

Cell and tissue specific functional genomics of psoriasis and psoriatic arthritis

Alicia Lledó Lara
Hertford College
University of Oxford

A thesis submitted in partial
fulfilment of the requirements for the degree of
Doctor of Philosophy
Hilary Term, 2019

Abstract

Cell and tissue specific functional genomics of psoriasis and psoriatic arthritis

Alicia Lledó Lara, Hertford College, Hilary Term 2019
A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy of the University of Oxford

Psoriasis and psoriatic arthritis (PsA) are chronic inflammatory diseases mainly affecting the skin and joints that result from the interaction of genetic and environmental factors. Despite the success of genome-wide association studies (GWAS) in uncovering genetic risk loci, the functional mechanisms underlying these associations remain largely unresolved. This thesis aims to establish genome-wide epigenetic and gene expression profiles for immune cells isolated from blood and disease-relevant tissues to inform the understanding of pathogenesis and GWAS in psoriasis and PsA.

The first results chapter establishes methodological and analytical pipelines for novel chromatin profiling techniques in challenging clinical samples. Importantly, Omni-ATAC, a variant of Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), demonstrated the best performance for skin biopsies. Moreover, the use of cryopreservation and fixation in blood-isolated immune cells showed cell-type specific impact on the chromatin accessibility landscape that should be taken into consideration when planning the experimental design.

The second results chapter compares chromatin accessibility, histone acetylation and gene expression between psoriasis patients (n=8) and controls (n=10) for blood monocytes, B cells, $CD4^+$ and $CD8^+$ T cells. Only $CD8^+$ T cells showed a substantial number of differentially accessible regions (DARs) (n=55, FDR< 0.05), and intersection with differential H3K27ac was only seen at an intron of *DTD1*. Monocytes and $CD8^+$ T cells showed highest numbers of differentially expressed genes (n=671 and 651 respectively, FDR<0.05) with enrichment of MAPK and IL-12 signalling (both cell types) and NF-κB, TNF and chemokine signalling ($CD8^+$ T cells). Overall 1,227 genes (FDR<0.05) were differentially expressed between uninvolved and lesional psoriasis epidermis (n=3) with enrichment of metabolic and immune-related pathways. Integration of GWAS fine-mapping data with epigenetic and gene expression profiles implicated a potentially functional variant in the 2p15 GWAS locus modulating *B3GNT2*.

The third results chapter analyses differences in chromatin accessibility, gene and protein expression of immune cells between synovial fluid and peripheral blood of PsA patients (n=3). The highest number of DARs were found in monocytes (5,285 FDR<0.01) for both tissues with synovial fluid monocytes specifically enriched for interleukin and NF-κB signalling pathways. Single-cell RNA-seq identified two functionally relevant synovial fluid monocyte subpopulations characterised by up-regulation of IFN signalling and *IL7R* genes, respectively. Mass-cytometry analysis (n=10) confirmed increased *CCL2* and *CXCL10* protein levels in monocytes from synovial fluid. Furthermore, statistical fine-mapping of PsA GWAS loci and integration with ATAC data suggested rs11249213 as a possible regulator of *RUNX3* in $CD8^+$ cells in the inflamed synovium.

Overall this thesis highlights the context-specificity of the epigenomic landscape in psoriasis and PsA, and the potential of a multi-omics approach to provide new insights into pathophysiology and interpretation of GWAS.

Acknowledgements

I can hardly believe that the time of writing the last page of this thesis has come. This is no doubt the most important one for myself, as it gives me the opportunity to be grateful to all the people who have helped me during this challenging time.

First, I would like to thank my supervisor Prof Julian Knight and my co-supervisor Dr Antonio Berlanga for giving me this opportunity to grow as a scientist and as a person, for their trust, patience and understanding. I am also very greatful to Dr Hussein Al-Mossawi, who has been advicing and helping throughout the project, and especially during the writing up.

