modeling framework, as a validation tool to strengthen statistical inference and confirm differential abundance patterns identified by proportion-based methods. Additionally, Milo (Dann et al., 2022) was utilized as a validation tool at a different resolution, assessing differential abundance across overlapping cell-state neighborhoods to detect subtle, internal shifts within the predefined clusters. As Milo relies on the original neighborgraph constructed using all samples, this approach corresponds to the extended cohort analysis.

To investigate program-based differences beyond individual cell-type changes, I additionally applied non-negative matrix factorization (NMF) to the per-sample cell-type frequency data to identify latent cellular abundance programs associated with disease status, capturing high-level compositional patterns not evident when examining single cell types alone.

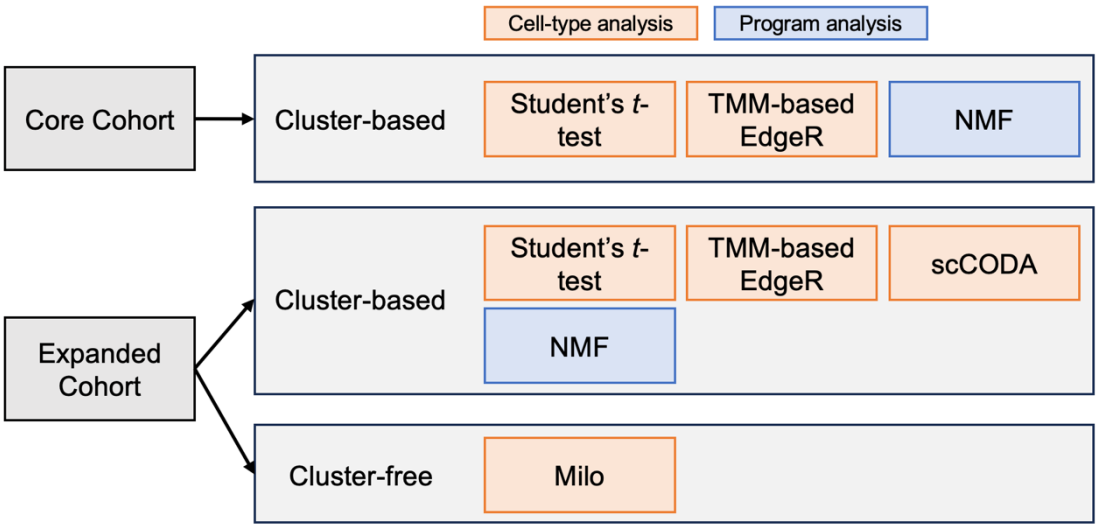


Figure 5.1 Analysis framework for compositional analysis across cohorts

Overview of the analytical methods applied to core and expanded cohorts. Analyses were categorized into two complementary approaches: individual cell-type analysis and program analysis, each further classified as cluster-based or cluster-free. Individual cell type analysis employed Student's t-test, edgeR, scCODA and Milo to identify disease-associated immune subsets. Program-level analyses utilized NMF to identify latent co-varying compositional signatures associated with disease status.

While utilizing each method, the cross-condition comparisons were performed in two main contexts: AxSpA versus healthy controls, and AxSpA versus RA, for which well-matched comparator groups were available. Although samples were also collected from patients with other spondyloarthritis conditions, such as psoriatic arthritis and monoarthritis, these had limited numbers so were not included as separate comparison groups.

5.3.2 EdgeR results comparing cell-type composition between AxSpA, HC, and RA groups

5.3.2.1 Comparison strategy in core and expanded cohorts

For each cell type, I calculated proportions and performed relative analysis within their respective compartments, which comprised of $CD4^+$ T cells, $CD8^+$ and other T cells, NK cells, myeloid cells, B and plasma cells, and other immune populations. Within each compartment, analyses were conducted at three annotation resolutions: "major subset," "minor subset," and "cell type." These compartments and annotation levels were defined in Chapter 4 based on lineage relationships and transcriptional similarity.

Using the balanced core cohort, I performed three main comparisons: AxSpA vs HC in peripheral blood (PB), and AxSpA vs RA in both synovial fluid (SF) and PB. To account for potential therapy effects, I also examined biologic-naïve patient subgroups within each comparison.

In the expanded cohort, I extended the comparisons to include additional samples from PB and SF for both AxSpA vs HC and AxSpA vs RA analyses, as well as introducing comparisons in other disease-relevant tissues, including ileum (ILM), peri-entheseal bone (PEB), and entheseal soft tissue (EST), specifically for AxSpA vs HC. These additional tissues were not included in the core cohort due to potential confounding arising from batch effects and differences in sample handling. Specifically, in PB and SF, some batches contained samples from only a single condition or combined fresh and frozen samples. In PEB and EST, AxSpA samples were processed from frozen tissues, whereas HC samples were fresh. In ileal tissues, samples from the Gut_Naples project were processed from frozen tissues, while those from the Gut_TGU project were fresh and underwent further FACS sorting into $CD45^+$ and $CD45^-$ subsets. Detailed sample handling procedures are provided in the **Methods**, and comprehensive metadata is available in **Appendix A.1**.

When examining proportions within the $CD4^+$ and $CD8^+$ T-cell compartments, I observed a strong correlation between naïve T cell proportions and donor age in PB samples (see **Appendix D.2**). This observation aligns with previous findings that proportions of naïve T cells within both $CD4^+$ and $CD8^+$ subsets sharply decrease with age, especially for the CD8 population (Li et al., 2019). Therefore, I evaluated compositional changes using both total T-cell proportions and a restricted analysis focused specifically on non-naïve T-cell subsets, to highlight compositional shifts within effector and memory compartments.

5.3.2.2 CD4-positive T Cell Compartment

The edgeR results for the non-naive $CD4^+$ T-cell compartment are shown in **Figure 5.2**, while corresponding analysis for the full $CD4^+$ T cell compartment provided in **Appendix D.3**. In synovial fluid, AxSpA exhibited a significantly higher frequency of $CD4^+$ TEMRA cells in biologic-naïve patient groups compared to RA condition (BH-adjusted $p < 0.05$), with a similar increasing trend observed in the overall patient groups (BH-adjusted $p < 0.1$), indicating a consistent change that was not influenced by biologic therapy. Notably, CD4.Tr1.act cells were significantly reduced in AxSpA compared with RA in both synovial fluid and peripheral blood (BH-adjusted $p < 0.05$), regardless of therapeutic stage. This CD4.Tr1.act subset is characterized by PD-1^+ surface expression and high *CXCL13* gene expression This population closely corresponds to the previously described CXCL13-producing PD-$1^{hi}CXCR5^-CD4^+$ T cells (Kobayashi et al., 2013), later termed T peripheral helper (Tph) cells (Rao et al., 2017). These cells have been repeatedly reported as expanded in RA peripheral blood and synovial fluid, likely representing the same cluster identified in this dataset.

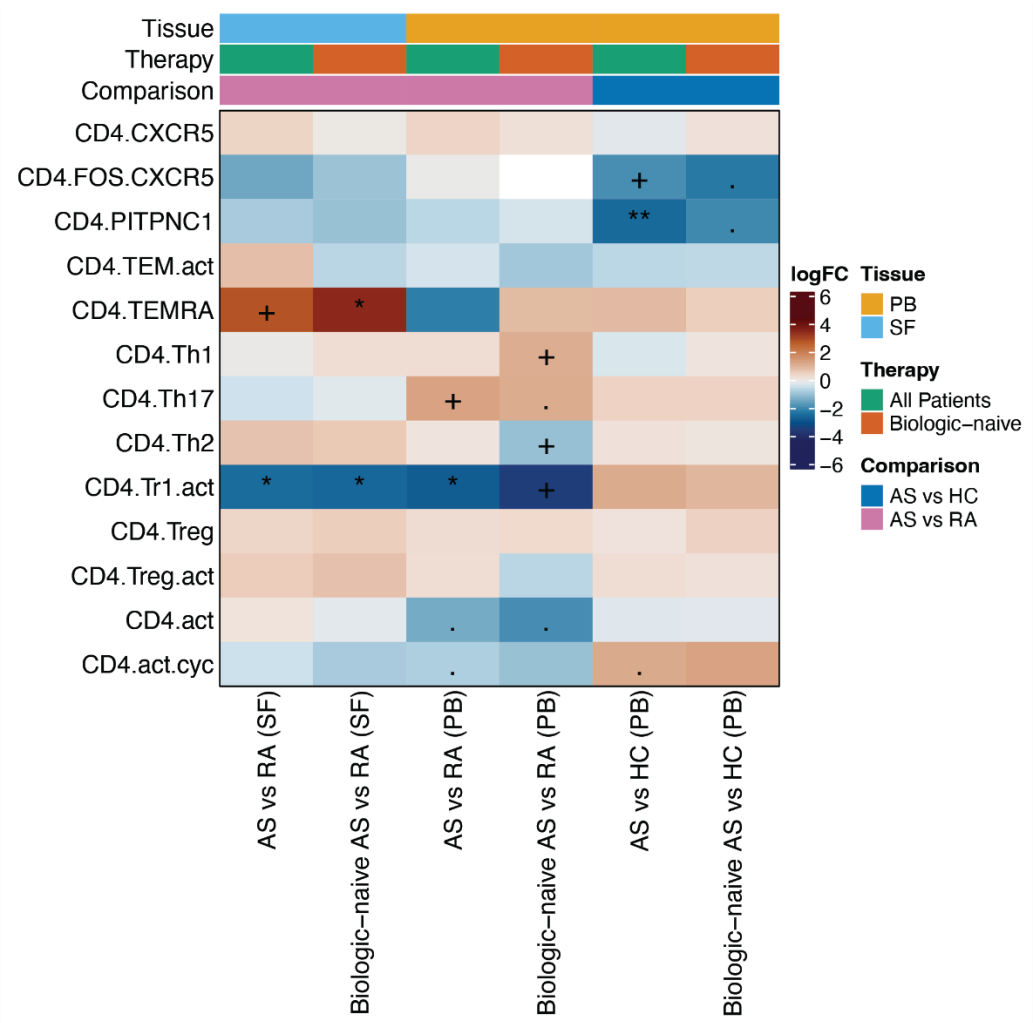


Figure 5.2 Differential abundance of CD4-positive non-naïve T-cell subsets across tissues and patient groups at the cell type level in the core cohort
Heatmap showing differential abundance analysis of non-naïve CD4-positive T cell subsets in the minimal cohort using edgeR with TMM-normalized cell type counts. Colour indicates log fold change (logFC) of each subset in AxSpA relative to the comparison group. Statistical significance was determined by edgeR with symbols indicating adjusted p-value thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1, . < 0.2, and no label for non-significant results.

In peripheral blood, AxSpA exhibited an increased frequency of Th1 (BH-adjusted p < 0.2) and Th17 subsets (BH-adjusted p < 0.1) in biologic naïve patients compared with RA, suggesting a broader pro-inflammatory T helper cell phenotype. However, these subsets were not significantly increased relative to healthy controls, indicating that the observed difference could also reflect a relative decrease in RA rather than a specific expansion in AxSpA. AxSpA

PBMCs showed a significant reduction in CD4.PITPNC1 cells (BH-adjusted p < 0.01) compared with healthy controls. Given that PITPNC1 has been reported to suppress $CD8^+$ T cell immune activity in some context (Liang et al., 2024), this decrease may reflect a loss of a regulatory and suppressive-like $CD4^+$ subset in AxSpA.

I further applied same analysis to the expanded cohort within the non-naïve $CD4^+$ T cell compartment, which is shown in **Figure 5.3**, while corresponding analysis with full $CD4^+$ T cell compartment is shown in **Appendix D.4** and boxplots illustrating each comparison are included in **Appendices D.5 - D.10**.

In the expanded cohort, the previously observed reduction of the CD4.Tr1.act subset in AxSpA compared to RA remained significant (BH-adjusted p < 0.001), further confirming this finding. Additionally, the CD4.TEMRA cluster exhibited an increased trend in AxSpA peri-entheseal bone relative to healthy controls, though this difference only reached nominal significance (BH-adjusted p < 0.2).

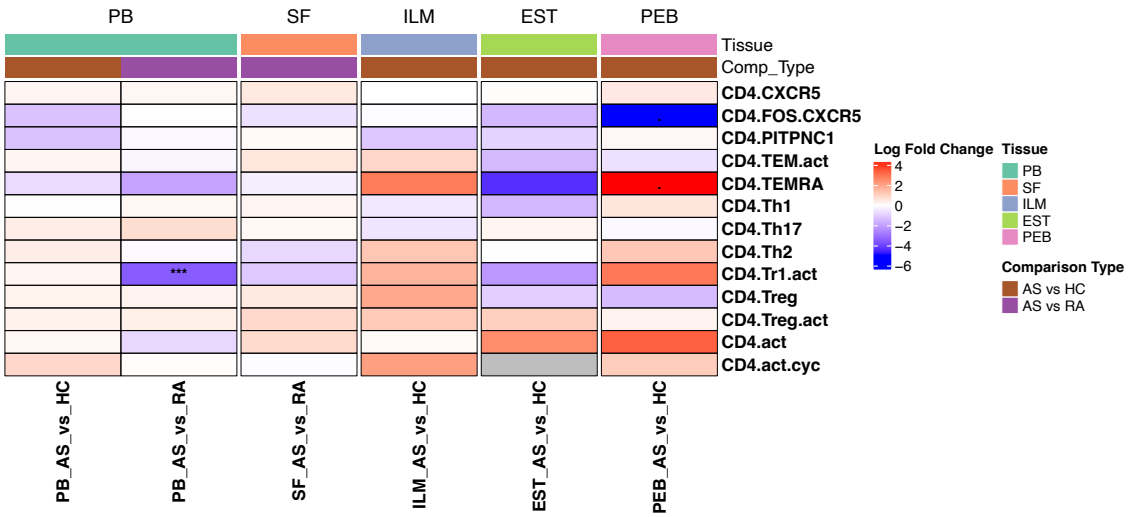


Figure 5.3 Differential abundance of CD4-positive non-naïve T-cell subsets at the cell type level in the expanded cohort
Heatmap showing differential abundance analysis of non-naïve CD4-positive T cell subsets in the expanded cohort using edgeR across peripheral blood (PB), synovial fluid (SF), ileum (ILM), entheseal-soft tissue (EST) and peri-entheseal bone (PEB). Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

5.3.2.3 CD8-positive T Cell Compartment

The edgeR results for the non-naïve $CD8^+$ T cell compartment are shown in **Figure 5.4**, with corresponding results for the full $CD8^+$ T-cell compartment provided in **Appendix D.13**. Significant differences were mainly observed in synovial fluid, where CD8.TEMRA.ZNF683 subsets were significantly increased in AxSpA compared with RA (BH-adjusted $p < 0.05$), while γδ T cells (UTC.GD.VD1) were significantly decreased (BH-adjusted $p < 0.05$). However, the γδ T-cell signal appears to be influenced by a single outlier sample within the RA group (**Appendix D.16**), and should therefore be interpreted with caution. In peripheral blood, AxSpA showed higher proportions of CD8.TCMRA.ZNF683.XCL1 cells compared with RA, particularly within the biologic-naïve subgroup, although this difference was not significant relative to healthy controls.

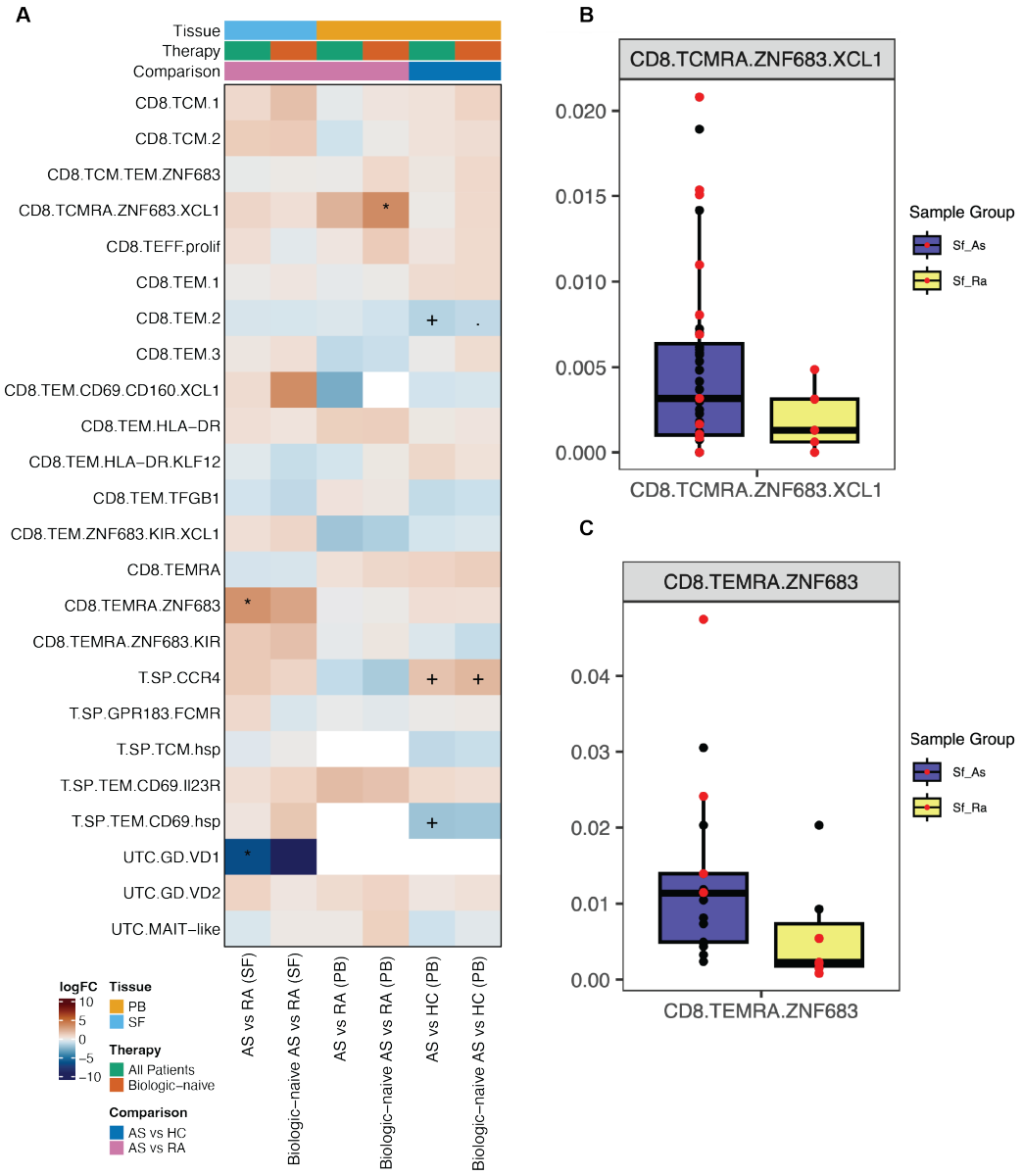


Figure 5.4 Differential abundance of CD8-positive non-naïve T-cell subsets at the cell type level in the core cohort
(A) Heatmap showing differential abundance analysis of non-naïve CD8-positive T cell subsets in the core cohort using edgeR with TMM-normalized cell type counts. Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s. (B, C) Boxplots illustrating the relative abundance of CD8.TCMRA.ZNF683.XCL1 in AxSpA versus RA in peripheral blood (B) and CD8.TEMRA.ZNF683 in AxSpA versus RA in synovial fluid (C) in the expanded cohort. Each dot represents an individual sample; red dots indicate samples included in the core cohort.

Furthermore, when compared with healthy controls, the T.SP.CCR4 subset was consistently enriched in both AxSpA and biologic-naïve cohorts (BH-adjusted $p < 0.1$). However, because this subset did not show a significantly

greater increase in AxSpA relative to RA, it likely reflects a shared inflammatory signature rather than a disease-specific feature. Consistently, both CD4-positive and CD8-positive $CCR4^+$ T cells have been observed in the peripheral blood from ankylosing spondylitis and rheumatoid arthritis patients, which aligns with our findings (Martini et al., 2024; Yang et al., 2004).

Figure 5.5 shows the edgeR results for the non-naïve $CD8^+$ T cell compartment in the expanded cohort. In peripheral blood, the previously identified increase in T.SP.CCR4 relative to healthy controls was replicated (BH-adjusted $p < 0.1$). Additionally, CD8.TEM.HLA-DR.KLF12 showed a significant increase in AxSpA peripheral blood (BH-adjusted $p < 0.05$), accompanied by a similar trend in ileum (BH-adjusted $p < 0.2$). Another cell cluster, CD8.TEM.3, exhibited increased abundance in peripheral blood compared to healthy controls (BH-adjusted $p < 0.2$), and also showed elevated levels in AxSpA synovial fluid compared to RA (BH-adjusted $p < 0.2$).

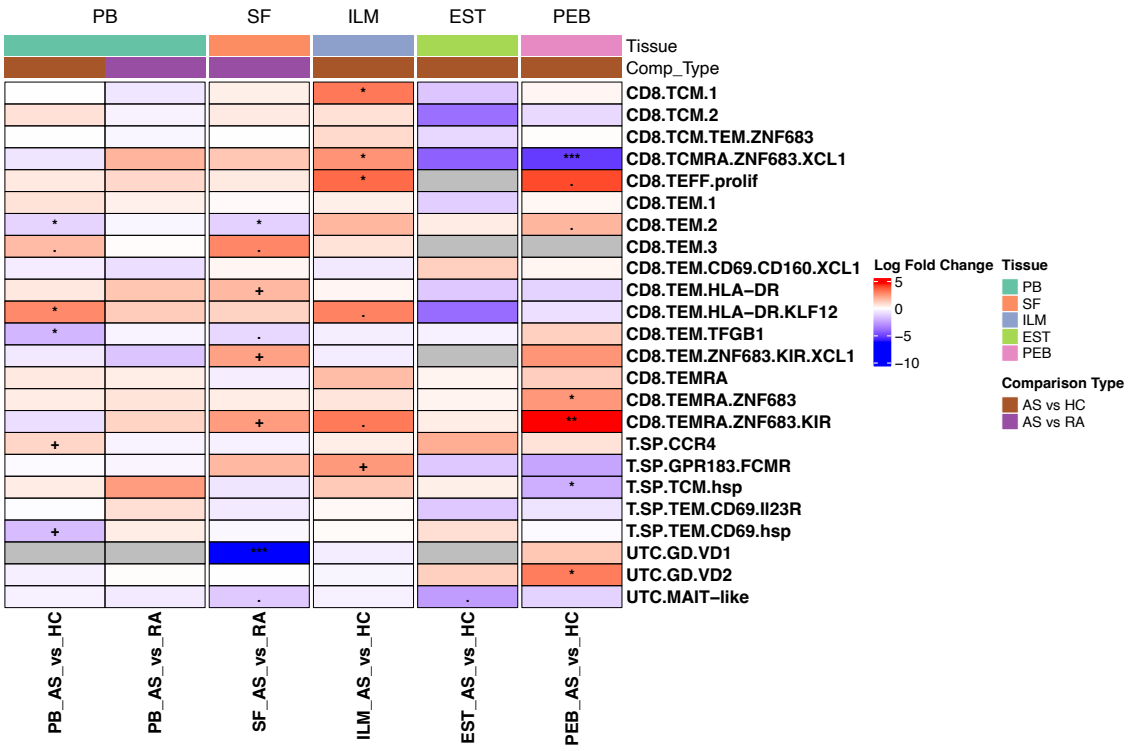


Figure 5.5 Differential abundance of CD8-positive non-naïve T-cell subsets at the cell type level in the expanded cohort
Heatmap showing differential abundance analysis of CD4-positive T cell subsets in the expanded cohort using edgeR across PB, SF, ILM, EST and PEB. Colour indicates logFC of each subset in AxSpA relative to the comparison group; subsets with no cells detected in that tissue for either condition are shown in grey. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

When examining additional tissue comparisons in the expanded cohort, the CD8.TEMRA.ZNF683.KIR cluster exhibited a consistent cross-tissue pattern of increased abundance in AxSpA. This cluster was significantly increased relative to healthy controls in PEB (BH-adjusted $p < 0.01$), showed a similar trend in ileum (BH-adjusted $p < 0.2$), and was also elevated in AxSpA compared to RA in synovial fluid (BH-adjusted $p < 0.1$), indicating an AxSpA-related cross-tissue signature. Similarly, the CD8.TEFF.prolif subset showed increased abundance in AxSpA patients within both PEB (BH-adjusted $p < 0.2$) and ileum (BH-adjusted $p < 0.05$). Additional subsets specifically enriched in AxSpA ileum included CD8.TCM.1 (BH-adjusted $p < 0.05$) and CD8.TCMRA.ZNF683.XCL1

(BH-adjusted $p < 0.05$), while PEB demonstrated significant enrichment of CD8.TEMRA.ZNF683 (BH-adjusted $p < 0.05$) and UTC.GD.VD2 (BH-adjusted $p < 0.05$). Overall $CD8^+$ T-cell comparisons including naïve cluster and corresponding box plots for each comparison are provided in **Appendices D.12-D.18**.

In the comparison to RA in synovial fluid across the core and expanded cohorts, the core cohort showed an increase in CD8.TEMRA.ZNF683, and the expanded cohort showed an increase in CD8.TEMRA.ZNF683.KIR and CD8.TEMRA.ZNF683.KIR.XCL1, all of which are *ZNF683* positive TEMRA cells. *ZNF683*, also known as *Hobit*, is a transcription factor that programs tissue-resident and effector functions in lymphocytes. It represses genes required for tissue egress and promotes transcriptional programs that sustain residency and cytotoxicity. In human studies, high *ZNF683* expression marks an intermediate effector/effector-memory $CD8^+$ T-cell population with strong activation and cytotoxic signatures, enriched in sites of local immune activity and in responders to PD-1 checkpoint blockade (Parry et al., 2023). In murine models, *Hobit* is up-regulated early in antigen-driven effector $CD8^+$ T cells within peripheral tissues and instructs their differentiation into TRM cells, characterized by down-regulation of tissue-exit receptors and acquisition of residency-associated molecules (Parga-Vidal et al., 2021). Together with the related transcription factor *Blimp1*, *Hobit* forms a core regulatory module essential for the development and maintenance of tissue-resident lymphocytes across multiple tissues (Mackay et al., 2016). Thus, the increased CD8.TEMRA.ZNF683 / CD8.TEMRA.ZNF683.KIR / CD8.TEMRA.ZNF683.KIR.XCL1 cluster in AxSpA synovial fluid likely indicates

a locally retained, cytotoxic and effector-biased $CD8^+$ T-cell population (TRM/TEMRA-like) exhibiting transcriptional programs favoring tissue residency and activation rather than recirculation, consistent with local antigen-driven or tissue-adapted immune activity in AxSpA.

5.3.2.4 NK cell compartment

The edgeR analysis for the NK cell compartment of the core cohort is shown in **Figure 5.6**. Among all comparisons, a modest but notable difference was observed in synovial fluid, where the NK.$CD16^{lo}$.LTB subset was decreased in AxSpA compared with RA (BH-adjusted $p < 0.1$). This cluster likely corresponds to the previously described NK2 subset reported in a published single-cell study of natural killer cells across tissue (Rebuffet et al., 2024), as both express characteristic marker genes such as *LTB, IL7R, XCL1,* and *XCL2*. The NK2 subset is known to exhibit cytokine-producing and immunoregulatory functions, and its reduction in AxSpA suggests a shift from a regulatory toward a more cytotoxic NK-cell pattern compared with RA. No other NK-cell subsets showed significant changes across tissues or treatment groups in the core cohort comparison.

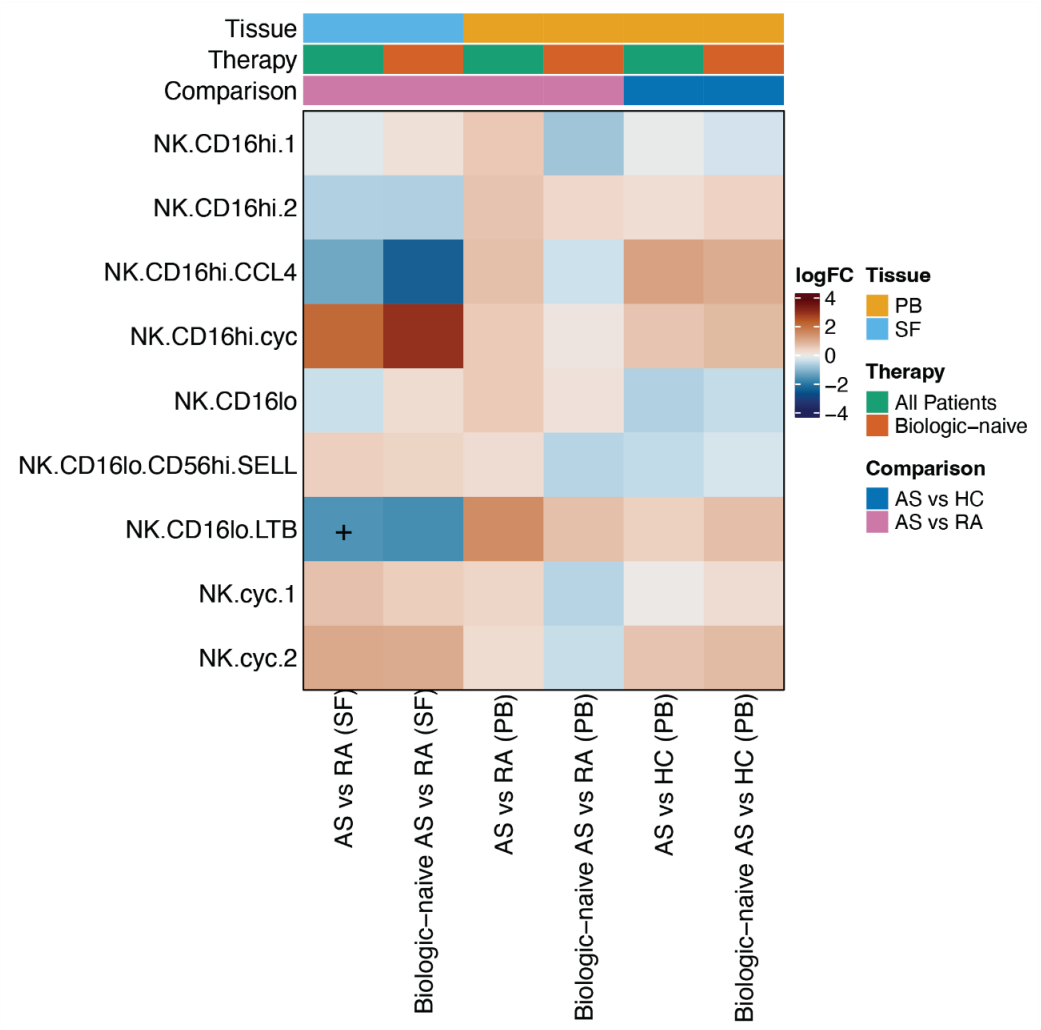


Figure 5.6 Differential abundance of NK cell subsets at the cell type level in the core cohort
Heatmap showing differential abundance analysis of NK cell subsets in the core cohort using edgeR with TMM-normalized cell type counts. Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

In the expanded cohort, the cycling NK cluster NK.CD16hi.cyc showed a trend towards increased abundance in AxSpA relative to RA in synovial fluid (BH-adjusted $p < 0.1$), although this did not reach significance in the core cohort (**Figure 5.7**). Related box plots for these comparisons can be found in **Appendices D.19-D.24**.

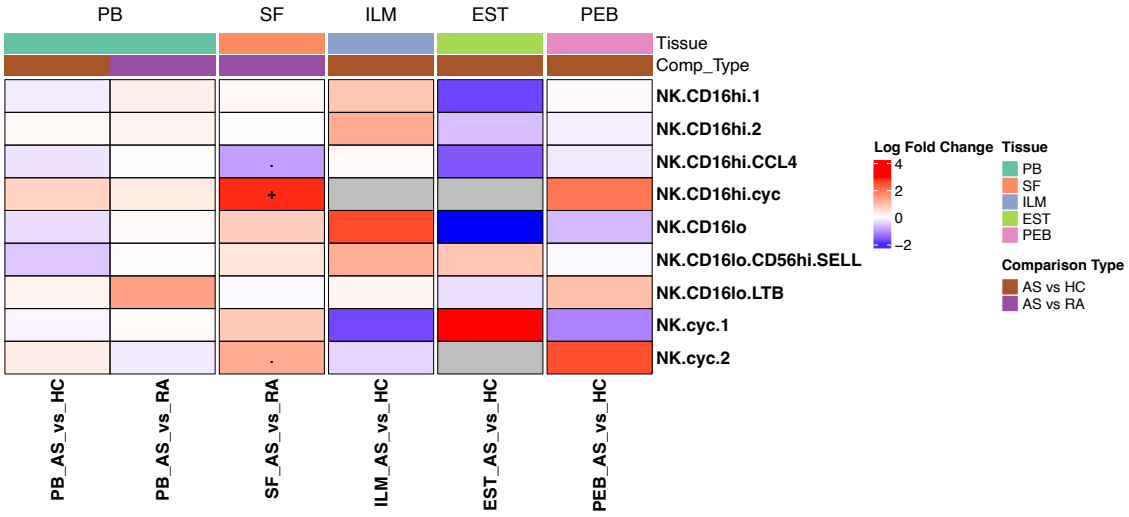


Figure 5.7 Differential abundance of NK cell subsets at the cell type level in the expanded cohort
Heatmap showing differential abundance analysis of NK cell subsets in the expanded cohort using edgeR across PB, SF, ILM, EST and PEB. Colour indicates logFC of each subset in AxSpA relative to the comparison group; subsets with no cells detected in that tissue for either condition are shown in grey. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

5.3.2.5 Myeloid cell compartment

The edgeR analysis for the myeloid cell compartment in the core cohort is shown in **Figure 5.8**, and the corresponding box plots for each cell type is shown in **Appendices D.26-D.31**. In synovial fluid, AxSpA exhibited enrichment of cMono.3 (BH-adjusted $p < 0.2$) and intMono.1 (BH-adjusted $p < 0.05$) compared with RA. Among the intermediate monocyte clusters, intMono.1 was marked by high surface expression of CD169 and increased *SIGLEC1* (CD169) gene expression, a feature known to be induced in monocytes by interferons and Toll-like receptor agonists (York et al., 2007). This suggests that the intMono.1 population enriched in AxSpA synovial fluid represents an IFN-responsive, activated monocyte subset.

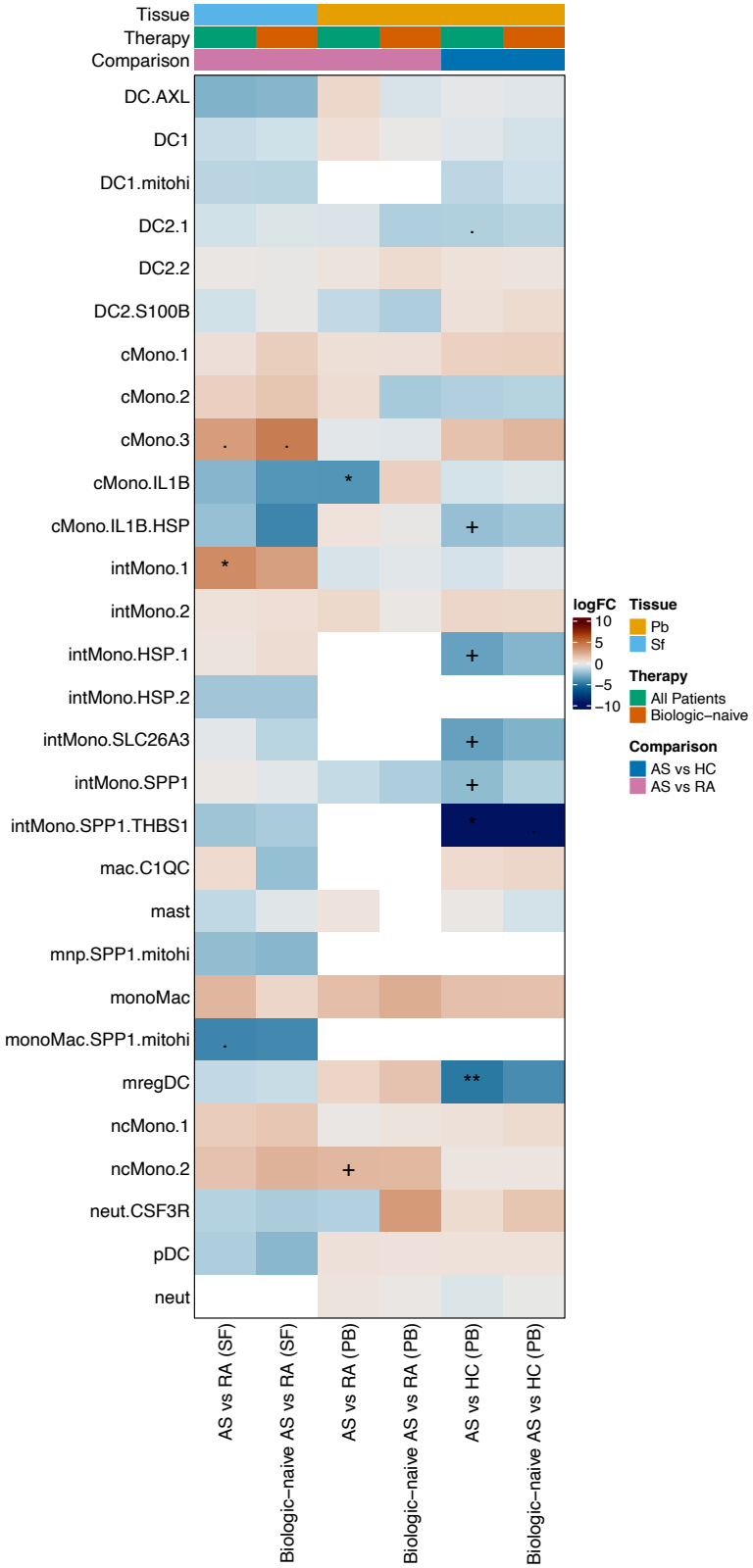


Figure 5.8 Differential abundance of myeloid cell subsets at the cell type level in the core cohort

Heatmap showing differential abundance analysis of myeloid cell subsets in the core cohort using edgeR with TMM-normalized cell type counts. Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

In peripheral blood, AxSpA showed higher proportions of ncMono.2 (BH-adjusted < 0.2) relative to RA, but not relative to HC, suggesting this may be an RA-specific signal. Compared with healthy controls, several intermediate monocyte subsets were reduced in AxSpA, indicating potential alterations in the intermediate monocyte phenotype, and the mregDC population was also significantly decreased.

To determine whether the observed differences reflected an overall abundance change or a phenotypic shift within myeloid populations, I examined the results at a higher annotation level which corresponds to minor subset (**Figure 5.9**). In synovial fluid, no significant minor subset level changes were detected at this broader classification level. In peripheral blood, a decrease in mregDCs was confirmed in AxSpA compared with healthy controls, and non-classical monocytes showed a modest increase in AxSpA relative to RA.

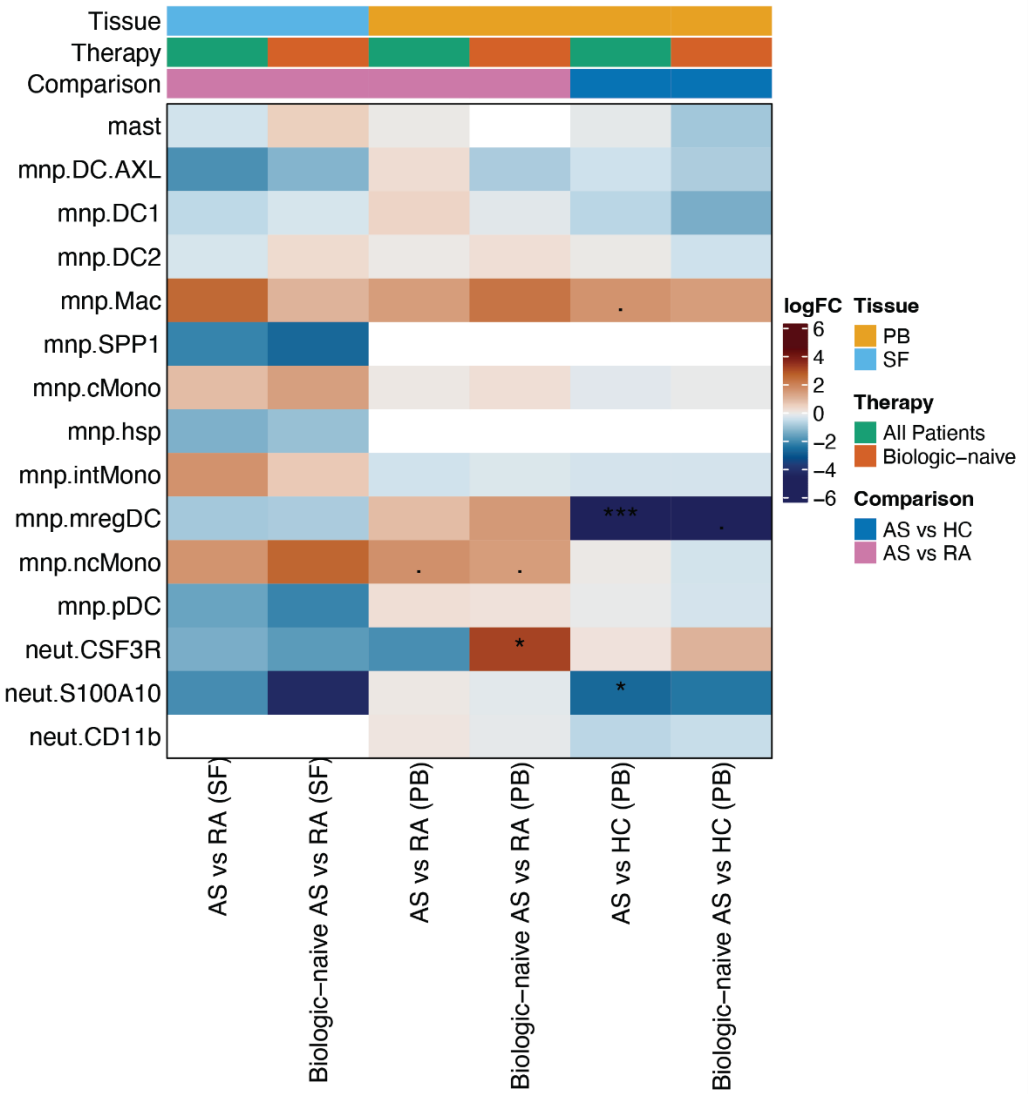


Figure 5.9 Differential abundance of myeloid cell subsets at the minor subset level in the core cohort
Heatmap showing differential abundance analysis of myeloid cell subsets in the core cohort using edgeR at the minor subset level. Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

Interestingly, a neutrophil cluster, characterised by CSF3R and higher IL1B gene expression, was significantly increased in the biologic-naïve AxSpA group compared with RA blood samples (BH-adjusted $p < 0.05$). Neutrophils in AxSpA have been shown to form IL-17A- and IL-1β-rich extracellular traps that drive mesenchymal stem-cell osteogenic differentiation and pathological bone formation (Papagoras et al., 2021). Along with clinical evidence linking higher

neutrophil-to-lymphocyte ratios to AxSpA disease activity (Al-Osami et al., 2020), the increased $CSF3R^{+}$ $IL1B^{hi}$ neutrophil population observed in blood may reflect enhanced myeloid activation or the spillover of activated, tissue-primed neutrophils into circulation, potentially mirroring inflammatory activity at disease sites in AxSpA. However, given the limited neutrophil coverage resulting from the sample collection procedures, these findings are based on a relatively small neutrophil proportion and should therefore be interpreted with caution.

When exploring the myeloid compartment in the expanded cohort, several findings remained consistent with results from the core cohort (**Figure 5.10**). Notably, the cluster cMono.3 showed a modest increase in AxSpA compared to healthy controls both in PBMC (BH-adjusted $p < 0.1$) and in ileum (BH-adjusted $p < 0.1$).

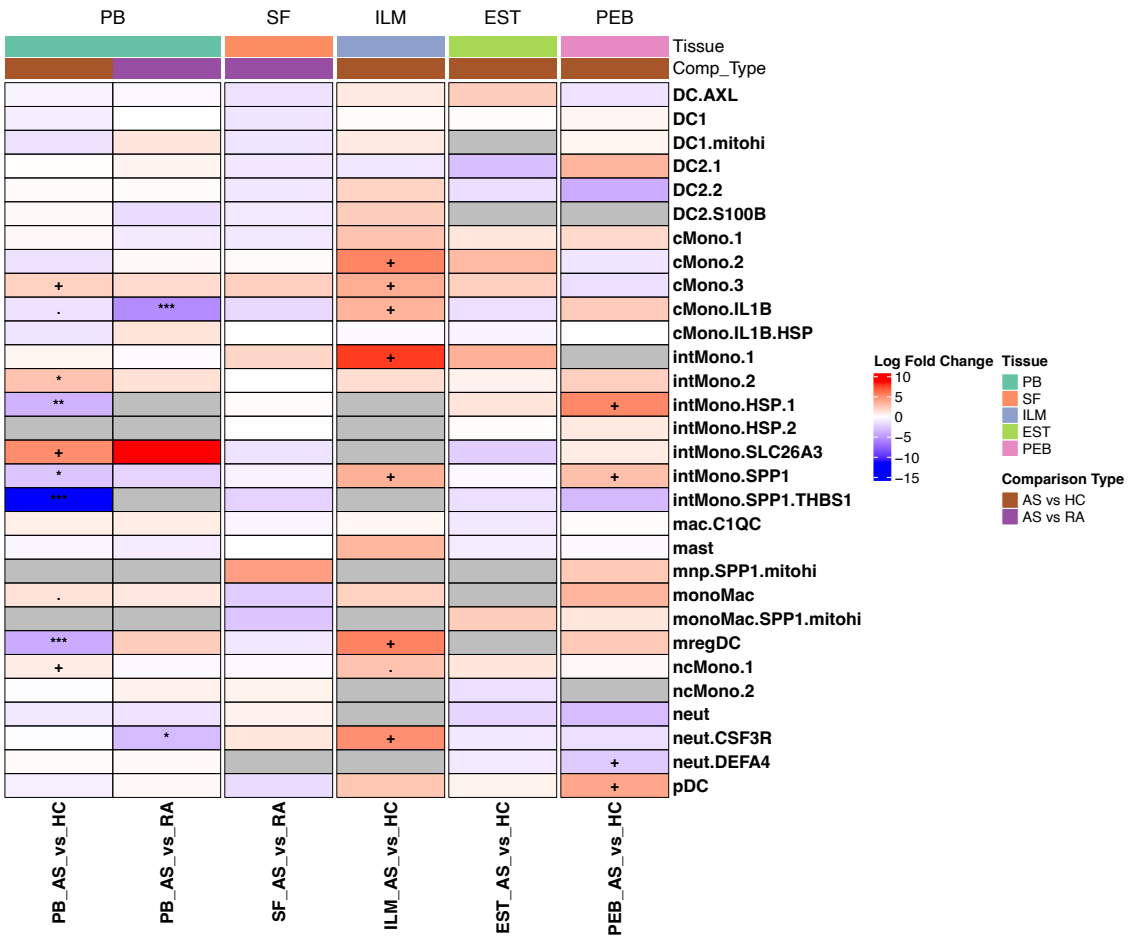


Figure 5.10 Differential abundance of myeloid cell subsets at the cell type level in the expanded cohort
Heatmap showing differential abundance analysis of myeloid cell subsets in the expanded cohort using edgeR across PB, SF, ILM, EST and PEB. Colour indicates logFC of each subset in AxSpA relative to the comparison group; subsets with no cells detected in that tissue for either condition are shown in grey. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

The decrease in mregDCs observed in AxSpA peripheral blood compared to healthy controls was confirmed in the expanded cohort (BH-adjusted $p < 0.01$), whereas an opposing, increasing trend was detected in AxSpA ileum (BH-adjusted $p < 0.1$). Similar patterns of opposing changes between peripheral blood and local tissues were observed for additional myeloid populations, including cMono.IL1B and intMono.SPP1, which decreased in peripheral blood but increased in the ileum. Likewise,

intMono.HSP.1 and intMono.SPP1 exhibited contrasting trends between peripheral blood and peri-entheseal bone. These observations collectively suggest possible tissue-specific redistribution or trafficking of these myeloid populations in AxSpA, though further validation is necessary.

5.3.2.6 B and plasma cell compartment

The edgeR results for the B and plasma cell compartment in the core cohort are shown in **Figure 5.11**, with corresponding box plots for each cell type provided in **Appendices D.32-D.37**.

In the B cell compartment, the most significant change observed was a decrease of a *PRDX1*-expressing activated unswitched memory B cells in AxSpA compared to healthy controls in the blood (BH-adjusted $p < 0.01$). However, this decrease was not observed relative to RA; instead, AxSpA patients showed a slight increase compared to RA (BH-adjusted $p < 0.2$). This suggests that the observed reduction relative to healthy controls may reflect a general inflammatory pattern rather than an AxSpA-specific signature.

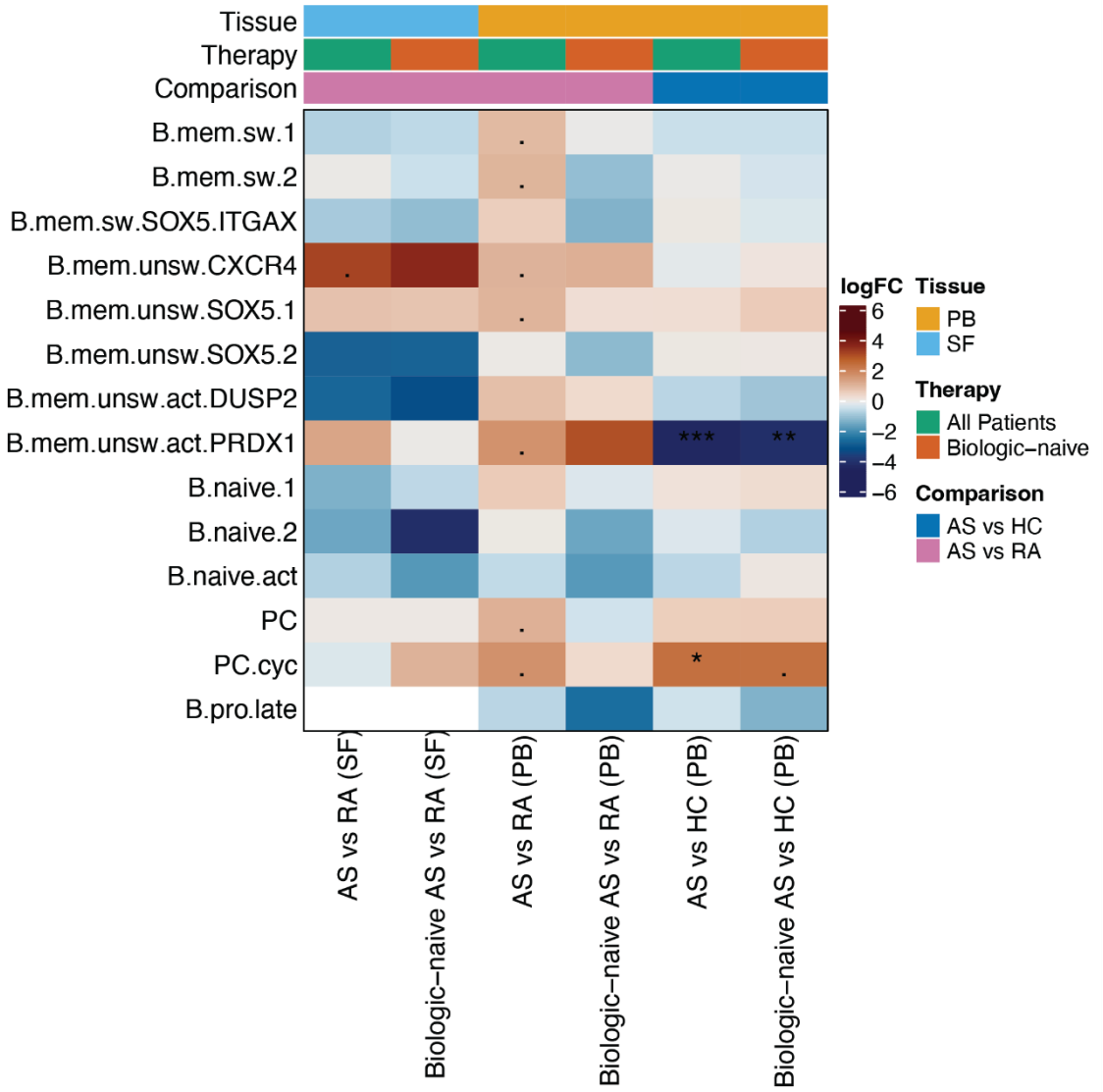


Figure 5.11 Differential abundance of B and plasma cell compartments at the cell type level in the core cohort
Heatmap showing differential abundance analysis of B and plasma cell subsets in the core cohort using edgeR. Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

Also, there was a significant increase in cycling plasma cells in AxSpA compared with both healthy controls and RA in peripheral blood. As described in Chapter 3, total plasma cells were elevated in AxSpA relative to HC, and the current finding suggests that this increase is primarily driven by the expansion of the cycling plasma cell subset. This increase in cycling plasma cells was also

observed when comparing AxSpA with RA in peripheral blood, though the significance was weaker (BH-adjusted $p < 0.2$). This suggests the expansion of cycling plasma cells may represent an AxSpA-specific feature rather than a general inflammatory response shared with RA.

In addition, several memory B-cell subclusters showed moderate increases in AxSpA compared with RA in peripheral blood, consistent with reports that AxSpA exhibits altered circulating memory B-cell subsets (Wilbrink et al., 2021). In synovial fluid, a subset of $CXCR4^{+}$ unswitched memory B cells were enriched in AxSpA compared with RA (BH-adjusted $p < 0.2$), potentially reflecting enhanced tissue retention via the CXCL12–CXCR4 axis and local antigen-experienced humoral activity within inflamed joints. Similar CXCL12–CXCR4–mediated B-cell recruitment and activation have also been described in other autoimmune diseases such as bullous pemphigoid (Fang et al., 2023), suggesting a shared mechanism of tissue-directed B-cell responses.

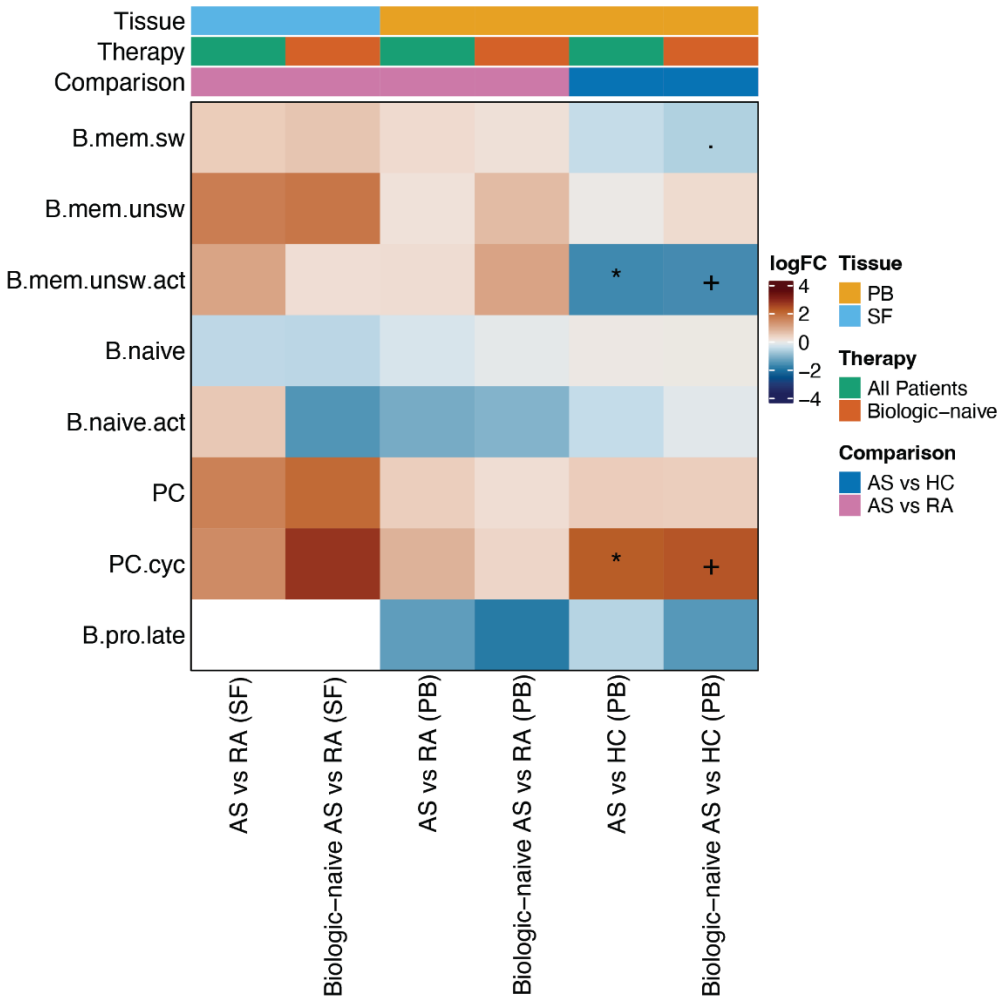


Figure 5.12 Differential abundance of B and plasma cell subsets at the minor subset level in the core cohort
Heatmap showing differential abundance analysis of B and plasma subsets in the core cohort using edgeR at the minor subset level. Colour indicates logFC of each subset in AxSpA relative to the comparison group. Significance: ***<0.001, **<0.01, *<0.05, +<0.1, .<0.2, blank = n.s.