I want to say a huge thank you to my colleague and friend Anna Sanniti. Without you I would have never got here. Thank you for making me a better person, for teaching me to believe in myself and for all the amazing memories during the past 3 years. I am also extremely grateful to everyone in the Knight group (present and past members), and especially to Katie, Andy, Ola, Evelyn, Hai, Justin, Giuseppe and Cyndi, for their support in the scientific and personal level, and for giving me so much love. You are an amazing group. Big thank you also to everyone in the WCHG that have made things easier (Core, IT, lab support and colleagues from other groups), with special mention to Moustafa Attar, Amy Trebes, Silvia Salatino and Ruth Porter.

I would like to thank my group of friends in Oxford: César, Lucía, Raquel, Esther, Jacob, Blanca and Adrián. Particularly, I want to thank César and Lucía for taking care of me as if they were family, and for being there day and night. You both mean the world to me. A mention also to my mentor Ruth McCalman for your care, love, support and advice. And of course thank you to the soul of Branca (Bobbie) for looking after me during the long hours of writing up there.

I would also like to say a big thank you to my Spanish girls Helena and Patri for always being supportive in the distance, and for being ready to visit me and cheer me up. I would also like to express my gratitude to my friends from Valencia and uni: Esther, Marta, Carlos, Regina, Edgar and, particularly, to Eva and David, who have always been there dealing with my insecurities for the last 15 years and pushing me forward every time I felt I couldn't do it. I love you lots.

I would have never got here without the love from my mum, my dad and my brother, who have supported me unconditionally at all the stages in my life. Thank you for believing in me and for always being proud of me. Thank you to mi yaya Lola, and to yayo Sento, yaya Mila and yayo Pepe (wherever you are) who haven't seen me crossing the finish line but have taught me many of the most important values in life. I miss you all.

And lastly, thank you to Javi for having walked alongside with me during the last past year, despite all the ups and downs. I know it hasn't been easy and I really appreciate it. At the end, as you always say, ''todo va a estar bien".

Declarations

I declare that, unless otherwise stated, all work presented in this thesis is my own. Some aspects of the thesis were a collaboration, with some of the work conducted with or by others.

All the healthy volunteers and psoriasis patients' samples were collected by myself and processing was part of a collaborative effort with past and current lab members Dr Anna Sanniti, Dr Andrew Brown and Giuseppe Scozzafava. Processing of skin biopsies for the adherent assay was conducted with help from Prof Graham Ogg and Dr Danuta Gutowska-Owsiak. The psoriatic arthritis samples processed for ATAC, qPCR array and mass cytometry were part of the Immune Function in Inflammatory Arthritis (IFIA) study established in 2006 and sample collection was a collaborative effort with Dr Hussein Al-Mossawi and Dr Nicole Yager.

The fixation protocol for sorted primary cells using DSP was optimised by Moustafa Attar. RNA extraction, ATAC and ChIPm processing for the healthy controls and psoriasis cohorts was carried out together with the ankylosing spondylitis samples in collaboration with Dr Anna Sanniti and Dr Andrew Brown. Advice for ATAC and ChIPm library indexing and sequencing was provided by Amy Trebes. RNA-seq and 10X Genomics technology Chromium single cell 3' expression library preparations and sequencing together with ATAC and ChIPm sequencing were performed by Oxford Genomics Centre at the Wellcome Centre for Human Genetics. Processing of the qPCR array and mass cytometry samples was conducted by UCB and measurement of synovial fluid cytokine and chemokine abundance was carried out by collaborators in Basle.

Regarding data analysis, mass cytometry data was analysed by Dr Nicole Yager. ATAC and ChIPm NGS data processing was conducted using in house pipelines developed by Dr Gabriele Migliorini towards which I actively contributed by performing literature review analysis of the most appropriate methods, facilitating code for some of the parts and performing additional analysis to test the approaches. Peak filtering strategy for ATAC using IDR was proposed and implemented by Dr Gabriele Migliorini and I conducted additional analysis for validation. The strategy to perform filtering of chromatin accessible regions based on an empirical cut-off to remove excessive noise was developed and implemented together with Dr Hai Fang. RNA-seq NGS data processing was performed using the in-house pipeline developed by Dr Katie Burnham. Resources for fine-mapping analysis were provided by Dr Adrián Cortés. Single-cell advice and scripts for some of the analysis was facilitated by Arcadio Rubio. The script to calculate enrichment across TSS was provided by Dr Silvia Salatino. The function for colour-coding KEGG pathways based on gene expression data was developed by Dr Hai Fang with contributions for pathway curation from Dr Anna Sanniti and myself. General advice on analysis were provided by Dr Silvia Salatino, Dr Hai Fang, Dr Katie Burnham, Dr Félicie Constantino, Dr Gabriele Migliorini and Enrique Váquez de Luis.

Contents

Abstract i

Acknowledgements ii

Declarations iii

Contents iv

List of Figures xiii

List of Tables xv

Abbreviations xvi

1 Introduction 1

1.1 Psoriasis and psoriatic arthritis . . . 1
1.1.1 Epidemiology and global impact . . . 2
1.1.2 Psoriasis and inflammatory dermatoses . . . 3
1.1.3 PsA and spondyloarthropathies . . . 3
1.2 Pathophysiology of psoriasis and psoriatic arthritis . . . 5
1.2.1 Clinical presentation and diagnosis . . . 5
1.2.2 Aetiology of psoriasis and PsA . . . 7
1.2.3 Cell types involved in psoriasis and PsA pathogenesis . . . 12
1.2.4 Therapeutic intervention . . . 16
1.3 Genetics of psoriasis and psoriatic arthritis . . . 18
1.3.1 Heritability . . . 18
1.3.2 Genome-wide association studies . . . 18
1.3.3 Relevance of non-coding variants in disease susceptibility . 25
1.3.4 The role of GWAS in highlighting immune-relevant cell types and pathways . . . 26
1.3.5 Psoriasis and PsA: the same or distinct disease entities . . . 31
1.3.6 Limitations of GWAS . . . 33
1.4 Functional interpretation of GWAS in complex diseases . . . 34
1.4.1 Overcoming GWAS limitations . . . 34
1.4.2 Understanding the epigenetic landscape in complex diseases 35
1.4.3 The chromatin landscape . . . 36
1.4.4 Transcriptional profiles in disease . . . 42
1.4.5 Transcriptional regulation in complex diseases . . . 45
1.4.6 The use of fine-mapping to prioritise functional causal variants . . . 46
1.4.7 Generation, integration and interpretation of genomic data 48
1.4.8 Aims and objectives . . . 49

2 Material and Methods **52**
2.1 Ethical approval and recruitment of study participants 52
2.1.1 Psoriasis patient recruitment 52
2.1.2 PsA patient recruitment 53
2.1.3 Healthy volunteer recruitment 53
2.2 Sample processing 54
2.2.1 PBMCs and synovial fluid cells isolation 54
2.2.2 Skin biopsies processing and adherence assays 55
2.2.3 Fixation, cryopreservation and cell culture 56
2.2.4 Primary cell isolation using magnetic-activated cell sorting 57
2.2.5 Primary cell isolation using fluorescence-activated cell sorting 57
2.3 Experimental protocols 59
2.3.1 ATAC - Chromatin Accessibility 59
2.3.2 Chromatin immunoprecipitation with sequencing library preparation by Tn5 transposase 61
2.3.3 RNA extraction and gene expression quantification 63
2.3.4 DNA extraction and rs4672405 genotyping 66
2.3.5 Mass cytometry using cytometry by time of flight (CyTOF) 66
2.4 Computational and statistical analysis 68
2.4.1 ATAC data analysis 68
2.4.2 ChIPm data analysis 70
2.4.3 Gene expression analysis 72
2.4.4 Genomic region annotation, enrichment analysis and pathway visualisation 75
2.4.5 Statistical fine-mapping 77
2.4.6 Mass cytometry data analysis 79

3 Establishing laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples **80**
3.1 Introduction 80
3.1.1 Principle of ATAC-seq and compatibility with clinical samples 80
3.1.2 ATAC-seq limitations and advances in optimisation 81
3.1.3 Challenges of ATAC data analysis 82
3.1.4 The challenge of working with clinical samples 85
3.1.5 Aims 86
3.2 Results 87
3.2.1 Establishment of an ATAC-seq data analysis pipeline 87
3.2.2 Assessment of ATAC-seq transposition times in relevant cell types 100
3.2.3 Comparison of ATAC-seq with Fast-ATAC protocol 102
3.2.4 Limitations of ATAC-seq and Fast-ATAC to assess chromatin accessibility in keratinocytes 104
3.2.5 Effect of cryopreservation and fixation in the chromatin landscape of immune primary cells 111
3.3 Discussion 121