I further explored the changes at a higher level of annotation in the B cell compartment (**Figure 5.12**). When comparing broader lineage-level proportions, the increase in cycling plasma cells in AxSpA compared with healthy controls was confirmed. The activated unswitched B cells were significantly decreased in AxSpA compared with healthy controls in peripheral blood. However, there were no significant differences in the proportions of major B-cell lineages between AxSpA and RA in blood, suggesting that the changes observed in **Figure 5.11**

reflect phenotypic alterations within the broader B-cell compartment rather than overall differences in cell abundance.

5.3.2.7 Technical validation of compositional changes with scCODA and Milo

Next, I used scCODA and Milo as additional validation tools to assess whether similar trends were observed across cell type clusters using different inference methods. Since scCODA requires a reference cell type, I followed the recommended approach of treating each cell type sequentially as the reference and counting the number of credible differences observed to evaluate the relative abundance changes of each cell type in a reference-independent manner.

In the $CD8^+$ T cell compartment, the differential abundance analysis performed using PEB revealed that the CD8.TEMRA.ZNF683.KIR subset exhibited a significantly higher frequency in AxSpA compared with healthy controls (BH-adjusted $p < 0.01$), whereas the CD8.TCMRA.ZNF683.XCL1 subset showed a significant decrease in AxSpA samples (BH-adjusted $p < 0.001$, **Figure 5.5**). In the corresponding scCODA analysis (**Figure 5.13**), both subsets showed strong evidence of differential abundance, being identified as credible 15 times at an BH-adjusted p-value threshold of < 0.2. These findings indicate that the two cell types consistently demonstrated differential abundance regardless of which reference cell type was used, in agreement with the results from the edgeR analysis.

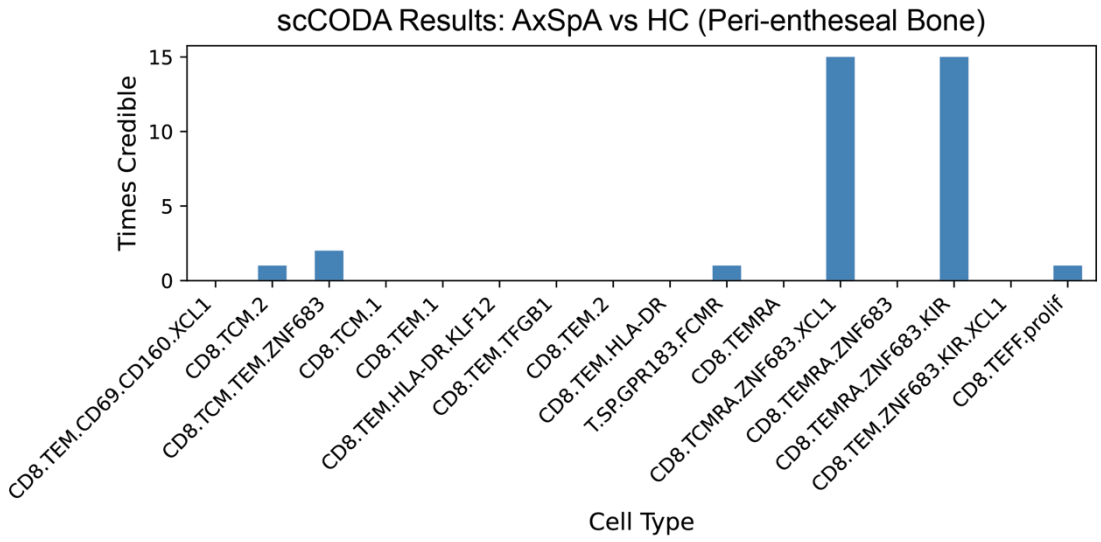


Figure 5.13 scCODA analysis of AxSpA versus healthy controls across CD8⁺ T-cell subsets in PEB in the expanded cohort
Number of cell types identified as credible compositional changes (adjusted p-value < 0.2) when each CD8⁺ T-cell subset was used as the reference cell type in the AxSpA vs HC comparison. Bars indicate the count of credible changes per reference subset, summarised for peri-entheseal bone samples.

I further performed differential abundance testing using Milo, modelling cell neighbourhoods on the neighbourgraph of our single-cell dataset and comparing the abundance of these neighbourhoods between AxSpA and healthy individuals (**Figure 5.14**). In the PBMC, clusters previously highlighted by edgeR (CD4.act.cyc, CD8.TEM.3, CD8.TEM.HLA-DR.KLF12) showed neighbourhoods exclusively enriched in AxSpA. Some clusters, such as UTC.MAIT-like cells, displayed significant neighbourhoods which are differentially abundant in both AxSpA and HC condition, indicating potential heterogeneity within these cell populations.

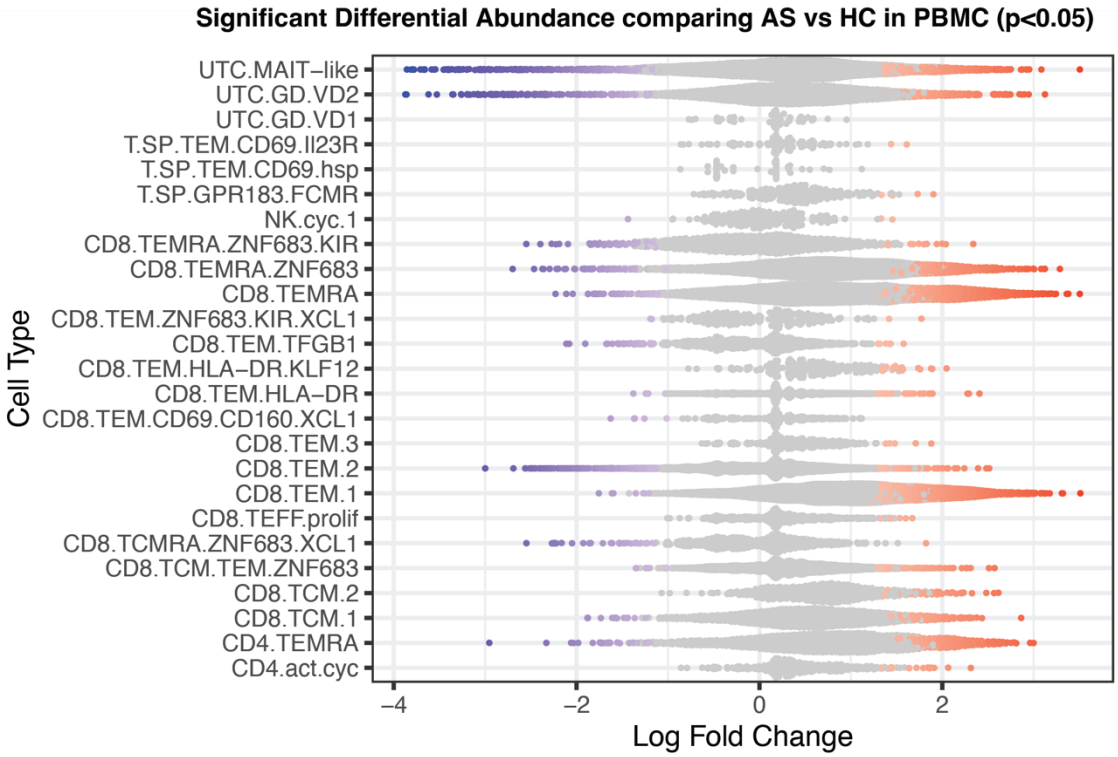


Figure 5.14 Milo differential abundance analysis of the CD8⁺ T-cell region in PBMCs between AxSpA and healthy controls in the expanded cohort
Beeswarm plot showing neighbourhood-level differential abundance results from Milo analysis comparing AxSpA and healthy controls within the CD8⁺ T-cell region. Each point represents a neighbourhood of transcriptionally similar cells, positioned along the x-axis by log fold change. Neighbourhoods showing significant changes (FDR < 0.05) are highlighted.

Consistent with the edgeR findings, the CD8.TEM.3 and CD8.TEM.ZNF683.KIR.XCL1 clusters which were previously shown to be enriched in AxSpA relative to RA in the expanded cohort exhibited significant neighbourhoods, all displaying higher abundance in AxSpA than in RA (**Figure 5.15**).

Among unconventional T cells, the UTC.GD.VD1 cluster, which showed a consistent and significant reduction in AxSpA compared with RA (adjusted p < 0.001) across both core and expanded cohorts in the edgeR analysis, was further supported by Milo validation. Nearly all differential neighbourhoods

within this cluster were significantly enriched in RA (**Figure 5.15**), confirming a robust increase of UTC.GD.VD1 cells under RA conditions.

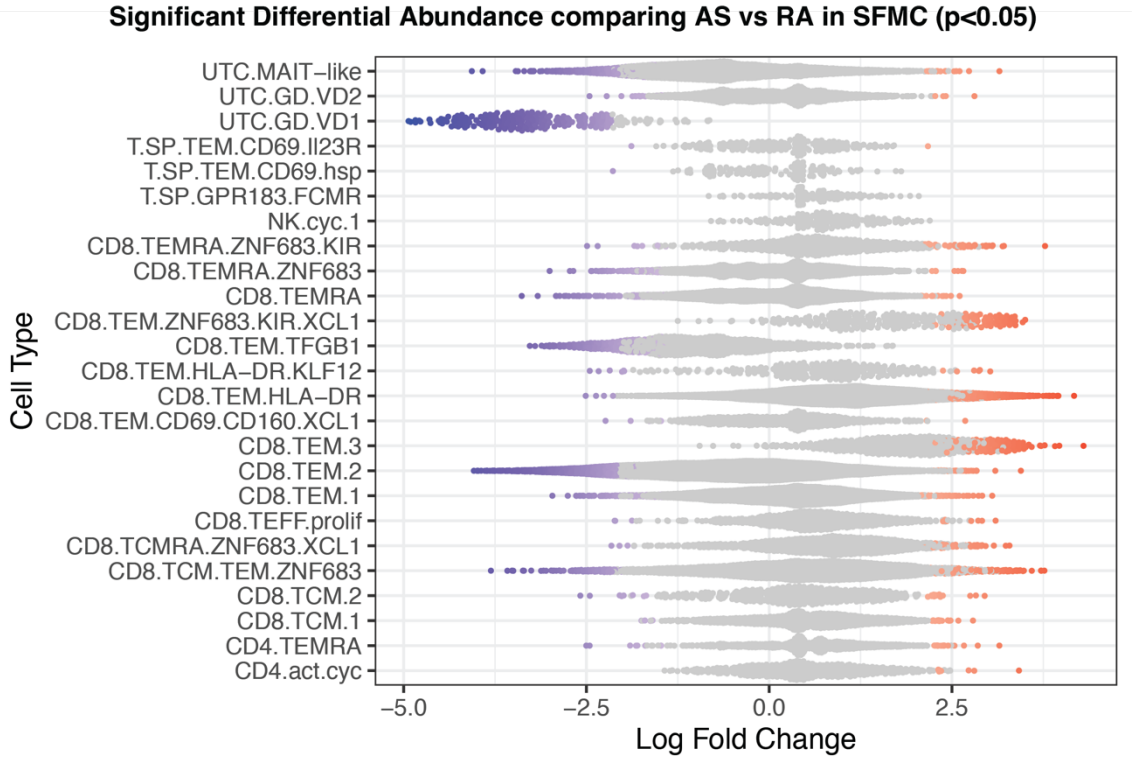


Figure 5.15 Milo differential abundance analysis of the CD8-positive T cell region in SFMCs between AxSpA and RA in the expanded cohort
Beeswarm plot showing neighbourhood-level differential abundance results from Milo analysis comparing AxSpA and healthy controls within the CD8-positive T cell region. Each point represents a neighbourhood of transcriptionally similar cells, positioned along the x-axis by log fold change. Neighbourhoods showing significant changes (FDR < 0.05) are highlighted.

In the myeloid compartment, the cMono.3 cluster showed increased abundance in AxSpA compared with RA within the SFMC samples (adj- $p < 0.2$) in the core cohort, and similarly increased abundance compared with healthy controls in PBMCs from the expanded cohort (BH-adjusted $p < 0.1$). This increase in PBMCs was supported by the scCODA analysis in the expanded cohort, where the cluster consistently demonstrated credible differences across all reference selections (**Figure 5.16**). Furthermore, the Milo analysis identified

representative neighbourhoods within this cluster that were significantly biased towards AxSpA (**Figure 5.17**). The increase observed in SFMCs relative to RA was again confirmed by scCODA results (**Figure 5.18**).

The consistent identification of cMono.3 across core and expanded cohorts, together with its detection in analyses involving both RA and healthy control comparators and validation by scCODA and Milo, strongly indicate that this is a robust and reliable AxSpA-associated monocyte cluster. Notably, cMono.3 also exhibited elevated abundance in ileal tissue compared to healthy controls, suggesting a potential link between peripheral blood and tissue-specific immune signatures.

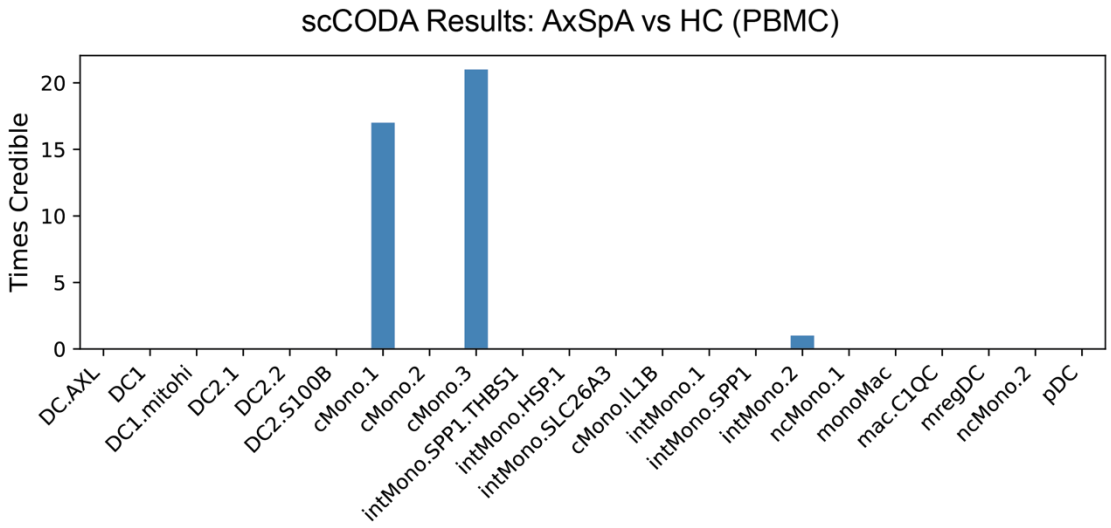


Figure 5.16 scCODA analysis of AxSpA versus healthy controls across myeloid cell components in the expanded cohort
Number of cell types identified as credible compositional changes when each myeloid cell subset was used as the reference cell type in the AxSpA vs HC comparison. Bars indicate the count of credible changes per reference subset in PBMC.

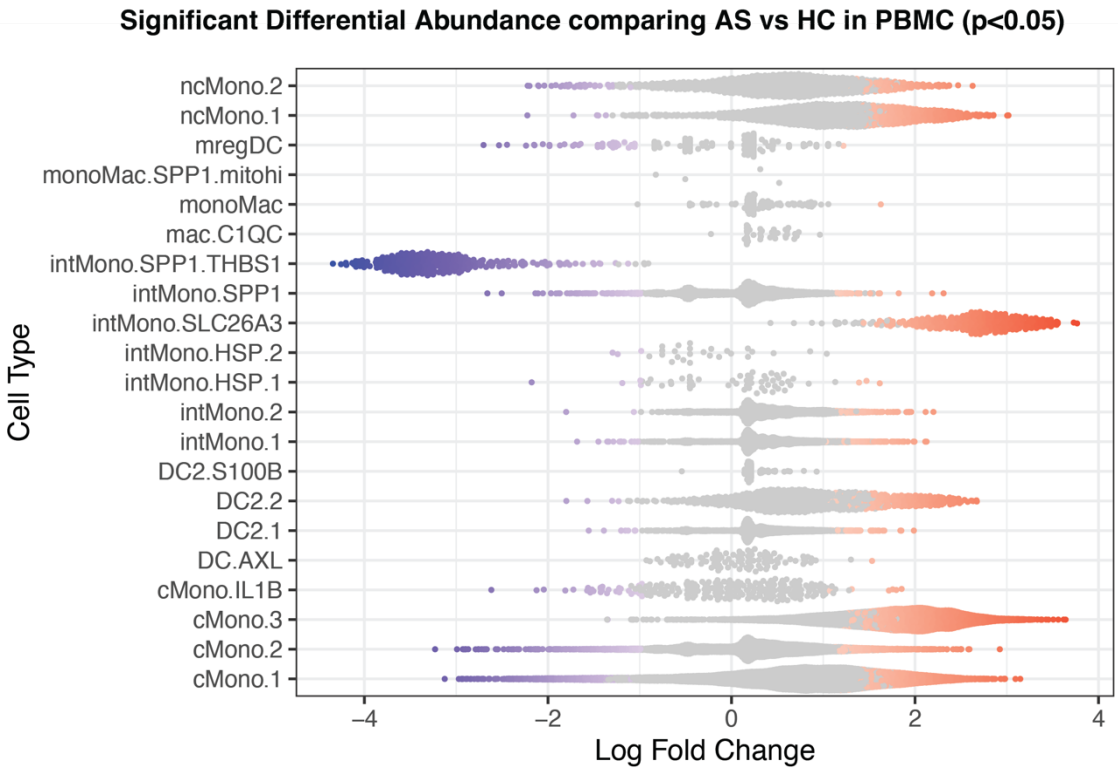


Figure 5.17 Milo differential abundance analysis of the myeloid cell region in PBMCs between AxSpA and healthy controls in the expanded cohort
Beeswarm plot showing neighbourhood-level differential abundance results from Milo analysis comparing AxSpA and healthy controls within the myeloid cell region. Each point represents a neighbourhood of transcriptionally similar cells, positioned along the x-axis by log fold change. Neighbourhoods showing significant changes (FDR < 0.05) are highlighted.

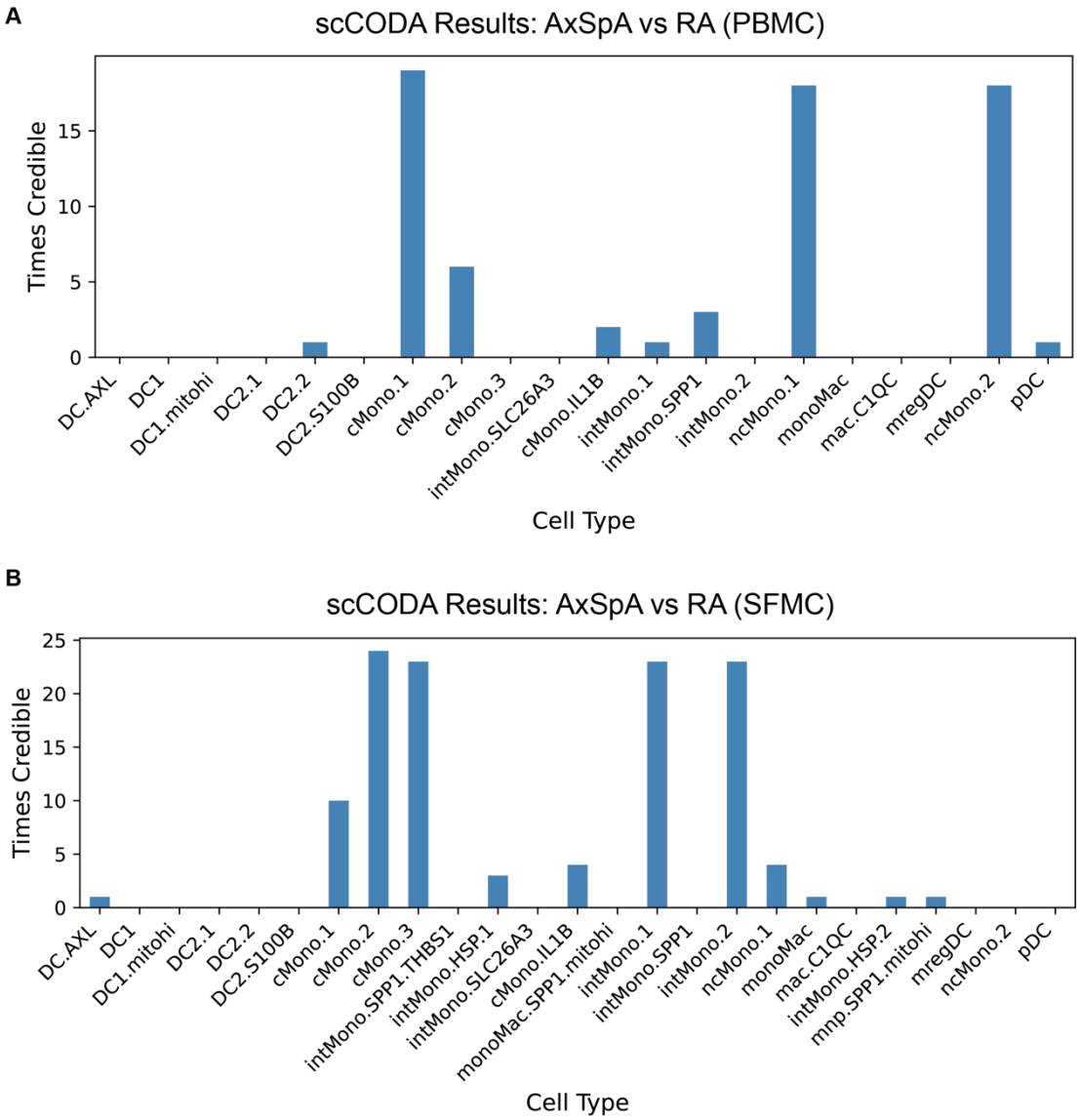


Figure 5.18 scCODA analysis of AxSpA versus RA across myeloid cell components in the expanded cohort
Number of cell types identified as credible compositional changes when myeloid cell subset was used as the reference cell type in the AxSpA vs RA comparison. Bars indicate the count of credible changes per reference subset, summarised for (A) PBMC and (B) SFMC samples.

In addition to cMono.3, another monocyte cluster, cMono.IL1B, was significantly decreased in AxSpA peripheral blood compared with RA in the core cohort (BH-adjusted $p < 0.05$; **Figure 5.8**) and in the expanded cohort (BH-adjusted $p < 0.01$; **Figure 5.10**). This finding was further supported by the Milo analysis, which identified significant neighbourhood differences between AxSpA

and RA in PBMCs (**Figure 5.19A**). However, this decrease was not clearly reproduced in the scCODA analysis (**Figure 5.18**). Notably, cMono.IL1B was previously found to be increased in AxSpA ileal tissue compared with healthy controls (BH-adjusted $p < 0.1$) in the expanded cohort, suggesting that the observed blood-tissue discrepancy may result from altered cellular trafficking in AxSpA.

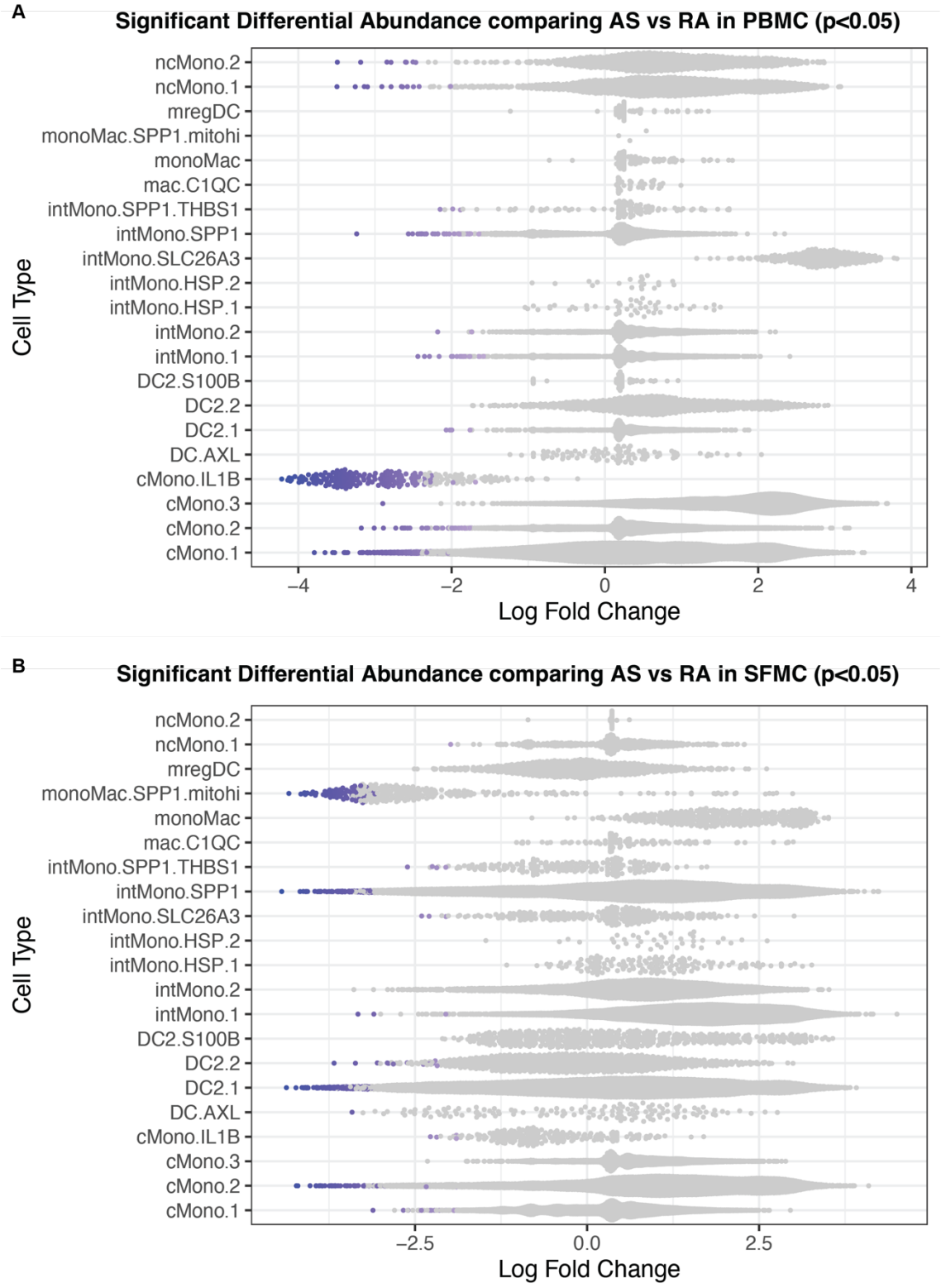


Figure 5.19 Milo differential abundance analysis of the myeloid cell region between AxSpA and RA in the expanded cohort
Beeswarm plot showing neighbourhood-level differential abundance results from Milo analysis comparing AxSpA and healthy controls within the myeloid cell region, in (A) PBMC and (B) SFMC. Each point represents a neighbourhood of transcriptionally similar cells, positioned along the x-axis by log fold change. Neighbourhoods showing significant changes (FDR < 0.05) are highlighted.

5.3.2.8 Co-varying pattern analysis

After examining changes at the single cell-type level, I next sought to identify co-varying patterns of T-cell subsets that distinguish disease from control conditions. Such patterns may reflect coordinated cellular programs or shared activation states among different T-cell clusters. To this end, I performed non-negative matrix factorization (NMF) analysis on synovial fluid T-cell data from expanded cohorts to uncover latent programs representing combinations of T-cell subsets whose abundances co-vary across samples to capturing disease-associated cellular signatures beyond individual cluster-level differences.

I put all non-naïve T cells together as the NMF matrix input, with each proportion still calculated within their own main region for biological meaning. I used multiple criteria to evaluate the optimal factorization rank, **Figure 5.20** shows the process of determination of optimal factorization rank for nmf analysis. There are multiple criteria and recommendations for selecting the optimal factorization rank in NMF (Brunet et al., 2004; Hutchins et al., 2008). I chose rank = 4 because the cophenetic coefficient remained high between ranks 2- 4, and ranks 4–6 showed relatively low dispersion, indicating stable clustering. At this rank, the explained variance was moderate, while improvements in residual sum of squares (RSS) and residuals beyond rank 4 became less pronounced.