3.3.1 ATAC: methodological aspects and pipeline establishment 122
3.3.2 The challenges of performing differential chromatin accessibility analysis . . . 125
3.3.3 Studying the chromatin landscape from psoriasis skin biopsies . . . 126
3.3.4 Characterisation of the effect of preservative techniques in the chromatin landscape . . . 127
3.3.5 Limitations . . . 130
3.4 Conclusions . . . 131

4 Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis 132
4.1 Introduction . . . 132
4.1.1 The systemic and skin-specific manifestations of psoriasis . 132
4.1.2 The personalised epigenome in disease . . . 133
4.1.3 Transcriptional profiles in psoriasis . . . 134
4.1.4 Chromatin accessibility, gene expression and genetic variability . . . 138
4.1.5 Fine-mapping using summary statistics . . . 139
4.2 Aims . . . 140
4.3 Results . . . 141
4.3.1 Psoriasis and healthy controls: cohort description and datasets . . . 141
4.3.2 Investigation of psoriasis-specific changes in the enhancer mark H3K27ac in peripheral blood immune cell populations 145
4.3.3 Interrogation of chromatin accessibility changes in psoriasis peripheral blood immune cells . . . 150
4.3.4 Differential gene expression analysis in circulating immune cells in psoriasis . . . 156
4.3.5 RNA-seq in epidermis from psoriasis patients . . . 168
4.3.6 Comparison of systemic and tissue specific gene expression signatures in psoriasis . . . 177
4.3.7 Integration of chromatin accessibility and expression data for peripheral blood immune cells in psoriasis . . . 178
4.3.8 Fine-mapping of psoriasis GWAS loci and functional interpretation . . . 179
4.3.9 Allele-specific differences in chromatin accessibility at the GWAS locus 2p15 . . . 187
4.4 Discussion . . . 194
4.4.1 Chromatin accessibility and H3K27ac landscape in psoriasis immune cells . . . 194
4.4.2 Dysregulation of gene expression in psoriasis circulating immune cells . . . 195
4.4.3 Integration of chromatin accessibility and gene expression data . . . 200
4.4.4 Trancriptomic profiles in lesional and uninvolved psoriatic epidermis . . . 201

4.4.5 LncRNAs in psoriasis . . . 205
4.4.6 Fine-mapping using summary statistics and integration with epigenetic data . . . 207
4.4.7 Limitations in the approach . . . 211
4.5 Conclusions . . . 212

5 Cross-tissue comparison of chromatin accessibility, gene expression signature and cytokine production in PsA 213
5.1 Introduction . . . 213
5.1.1 The relevance of cell type and tissue specificity in the study of PsA . . . 213
5.1.2 Bulk transcriptomic studies in PsA and their limitations . . 214
5.1.3 Transcriptomics and proteomics at the single cell resolution 215
5.1.4 Multi-omics approach in the study of complex diseases . . 216
5.1.5 Integration of fine-mapping GWAS SNPs and functional data in PsA . . . 218
5.2 Aims . . . 219
5.3 Results . . . 220
5.3.1 PsA patients cohort description and datasets . . . 220
5.3.2 Immune cellular composition of blood and synovial fluid in the PsA cohort . . . 222
5.3.3 Differential chromatin accessibility analysis in immune cells reveals differences between synovial fluid and peripheral blood . . . 223
5.3.4 Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility . . . 232
5.3.5 Differential gene expression analysis in paired circulating and synovial immune cells . . . 234
5.3.6 Characterisation of $CD14^+$ monocyte heterogeneity in PsA using scRNA-seq . . . 244
5.3.7 Identification of additional pathways dysregulated between synovial fluid and peripheral blood $CD14^+$ monocytes using scRNA-seq data . . . 251
5.3.8 Integration of mass cytometry differences in protein expression with chromatin accessibility and gene expression data in PsA $CD14^+$ monocytes. . . . 254
5.3.9 Prioritisation and interpretation of PsA GWAS SNPs . . . 259
5.4 Discussion . . . 271
5.4.1 Characterising chromatin accessibility in PsA samples . . . 271
5.4.2 Bulk qPCR expression profiling and integration with paired chromatin accessibility data . . . 272
5.4.3 The relevance of monocytes in the PsA synovial fluid . . . 275
5.4.4 Characterisation of synovial and peripheral blood $CD14^+$ monocytes by scRNA-seq . . . 276
5.4.5 Integrating mass cytometry data with chromatin accessibility and gene expression changes in PsA $CD14^+$ monocytes . . 280
5.4.6 Use of PsA functional data to inform GWAS fine-mapping . 283

5.4.7 Limitations of the study 285
5.5 Conclusions 286

6 General discussion **288**
6.1 The importance of cell type and context specificity in disease . . . 289
6.2 The epigenetic landscape in psoriasis and PsA clinical samples: systemic vs affected tissue 290
6.3 Cell type and tissue specific transcriptional profiles in clinical cohorts 291
6.4 Data integration in multi-omic studies 294
6.5 GWAS fine-mapping and integration with patients-derived epigenetic data 296
6.6 Clinical relevance and translation 297
6.7 Future work 298
6.8 Concluding remarks 302

Appendices **304**

A Tables **307**
A.1 Chapter 3 Tables 307
A.2 Chapter 4 Tables 308
A.3 Chapter 5 Tables 312

B Additional figures **315**
B.1 Chapter 3 Figures 315
B.2 Chapter 4 Figures 318
B.3 Chapter 5 Figures 323

Bibliography **369**

List of Figures

1.1 Environmental triggers and genetic predisposition leading to psoriasis and PsA (adapted from Nestle *et al.* 2009). 12
1.2 The epigenetic landscape and its role in regulation of gene expression. 37

3.1 Measurements for quality control assessment in ATAC-seq samples 89
3.2 FRiP and peak calling at different sequencing depths in ATAC-seq libraries. 92
3.3 Peak calling filtering using IDR analysis in ATAC-seq samples. . . 94
3.4 Work flow illustrating the strategy to account for ATAC background noise prior to differential analysis. 96
3.5 Normalisation and differential chromatin accessibility analysis for different cut-offs using quantile normalisation limma voom and DESeq2. 98
3.6 Comparison of the DARs identified by differential analysis using limma voom or DESeq for the consensus peak list filtered at the optimal 80% cut-off. 99
3.7 Enrichment analysis for the significant DARs identified by DESeq2 between $CD14^+$ monocytes and $CD4^+$ cells. 100
3.8 Assessment of the effect of transposition times on the ATAC-seq measurements 102
3.9 Differences in mitochondrial DNA abundance and TSS enrichment between ATAC-seq and Fast-ATAC protocols. 103
3.10 Quality control assessment of different ATAC protocols in keratinocytes from a psoriatic lesion and NHEKs. 106
3.11 Quality control assessment of Fast-ATAC and Omni-ATAC in cultured NHEK 108
3.12 Comparison of 4 different ATAC protocols applied to psoriasis and/or healthy keratinocytes and NHEKs cells with published ENCODE DNase-seq data. 110
3.13 Experimental design to assess the impact of cryopreservation and fixation in the chromatin accessibility of immune primary cells. . . 112
3.14 Total number of ATAC-seq reads for the fresh, frozen and fixed $CD14^+$ monocytes and $CD4^+$ samples for three volunteers (CTL1-3).114
3.15 Fragment size density distribution for ATAC-seq fresh, fixed and frozen in $CD14^+$ monocytes and $CD4^+$ cells. 115
3.16 ATAC-seq enrichment of nucleosome-free and di-nucleosome fragments at the TSS and surroundings in $CD14^+$ monocytes and $CD4^+$ samples for the three conditions. 116
3.17 Genomic features annotation for the ATAC-seq peaks called in each of the fresh, frozen and fixed samples from $CD14^+$ monocytes and total $CD4^+$. 117