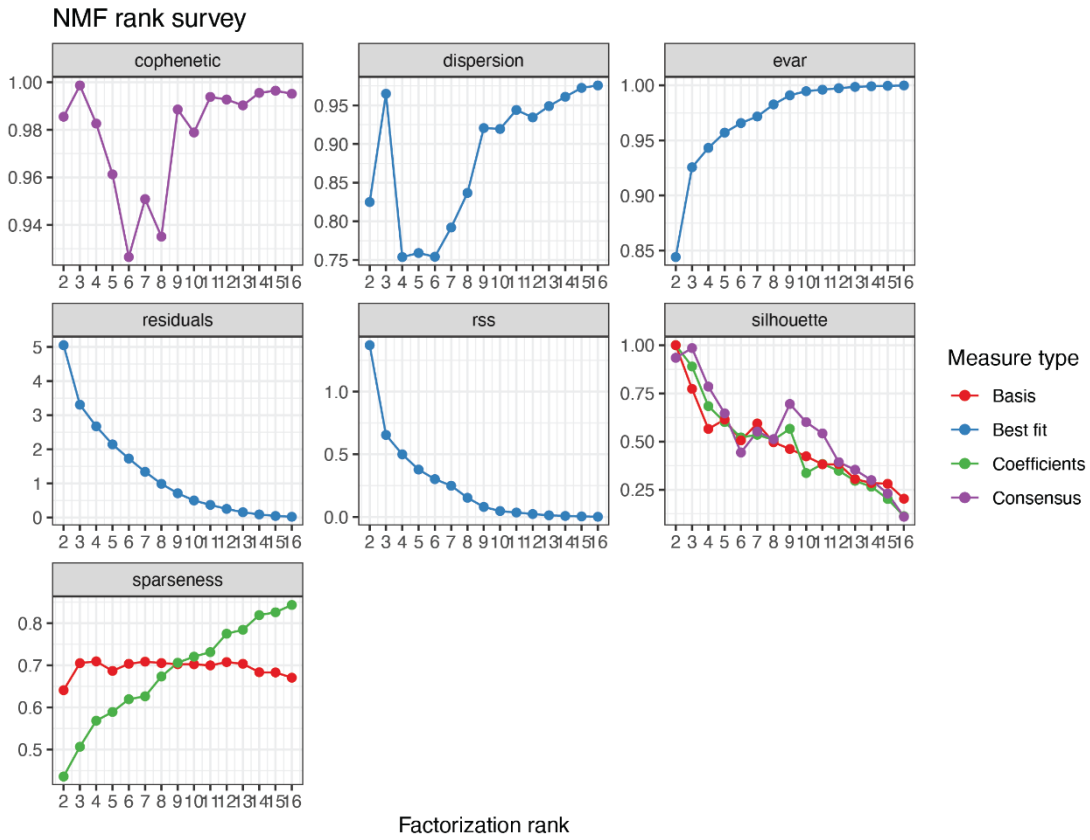


Figure 5.20 Determination of optimal factorization rank for NMF analysis in synovial fluid T cell data
Evaluation of NMF performance across factorization ranks (k = 2–16). Multiple quality metrics were assessed, including cophenetic correlation, dispersion, explained variance (evar), residuals, residual sum of squares (rss), silhouette width, and sparseness of basis and coefficient matrices.

After determining the optimal rank, I performed NMF analysis on synovial fluid T-cell data from AxSpA and RA samples. This analysis identified 4 distinct programs, each representing a unique co-varying pattern of T-cell subsets with varying contribution weights, reflecting coordinated rather than isolated changes in subset abundance (**Figure 5.21A**).

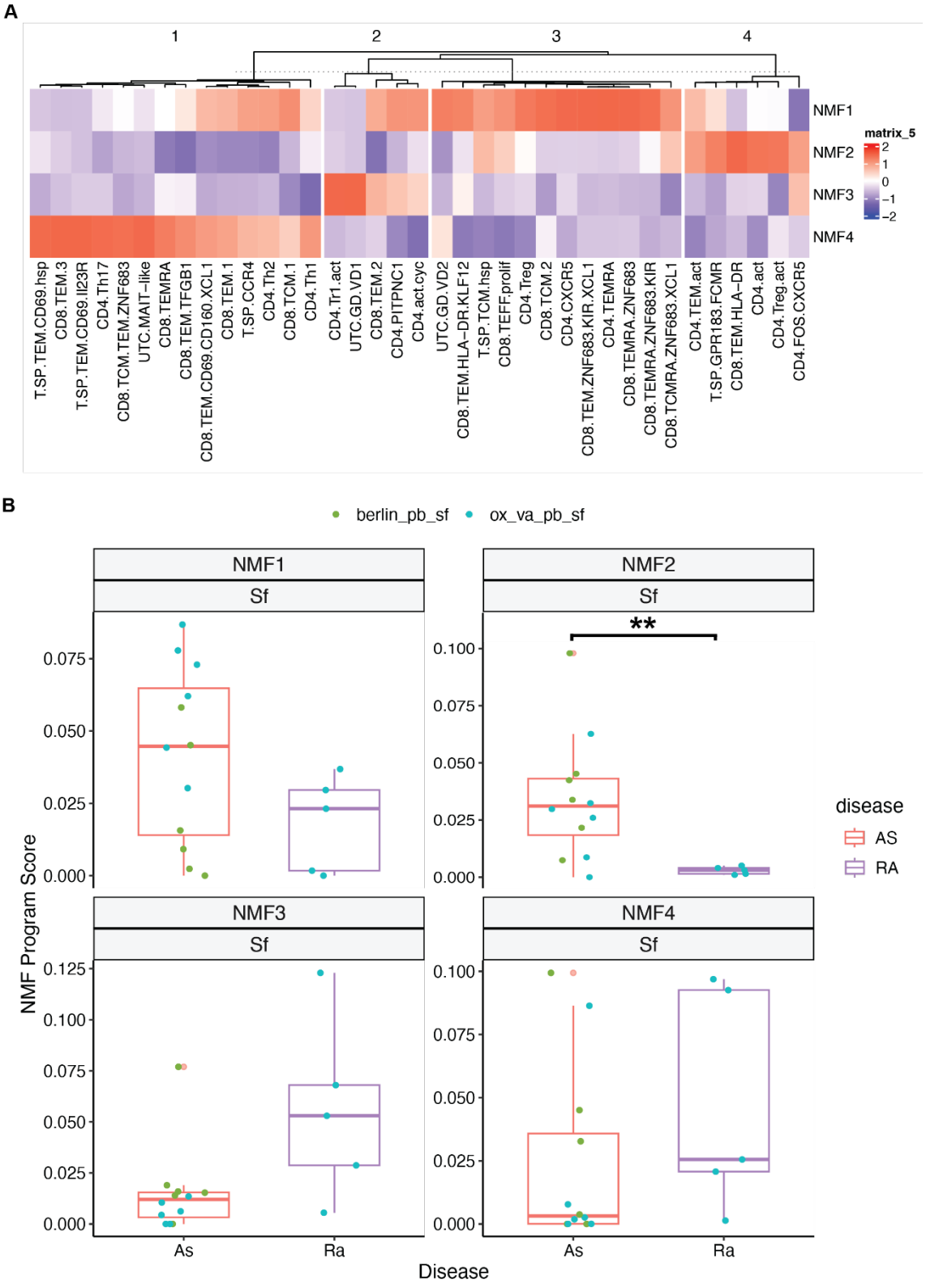


Figure 5.21 NMF identifies disease-associated T-cell programs in synovial fluid in the expanded cohort

(A) Heatmap showing the contribution of non-naïve individual T-cell subsets to each NMF-derived program (NMF1–NMF4) in synovial fluid. Hierarchical clustering grouped correlated subsets into 4 co-varying modules. Each program represents a distinct pattern of subset co-variation across samples, with warm colours indicating stronger positive contributions. (B) Comparison of NMF program scores between AS and RA

synovial fluid samples. Among the identified programs, NMF2 showed a significantly higher program score in AxSpA compared with RA (Student's t-test, BH-adjusted p < 0.01; **). Dot is colored by different projects, with 'ox_va_pb_sf' corresponds to the samples from core cohort.

I then examined how the NMF programs differed between AxSpA and RA in the extended cohort (**Figure 5.21B**). When comparing program scores, NMF2 showed a significantly higher score in AxSpA than in RA (BH-adjusted-p < 0.01). This program was composed of CD8.TEM.HLA-DR, CD4.TEM.act, CD4.act, CD4.Treg.act, CD4.FOS.CXCR5, and T.SP.GPR183.FCMR subsets contributing positively to the program. Comparisons of each NMF program in the core cohort are shown separately in **Appendix D.41**. In the core cohort, this program showed a similar trend (unadjusted p = 0.044, BH-adjusted p = 0.088).

The NMF1 program did not show a statistically significant difference in the extended cohort; however, within the core cohort, the t-test yielded a significant result (unadjusted p = 0.004, BH-adjusted p = 0.016). This program comprises several ZNF683-positive TEM/TEMRA clusters previously identified in the edgeR analysis, including CD8.TEMRA.ZNF683, CD8.TEMRA.ZNF683.KIR, CD8.TCMRA.ZNF683.XCL1, as well as the CD4.TEMRA population, all of which were significantly enriched in AxSpA synovial fluid compared with RA in either the core or expanded cohorts. Together, these findings indicate a coordinated enrichment of late-differentiated effector T-cell states in AxSpA, reflecting a shared activation pattern rather than isolated changes in individual subsets.

The NMF3 program comprises positively contributing CD4.Tr1.act cells and UTC.GD.VD1 cells, both of which were identified in earlier analyses as RA-associated populations. In the core cohort, the NMF3 program score was higher

in RA than in AxSpA (unadjusted p = 0.066, BH-adjusted p = 0.088), consistent with this pattern although not reaching statistical significance. These results indicate that NMF3 captures a regulatory and innate-like module that is more characteristic of RA synovial inflammation.

The NMF4 program contains CD8.TEM.3, CD4.Th17, T.SP.TEM.CD69.HSP, and UTC.MAIT-like subsets. These populations represent activated effector and mucosal-associated T-cell states, several of which showed signals of relevance to AxSpA in tissue-level GWAS-associated analyses. The combined behavior of these subsets within NMF4 suggests a program capturing mucosal-influenced effector differentiation, potentially linking Th17-associated activity, MAIT-like responses, and activated $CD8^+$ TEM states into a coordinated module. When evaluated in either the extended or core cohort, NMF4 did not show significant differences between AxSpA and RA, indicating that this program may reflect shared activation pathways or tissue-imprinted effector states rather than disease-specific compositional shifts in synovial fluid. Nonetheless, its structure highlights a biologically meaningful axis of effector–mucosal signaling that may warrant functional investigation beyond simple abundance differences.

5.3.2.9 Cross-tissue comparison

Next, I investigated the cellular composition across tissues affected in AxSpA to identify common pathological cell phenotypes among inflamed sites, while also exploring tissue-specific features that may underlie local inflammatory responses. For this analysis, I included sacroiliac joint samples which are available only from AxSpA patients, which were available only from AxSpA patients, as this tissue is typically the first site where disease is clinically

evident and plays a key role in disease diagnosis. I compared cellular compositions across AxSpA-affected tissues by fitting Poisson regression models to per-sample count data, adjusting for donor-specific effects, and identified tissue-specific enrichment based on significant positive coefficients (adjusted $p < 0.05$). A significant coefficient for a given cell type indicates that this population is relatively enriched in that tissue compared with all other tissues included in the model.

In the CD8-positive T-cell compartment, several clusters were enriched in the sacroiliac joint (SIJ), including cell types previously identified as significantly altered in other comparisons (**Figure 5.22**). Specifically, CD8.TCMRA.ZNF683.XCL1, which was previously observed to be increased in biologic-naïve AxSpA blood and ileum relative to healthy controls, was significantly enriched in SIJ samples. Similarly, CD8.TEM.HLA-DR.KLF12, found to be increased in peripheral blood and ileum from AxSpA patients compared to controls, showed enrichment in both SIJ and synovial fluid. These observations suggest that certain CD8-positive T-cell populations enriched in SIJ may also exhibit elevated abundance across peripheral blood and ileal tissues in AxSpA.

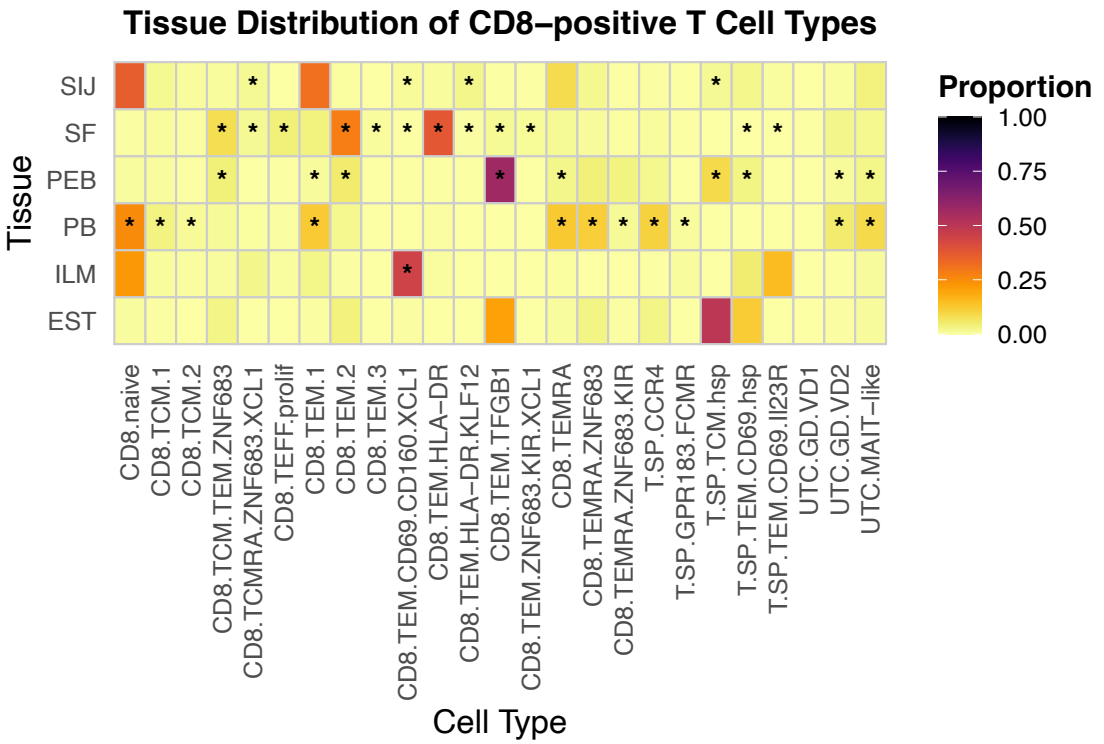


Figure 5.22 Tissue-associated enrichment of CD8⁺ T-cell compartment across AxSpA tissues
Heatmap showing differences in the relative abundance of CD8⁺ T-cell subsets across AxSpA-affected tissues. Significance was assessed using Poisson regression with donor adjustment, and subsets with positive coefficients (adjusted $p < 0.05$) were considered enriched in the corresponding tissue.

In the myeloid compartment, several notable clusters were enriched in the sacroiliac joint (SIJ), including cMono.IL1B.HSP, IntMono.2, and Mac.C1QC (**Figure 5.23**). Additionally, a neutrophil cluster neut.CSF3R with elevated expression of *IL1B* was significantly enriched in the SIJ. This neutrophil cluster had previously shown increased abundance in biologic-naïve AxSpA peripheral blood compared to RA, as well as enrichment in AxSpA ileum relative to healthy ileum, highlighting its consistent association across multiple sites.

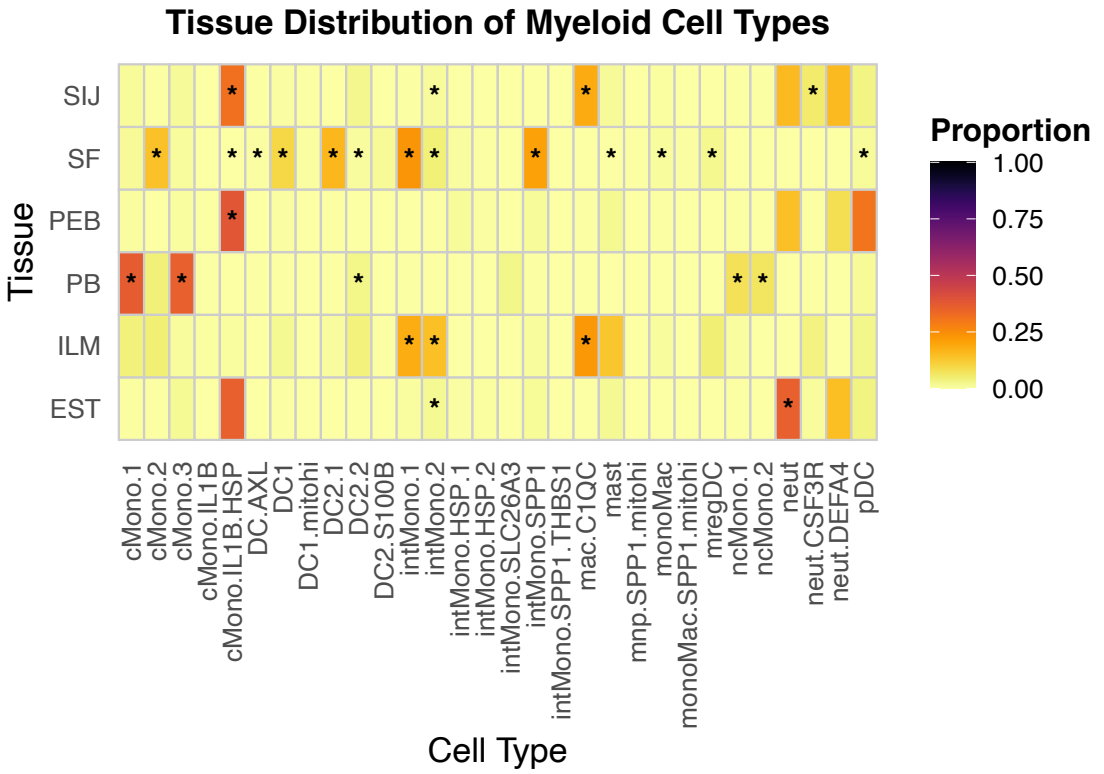


Figure 5.23 Tissue-associated compositional differences in the myeloid cell compartment
Heatmap showing differences in the relative abundance of myeloid cell subsets across AxSpA-affected tissues. Significance was assessed using Poisson regression with donor adjustment, and subsets with positive coefficients (adjusted $p < 0.05$) were considered enriched in the corresponding tissue.

Within the B and plasma cell compartment, plasma cells were enriched in the sacroiliac joint, peri-entheseal bone, and ileum (**Figure 5.24**). In addition, cycling plasma cells were increased in AxSpA peripheral blood compared with healthy controls. These observations together show a higher representation of plasma cells across multiple AxSpA tissues and in circulation.

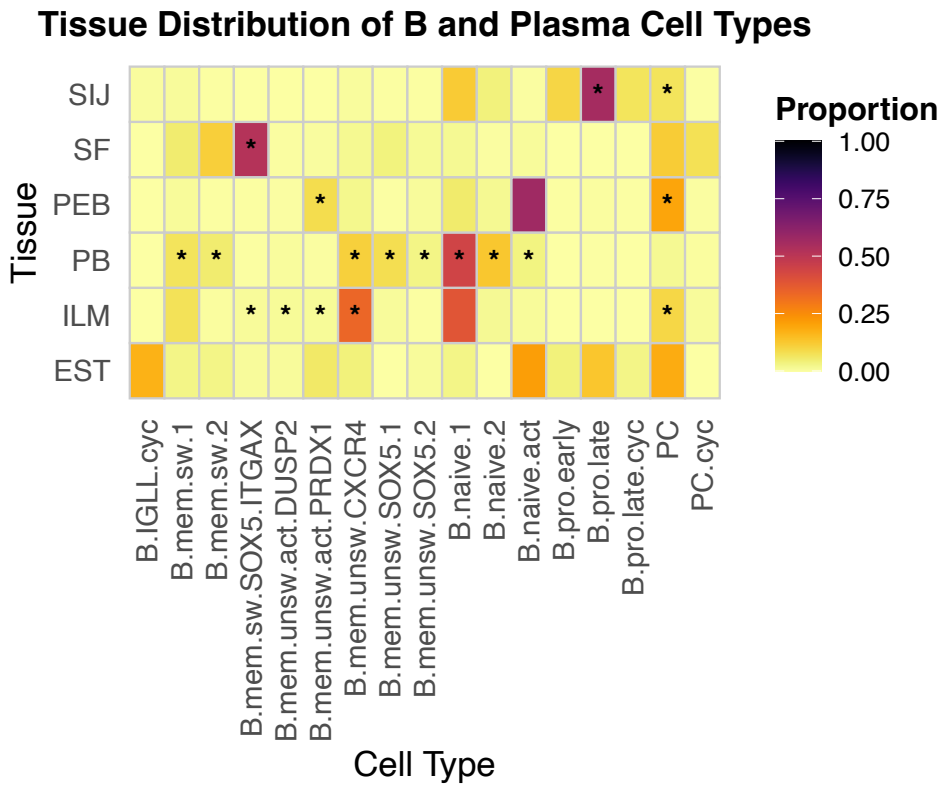


Figure 5.24 Tissue-associated compositional differences in the B cell compartment
Heatmap showing differences in the relative abundance of B cell subsets across AxSpA-affected tissues. Significance was assessed using Poisson regression with donor adjustment, and subsets with positive coefficients (adjusted $p < 0.05$) were considered enriched in the corresponding tissue.

The remaining cell compartments, including $CD4^+$ T cells, NK cells, and other populations, are shown in **Appendices D.38-D.40.**

5.4 Discussion

In this section, I integrate compositional analyses performed across multiple AxSpA-associated tissues to explore the differential contribution of each immune cell type within distinct tissue environments. By combining several analytical frameworks including edgeR, scCODA, Milo, NMF, and cross-tissue Poisson regression, applied to both the core and expanded cohorts, this chapter identifies reproducible AxSpA-associated signatures, tissue-restricted features,

and compartments without consistent disease-specific patterns. Among these, ZNF683^{+} CD8^{+} TEM/TEMRA populations and the cMono.3 monocyte subset emerged as the most robust cross-tissue and cross-method signatures; the neut.CSF3R^{+}IL1B^{+} cluster and plasma cells showed concordant cross-tissue expansion; tissue-restricted features included γδ.Vδ2 T cells and intMono.SPP1 monocytes in peri-entheseal bone, and a CXCR4^{+} unswitched memory B-cell subset in synovial fluid; the NK cell and CD4^{+} T-cell compartments showed no widespread AxSpA-specific abundance changes. The following sections discuss each of these findings in detail, alongside RA-associated signals consistent with prior literature, methodological caveats, and future validation strategies.

Several of the observations in RA aligned with established reports, providing confidence in the validity of our analyses. The CD4.Tr1.act cluster was elevated in RA synovial fluid and peripheral blood compared to AxSpA, consistent with previous reports describing increased PD-1^{hi} CD4 positive T cells in RA (Kobayashi et al., 2013). The T.SP.CCR4 population was upregulated in AxSpA compared with healthy controls but not compared with RA, supporting the observation that this subset is increased in both diseases and may reflect a shared inflammatory signature rather than an AxSpA-specific feature (Yang et al., 2004). The γδ1 T-cell population was significantly higher in RA synovial fluid compared with AxSpA, again consistent with previous evidence showing that γδ T cells with a repertoire biased towards Vδ1 preferentially accumulate in the rheumatoid arthritis synovium (Olive et al., 1992). These disease-specific differences suggest that, although both RA and AxSpA involve synovial inflammation, they are characterized by distinct T-cell phenotypes and effector responses within the joint environment.

Among T-cell populations, CD8.TEM.3 showed a mild increase in AxSpA peripheral blood (adj. $p < 0.2$) and consistent neighborhood-level upregulation in Milo analysis. This cluster was also more abundant in AxSpA synovial fluid compared with RA and was uniquely detected in the synovial fluid among tissue comparisons, suggesting tissue-specific effector enrichment.

Another notable T-cell subset, CD8.TEMRA.ZNF683.KIR, was significantly increased in AxSpA PEB and showed trends toward upregulation in ileum and synovial fluid, representing a potential cross-tissue signal. The enrichment of these ZNF683$^+$ CD8$^+$ TEMRA populations across tissues is particularly interesting in light of the known role of *ZNF683*, also known as *Hobit*, in programming tissue-resident and cytotoxic effector functions. Consistent with the findings in both core and expanded cohorts - where ZNF683$^+$ subsets such as CD8.TEMRA.ZNF683, CD8.TEMRA.ZNF683.KIR, and CD8.TEMRA.ZNF683.KIR.XCL1 were increased in AxSpA compared with RA - this pattern likely reflects an accumulation of locally retained, effector-biased CD8$^+$ T cells exhibiting transcriptional programs that favour tissue residency and activation rather than recirculation. Such a phenotype is in line with the concept of tissue-resident memory-like (TRM/TEMRA-like) CD8$^+$ T cells engaged in chronic or tissue-adapted immune responses in AxSpA. KIR$^+$ CD8$^+$ T cells have been reported to possess regulatory potential and are increased across several autoimmune diseases (Li et al., 2022). Their expansion may therefore represent a compensatory regulatory or exhaustion-prone response to persistent antigenic stimulation within the inflamed microenvironment, possibly coexisting with cytotoxic activity. Together, these findings suggest a dynamic equilibrium between effector and regulatory functions within the CD8$^+$

compartment in AxSpA, shaped by tissue adaptation and chronic inflammatory signalling.

Within unconventional T cells, the UTC.γδ.Vδ2 population was enriched specifically in PEB in AxSpA. Vδ2^{+} γδ T cells in the enthesis have been reported to express IL23R, RORC, and CCR6, and to produce IL-17A/F upon stimulation (Cuthbert et al., 2019). Their selective expansion in AxSpA PEB may therefore reflect proliferation of IL-23/IL-17-responsive, tissue-resident γδ T cells, consistent with the IL-23/IL-17 axis as a key pathogenic pathway in spondyloarthritis.

Regarding NMF analysis, the decomposition revealed several coherent co-varying programs that captured biologically meaningful T-cell modules across synovial fluid samples. NMF2 emerged as the most AxSpA-associated program in the extended cohort, driven by coordinated increases in activated CD4^{+} and CD8^{+} effector populations, and showed significant elevation in AxSpA compared with RA in the extended cohort. NMF1 reflected a late-differentiated ZNF683^{+} TEM/TEMRA module that was enriched in AxSpA in the core cohort, consistent with compositional findings from edgeR. In contrast, NMF3 represented a regulatory and innate-like module enriched in RA, while NMF4 encompassed mucosal-associated effector states, suggesting a shared activation axis rather than disease-specific abundance changes. Together, these programs highlight that AxSpA is characterised not only by alterations in individual subsets but also by coordinated shifts in broader effector and regulatory T-cell modules.

Myeloid populations also showed tissue- and disease-specific associations. A consistent and AxSpA-specific signature identified across both

the core and expanded cohorts was the cMono.3 subset. This population was increased in AxSpA peripheral blood compared with RA controls in the core cohort (including both biologic-naïve and overall groups) and similarly elevated in peripheral blood relative to healthy controls in the expanded cohort. Concordant upregulation of this subset was also observed in the ileum, supporting its potential involvement in both systemic and tissue-level inflammatory processes. The defining feature of cMono.3 was the elevated expression of *MNDA*, a key regulator of the type I interferon cascade in human myeloid cells (Gu et al., 2022). The increased *MNDA* expression suggests that cMono.3 monocytes may exist in an IFN-stimulated, transcriptionally activated state, potentially amplifying IFNα responses through the IRF7 pathway in both blood and gut compartments.

In peri-entheseal bone and ileum, intMono.SPP1 and mregDC clusters exhibited significant AxSpA genetic correlation signals, while these associations were absent in synovial fluid and blood, suggesting shared inflammatory transcriptional programs between gut and enthesis. The neut.CSF3R cluster, characterized by high expression of *CSF3R* and *IL1B*, was significantly increased in AxSpA compared with RA in peripheral blood and elevated in ileum relative to healthy controls. This subset was also enriched in the sacroiliac joint, suggesting that neutrophil activation may bridge systemic and local inflammation. Previous studies have proposed that neutrophils play a role in the initiation of pathogenic IL-23/IL-17 responses and disease development in spondyloarthritis which supports this interpretation (Coletto et al., 2023; Stavre et al., 2022). However, the number of cells within this cluster was relatively low,

and the observed trends should therefore be interpreted with caution until validated in larger datasets.

When discussing B and plasma cell compartments, plasma cells were enriched in the sacroiliac joint, peri-entheseal bone, and ileum, and cycling plasma cells were increased in peripheral blood compared with healthy controls. This widespread presence of plasma cell subsets suggests sustained humoral activation across systemic and local compartments in AxSpA.

Each compositional analysis method has inherent limitations. While edgeR provides a powerful framework for modelling compositional differences, it can overestimate fold changes in rare populations. In contrast, scCODA captures relative abundance shifts but lacks direct abundance scaling, and Milo detects neighbourhood-level variations that may not always translate to sample-level differences. Therefore, scCODA and Milo were employed primarily as technical validation frameworks to support the edgeR discoveries rather than as independent discovery methods. Concordant results across these complementary approaches were considered the most robust. Nevertheless, compositional biases remain intrinsic to single-cell data and should be interpreted with caution.

Cross-tissue comparisons also carry methodological caveats due to differences in tissue processing. Most PBMCs and SFMCs were isolated from frozen samples, whereas ileal, enthesis, and sacroiliac joint tissues required enzymatic digestion and distinct dissociation protocols. These technical variations likely contribute to compositional bias, and absolute comparisons between tissues should therefore be interpreted qualitatively rather than quantitatively.

A further methodological consideration is the potential influence of treatment differences on the observed compositional patterns. Although biologic-naïve subgroups were analysed separately for the main AxSpA versus RA comparisons to mitigate the effect of biologic therapy, detailed information on other concomitant medications, including corticosteroids and non-biologic DMARDs, was not consistently available across samples. Some of the differences observed between AxSpA and RA in this chapter may therefore partially reflect treatment-related effects on circulating immune populations rather than disease-specific compositional shifts. For example, the increased Th1 and Th17 subsets observed in AxSpA peripheral blood relative to RA but not relative to healthy controls could be consistent with such a confounder. Cohorts with more detailed treatment annotation will be needed to distinguish disease-specific signatures from treatment-related effects.

Beyond these methodological caveats, the observations made here will be further explored through complementary analyses performed by other team members, including differential expression profiling, TCR repertoire characterization and deeper tissue-specific investigations. Such analyses will further refine mechanistic insights and facilitate the validation of key cellular findings identified in this study.

To strengthen these conclusions, future validation could involve analysing independent patient cohorts using orthogonal technologies such as single-cell RNA sequencing, CyTOF, or flow cytometry to confirm the identified cell populations and their abundance patterns. In addition, spatial transcriptomics or imaging-based approaches could be employed to assess the localization and cellular interactions of these subsets within inflamed tissues,

providing spatial context to the molecular and compositional signatures observed here.

Collectively, these integrated approaches have revealed both shared and tissue-specific immune programs in AxSpA, highlighting candidate cellular mediators that connect inflammation across gut, bone, joint, and blood through IL-, TNF-, and IFN-driven immune circuits, thus providing potential therapeutic targets and pathways underlying disease propagation and regulation.

5.5 References

Al-Osami, M. H., Awadh, N. I., Khalid, K. B., & Awadh, A. I. (2020). Neutrophil/lymphocyte and platelet/lymphocyte ratios as potential markers of disease activity in patients with Ankylosing spondylitis: a case-control study. *Adv Rheumatol*, *60*(1), 13. https://doi.org/10.1186/s42358-020-0113-5

Brunet, J. P., Tamayo, P., Golub, T. R., & Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A*, *101*(12), 4164–4169. https://doi.org/10.1073/pnas.0308531101

Buttner, M., Ostner, J., Muller, C. L., Theis, F. J., & Schubert, B. (2021). scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun*, *12*(1), 6876. https://doi.org/10.1038/s41467-021-27150-6

Ciccia, F., Guggino, G., Rizzo, A., Alessandro, R., Luchetti, M. M., Milling, S., Saieva, L., Cypers, H., Stampone, T., Di Benedetto, P., Gabrielli, A., Fasano, A., Elewaut, D., & Triolo, G. (2017). Dysbiosis and zonulin upregulation alter gut epithelial and vascular barriers in patients with ankylosing spondylitis. *Ann Rheum Dis*, *76*(6), 1123–1132. https://doi.org/10.1136/annrheumdis-2016-210000

Coletto, L. A., Rizzo, C., Guggino, G., Caporali, R., Alivernini, S., & D'Agostino, M. A. (2023). The Role of Neutrophils in Spondyloarthritis: A Journey across the Spectrum of Disease Manifestations. *Int J Mol Sci*, *24*(4). https://doi.org/10.3390/ijms24044108

Costello, M. E., Ciccia, F., Willner, D., Warrington, N., Robinson, P. C., Gardiner, B., Marshall, M., Kenna, T. J., Triolo, G., & Brown, M. A. (2015). Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis. *Arthritis Rheumatol*, *67*(3), 686–691. https://doi.org/10.1002/art.38967

Cuthbert, R. J., Watad, A., Fragkakis, E. M., Dunsmuir, R., Loughenbury, P., Khan, A., Millner, P. A., Davison, A., Marzo-Ortega, H., Newton, D., Bridgewood, C., & McGonagle, D. G. (2019). Evidence that tissue resident human enthesis gammadeltaT-cells can produce IL-17A independently of IL-23R transcript expression. *Ann Rheum Dis*, *78*(11), 1559–1565. https://doi.org/10.1136/annrheumdis-2019-215210

Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D., & Marioni, J. C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol*, *40*(2), 245–253. https://doi.org/10.1038/s41587-021-01033-z

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., Park, Y. R., Raychaudhuri, S., Pouget, J. G., Hubenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H. J., Franke, L., Pers, T. H., Weersma, R. K., Collij, V.,…Franke, A. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*, *48*(5), 510–518. https://doi.org/10.1038/ng.3528

Fang, H., Xue, K., Cao, T., Li, Q., Dang, E., Liu, Y., Zhang, J., Qiao, P., Chen, J., Ma, J., Shen, S., Pang, B., Bai, Y., Qiao, H., Shao, S., & Wang, G. (2023). CXCL12/CXCR4 Axis Drives the Chemotaxis and Differentiation of B Cells in Bullous Pemphigoid. *J Invest Dermatol*, *143*(2), 197–208 e196. https://doi.org/10.1016/j.jid.2022.08.041

Gracey, E., Vereecke, L., McGovern, D., Frohling, M., Schett, G., Danese, S., De Vos, M., Van den Bosch, F., & Elewaut, D. (2020). Revisiting the gut-joint axis: links between gut inflammation and spondyloarthritis. *Nat Rev Rheumatol*, *16*(8), 415–433. https://doi.org/10.1038/s41584-020-0454-9

Gu, L., Casserly, D., Brady, G., Carpenter, S., Bracken, A. P., Fitzgerald, K. A., Unterholzner, L., & Bowie, A. G. (2022). Myeloid cell nuclear differentiation antigen controls the pathogen-stimulated type I interferon cascade in human monocytes by transcriptional regulation of IRF7. *Nat Commun*, *13*(1), 14. https://doi.org/10.1038/s41467-021-27701-x

Guggino, G., Rizzo, A., Mauro, D., Macaluso, F., & Ciccia, F. (2021). Gut-derived CD8(+) tissue-resident memory T cells are expanded in the peripheral blood and synovia of SpA patients. *Ann Rheum Dis*, *80*(11), e174. https://doi.org/10.1136/annrheumdis-2019-216456

Hecquet, S., Totoson, P., Martin, H., Prati, C., Wendling, D., Demougeot, C., & Verhoeven, F. (2021). Intestinal permeability in spondyloarthritis and rheumatoid arthritis: A systematic review of the literature. *Semin Arthritis Rheum*, *51*(4), 712–718. https://doi.org/10.1016/j.semarthrit.2021.04.015

Hutchins, L. N., Murphy, S. M., Singh, P., & Graber, J. H. (2008). Position-dependent motif characterization using non-negative matrix factorization. *Bioinformatics*, *24*(23), 2684–2690. https://doi.org/10.1093/bioinformatics/btn526

Kobayashi, S., Murata, K., Shibuya, H., Morita, M., Ishikawa, M., Furu, M., Ito, H., Ito, J., Matsuda, S., Watanabe, T., & Yoshitomi, H. (2013). A distinct human CD4+ T cell subset that secretes CXCL13 in rheumatoid synovium. *Arthritis Rheum*, *65*(12), 3063–3072. https://doi.org/10.1002/art.38173

Li, J., Zaslavsky, M., Su, Y., Guo, J., Sikora, M. J., van Unen, V., Christophersen, A., Chiou, S. H., Chen, L., Li, J., Ji, X., Wilhelmy, J., McSween, A. M., Palanski, B. A., Mallajosyula, V. V. A., Bracey, N. A., Dhondalay, G. K. R., Bhamidipati, K., Pai, J.,…Davis, M. M. (2022). KIR(+)CD8(+) T cells suppress pathogenic T cells and are active in autoimmune diseases and COVID-19. *Science*, *376*(6590), eabi9591. https://doi.org/10.1126/science.abi9591

Li, M., Yao, D., Zeng, X., Kasakovski, D., Zhang, Y., Chen, S., Zha, X., Li, Y., & Xu, L. (2019). Age related human T cell subset evolution and senescence. *Immun Ageing*, *16*, 24. https://doi.org/10.1186/s12979-019-0165-8

Liang, J., Liao, L., Xie, L., Tang, W., Yu, X., Lu, Y., Chen, H., Xu, J., Sun, L., Wu, H., Cui, C., & Tan, Y. (2024). PITPNC1 Suppress CD8(+) T cell immune function and promote radioresistance in rectal cancer by modulating FASN/CD155. *J Transl Med*, *22*(1), 117. https://doi.org/10.1186/s12967-024-04931-3

Mackay, L. K., Minnich, M., Kragten, N. A., Liao, Y., Nota, B., Seillet, C., Zaid, A., Man, K., Preston, S., Freestone, D., Braun, A., Wynne-Jones, E., Behr, F. M., Stark, R., Pellicci, D. G., Godfrey, D. I., Belz, G. T., Pellegrini, M., Gebhardt, T.,…van Gisbergen, K. P. (2016). Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science*, *352*(6284), 459–463. https://doi.org/10.1126/science.aad2035

Martini, V., Silvestri, Y., Ciurea, A., Moller, B., Danelon, G., Flamigni, F., Jarrossay, D., Kwee, I., Foglierini, M., Rinaldi, A., Cecchinato, V., & Uguccioni, M. (2024). Patients with ankylosing spondylitis present a distinct CD8 T cell subset with osteogenic and cytotoxic potential. *RMD Open*, *10*(1). https://doi.org/10.1136/rmdopen-2023-003926

Olive, C., Gatenby, P. A., & Serjeantson, S. W. (1992). Variable gene usage of T cell receptor gamma- and delta-chain transcripts expressed in synovia and peripheral blood of patients with rheumatoid arthritis. *Clin Exp Immunol*, *87*(2), 172–177. https://doi.org/10.1111/j.1365-2249.1992.tb02970.x

Papagoras, C., Chrysanthopoulou, A., Mitsios, A., Ntinopoulou, M., Tsironidou, V., Batsali, A. K., Papadaki, H. A., Skendros, P., & Ritis, K. (2021). IL-17A expressed on neutrophil extracellular traps promotes mesenchymal stem cell differentiation toward bone-forming cells in ankylosing spondylitis. *Eur J Immunol*, *51*(4), 930–942. https://doi.org/10.1002/eji.202048878

Parga-Vidal, L., Behr, F. M., Kragten, N. A. M., Nota, B., Wesselink, T. H., Kavazovic, I., Covill, L. E., Schuller, M. B. P., Bryceson, Y. T., Wensveen, F. M., van Lier, R. A. W., van Dam, T. J. P., Stark, R., & van Gisbergen, K. (2021). Hobit identifies tissue-resident memory T cell precursors that are regulated by Eomes. *Sci Immunol*, *6*(62). https://doi.org/10.1126/sciimmunol.abg3533

Parry, E. M., Lemvigh, C. K., Deng, S., Dangle, N., Ruthen, N., Knisbacher, B. A., Broseus, J., Hergalant, S., Guieze, R., Li, S., Zhang, W., Johnson, C., Long, J. M., Yin, S., Werner, L., Anandappa, A., Purroy, N., Gohil, S., Oliveira, G.,…Wu, C. J. (2023). ZNF683 marks a CD8(+) T cell population associated with anti-tumor immunity following anti-PD-1 therapy for Richter syndrome. *Cancer Cell*, *41*(10), 1803–1816 e1808. https://doi.org/10.1016/j.ccell.2023.08.013

Qaiyum, Z., Lim, M., & Inman, R. D. (2021). The gut-joint axis in spondyloarthritis: immunological, microbial, and clinical insights. *Semin Immunopathol*, *43*(2), 173–192. https://doi.org/10.1007/s00281-021-00845-0

Rao, D. A., Gurish, M. F., Marshall, J. L., Slowikowski, K., Fonseka, C. Y., Liu, Y., Donlin, L. T., Henderson, L. A., Wei, K., Mizoguchi, F., Teslovich, N. C., Weinblatt, M. E., Massarotti, E. M., Coblyn, J. S., Helfgott, S. M., Lee,

Y. C., Todd, D. J., Bykerk, V. P., Goodman, S. M.,…Brenner, M. B. (2017). Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature*, *542*(7639), 110–114. https://doi.org/10.1038/nature20810

Rebuffet, L., Melsen, J. E., Escaliere, B., Basurto-Lozada, D., Bhandoola, A., Bjorkstrom, N. K., Bryceson, Y. T., Castriconi, R., Cichocki, F., Colonna, M., Davis, D. M., Diefenbach, A., Ding, Y., Haniffa, M., Horowitz, A., Lanier, L. L., Malmberg, K. J., Miller, J. S., Moretta, L.,…Vivier, E. (2024). High-dimensional single-cell analysis of human natural killer cell heterogeneity. *Nat Immunol*, *25*(8), 1474–1488. https://doi.org/10.1038/s41590-024-01883-0

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rudwaleit, M., & Baeten, D. (2006). Ankylosing spondylitis and bowel disease. *Best Pract Res Clin Rheumatol*, *20*(3), 451–471. https://doi.org/10.1016/j.berh.2006.03.010

Stavre, Z., Bridgewood, C., Zhou, Q., Maeda, Y., Huang, T. T., Karman, J., Khan, A., Giryes, S., Sharif, K., McGonagle, D., & Gravallese, E. M. (2022). A role for neutrophils in early enthesitis in spondyloarthritis. *Arthritis Res Ther*, *24*(1), 24. https://doi.org/10.1186/s13075-021-02693-7

Thiran, A., & Vereecke, L. (2025). New thoughts on the gut-immune axis of arthritis. *Trends Immunol*, *46*(3), 206–218. https://doi.org/10.1016/j.it.2025.01.006

Van Praet, L., Van den Bosch, F. E., Jacques, P., Carron, P., Jans, L., Colman, R., Glorieus, E., Peeters, H., Mielants, H., De Vos, M., Cuvelier, C., & Elewaut, D. (2013). Microscopic gut inflammation in axial spondyloarthritis: a multiparametric predictive model. *Ann Rheum Dis*, *72*(3), 414–417. https://doi.org/10.1136/annrheumdis-2012-202135

Wilbrink, R., Spoorenberg, A., Arends, S., van der Geest, K. S. M., Brouwer, E., Bootsma, H., Kroese, F. G. M., & Verstappen, G. M. (2021). CD27(-)CD38(low)CD21(low) B-Cells Are Increased in Axial Spondyloarthritis. *Front Immunol*, *12*, 686273. https://doi.org/10.3389/fimmu.2021.686273

Yang, P. T., Kasai, H., Zhao, L. J., Xiao, W. G., Tanabe, F., & Ito, M. (2004). Increased CCR4 expression on circulating CD4(+) T cells in ankylosing spondylitis, rheumatoid arthritis and systemic lupus erythematosus. *Clin Exp Immunol*, *138*(2), 342–347. https://doi.org/10.1111/j.1365-2249.2004.02617.x

Yang, X., Garner, L. I., Zvyagin, I. V., Paley, M. A., Komech, E. A., Jude, K. M., Zhao, X., Fernandes, R. A., Hassman, L. M., Paley, G. L., Savvides, C. S., Brackenridge, S., Quastel, M. N., Chudakov, D. M., Bowness, P., Yokoyama, W. M., McMichael, A. J., Gillespie, G. M., & Garcia, K. C. (2022). Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature*, *612*(7941), 771–777. https://doi.org/10.1038/s41586-022-05501-7

York, M. R., Nagai, T., Mangini, A. J., Lemaire, R., van Seventer, J. M., & Lafyatis, R. (2007). A macrophage marker, Siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by

type I interferons and toll-like receptor agonists. *Arthritis Rheum*, *56*(3), 1010–1020. https://doi.org/10.1002/art.22382

Chapter 6: Transcriptional and Genetic Mechanisms Underlying AxSpA Pathogenesis

6.1 Introduction

AS has a strong genetic component, and its heritability has been extensively investigated. Twin studies estimate the heritability of AS to exceed 90% (Brown et al., 1997; Pedersen et al., 2008), indicating that genetic factors are the predominant determinants of disease risk. While the HLA-B*27 allele alone accounts for approximately 20% of the variance in susceptibility (Ellinghaus et al., 2016), the remaining genetic contribution arises from thousands of common variants, each exerting a modest effect (Li et al., 2021). Polygenic risk scores (PRS) aggregate these variants into a single quantitative measure of genetic liability and have shown promising discriminatory power for AS in European cohorts when compared with C-reactive protein (CRP), sacroiliac MRI findings, or HLA-B*27 status alone; whether PRS provides incremental value in routine clinical practice remains to be established. More broadly, the development of PRS for AS highlights the value of considering combined effects of multiple genetic variants when investigating disease pathogenesis.

Following the discovery of disease susceptibility loci through GWAS, increasing efforts have focused on understanding the regulatory mechanisms underlying these variants. An increasing number of studies demonstrate that genetic regulation is highly tissue- and cell-type-specific (Kim-Hellmuth et al., 2020; Natri et al., 2024), underscoring the need to connect genetic risk loci with their precise cellular and tissue contexts. Identifying the specific cell types

influenced by genetic variation is essential for elucidating the molecular pathways that link genetic risk to disease phenotypes.

Recent methodological advances have leveraged single-cell transcriptomic data to address this challenge by integrating GWAS findings with cell-type-resolved gene expression. Approaches such as sc-linker (Jagadeesh et al., 2022) and scGWAS (Jia et al., 2022) link human traits to specific cell types or gene programs, providing insight into how genetic risk converges on particular cellular processes. Among these, the single-cell disease relevance score (scDRS) (Zhang et al., 2022) represents a major step forward, as it directly associates individual cells in single-cell RNA-seq datasets with disease GWAS signals. By quantifying disease-association scores at the single-cell level, scDRS enables the identification of cell populations most transcriptionally aligned with genetic risk, offering a powerful framework for dissecting cell-type-specific mechanisms of complex diseases such as AxSpA.

In AxSpA, the integration of genetic information with single-cell transcriptomic data remains limited but provides a valuable opportunity to define the cellular basis of genetic susceptibility. Previous GWAS have identified over one hundred AxSpA-associated loci, including numerous non-HLA genome-wide significant variants, many of which are shared with other immune-mediated diseases such as inflammatory bowel disease (Ellinghaus et al., 2016; International Genetics of Ankylosing Spondylitis et al., 2013). Most of these variants lie in non-coding regions, suggesting that they act through transcriptional regulation rather than direct alteration of protein structure. Genetic and epigenetic studies have implicated immune pathways like IL-17/IL-23 in the pathogenesis of AxSpA but lack the resolution to pinpoint the specific

immune cell subsets affected (Brown et al., 2016). Leveraging our AxSpA single-cell atlas, I applied scDRS to integrate GWAS summary statistics with single-cell RNA-seq data, enabling the quantification of disease-relevance scores across immune cell types and tissues. This analysis aims to identify the cell populations and pathways most transcriptionally aligned with genetic risk, thereby linking genetic susceptibility to cell-type-specific transcriptional programs in AxSpA.

In addition, I applied the same framework to manually selected disease-associated pathways to assess their relevance across cell types in AxSpA and to determine which populations showed pathway-level enrichment. This approach provides insight into both shared and tissue-specific disease mechanisms. Together, these analyses aim to elucidate how genetic susceptibility and transcriptional dysregulation collectively contribute to AxSpA pathogenesis, thereby extending the mechanistic understanding established in the compositional analysis.

6.2 Aim

This chapter aims to evaluate associations between GWAS-identified genetic risk variants and cell-type-specific transcriptional states in AxSpA. I further explore how known disease-related pathways are enriched within distinct immune cell populations in AxSpA condition. The overall goal is to identify immune cell types and transcriptional phenotypes associated with AxSpA associated genetic variants, and to characterize pathway enrichment patterns within these cells, thereby providing deeper insight into disease pathogenesis and potential therapeutic targets.

6.3 Results

6.3.1 Genetic associations at the cell type level

To identify cell types most associated with polygenic disease risk, I applied scDRS (Zhang et al., 2022) to the extended AxSpA single-cell cohort described in Chapter 5. This method integrates GWAS summary statistics to calculate disease-association scores for individual cells. Using data from the 2013 AS GWAS (International Genetics of Ankylosing Spondylitis et al., 2013), I computed AS disease-correlation scores across cell types within each tissue after excluding variants in the MHC region. In addition to AS, three comparator traits: RA, Crohn's disease (CD), and ulcerative colitis (UC), were included to contextualise the genetic associations: the enrichment of RA GWAS signals served as a comparator, representing another form of inflammatory arthritis with matched peripheral blood and synovial fluid datasets; while CD and UC GWAS enrichments were examined to evaluate the contribution of shared genetic architecture with IBD, given the known well-established stronger genetic overlap between AxSpA and CD. To reduce noise from sparsely represented populations, cell types containing fewer than ten cells within the disease condition were excluded from the analysis.

6.3.1.1 Genetic associations in the sacroiliac joint

I first examined cells from sacroiliac joint (SIJ) to assess their association with genetic risks of AxSpA. In this tissue, no cell type reached significance at an adjusted $p < 0.05$ threshold; therefore, a relaxed cutoff of BH-adjusted $p < 0.1$ was applied to identify suggestive signals. Under this threshold, several cell types showed notable disease associations, including CD4.TEMRA, CD4.Th1,

and CD4.Th17 subsets within the CD4⁺ T-cell compartment; CD8.TEMRA.HLA-DR.KLF12 cells within the CD8⁺ T-cell compartment; IL23R-expressing effector memory T cells and MAIT-like clusters; as well as some NK and myeloid populations. When compared with RA, CD, and UC, the enrichment of Th17, MAIT-like αβ T cells, and UTC γδ T-cell populations appeared specific to AxSpA, as these signals were not observed for the comparator traits.

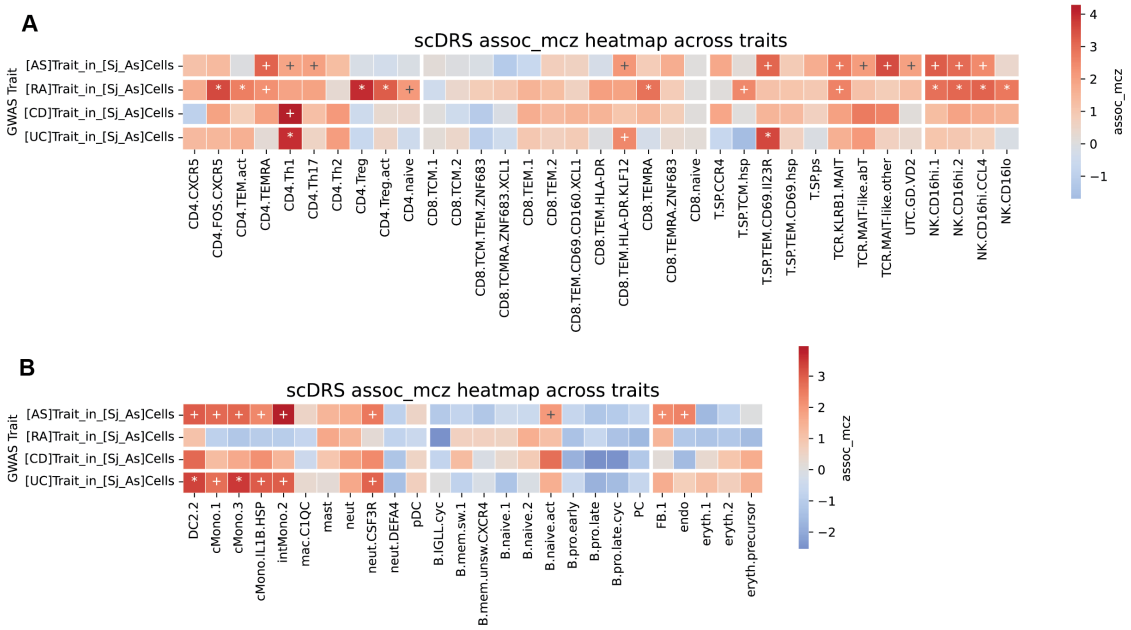


Figure 6.1 scDRS-based disease association across cell types in the SIJ
Heatmap showing scDRS-derived disease association scores across (A) T and NK cell subsets and (B) other immune populations, including myeloid cells, B cells, and additional minor subsets. Only cell types with more than 10 cells in total within the condition were included in the analysis. Colour scale represents the association Z-score, with higher values indicating stronger positive association with AxSpA. Statistical significance is denoted by BH-adjusted p-values: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1.

These findings suggest that effector and cytotoxic T-cell subsets, together with innate-like lymphocytes, may contribute to disease processes within the sacroiliac joint which is one of the earliest and most characteristic sites of inflammation in AxSpA. The GWAS enrichment observed in IL23R⁺ and

Th17-related populations is consistent with the established involvement of the IL-23/IL-17 axis in AxSpA pathogenesis (Smith & Colbert, 2014). Moreover, the enrichment signal in MAIT-like subsets indicates that disease-associated variants may influence pathways characteristic of mucosal-type, IL-17-producing T cells. Together, these results support the concept that genetic risk in AxSpA converges on effector programs integrating mucosal immunity and joint inflammation.

6.3.1.2 Genetic associations in the spine enthesis

The scDRS results for PEB and EST are shown in **Figures 6.2** and **Figure 6.3**, respectively.

In PEB, the overall strongest genetic associations were observed in NK cells from both AxSpA samples and healthy controls, consistent with previous findings showing significant enrichment of AxSpA GWAS variants in NK cells and suggesting that these cells may contribute to disease pathogenesis (Chinas et al., 2025). Among T-cell subsets, disease-association signals were observed in CD4.TEMRA, CD4.Th1, CD8.TCM.2, MAIT-like, and unconventional T-cell clusters, indicating that cytotoxic and innate-like T-cell programs may be particularly enriched for the expression of genes associated with AxSpA-related genetic variation.

Within the myeloid lineage, enrichment was detected in IL1B$^+$ classical monocytes, SPP1$^+$ intermediate monocytes, and pDCs, together with a switched memory B-cell subset (B.mem.sw.SOX5.ITGAX). The prominence of IL1B$^+$ and SPP1$^+$ monocytes support their established roles in inflammatory amplification, while enrichment in pDCs and memory B cells may reflect the contribution of antigen presentation and type I interferon–related processes.

Collectively, these data suggest that AxSpA genetic risk converges on immune networks spanning both innate and adaptive compartments within the peri-entheseal environment, emphasizing pathways of cytotoxicity, IL-17 signalling, myeloid activation and MAIT-cell associated inflammatory programs.