3.18 PCA based on the ATAC-seq chromatin accessibility landscape in fresh, fixed and frozen samples. 118
3.19 Comparison of the $\log_2$ normalised ATAC-seq counts at the consensus list of peaks in fresh and frozen conditions. 119
3.20 Top fourteen Gene Ontology biological processes enriched for DARs between ATAC-seq fresh and ATAC-seq frozen samples in $CD14^+$ monocytes. 120
3.21 Differential chromatin accessibility at the *TNFSF14* gene between ATAC-seq fresh and ATAC-seq frozen in $CD14^+$ monocytes. 121

4.1 Quality control evaluation of the H3K27ac ChIPm libraries in immune cells isolated from psoriasis and control samples. 146
4.2 PCA and chromatin annotation states of the consensus list of H3K27ac called peaks in four immune primary cell types from psoriasis and healthy control samples. 147
4.3 Differential H3K27ac modification at a putative intergenic enhancer region in circulating $CD14^+$ monocytes between psoriasis patients and healthy controls. 149
4.4 Quality control assessment of the ATAC libraries generated from circulating immune cells in psoriasis and control samples. 151
4.5 Clustered heatmap and conserved TF binding sites enrichment analysis in the consensus list of called ATAC peaks in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells from the patients and controls cohort. 152
4.6 Epigenetic landscape at two ATAC differential accessible regions between patients and controls in $CD8^+$ cells. 155
4.7 Epigenetic landscape at *DTD1* locus showing differential H3K27ac and chromatin accessibility between psoriasis patients and controls in $CD8^+$ cells. 157
4.8 PCA illustrating the sample variability based on the gene expression profiles for all 72 samples. 158
4.9 Magnitude and significance of gene expression changes between psoriasis patients and healthy controls in four immune cell types. 160
4.10 RNA-seq expression levels of the lncRNA *HOTAIRM1* and its experimentally validated target *UPF1* in psoriasis and healthy controls $CD14^+$ monocytes. 162
4.11 Differential expression between psoriasis and healthy controls of *MAPK* and *DUSP* genes contributing to MAPK signalling enrichment in $CD14^+$ monocytes and $CD8^+$ T cells. 164
4.12 Mapping of the DEGs identified in $CD8^+$ cells between psoriasis patients and healthy controls onto the TNF-α and the chemokine signalling pathways. 167
4.13 PCA for RNA-seq in the uninvolved and lesional epidermis from psoriasis patients. 169
4.14 Magnitude and significance of the gene expression changes between matched lesional and uninvolved epidermal biopsies from three psoriasis patients. 170

4.15 Overlap of the significant differentially expressed genes between lesional and uninvolved epidermal sheets, split epidermis and full-thickness skin biopsies. 171
4.16 Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the HIF-1 and IL-17 signalling pathways. 175
4.17 Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the NOD-like signalling pathway. . . 176
4.18 Differential gene expression and chromatin accessibility landscape for *TIAM1* and *TRAT1* genes in $CD8^+$ cells. 180
4.19 Association plot for the fine-mapping using summary statistics from the psoriasis GAPC cohort Immunochip GWAS results of the *SLC45A1/TNFRSF9* locus. 186
4.20 Epigenetic landscape at the location of the SNPs in the four SNPs in the 90% credible set for the *SLC45A1/TNFRSF9* psoriasis GWAS locus. 188
4.21 Epigenetic landscape at the location of the 95% credible set of AS fine-mapped SNPs for the chr2p15. 191
4.22 rs4672505 genotype and chromatin accessibility at chr2:62,559,749-62,561,442 in $CD8^+$ cells. 192
4.23 Potential role of rs4672505 in regulating *B3GNT2* gene expression. 193

5.1 Comparative percentages of peripheral blood and synovial fluid immune cellular composition from the PsA cohort. 223
5.2 Quality control assessment of ATAC data generated in four immune cell types isolated from peripheral blood and synovial fluid of PsA patients samples. 225
5.3 PCA based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and synovial fluid. 226
5.4 Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states. 228
5.5 Enrichment of PsA DARs for the FANTOM5 eRNA dataset. 229
5.6 Differentially accessible regions located within gene bodies in $CD14^+$ monocytes and NK cells from PsA patients. 231
5.7 Functional pathways significantly enriched and unique for either synovial fluid open or peripheral blood open DARs in $CD14^+$ monocytes, $CD4m^+$,$CD8m^+$ and NK cells. 233
5.8 Heatmap illustrating fold changes in gene expression between synovial fluid and peripheral blood across $CD14^+$ monocytes and $mCD4^+$ and $mCD8^+$ T cells. 235
5.9 Gene expression changes in immune-relevant genes between synovial fluid and peripheral blood in $CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ cells. 237
5.10 Comparison of immune-relevant gene expression modulation across PsA tissues (synovial fluid vs peripheral blood) and in PsA patients vs healthy controls. 239
5.11 Chromatin accessibility landscape at the qPCR differentially expressed *FN1* gene in $CD14^+$ monocytes. 241