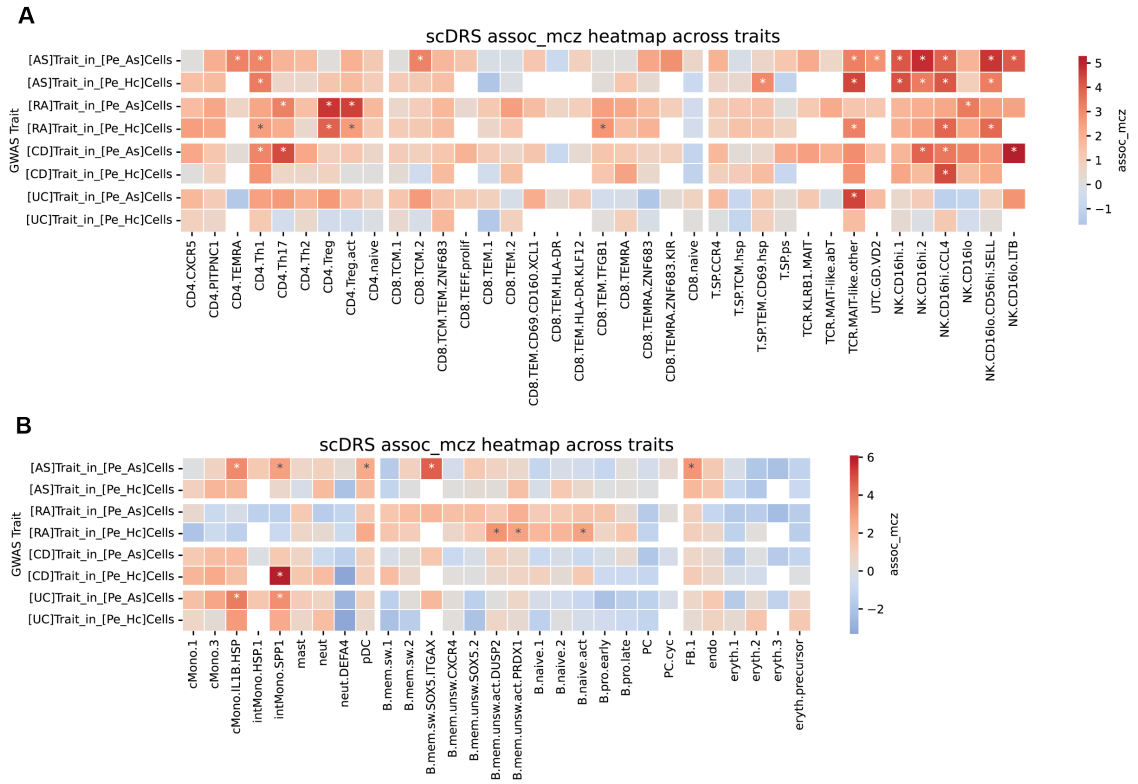


Figure 6.2 scDRS-based disease association across cell types in PEB
Heatmap showing scDRS-derived disease association scores across (A) T and NK cell subsets and (B) other immune populations, including myeloid cells, B cells, and additional minor subsets. Only cell types with more than 10 cells in total within the condition were included in the analysis. Colour scale represents the association Z-score, with higher values indicating stronger positive association with AxSpA. Statistical significance is denoted by BH-adjusted p-values: *** < 0.001, ** < 0.01, * < 0.05.

Moving from PEB to EST, the pattern of disease-associated cell types shifted, suggesting that genetic risk may influence immune activity differently across these adjacent microenvironments. The number of immune cells captured from EST was relatively low, and after applying a minimum threshold

of 10 cells per condition, only a subset of cell types remained for analysis. Among these, several T-cell populations showed AxSpA-specific enrichment of the AS genetic trait, with associations detected exclusively in cells from AxSpA patients but not in healthy controls. CD4.PITPNC1, CD4.Th1, CD4.Treg.act, and CD8.TEMRA.ZNF683 subsets exhibited significant disease associations, and notably, CD8.TEMRA.ZNF683 was previously identified as an AxSpA-enriched subset in the compositional analysis from synovial fluid compared to RA condition in the core cohort.

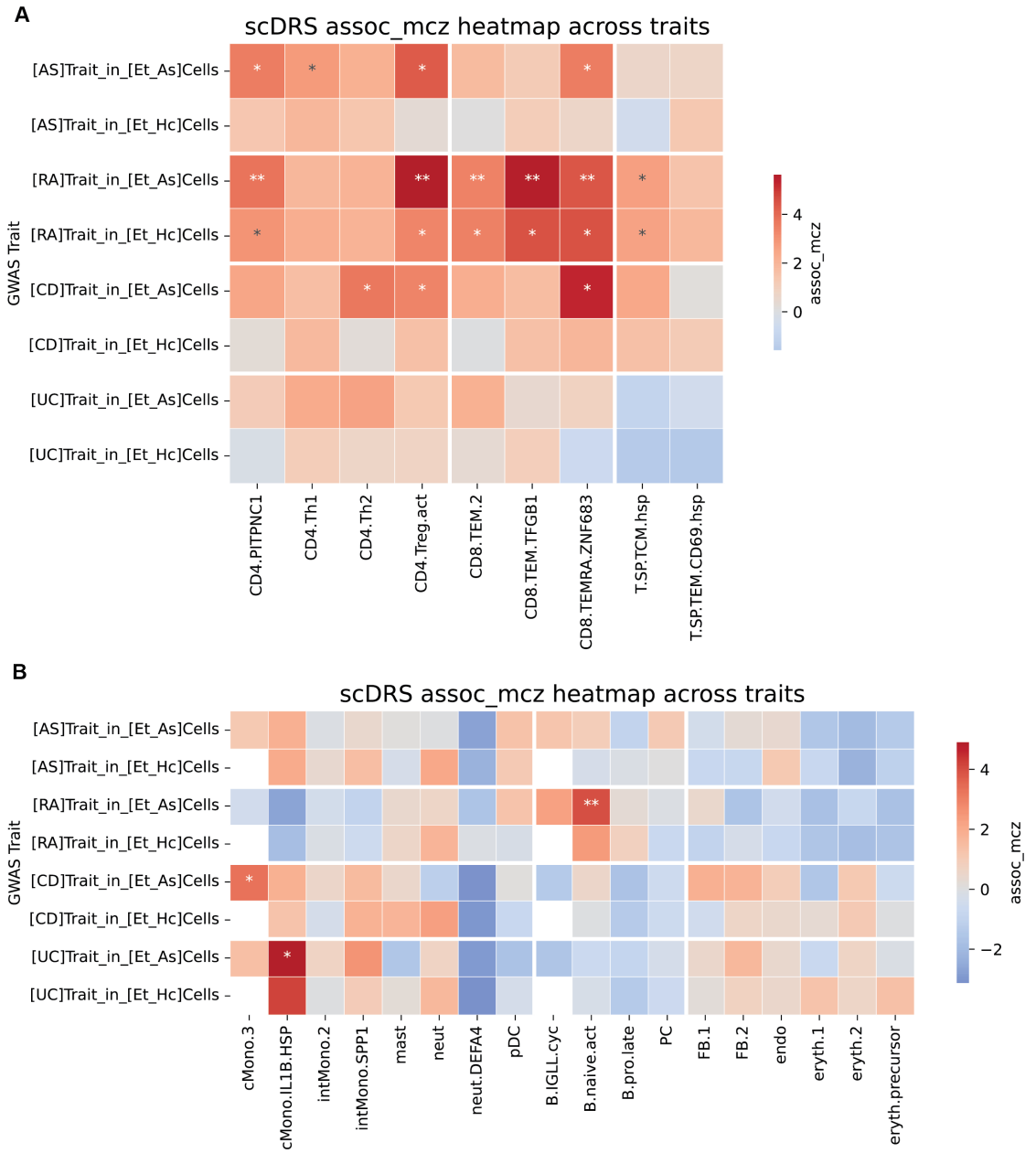


Figure 6.3 scDRS-based disease association across cell types in EST
Heatmap showing scDRS-derived disease association scores across (A) T and NK cell subsets and (B) other immune populations, including myeloid cells, B cells, and additional minor subsets. Only cell types with more than 10 cells in total within the condition were included in the analysis. Colour scale represents the association Z-score, with higher values indicating stronger positive association with AxSpA. Statistical significance is denoted by adjusted p-values: *** < 0.001, ** < 0.01, * < 0.05.

No significant associations were detected within the myeloid or B cell compartment in EST. When examining associations with other disease traits, CD4.Treg.act and CD8.TEMRA.ZNF683 cells from AxSpA patients, both of

which showed significant associations with AS, also exhibited strong genetic correlations with CD. This observation is consistent with the known genetic overlap between AxSpA and CD (Ellinghaus et al., 2016), and suggests that shared pathogenic mechanisms, particularly those involving ZNF683^{+} effector memory T cells and regulatory T-cell programs, may contribute to the immune dysregulation common to both conditions.

6.3.1.3 Genetic associations in ileum

In the ileum, several immune cell subsets exhibited significant disease associations, as shown in **Figure 6.4**. Among CD4^{+} T cells, the CD4.Th1, CD4.Tr1.act, and CD4.Treg.act populations were enriched in AxSpA, while within the CD8^{+} compartment, CD8.TEM.HLA-DR and CD8.TEMRA subsets showed strong associations with disease. Tissue-specific populations, including T.SP.TEM.CD69.IL23R cells and those within the MAIT-like cluster, also displayed significantly higher disease-related signatures.

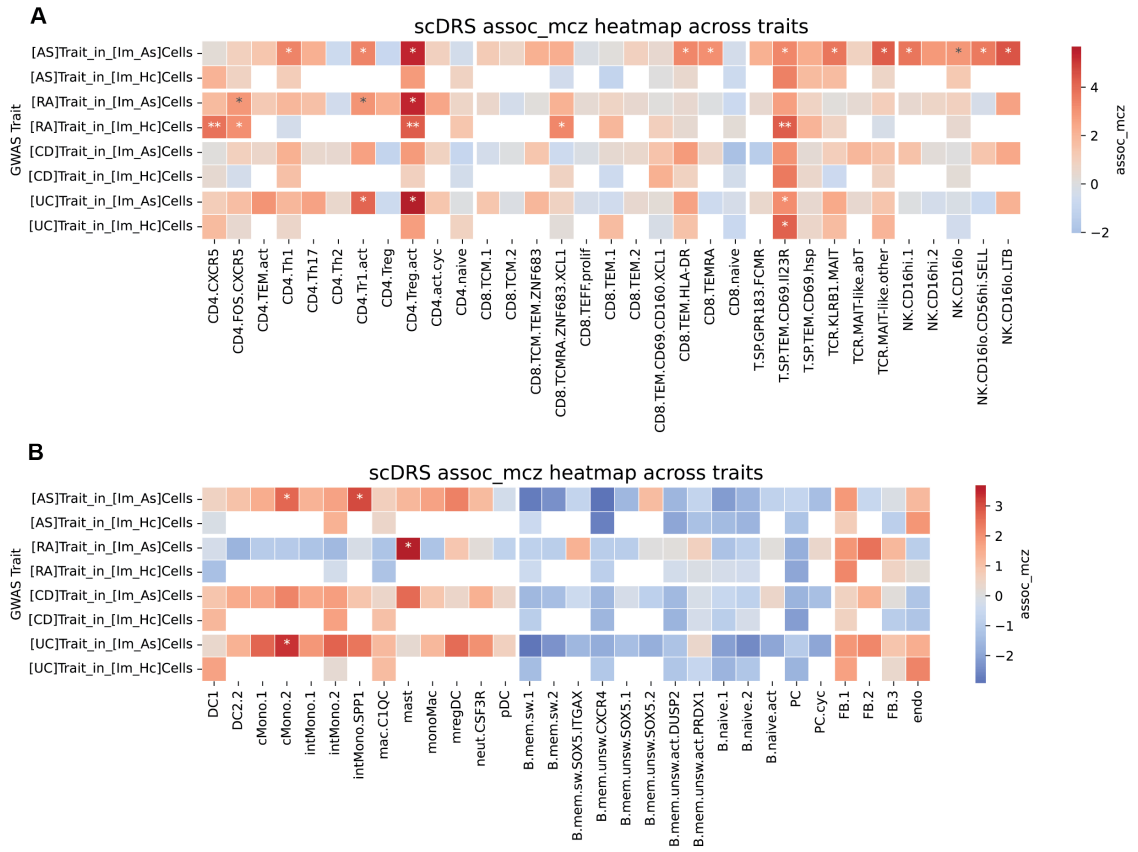


Figure 6.4 scDRS-based disease association across cell types in ileum
Heatmap showing scDRS-derived disease association scores across (A) T and NK cell subsets and (B) other immune populations, including myeloid cells, B cells, and additional minor subsets. Only cell types with more than 10 cells in total within the condition were included in the analysis. Colour scale represents the association Z-score, with higher values indicating stronger positive association with AxSpA. Statistical significance is denoted by adjusted p-values: *** < 0.001, ** < 0.01, * < 0.05.

Within the myeloid compartment, significant associations with AxSpA were observed for cMono.2 and IntMono.SPP1. cMono.2 also showed significant associations with UC GWAS traits, along with CD4.Tr1.act and CD4.Treg.act subsets in the T-cell compartment, which showed significantly higher scDRS scores for both AS and UC GWAS traits. These overlaps suggest that specific regulatory T-cell and monocyte populations may contribute to immune pathways shared between the two diseases, particularly those linked to mucosal and barrier-associated inflammation. Notably, IntMono.SPP1 was also

more abundant in the ileum of AxSpA patients compared with healthy controls in the extended cohort (BH-adjusted $p < 0.1$).

6.3.1.4 Cross-tissue comparison of genetic associations

Only a limited number of cell types from synovial fluid and peripheral blood showed detectable associations with AS GWAS signals, and none reached the BH-adjusted $p < 0.05$ threshold for significant disease association within these tissues. The scDRS results for SFMC and PBMC are provided in **Appendices E.2-E.3**. The absence of strong significance in these compartments may reflect that circulating and synovial-fluid cells do not fully capture the tissue-embedded immune states present in other disease-relevant tissues including enthesis and gut, which may be more representative of AxSpA pathogenesis. **Figure 6.5** highlights selected cell types that exhibited significant associations with at least one disease trait (AS, RA, CD, or UC) and are representative of the overall scDRS analysis. This cross-tissue comparison illustrates how disease-associated genetic signatures are distributed across immune cell populations and tissues, providing an integrative overview of the findings.

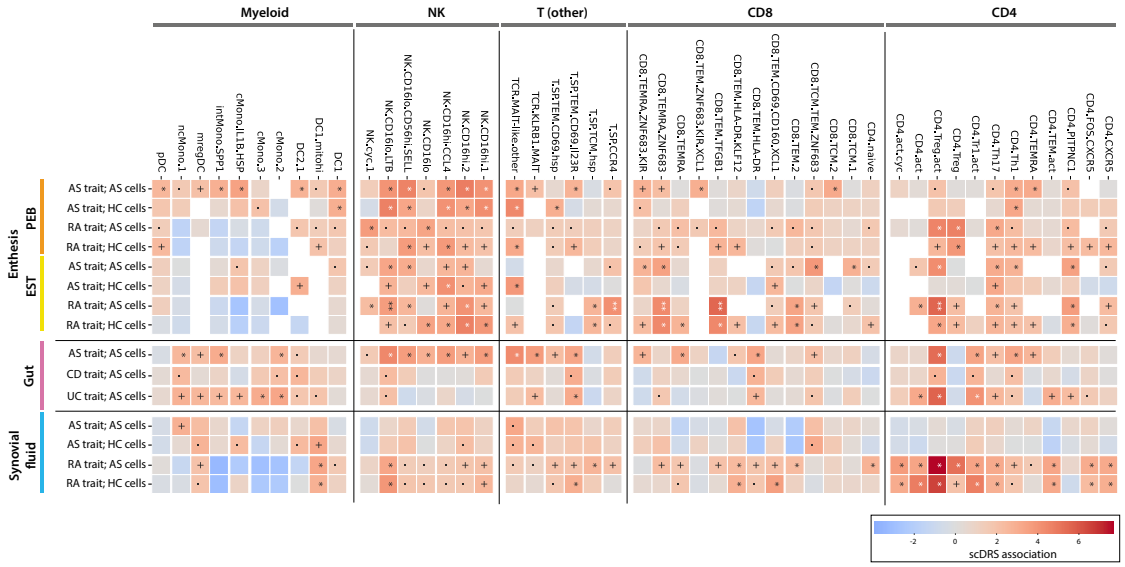


Figure 6.5 Summary of scDRS-based disease association across cell types and tissue
Heatmap showing scDRS disease association scores across selected cell types in enthesis, gut and synovial fluid, comparing AxSpA, RA and healthy controls. Colour indicates the association Z-score, with warmer colours representing stronger positive disease associations. Significance is denoted by adjusted P-values: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1, . < 0.2, and no label for non-significant results.

In the ileum, most NK cells showed a positive correlation with AS-associated genetic signals but not with CD or UC traits, suggesting that their transcriptional profiles are specifically linked to AS rather than to other ileum-associated inflammatory diseases. In contrast, NK cells from PEB and EST displayed high genetic associations with both AS and RA, implying the presence of shared inflammatory or effector programs between these two forms of arthritis.

Within the myeloid compartment, PEB and ileum shared several cell types that showed significant association with AS genetic signals, including intMono.SPP1. This cluster exhibited significant disease-related enrichment in both tissues and was also more abundant in the PEB and ileum of AxSpA patients compared with healthy controls in the extended cohort (adjusted p <

0.1), as shown in the compositional analysis from the previous chapter. These findings suggest that this monocyte subset, whose transcriptional profile is correlated with AS genetic risk, may be preferentially expanded or recruited within the ileal environment.

Within the T-cell compartment, several populations showed significant AS-related cross-tissue associations in both the spinal enthesis and gut ileum, including the MAIT-like cluster, IL23R^{+} CD69^{+} tissue-resident T cells, CD4.Treg.act, and CD4.Th1 subsets. Among these, the tissue-resident cluster T.SP.TEM.CD69.IL23R from PEB was significantly associated with AS only in cells derived from AxSpA patients, whereas cells from the same cluster in healthy controls exhibited negative Z-scores. This contrast further emphasizes that, even within the same cell type, tissue-resident T cells in AxSpA exhibit transcriptional states distinct from those in healthy individuals.

6.3.2 Disease-associated pathway analysis

Next, I examined how disease-associated and hypothesized pathogenic pathways were linked to specific cell types across tissues. Using scDRS*,* I computed pathway-level association scores at the single-cell level and then aggregated these scores by cell type. This approach enabled the identification of cell types within each tissue that showed significant enrichment for genes involved in disease-relevant pathways. The manually curated list of pathways, including their names and constituent genes, is provided in **Appendix E.1**, which were selected based on prior literature and relevance to established theories of AxSpA pathogenesis covering major immune and inflammatory processes such as IL-23/IL-17 signaling, TNF and IFN pathways, antigen presentation, and cellular stress responses. A complementary de novo pathway

discovery analysis using cNMF was performed by another member of the group (results not shown).

6.3.2.1 Targeted pathway analysis in the sacroiliac joint

The scDRS pathway analysis of cells from sacroiliac joint is shown in **Figure 6.6**. For the IL-23 signaling pathway, the MAIT-like.αβT and KLRB1.MAIT clusters showed the strongest enrichment (BH-adjusted $p < 0.1$) compared with other immune cell populations. The MAIT-like.other subset displayed a similar Z-score but did not reach significance, likely because it represents a mixture of MAIT and MAIT-like αβ T cells with higher transcriptional heterogeneity and variance. In contrast, for the IL-17 signaling pathway, only the MAIT-like αβ T subset, rather than conventional MAIT populations, showed a significant association (BH-adjusted $p < 0.05$). A similar pattern was observed for the TCR signaling pathway, where MAIT-like αβ T cells reached significance (BH-adjusted $p < 0.01$), while MAIT.KLRB1 cells did not. Considering the involvement of the IL-17/IL-23 axis in AxSpA pathogenesis, the MAIT-like αβ T and MAIT.KLRB1 subsets, while both part of the broader MAIT-like cluster and sharing related transcriptional features, show distinct pathway-level associations and may contribute differently to this signaling axis.

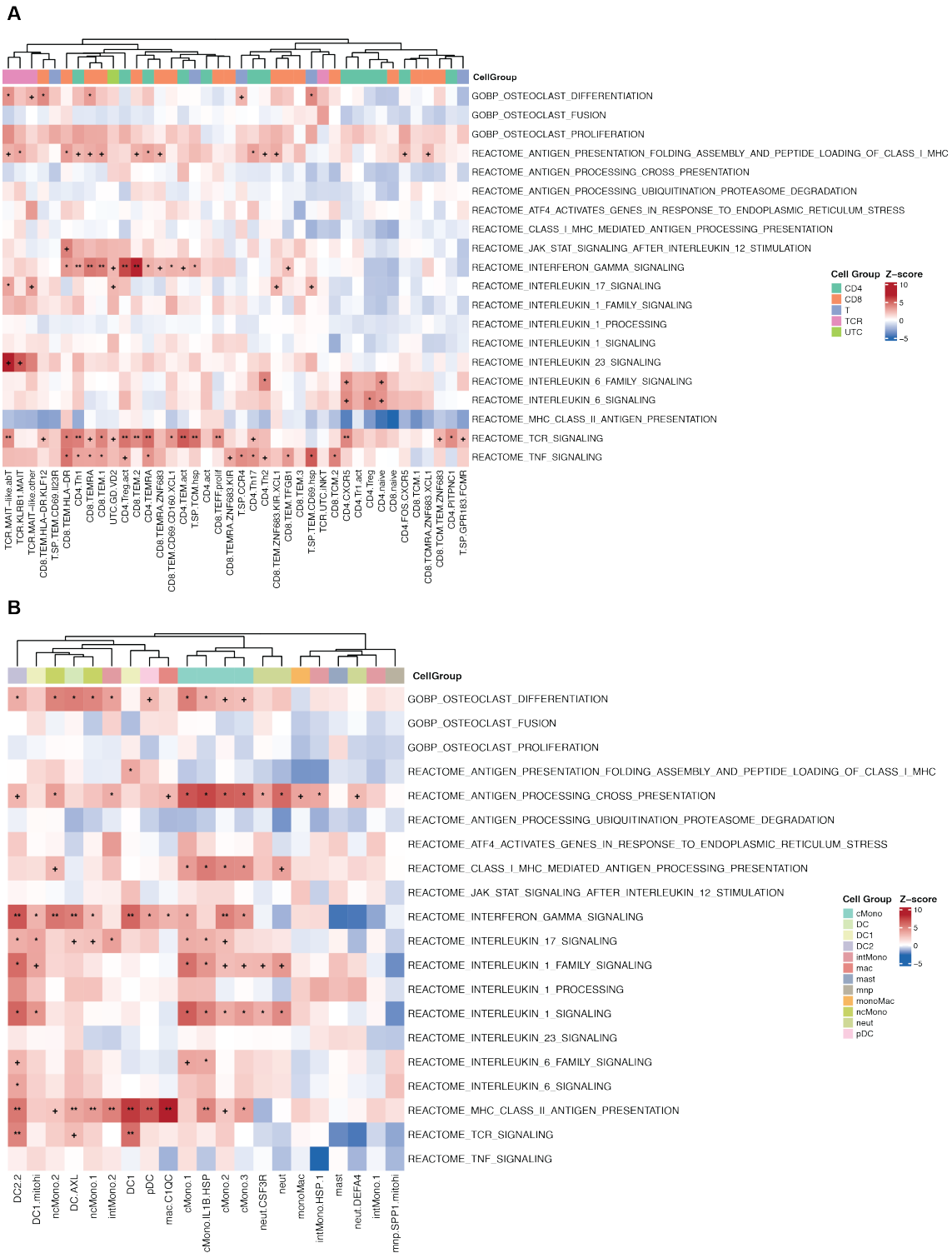


Figure 6.6 Disease-associated pathway enrichment analysis in the AxSpA sacroiliac joint samples

Heatmap showing selected disease-associated pathways and their scDRS-derived enrichment Z-scores in the sacroiliac joint. (A) T cell compartments. (B) Myeloid-cell compartments. Colour represents the pathway enrichment Z-score, with higher values indicating stronger positive association with the selected pathway. Major cell-type groups are annotated. Significance thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1.

Additionally, for the IL-17 signaling pathway, several myeloid subsets, including specific monocyte and DC populations, showed significant associations, which may indicate their potential role in amplifying or modulating IL-17-driven immune responses in AxSpA. When examining the osteoclast differentiation pathway, multiple T-cell clusters, including MAIT-like αβ T, CD8.TEMRA, CD8.TEM.HLA-DR.KLF12, and T.SP.TEM.CD69.HSP showed significant associations. These findings suggest that beyond classical myeloid involvement, activated effector T-cell subsets may also contribute to osteoimmune interactions and bone remodeling processes implicated in disease pathogenesis.

6.3.2.2 Targeted pathway analysis in spine enthesis

Within the spinal enthesis, PEB and EST exhibited distinct pathway association patterns, as shown in **Figures 6.7** and **Figure 6.8**, respectively. In PEB, several T-cell subsets including CD4.TEMRA, MAIT-like αβ T, CD4.Th1, and CD4.Th17 cells were associated with the TNF signaling pathway, while CD8.TEM.HLA-DR and T.SP.TEM.CD69.HSP clusters showed significant associations with the osteoclast differentiation pathway. In contrast, in EST, CD4.PITPNC1, CD4.Treg.act, CD4.CD17, and $CCR4^{+}$ T cells were enriched for TNF signaling, indicating tissue-specific differences in the activation of inflammatory pathways.

In the myeloid compartment, IL-1 signaling and IL-1 family signaling were specifically enriched in the cMono.IL1B.HSP cluster, and IL-1 signaling is also associated with DC2.1. By contrast, these IL-1 related pathways were not significantly associated with any cell type in EST (adjusted $p < 0.05$), suggesting

that IL-1 driven inflammatory activity may represent a site-specific feature of peri-entheseal bone rather than a general characteristic of the enthesis.

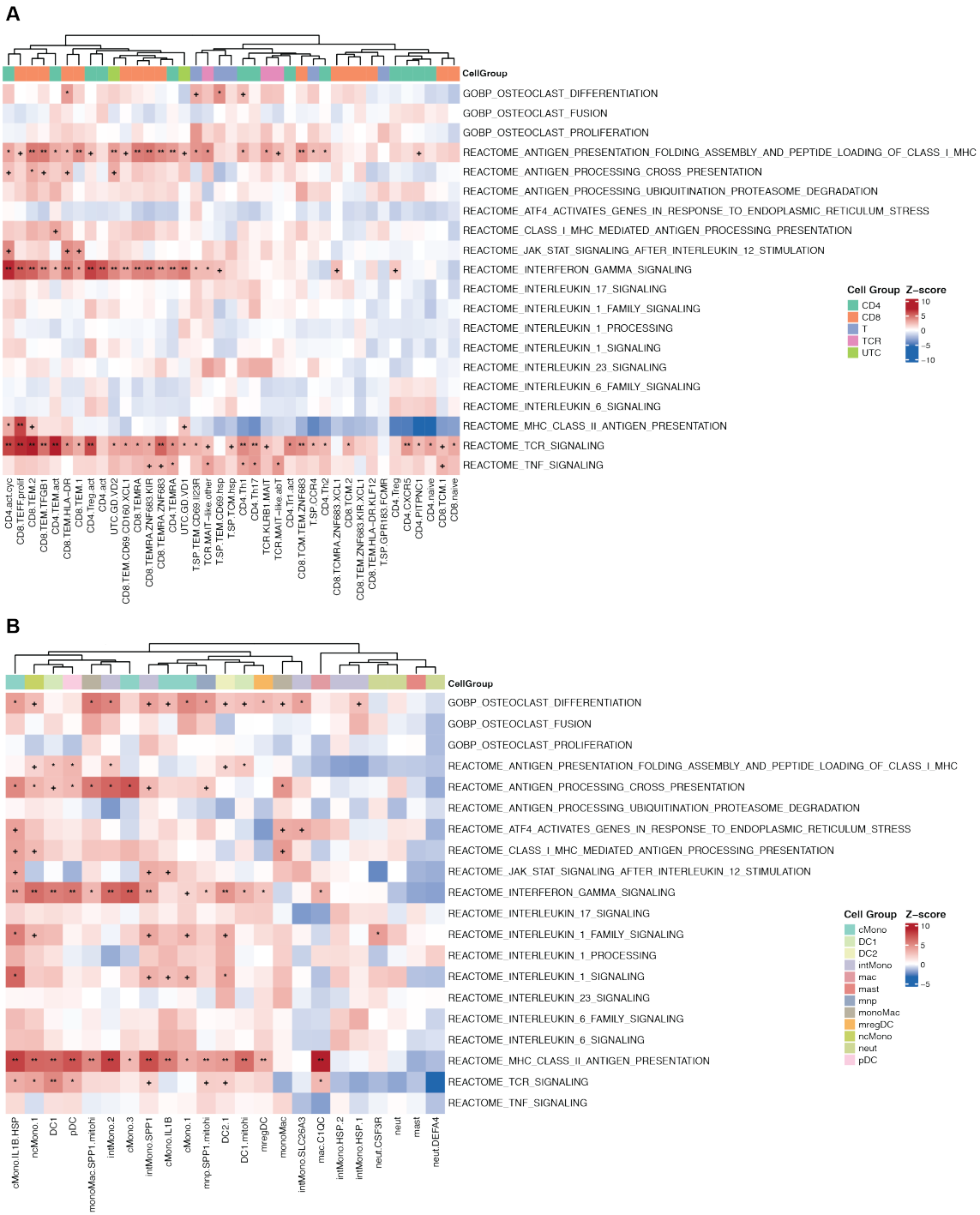


Figure 6.7 Disease-associated pathway enrichment analysis in the AxSpA PEB samples
Heatmap showing selected disease-associated pathways and their scDRS-derived enrichment Z-scores in PEB. (A) T cell compartments. (B) Myeloid-cell compartments. Colour represents the pathway enrichment Z-score, with higher values indicating

stronger positive association with the selected pathway. Major cell-type groups are annotated. Significance thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1.

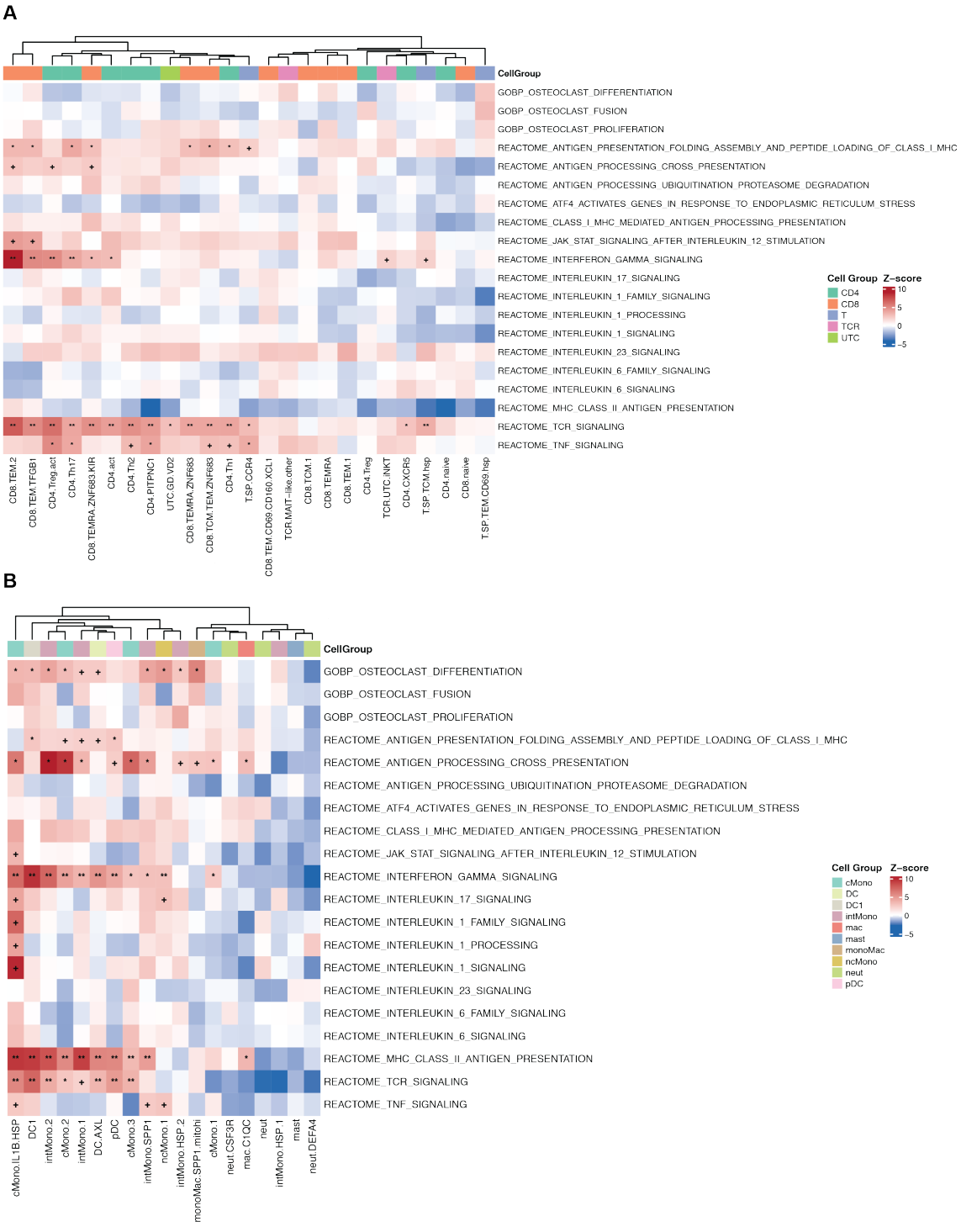


Figure 6.8 Disease-associated pathway enrichment analysis in the AxSpA EST samples

Heatmap showing selected disease-associated pathways and their scDRS-derived enrichment Z-scores in EST. (A) T cell compartments. (B) Myeloid-cell compartments. Colour represents the pathway enrichment Z-score, with higher values indicating

stronger positive association with the selected pathway. Major cell-type groups are annotated. Significance thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1.

6.3.2.3 Targeted pathway analysis in ileum

In the ileum, pathway analysis was performed on both all AS samples and on the subset of fresh AS samples. Results from all AS samples are shown in **Figure 6.9**, and those only from the fresh samples are presented in **Appendix E.4**. Several T-cell and monocyte clusters showed significant associations with interleukin signaling pathways, including both $CD4^+$ and $CD8^+$ T-cell subsets as well as conventional and intermediate monocytes. Although only reaching BH-adjusted $p < 0.1$, for the IL-23 signaling pathway, the $IL23R^+$ tissue-resident T-cell cluster (T.SP.TEM.CD69.IL23R) and CD4.Th17 cells showed the strongest associations among all immune populations. Tissue-resident $CD69^+$ clusters, together with MAIT-like αβ T and Th1 cells, also showed significant enrichment for TNF signaling, indicating potential convergence of TNF- and IL-23-related inflammatory programs in the ileal T-cell compartment.

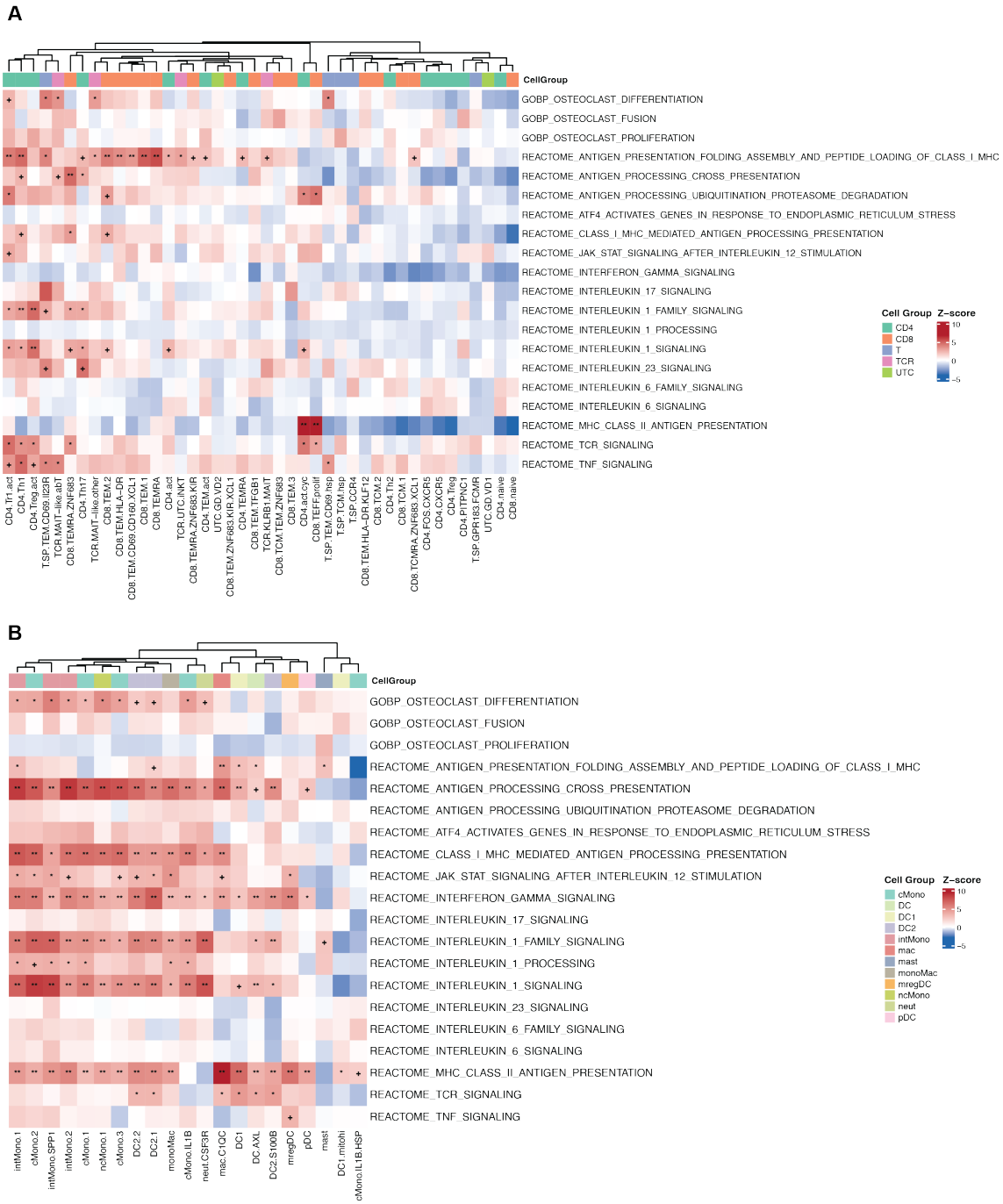


Figure 6.9 Disease-associated pathway enrichment analysis in the AxSpA ileum samples

Heatmap showing selected disease-associated pathways and their scDRS-derived enrichment Z-scores in the ileum. (A) T cell compartments. (B) Myeloid-cell compartments. Colour represents the pathway enrichment Z-score, with higher values indicating stronger positive association with the selected pathway. Major cell-type groups are annotated. Significance thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1.

Compared with other tissues, ileal monocytes demonstrated the strongest and most consistent associations with IL-1 and IL-1 family signaling, a pattern not observed in other tissue analyses. This suggests that myeloid cells in the ileal microenvironment are more transcriptionally engaged in cytokine-driven inflammatory responses, consistent with intestinal involvement in AxSpA. In addition, IFN-γ signaling was broadly enriched across multiple myeloid clusters, suggesting activation of antigen-processing and antimicrobial programs that could amplify downstream inflammation.

Together, these findings indicate a cytokine-rich and immune-activated state in the AxSpA ileum, characterized by IL-1- and IFN-γ-responsive myeloid cells and TNF/IL-23-driven T-cell populations.

6.3.2.4 Targeted pathway analysis in synovial fluid

In the synovial fluid, pathway analysis was performed on all AS samples (**Figure 6.10**), with results from the fresh sample subset shown in **Appendix E.5**. Compared with the ileum, T-cell subsets in synovial fluid showed relatively limited enrichment for cytokine signaling pathways, and no population reached significance (BH-adjusted $p < 0.05$) in the IL-17/IL-23, IL-1, IL-6, IL-12, or TNF pathways. The CD8.TEM.HLA-DR cluster showed a significant association with IFN-γ signaling. This subset was also upregulated in synovial fluid compared with RA in the expanded cohort (BH-adjusted $p < 0.1$), suggesting that expanded $CD8^+$ effector populations in AxSpA synovial joints may engage transcriptional programs responsive to IFN-γ, contributing to the local inflammatory milieu within the synovial environment.

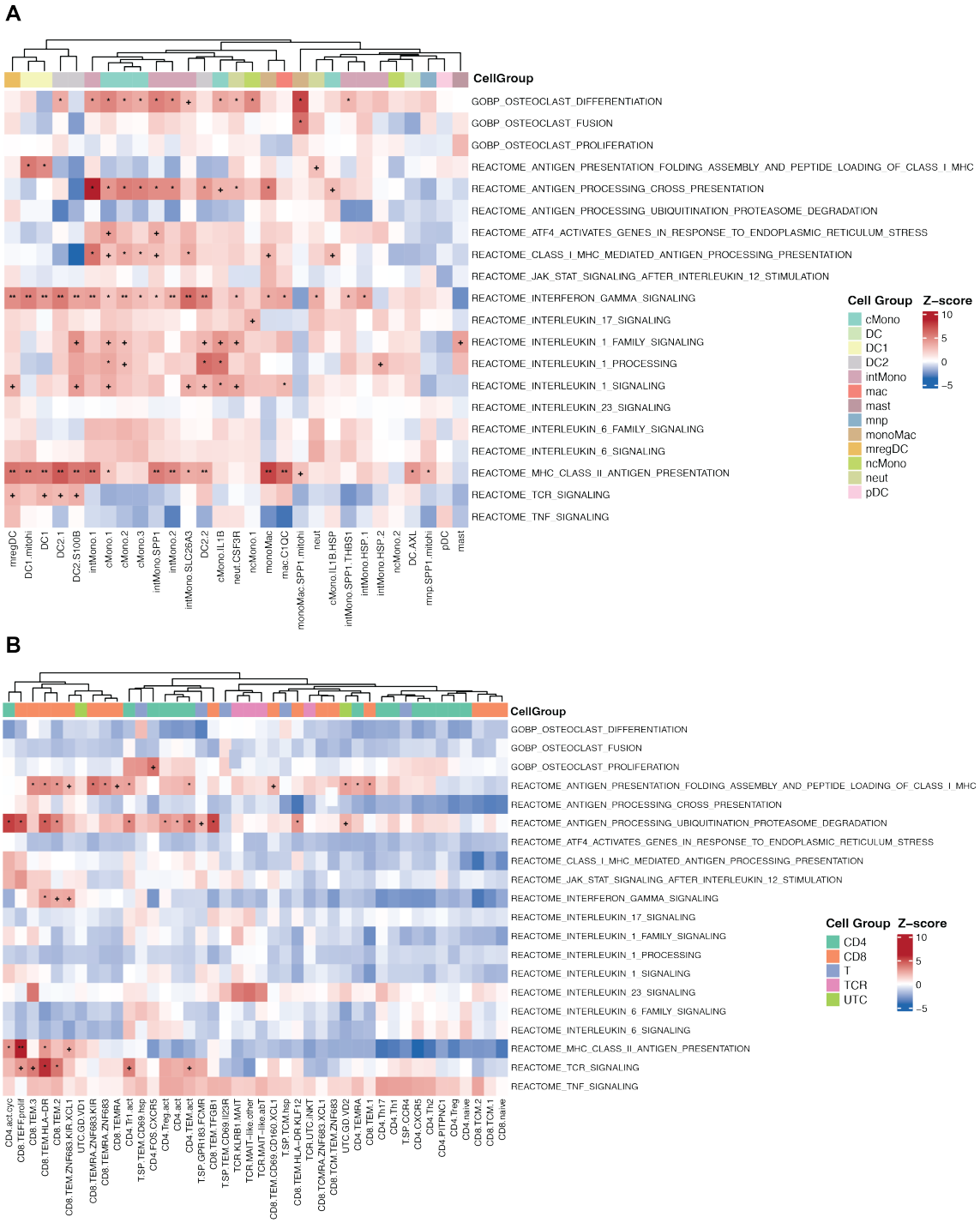


Figure 6.10 Disease-associated pathway enrichment analysis in the AxSpA SFMC samples

Heatmap showing scDRS-derived pathway enrichment Z-scores in the SFMC. (A) T-cell compartments. (B) Myeloid-cell compartments. Colour represents the pathway enrichment Z-score, with higher values indicating stronger disease association. Major cell-type groups are annotated. Significance thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1.

In the myeloid compartment, ncMono.1 showed a relatively strong association with the IL-17 signaling pathway (BH-adjusted $p < 0.1$). Multiple myeloid clusters, including cMono.1, cMono.IL1B, DC2.2, and Mac.C1QC, displayed significant associations with interleukin processing and interleukin signaling pathways. The MonoMac.SPP1.mitohi cluster showed significant enrichment for osteoclast differentiation and osteoclast fusion pathways, consistent with its transcriptional profile marked by SPP1 expression, which has been linked to osteoclast differentiation and bone remodeling processes that may contribute to joint pathology (Cai et al., 2022).

In addition, the DC1 and DC1.mitohi clusters were significantly associated with pathways related to antigen presentation, assembly, and loading of MHC class I. Although MHC I related antigen processing appears broadly represented across myeloid populations, these DC subsets showed stronger enrichment for this pathway based on scDRS analysis, suggesting that MHC I-associated processes may be more transcriptionally represented in these subsets, potentially relevant to the hypothesized arthritogenic peptide presentation in AxSpA.

6.4 Discussion

In this chapter, I performed scDRS-based genetic association analysis to integrate GWAS summary statistics with single-cell transcriptomic data from AxSpA-relevant tissues and conducted pathway analyses to examine disease-associated transcriptional programs across immune cell types. Together, these analyses provide complementary insights into how genetic susceptibility and

transcriptional dysregulation contribute to AxSpA pathogenesis at the cell-type level.

The interpretation of scDRS results depends on the distribution of disease-relevance scores across conditions. High scores in both AxSpA and healthy controls suggest that genetic risk is broadly expressed in those cell types, reflecting immune programs active in both physiological and disease states. High association restricted to AxSpA indicates disease-specific transcriptional activation of risk-related pathways, while high association only in healthy controls may represent essential immune programs that become dysregulated or lost in disease. Although scDRS does not establish direct causality, these comparisons provide suggestive evidence of how genetic predisposition manifests across different immune cell types and tissue contexts.

The genetic association analyses revealed that NK cells displayed the strongest AxSpA-related disease-association signals in both the enthesis and gut, consistent with previous studies emphasizing the role of innate immune activation in these tissues (Chinas et al., 2025). In contrast, genetic associations detected in synovial fluid and peripheral blood were relatively limited. This likely reflects that these compartments do not fully represent the tissue environments in which AxSpA-related immune programs arise or are most pronounced. Peripheral blood contains circulating cells that are not shaped by the same microenvironmental cues as tissue-resident populations, and synovial fluid comprises cells that have migrated out of tissue and therefore no longer retain the full context of their original niche. As a result, disease-relevant transcriptional states present within enthesis or gut tissues may not be

fully captured in these compartments. Together, these observations reinforce the importance of examining disease-relevant tissues directly.

A central finding is that the genetic risk of AxSpA preferentially maps onto effector and cytotoxic T-cell subsets, particularly HLA-DR$^+$ and IL23R$^+$ tissue-resident populations, as well as innate-like lymphocytes such as MAIT cell and MAIT-like αβ T cells. These results are consistent with previous studies implicating the IL-23/IL-17 axis in AxSpA and suggest that IL-23-responsive and mucosal-type T-cell compartments in the sacroiliac joint, spine, and gut may be involved in mediating genetic risk. The enrichment of genetic risk in ZNF683$^+$ and CD69$^+$ resident T cells suggests that long-lived, tissue-adapted T cells represent key effectors of chronic inflammation in AxSpA. Such populations likely integrate signals from local cytokine milieus and microbial exposure, linking systemic genetic risk with persistent tissue-level immune activation.

The myeloid compartment also showed significant genetic associations. IL1B$^+$ and SPP1$^+$ monocytes exhibited consistent enrichment across enthesis and ileum, supporting their roles in cytokine-driven inflammation and bone remodeling. The repeated detection of SPP1$^+$ monocytes in multiple tissues suggests that this subset may represent a shared cellular mediator linking mucosal and skeletal inflammation in AxSpA.

Although previous studies have reported stronger overall genetic correlations between AS and CD than between AS and UC (Brown et al., 2020), the present analysis revealed a more complex, cell-type-specific pattern. Within the myeloid and CD4-positive T cell compartment, the genetic association of AxSpA more closely resembled that of UC than CD. This may reflect that, although AxSpA and CD share a broader genetic architecture, the downstream

effects of these variants are mediated through different cellular contexts. CD-associated variants are more linked to autophagy and microbial sensing pathways, whereas UC-associated variants more strongly involve cytokine signaling and immune regulations (Caba et al., 2025; Jostins et al., 2012; Liu & Anderson, 2014). The similarity between AxSpA and UC GWAS enrichment at the transcriptional level in the immune cells therefore suggests that overlapping inflammatory programs, particularly those involving IL-1 and TNF signaling in myeloid cells and cytokine-responsive $CD4^+$ T cells, may represent a common effector axis, even if the primary genetic overlap is greater between AS and CD at the population level.

Pathway-level analyses further highlighted tissue-specific inflammatory landscapes. In the PEB, multiple T-cell subsets were associated with TNF and IL-17 signaling, while both T and myeloid populations were linked to osteoclast differentiation pathways, consistent with the bone-remodeling pathology that characterizes AxSpA. In the ileum, several T-cell and monocyte subsets were enriched for IL-1 and IL-1 family signaling, reflecting the cytokine-rich and microbially responsive intestinal environment. Myeloid populations in this tissue also showed broad enrichment for interferon-γ signaling, indicating a heightened antimicrobial and antigen-presenting state that may amplify mucosal inflammation. Within synovial fluid, cMono.1, cMono.2, and cMono.IL1B subsets were associated with interleukin-related pathways, while the AxSpA-enriched intMono.1 population demonstrated a strong association with class I MHC mediated antigen processing. This pattern points to potential roles for monocyte-derived cells in presenting arthritogenic peptides and supporting $CD8^+$ T-cell activation within the joint microenvironment.

Despite the strengths of this integrative approach, several limitations should be acknowledged. First, the publicly available GWAS used in this analysis may not be perfectly matched to the disease samples profiled in this thesis. The IGAS 2013 GWAS for AS was based on cases defined by the modified New York criteria, while the AxSpA samples analysed in this thesis were recruited over an extended period spanning the introduction of the ASAS axial spondyloarthritis classification criteria (Rudwaleit et al., 2009), and may therefore include cases classified under either framework. As the modified New York and ASAS criteria capture overlapping but not identical patient populations, subtle differences in the spectrum of cases between the GWAS reference and the samples profiled here may affect the direct comparability of genetic enrichment signals. The rheumatoid arthritis (Okada et al., 2014) and IBD (de Lange et al., 2017) GWAS were derived from large, broadly defined disease cohorts and may not perfectly mirror the specific clinical and demographic characteristics of the disease samples profiled in this thesis. Such cohort-level differences may partially obscure cell-type associations that are most relevant to the patient populations represented in the single-cell data. Second, the scDRS framework itself has inherent limitations that should be considered when interpreting the results. The enrichment signals observed in this chapter were generally modest, which is consistent with the polygenic architecture of these diseases and the fact that risk variants exert small effects distributed across many genes. The scDRS scores are also sensitive to the choice of GWAS input, the gene-mapping window used by MAGMA, and the granularity of cell-type definitions; alternative parameter choices or cell-type annotations may shift the apparent strength and distribution of enrichment signals. Third, because this

analysis relied on public GWAS summary statistics, it was not possible to directly link genetic variants to the transcriptomes of the individual samples analyzed here, which limits causal inference and confines the findings to correlational interpretations. Fourth, sample and cell numbers from PEB and EST was limited by clinical availability and tissue size, and the use of frozen AxSpA samples versus fresh healthy controls introduced batch effects that may have affected comparability and reduced statistical power. Although these issues could not be fully resolved, some analyses were mitigated by including additional assessments of the fresh samples separately, which are provided in the Appendix. Finally, the spatial and temporal heterogeneity of inflammation cannot be fully captured by cross-sectional sampling, and the dynamic transitions between resting and activated immune states remain to be characterized. Nonetheless, by leveraging computational integration of GWAS and single-cell data, this analysis provides valuable insight into how genetic risk may be functionally expressed across immune cell populations and tissues relevant to AxSpA.

In summary, this chapter presents a comprehensive view of how polygenic susceptibility intersects with transcriptional activity at single-cell resolution. The findings highlight NK cells, effector and tissue-resident T cells, and inflammatory monocytes as key mediators of genetic risk, acting through cytokine and osteoimmune signaling pathways that underlie the chronic inflammation of AxSpA. While causal links between genetic variants and specific transcriptomic phenotypes remain to be established, these results lay the groundwork for future studies integrating genotype-resolved single-cell data

to uncover the precise molecular routes by which genetic predisposition translates into disease expression.

6.5 References

Brown, M. A., Kenna, T., & Wordsworth, B. P. (2016). Genetics of ankylosing spondylitis--insights into pathogenesis. *Nat Rev Rheumatol*, *12*(2), 81–91. https://doi.org/10.1038/nrrheum.2015.133

Brown, M. A., Kennedy, L. G., MacGregor, A. J., Darke, C., Duncan, E., Shatford, J. L., Taylor, A., Calin, A., & Wordsworth, P. (1997). Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. *Arthritis Rheum*, *40*(10), 1823–1828. https://doi.org/10.1002/art.1780401015

Brown, M. A., Xu, H., & Li, Z. (2020). Genetics and the axial spondyloarthritis spectrum. *Rheumatology (Oxford)*, *59*(Suppl4), iv58–iv66. https://doi.org/10.1093/rheumatology/keaa464

Caba, L., Florea, A., Cianga, P., Drug, V., Popescu, R., Mihai, C., Ciobanu, C. G., Iacob, V. V., Florea, L., & Gorduza, E. V. (2025). Genetic and Epigenetic Factors in Ulcerative Colitis: A Narrative Literature Review. *Genes (Basel)*, *16*(9). https://doi.org/10.3390/genes16091085

Cai, X., Zheng, Y., Ren, F., Zhang, S., Wu, L., & Yao, Y. (2022). Secretory phosphoprotein 1 secreted by fibroblast-like synoviocytes promotes osteoclasts formation via PI3K/AKT signaling in collagen-induced arthritis. *Biomed Pharmacother*, *155*, 113687. https://doi.org/10.1016/j.biopha.2022.113687

Chinas, M., Fernandez-Salinas, D., Aguiar, V. R. C., Nieto-Caballero, V. E., Lefton, M., Nigrovic, P. A., Ermann, J., & Gutierrez-Arcelus, M. (2025). Functional genomics implicates natural killer cells in the pathogenesis of ankylosing spondylitis. *HGG Adv*, *6*(1), 100375. https://doi.org/10.1016/j.xhgg.2024.100375

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S. G., Heap, G., Nimmo, E. R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D. C.,…Barrett, J. C. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*, *49*(2), 256–261. https://doi.org/10.1038/ng.3760

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., Park, Y. R., Raychaudhuri, S., Pouget, J. G., Hubenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H. J., Franke, L., Pers, T. H., Weersma, R. K., Collij, V.,…Franke, A. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*, *48*(5), 510–518. https://doi.org/10.1038/ng.3528

International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., Joo, K. B., Shim, S. C., Weisman, M., Ward, M.,

Zhou, X., Garchon, H. J., Chiocchia, G., Nossent, J.,…Brown, M. A. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*, *45*(7), 730–738. https://doi.org/10.1038/ng.2667

Jagadeesh, K. A., Dey, K. K., Montoro, D. T., Mohan, R., Gazal, S., Engreitz, J. M., Xavier, R. J., Price, A. L., & Regev, A. (2022). Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet*, *54*(10), 1479–1492. https://doi.org/10.1038/s41588-022-01187-9

Jia, P., Hu, R., Yan, F., Dai, Y., & Zhao, Z. (2022). scGWAS: landscape of trait-cell type associations by integrating single-cell transcriptomics-wide and genome-wide association studies. *Genome Biol*, *23*(1), 220. https://doi.org/10.1186/s13059-022-02785-w

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, L. P., Sharma, Y., Anderson, C. A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S. L., Raychaudhuri, S., Goyette, P., Wei, Z.,…Cho, J. H. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, *491*(7422), 119–124. https://doi.org/10.1038/nature11582

Kim-Hellmuth, S., Aguet, F., Oliva, M., Munoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S. E., Hamel, A. R., Vinuela, A., Roberts, A. L., Mangul, S., Wen, X., Wang, G., Barbeira, A. N., Garrido-Martin, D., Nadel, B. B., Zou, Y., Bonazzola, R., Quan, J.,…Lappalainen, T. (2020). Cell type-specific genetic regulation of gene expression across human tissues. *Science*, *369*(6509). https://doi.org/10.1126/science.aaz8528

Li, Z., Wu, X., Leo, P. J., De Guzman, E., Akkoc, N., Breban, M., Macfarlane, G. J., Mahmoudi, M., Marzo-Ortega, H., Anderson, L. K., Wheeler, L., Chou, C. T., Harrison, A. A., Stebbings, S., Jones, G. T., Bang, S. Y., Wang, G., Jamshidi, A., Farhadi, E.,…Group, T. A. (2021). Polygenic Risk Scores have high diagnostic capacity in ankylosing spondylitis. *Ann Rheum Dis*, *80*(9), 1168–1174. https://doi.org/10.1136/annrheumdis-2020-219446

Liu, J. Z., & Anderson, C. A. (2014). Genetic studies of Crohn's disease: past, present and future. *Best Pract Res Clin Gastroenterol*, *28*(3), 373–386. https://doi.org/10.1016/j.bpg.2014.04.009

Natri, H. M., Del Azodi, C. B., Peter, L., Taylor, C. J., Chugh, S., Kendle, R., Chung, M. I., Flaherty, D. K., Matlock, B. K., Calvi, C. L., Blackwell, T. S., Ware, L. B., Bacchetta, M., Walia, R., Shaver, C. M., Kropski, J. A., McCarthy, D. J., & Banovich, N. E. (2024). Cell-type-specific and disease-associated expression quantitative trait loci in the human lung. *Nat Genet*, *56*(4), 595–604. https://doi.org/10.1038/s41588-024-01702-0

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R. R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J. C., Carroll, R. J., Eyler, A. E., Greenberg, J. D., Kremer, J. M.,…Plenge, R. M. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, *506*(7488), 376–381. https://doi.org/10.1038/nature12873

Pedersen, O. B., Svendsen, A. J., Ejstrup, L., Skytthe, A., Harris, J. R., & Junker, P. (2008). Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the relative importance of genetic vs. environmental

effectors in disease causation. *Scand J Rheumatol*, *37*(2), 120–126. https://doi.org/10.1080/03009740701824613

Rudwaleit, M., Landewe, R., van der Heijde, D., Listing, J., Brandt, J., Braun, J., Burgos-Vargas, R., Collantes-Estevez, E., Davis, J., Dijkmans, B., Dougados, M., Emery, P., van der Horst-Bruinsma, I. E., Inman, R., Khan, M. A., Leirisalo-Repo, M., van der Linden, S., Maksymowych, W. P., Mielants, H.,…Sieper, J. (2009). The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis*, *68*(6), 770–776. https://doi.org/10.1136/ard.2009.108217

Smith, J. A., & Colbert, R. A. (2014). Review: The interleukin-23/interleukin-17 axis in spondyloarthritis pathogenesis: Th17 and beyond. *Arthritis Rheumatol*, *66*(2), 231–241. https://doi.org/10.1002/art.38291

Zhang, M. J., Hou, K., Dey, K. K., Sakaue, S., Jagadeesh, K. A., Weinand, K., Taychameekiatchai, A., Rao, P., Pisco, A. O., Zou, J., Wang, B., Gandal, M., Raychaudhuri, S., Pasaniuc, B., & Price, A. L. (2022). Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet*, *54*(10), 1572–1580. https://doi.org/10.1038/s41588-022-01167-z

Chapter 7: Overall Discussion and Future Directions

7.1 Alterations in Peripheral B Cell Lineages Following TNF Inhibition in AxSpA

The first aim of this thesis was to investigate how TNF inhibitor therapy shapes the peripheral immune landscape in AxSpA. In the initial PBMC atlas (Chapter 3), I examined responders to TNF inhibition to identify cellular and molecular changes that may accompany clinical improvement and provide insight into disease mechanisms. Using scCODA, I found that the proportions of overall B cell lineages were significantly increased in TNFi treated AxSpA patients compared with biologic naïve patients. This increase following TNFi treatment was further validated by an external bulk RNA-seq dataset (Wang et al., 2017), in which demultiplexing revealed a similar upregulation of circulating B cells and plasma cells. A study in AS has also reported an expansion of $CD19^{+}CD24^{high}CD38^{high}$ regulatory B cells (Chen et al., 2021), which may represent one component of the broader B cell lineage increase observed in my dataset, although the specific subsets contributing to this shift remain to be fully clarified.