5.12 Visualisation using UMAP dimensional reduction of the synovial fluid and peripheral blood $CD14^+$ monocytes. 246
5.13 Identification and characterisation of $CD14^+$ monocytes subpopulations in synovial fluid and peripheral blood. 247
5.14 Heatmap and hierarchical clustering for the top eight marker genes of the $CD14^+$ monocyte subpopulations identified in synovial fluid and peripheral blood from three PsA patients. . . . 249
5.15 Functional characterisation of the MC-1 and MC-4 $CD14^+$ PsA monocyte subpopulations. 252
5.16 Biaxial representation of $CD14^+$ monocytes expressing TNF-α, MCP-1, osteopontin and IP-10 in synovial fluid and peripheral blood before and after protein transport blockade with BFA using mass cytometry. 255
5.17 Increment in the percentage of $CD14^+$ monocytes producing TNF-α, osteopontin, MCP-1 and IP-10 after BFA treatment in synovial fluid and peripheral blood. 257
5.18 Chromatin landscape in $CD14^+$ monocytes in the proximity to *CCL2* and *CXCL10*. 258
5.19 Enrichment of PsA GWAS fine-mapped SNPs for DARs and ATAC peaks in $CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells. 263
5.20 Association plot for the Bayesian fine-mapping results of the 5q31 (*CSF2/P4HA2*) and *RUNX3/SYF3* PsA GWAS signal. 265
5.21 Epigenetic landscape at the genomic location of the fine-mapped SNP rs11249213 at the *RUNX3* PsA GWAS signal. 266
5.22 Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal. 270

B.1 Distribution of the background read counts from all the master list peaks absent in each sample. 315
B.2 Pre-sequencing profiles of relative abundance of DNA library fragment sizes for a psoriatic lesional keratinocytes ATAC 1 library generated using 40 min of transposition. 315
B.3 Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples. . . . 316
B.4 Fast-ATAC and Omni-ATAC NHEK pre-sequencing profiles of relative abundance of DNA library fragment sizes. 316
B.5 Differences in TSS fold-enrichment and percentage of mitochondrial between the ATAC-seq and Fast-ATAC libraries from all the samples generated in the thesis. 317
B.6 Mitochondrial reads and called peaks after IDR p-value filtering in the ATAC libraries generated in psoriasis patients and healthy control samples. 318
B.7 PCA analysis illustrating batch effect in ATAC and RNA-seq samples.319
B.8 Mapping rate and total reads after filtering (million) mapping to Ensembl genes in all the RNA-seq samples from psoriasis patients and controls in four cell types. 319

B.9 Mapping quality control for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. 320
B.10 Distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles for all 72 samples. 320
B.11 Correlation in gene expression between the miRs *MIR31HG* and the target gene *HIF-1A* in lesional and uninvolved skin. 321
B.12 Epigenetic landscape at the location of the fine-mapped SNP rs12223943 at the *ETS1* psoriasis GWAS locus. 322
B.13 Assessment of the percentage of mitochrondrial and duplicated reads and the number of called peaks after filtering in the ATAC data from peripheral blood and synovial fluid of PsA patients samples. 323
B.14 Permutation analysis for the DARs found in synovial vs peripheral blood in $CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells. 324
B.15 Enrichment of DARs in the 15 chromatin states of the cell type specific Roadmap Epigenomic Project segmentation maps. 325
B.16 Identification of the $CD14^+$ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. 325
B.17 Expression levels of antigen presentation and phagocytosis characteristic genes in synovial fluid and peripheral blood $CD14^+$ monocytes. 326
B.18 Differential gene expression between synovial fluid and peripheral blood PsA $CD14^+$ monocytes and correlation with qPCR and chromatin accessibility 327