The effects of TNF inhibition on B cell homeostasis have been more extensively characterized in RA, where TNF antagonists have been shown to influence B cell subset composition, differentiation and trafficking, although the direction and magnitude of these changes vary across studies (Anolik et al., 2008; Daien et al., 2014; Souto-Carneiro et al., 2009). This variability likely reflects differences in disease context, cohort characteristics, treatment type and analytical approaches, and highlights that TNFi induced alterations in

circulating B cell populations may not follow a uniform pattern. In AxSpA, the observed increase in B cell lineages among responders suggests that B cell associated pathways may be modulated during effective therapy; however, whether these changes reflect restoration toward a healthier immune state, a therapy specific effect or a combination of both remains uncertain. Continued investigation of B cell subset dynamics, including regulatory and antibody secreting populations, would be valuable for clarifying their potential contributions to successful TNFi treatment response and disease pathophysiology.

7.2 Methods and Challenges in Building the AxSpA Atlas

A major component of this thesis was the construction of a multi-omics single-cell cross-tissue atlas of AxSpA and relevant comparators integrating samples from blood, synovial fluid, sacroiliac joint, spine, ileum and eye (Chapter 4). To our knowledge, this is the first single cell resource that brings together these AxSpA related tissues within a unified framework. Such a dataset enables systematic examination of shared and tissue specific immune states and provides a reference for interpreting AxSpA associated immune alterations across multiple anatomical sites.

By combining gene expression, surface protein measurements and TCR information, I aimed to provide more accurate annotation and interpretation of immune populations. The TCR modality assists in distinguishing MAIT and iNKT cells, and ADT markers such as CD45RA and CD45RO improve resolution of T cell differentiation states. This multimodal annotation underpins the downstream

analyses presented in this thesis and supports ongoing collaborative work using this atlas to identify cell populations potentially involved in AxSpA pathogenesis.

A central challenge in generating this resource was managing integration across multiple sources of heterogeneity, including technical variation, tissue specific differences and distinct omics layers. Integrating data from multiple sites with differing technical biases required methods that correct batch effects while preserving biological variation, which has been evaluated extensively in the benchmark study (Luecken et al., 2022). After evaluating alternative strategies, I selected scVI, which models technical noise while learning a latent representation that retains meaningful biological structure (Lopez et al., 2018). For cross-tissue integration, I applied a bespoke highly variable gene (HVG) selection procedure that identified HVGs within each tissue condition combination and used their union to retain both within tissue and within disease variation while resolving major immune lineages.

Integration of multiple omics layers posed an additional challenge. For CITE-seq data, I evaluated several normalization approaches and ultimately used totalVI, which jointly model gene expression and ADT counts within a shared latent space while accounting for their distinct noise distributions (Gayoso et al., 2021). Although totalVI generally improved ADT performance, some undesired imputation was observed when gene expression profiles were highly similar (Section 4.3.3.3), highlighting a limitation of co-modelling approaches of this type and reflecting current limitations of AI-based cellular prediction.

Recent advances in computational and AI-based methods including deep generative models for single-cell omics and the emergence of foundation

models trained on increasingly large and diverse single-cell datasets have greatly expanded our ability to integrate, annotate, and interpret cellular heterogeneity. Deep generative frameworks such as variational autoencoders and diffusion-based models continue to improve batch correction, multimodal integration and the recovery of rare or transitional cell states (Rivero-Garcia et al., 2024). At the same time, emerging foundation models (Cui et al., 2024; Hao et al., 2024) demonstrate how large scale pretraining can provide general purpose representations that transfer across datasets, cohorts and analytical tasks. These advances may further enhance future atlas work by supporting more stable integration across cohorts and by enabling more sensitive refinement of intermediate or disease associated cell states.

Together, the atlas developed in this thesis provides a comprehensive framework for dissecting AxSpA related immune states, and future applications of emerging AI based models and spatial transcriptomic approaches hold promise for integrating newly generated datasets from other studies, refining the existing findings and uncovering deeper mechanistic insights.

7.3 Shared T Cell and Myeloid Alterations Support Arthritogenic Peptide Hypothesis and Gut-Joint Axis in AxSpA

One aim of this thesis was to identify both tissue specific and tissue shared immune cell populations within the cross-tissue atlas. Tissue specific signals were analyzed in detail in the results sections (Chapters 5 and 6). From a mechanistic perspective, examining cross tissue signals is essential, as such patterns may indicate immune populations that operate across multiple anatomical sites and provide insight into potential pathways of inflammatory

propagation in AxSpA. I summarized the tissue shared signals in **Table 7.1**, including cell populations whose compositional changes were observed in more than one tissue.

LINEAGE	CELL TYPE	TISSUE ENRICHMENT	↑ IN AXSPA	↓ IN AXSPA	AS GENETIC ASSOCIATION
T CELL	CD4.TEMRA	PBMC	SFMC$^{+, cr}$ PEB$^{., eh}$		SIJ^{+} PEB*
	CD4.Tr1.act	SFMC		PBMC$^{*, cr}$ SFMC$^{*, cr}$ SFMC$^{***, er}$	ILM*
	CD8.TEMRA.ZNF683	PBMC	SFMC$^{*, cr}$ PEB$^{*, eh}$		EST* PEB^{+}
	CD8.TEMRA.ZNF683.KIR	PBMC	**PEB**$^{**, eh}$ SFMC$^{+, er}$		PEB^{+} ILM^{+}
	CD8.TCMRA.ZNF683.XCL1	SFMC, SIJ	*PBMC*$^{*, cr}$ ILM$^{*, eh}$	**PEB**$^{***, eh}$	
	CD8.TEM.HLA-DR.KLF12	SFMC	PBMC$^{*, eh}$ ILM$^{., eh}$		SIJ^{+}
	CD8.TEM.2	PEB, SFMC	PEB$^{., eh}$	PBMC$^{*, eh}$ SFMC$^{*, er}$	
	CD8.TEM.3	SFMC	PBMC$^{., eh}$ **SFMC**$^{., er}$		
	UTC.MAIT-like	PBMC, PEB		SFMC$^{., er}$ EST$^{., eh}$	SIJ^{+} PEB* ILM*
MYELOID	neut.CSF3R	SIJ	*PBMC*$^{*, ch}$ ILM$^{+, eh}$	PBMC$^{*, er}$	SIJ^{+}
	cMono.3	PBMC	**SFMC**$^{., cr}$ **PBMC**$^{+, eh}$ ILM$^{+, eh}$		SIJ^{+}
	cMono.IL1B		ILM$^{+, eh}$	**PBMC**$^{*, cr}$	
	intMono.1	ILM, SFMC	**SFMC**$^{*, cr}$ ILM$^{+, eh}$		
	intMono.SPP1	SFMC	ILM$^{+, eh}$ PEB$^{+, eh}$	PBMC$^{*, eh}$	PEB* ILM*
	ncMono.1	PBMC	ILM$^{., eh}$ PBMC$^{+, eh}$		SFMC^{+}
	mregDC	SFMC	ILM$^{+, eh}$	PBMC$^{**, ch}$ PBMC$^{***, eh}$	
B CELL	B.mem.unsw.CXCR4	ILM, PBMC	SFMC$^{., cr}$ PBMC$^{., cr}$		
	PC.cyc		PBMC$^{., cr}$ PBMC$^{*, ch}$ PEB$^{+, eh}$		

Table 7.1 Summary of cell-type compositional differences observed across tissues

Summary of cell types showing differential abundance across more than one tissue. Superscripts indicate the evidence source (^{c} core cohort; ^{e} expanded cohort; ^{r} compared with RA; ^{h} compared with HC). Italic tissues indicate findings in the biologic-

naïve subgroup, and bold entries indicate technical validation by scCODA or Milo. AS genetic association is derived from scDRS scores; “MAIT-like” includes both KLRB1.MAIT and MAIT-like αβT cells. BH-adjusted p-value thresholds: *** < 0.001, ** < 0.01, * < 0.05, + < 0.1, . < 0.2.

As shown in the table, the shared differences observed in the T cell compartment were most prominent within the TEMRA, TCMRA, and TEM populations, indicating that memory T cell differentiation states are consistently altered across AxSpA-relevant tissues compared with disease and healthy comparators. Notably, $CD8^+$ and $CD4^+$ TEMRA populations, including the $ZNF683^+$ and KIR^+ CD8 TEMRA subsets were significantly enriched in both the core cohort of SFMC compared with RA, and similarly elevated in extended cohort of PEB compared with healthy controls. When examining the minor T cell-type annotations (**Appendix D.13**), total CD8 TERMA frequencies were also increased in both ILM and PEB, reaching BH-adjusted significance levels < 0.2 in the extended cohort. Although this cohort contains covariates, the directional consistency across tissues remains notable.

TEMRA cells represent terminally differentiated effector-memory T cells, typically arising after repeated or persistent antigenic stimulation and progressing toward cytotoxic phenotypes (Sallusto et al., 1999; Sathaliyawala et al., 2013). The presence of *ZNF683* (Hobit) expression of the $CD8^+$ TEMRA subsets further supports a profile associated with tissue retention, chronic stimulation and late effector differentiation (Mackay et al., 2016; Parry et al., 2023).

In addition to TEMRA populations, several TEM subsets (CD8.TEM.HLA-DR.KLF12, CD8.TEM.2, CD8.TEM.3) were also significantly increase cross AxSpA-related tissues, although observed in the extended cohort. Antigen-

specific $CD8^+$ T cells detected by HLA-peptide multimer staining are typically enriched within TEM and TEMRA subsets, while the relative distribution may vary depending on the antigen involved (Appay et al., 2002; Champagne et al., 2001; Voehringer et al., 2001). These phenotypes have been repeatedly associated with antigen driven expansion and persistence of cytotoxic T cell clones within inflamed tissues, including in autoimmune and chronic inflammatory conditions. These findings are conceptually aligned with the arthritogenic peptide hypothesis, which proposes that HLA-B*27-restricted arthritogenic peptides drive recurring TCR engagement and generate sustained cytotoxic effector responses in joint and enthesis-associated tissues. In this context, the observed expansion of TEM and TEMRA cells might reflect repeated TCR triggering, persistent presentation of candidate arthritogenic peptides, and local tissue inflammation. The parallel increase across disease-relevant tissues suggests a unified mechanism of chronic antigen exposure, reinforcing the arthritogenic hypothesis.

In the myeloid compartment, the most consistent tissue shared differences were observed between the ileum and other tissues, including PBMC, SFMC and PEB. All myeloid subsets showing shared changes were significantly increased in the ileum within the extended cohort. This pattern may suggest that mucosal myeloid activation propagates systemically through trafficking cells or soluble mediators, raising the possibility that ileal myeloid activity represents an upstream immune alteration in AxSpA, although direct causality cannot be inferred from current data. Mucosal dendritic cells and macrophages respond sensitively to microbial, metabolic and barrier related cues, and shifts in these populations can influence local cytokine production

and antigen presentation within the gut environment. Such ileal alterations could plausibly shape downstream immune responses by affecting circulating immune cells, modulating T cell differentiation or contributing to broader systemic inflammatory tone. While functional validation is needed, the observation that ileal myeloid signatures are shared with other AxSpA relevant tissues is consistent with the emerging view that gut derived antigens may contribute to systemic immune activation in AxSpA (Qaiyum et al., 2021). This interpretation aligns with the arthritogenic antigen hypothesis and proposes that HLA-B*27 restricted microbial or mucosal antigens originating in the gut can subsequently prime immune responses that contribute to the pathogenic T cell activation.

It's important to note the findings presented in this thesis primarily provide cellular correlates consistent with the arthritogenic peptide hypothesis rather than direct mechanistic evidence for it. The expansion of TEM and TEMRA populations and the cross-tissue convergence of cytotoxic effector signature are compatible with persistent HLA-B*27 restricted antigen presentation, although the antigen identities themselves were not identified here and represent a key direction of our future work. The two alternative hypotheses: ER stress / UPR and HLA-B27 free heavy chain formation were not directly tested in this thesis. ER stress and UPR pathways were examined only indirectly through targeted transcriptional pathway analyses and were not the focus of the cell-state findings described above. Direct surface staining for HLA-B*27 homodimers was attempted as part of the CITE-seq panel but did not yield reliable data, and homodimer-mediated mechanisms therefore remain outside the scope of the current cellular analyses. Differentiating between these three mechanisms and identifying their respective contributions to AxSpA

pathogenesis will ultimately require orthogonal approaches such as HLA-peptide multimer staining, surface biochemistry, or in vitro UPR assays beyond the single-cell measurements employed here.

7.4 Limitations and Future Analytical Directions

Although a substantial number of samples were incorporated into the atlas, several key constraints should be kept in mind when interpreting the results. In particular, the ileum and spine datasets are affected by limited sample numbers and potential covariates, including fresh versus frozen processing and FACS-sorted versus unfractionated input, which resulted in imbalances in the comparison design that cannot be fully controlled for. For the sacroiliac joint samples, no direct comparator group was available. As a result, observations from these tissues depend on the extended cohort rather than the core cohort with a fully matched design, and their interpretation requires caution. These sample size limitations also have direct implications for the detection power of cell-state analyses: rare cell populations within the sacroiliac joint and spine cohorts are particularly vulnerable to be under-detected, and tissue-specific differential signals in the extended cohort may not reach significance even when biologically meaningful trends are present. Sample acquisition is particularly challenging for the sacroiliac joint, which requires invasive surgical access and is difficult to obtain in sufficient numbers for a comparator-balanced study. In this thesis, sacroiliac joint tissue was obtained by CT-guided percutaneous bone marrow biopsy. Because CT does not provide real-time visualisation of active bone marrow oedema, needle targeting relied on spatial correlation between pre-procedural MRI and CT-visible bony

landmarks, without accounting for any interval change in disease activity. The cellular content of subchondral bone marrow samples is also inherently variable, potentially containing a mixture of hematopoietic marrow, fibrotic tissue, cartilage, and cortical bone fragments; this heterogeneity was considered during cell-type annotation.

Building on the current findings from this atlas, several important downstream analyses remain to be undertaken. A key next step will be to define the transcriptional and ADT phenotypes of the AxSpA-associated populations identified here and determine how these states differ from their counterparts in RA and healthy controls. These AxSpA-associated populations include those that are compositionally increased in AxSpA, genetically enriched for AS GWAS variants, transcriptomically enriched for disease-correlated pathways, or that contains public AxSpA-associated TRBV9^{+} TCR clones. For example, **Figure 7.1** shows the UMAP locations and cell-type annotations of cells carrying the public TRBV9^{+} clones, and further work is needed to assess whether the MAIT-like, CD8.TEM.3, and T.SP.TEM.CD69.IL23R populations containing these clones share common transcriptomic or ADT-defined phenotypes. Another priority will be a detailed comparison of the GZMK-expressing CD8^{+} T cells and the ZNF683^{+} CD8^{+} T cells between AxSpA and RA, as these subsets repeatedly emerged as AxSpA-related populations across tissues. The atlas also provides a foundation for testing observations from the wider literature, including the proposed involvement of Th17-related populations in spondyloarthritis, by examining whether such signatures are consistently detectable across tissues and cohorts.

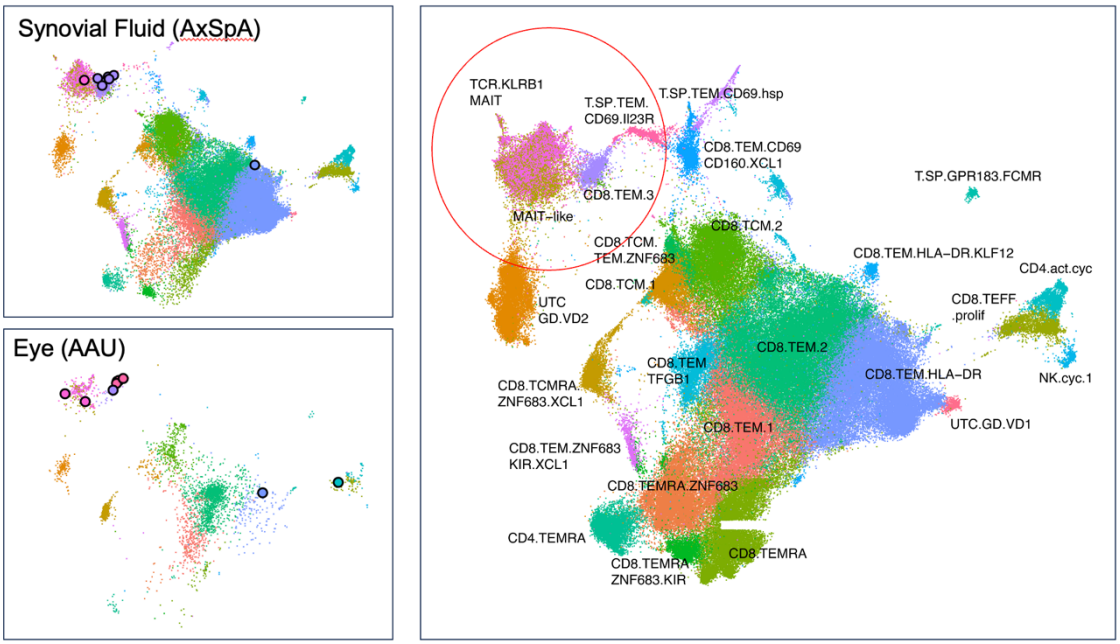


Figure 7.1 Public AxSpA-associated TRBV9⁺ T-cell clones detected across AxSpA synovial fluid and AAU ocular tissue
UMAP visualisation of the CD8⁺ T-cell region from the AxSpA atlas (right), shown alongside the distribution of public AxSpA-associated TRBV9⁺ TCR clones identified in AxSpA synovial fluid (top left) and acute anterior uveitis (AAU) ocular tissue (bottom left). Public clones are highlighted to illustrate their localization within specific cell populations. TCR data were processed by Dr Frank Penkava, and the figure was generated by Dr Stephen Sansom.

Beyond these specific analyses, broader future work will integrate atlas-derived findings with complementary approaches undertaken by other team members, including differential expression profiling on different cell subsets, further TCR repertoire characterisation, unbiased pathway-level analysis, and deeper tissue-specific investigations. Validation of the identified cell states and compositional patterns will be important for strengthening confidence in the initial observations, and several complementary strategies could be pursued. First, the candidate AxSpA-associated populations identified here, including ZNF683⁺ CD8⁺ TEM/TEMRA, cMono.3, and neut.CSF3R⁺IL1B⁺ subsets, could be re-examined in independent single-cell datasets from AxSpA cohorts. Second, flow cytometry and mass cytometry (CyTOF) panels could be designed

around the defining surface markers of these populations to permit higher-throughput quantification in larger patient cohorts. Third, FACS-isolation of these populations followed by bulk transcriptomic or epigenomic profiling could provide deeper characterisation of their regulatory programs and lineage relationships than is achievable from single-cell data alone. In addition, spatial transcriptomics or imaging-based approaches could clarify the localization, microenvironmental context and putative cellular interactions of these subsets within inflamed tissues. As spatial technologies continue to advance, integrating spatially resolved data with this atlas will be valuable for determining whether AxSpA associated cell states reside in specific tissue niches.

Together, these directions provide a foundation for translating atlas-level observations into a more detailed understanding of AxSpA pathophysiology. At the patient level, several translational opportunities could emerge from this work. First, the AxSpA-associated cell populations identified in this thesis, including the ZNF683^{+} CD8^{+} TEM/TEMRA subsets and cMono.3 monocyte subset, could serve as candidate biomarkers for monitoring disease activity or treatment response. Second, the heterogeneity captured in this atlas could inform stratification of AxSpA patients, which has potential implications for treatment selection. Third, the cellular pathways highlighted across multiple tissues identify candidate molecular nodes for targeted therapeutic intervention beyond TNF inhibition. In the longer term, delineating the key T cell and myeloid populations, their upstream drivers, and their tissue specific functions may help guide the development of more targeted therapeutic strategies, including approaches aimed at modulating chronic antigen stimulation, attenuating mucosal myeloid activation, or disrupting pathogenic effector programs.

7.5 References

Anolik, J. H., Ravikumar, R., Barnard, J., Owen, T., Almudevar, A., Milner, E. C., Miller, C. H., Dutcher, P. O., Hadley, J. A., & Sanz, I. (2008). Cutting edge: anti-tumor necrosis factor therapy in rheumatoid arthritis inhibits memory B lymphocytes via effects on lymphoid germinal centers and follicular dendritic cell networks. *J Immunol*, *180*(2), 688–692. https://doi.org/10.4049/jimmunol.180.2.688

Appay, V., Dunbar, P. R., Callan, M., Klenerman, P., Gillespie, G. M., Papagno, L., Ogg, G. S., King, A., Lechner, F., Spina, C. A., Little, S., Havlir, D. V., Richman, D. D., Gruener, N., Pape, G., Waters, A., Easterbrook, P., Salio, M., Cerundolo, V.,…Rowland-Jones, S. L. (2002). Memory CD8+ T cells vary in differentiation phenotype in different persistent virus infections. *Nat Med*, *8*(4), 379–385. https://doi.org/10.1038/nm0402-379

Champagne, P., Ogg, G. S., King, A. S., Knabenhans, C., Ellefsen, K., Nobile, M., Appay, V., Rizzardi, G. P., Fleury, S., Lipp, M., Forster, R., Rowland-Jones, S., Sekaly, R. P., McMichael, A. J., & Pantaleo, G. (2001). Skewed maturation of memory HIV-specific CD8 T lymphocytes. *Nature*, *410*(6824), 106–111. https://doi.org/10.1038/35065118

Chen, R., Qian, H., Yuan, X., Chen, S., Liu, Y., Wang, B., & Shi, G. (2021). Immunological Changes in Peripheral Blood of Ankylosing Spondylitis Patients during Anti-TNF-alpha Therapy and Their Correlations with Treatment Outcomes. *J Immunol Res*, *2021*, 1017938. https://doi.org/10.1155/2021/1017938

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., & Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods*, *21*(8), 1470–1480. https://doi.org/10.1038/s41592-024-02201-0

Daien, C. I., Gailhac, S., Mura, T., Combe, B., Hahne, M., & Morel, J. (2014). High levels of memory B cells are associated with response to a first tumor necrosis factor inhibitor in patients with rheumatoid arthritis in a longitudinal prospective study. *Arthritis Res Ther*, *16*(2), R95. https://doi.org/10.1186/ar4543

Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., & Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat Methods*, *18*(3), 272–282. https://doi.org/10.1038/s41592-020-01050-x

Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., Wang, T., Ma, J., Zhang, X., & Song, L. (2024). Large-scale foundation model on single-cell transcriptomics. *Nat Methods*, *21*(8), 1481–1491. https://doi.org/10.1038/s41592-024-02305-7

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat Methods*, *15*(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2

Luecken, M. D., Buttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colome-Tatche, M., & Theis, F. J. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods*, *19*(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8

Mackay, L. K., Minnich, M., Kragten, N. A., Liao, Y., Nota, B., Seillet, C., Zaid, A., Man, K., Preston, S., Freestone, D., Braun, A., Wynne-Jones, E., Behr, F. M., Stark, R., Pellicci, D. G., Godfrey, D. I., Belz, G. T., Pellegrini, M., Gebhardt, T.,…van Gisbergen, K. P. (2016). Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science*, *352*(6284), 459–463. https://doi.org/10.1126/science.aad2035

Parry, E. M., Lemvigh, C. K., Deng, S., Dangle, N., Ruthen, N., Knisbacher, B. A., Broseus, J., Hergalant, S., Guieze, R., Li, S., Zhang, W., Johnson, C., Long, J. M., Yin, S., Werner, L., Anandappa, A., Purroy, N., Gohil, S., Oliveira, G.,…Wu, C. J. (2023). ZNF683 marks a CD8(+) T cell population associated with anti-tumor immunity following anti-PD-1 therapy for Richter syndrome. *Cancer Cell*, *41*(10), 1803–1816 e1808. https://doi.org/10.1016/j.ccell.2023.08.013

Qaiyum, Z., Lim, M., & Inman, R. D. (2021). The gut-joint axis in spondyloarthritis: immunological, microbial, and clinical insights. *Semin Immunopathol*, *43*(2), 173–192. https://doi.org/10.1007/s00281-021-00845-0

Rivero-Garcia, I., Torres, M., & Sanchez-Cabo, F. (2024). Deep generative models in single-cell omics. *Comput Biol Med*, *176*, 108561. https://doi.org/10.1016/j.compbiomed.2024.108561

Sallusto, F., Lenig, D., Forster, R., Lipp, M., & Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature*, *401*(6754), 708–712. https://doi.org/10.1038/44385

Sathaliyawala, T., Kubota, M., Yudanin, N., Turner, D., Camp, P., Thome, J. J., Bickham, K. L., Lerner, H., Goldstein, M., Sykes, M., Kato, T., & Farber, D. L. (2013). Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity*, *38*(1), 187–197. https://doi.org/10.1016/j.immuni.2012.09.020

Souto-Carneiro, M. M., Mahadevan, V., Takada, K., Fritsch-Stork, R., Nanki, T., Brown, M., Fleisher, T. A., Wilson, M., Goldbach-Mansky, R., & Lipsky, P. E. (2009). Alterations in peripheral blood memory B cells in patients with active rheumatoid arthritis are dependent on the action of tumour necrosis factor. *Arthritis Res Ther*, *11*(3), R84. https://doi.org/10.1186/ar2718

Voehringer, D., Blaser, C., Brawand, P., Raulet, D. H., Hanke, T., & Pircher, H. (2001). Viral infections induce abundant numbers of senescent CD8 T cells. *J Immunol*, *167*(9), 4838–4843. https://doi.org/10.4049/jimmunol.167.9.4838

Wang, X. B., Ellis, J. J., Pennisi, D. J., Song, X., Batra, J., Hollis, K., Bradbury, L. A., Li, Z., Kenna, T. J., & Brown, M. A. (2017). Transcriptome analysis of ankylosing spondylitis patients before and after TNF-alpha inhibitor therapy reveals the pathways affected. *Genes Immun*, *18*(3), 184–190. https://doi.org/10.1038/gene.2017.19