List of Tables

1.1 Variables and scoring used in the Psoriasis Area and Severity Index (PASI) 6
1.2 Variables and scoring used in the Psoriatic Arthritis Response Criteria (PsARC) 7
1.3 Main GWAS in psoriasis and PsA 23

2.1 Summary table of samples/cohorts recruited in this thesis for generation of various datasets. 54
2.2 Antibody panel used for FACS separation of primary cell populations in Chapter 3 controls and Chapter 5 PsA samples. . . 58
2.3 Molecules targeted by the mass-cytometry antibody panel for peripheral blood and synovial fluid samples. 67

3.1 Summary table of ATAC analysis methodology for peak calling, filtering and differential analysis. 83
3.2 ATAC-seq percentage of mitochondrial reads and fraction of reads in called peaks (FRiP). 90
3.3 Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies. . . 107
3.4 Summary results from the differential chromatin accessibility analysis comparing ATAC-seq frozen or fixed chromatin landscape to the reference ATAC-seq fresh. 120

4.1 Summary table of the most comprehensive transcriptional studies in psoriasis skin and blood. 136
4.2 Description and metadata of the psoriasis patients cohort. 143
4.3 Description of the healthy control cohort. 144
4.4 Summary results from the differential H3K27ac analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. 148
4.5 Summary results from the differential chromatin accessibility analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. 154
4.6 Summary results from the differential gene expression analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. 159
4.7 Overlap between putative psoriasis GWAS genes and the reported significantly DEGs in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. 161
4.8 Pathways enriched for DEGs between psoriasis patients and healthy controls in $CD14^+$ monocytes and $CD8^+$ cells. 165
4.9 Summary of the differential gene expression analysis between uninvolved and lesional psoriatic epidermal biopsies. 169

4.10 Most relevant pathways enriched for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies. 174
4.11 Overlap between the DEGs in the four circulating immune cell types (psoriasis patients vs controls) and the DEGs in psoriasis patients skin biopsies (lesional vs uninvolved). 178
4.12 Summary results for the fine-mapping analysis of 26 psoriasis Immunochip GWAS loci. 182
4.13 SNPs from the 90% credible set of the successfully fine-mapped psoriasis loci overlapping ATAC accessible chromatin in four cell types. 185

5.1 Description and metadata of the PsA patients cohort. 221
5.2 Datasets generated for each sample in the PsA cohort. 222
5.3 Summary results of the differential chromatin accessibility analysis between synovial fluid and peripheral blood in PsA samples. 227
5.4 Characterisation of the DARs located within genes in each of the four cell types from PsA samples. 230
5.5 Immune genes with significant modulated expression in synovial fluid and proximal to a DAR in ATAC. 240
5.6 Pathway enrichment analysis for the modulated genes between synovial fluid and peripheral blood in $CD14^+$ and $mCD4^+$. 243
5.7 Summary table of the PsA GWAS loci presenting $\log_{10}BF \geq 3$ for the fine-mapping lead SNP. 260
5.8 PsA fine-mapped SNPs from the 90% credible sets overlapping accessible chromatin identified by ATAC in four cell types. 264
5.9 Publicly available *cis*-eQTL datasets reporting an effect for the PsA 5q31 GWAS locus fine-mapped SNPs (90% credible set) overlapping ATAC accessible regions. 268

A.1 Enrichment of ATAC-seq reads across the TSS for the $CD14^+$ monocytes and $CD4^+$ samples fresh, frozen and fixed. 307
A.2 Size of the master lists generated by DiffBind for the H3K27ac differential analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. . . . 308
A.3 Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChIPm assay. 309
A.4 Functional interactions and overlap with another study for the differentially expressed lncRNAs in each cell type. 309
A.5 Additional enriched pathways for DEGs between psoriasis and healthy controls in $CD14^+$ monocytes and $CD8^+$ cells. 310
A.6 Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies. 311
A.7 Enriched pathways for the scRNA-seq DEGs between synovial fluid PsA $CD14^+$ monocytes. 312
A.8 PsA GWAS Immunochip loci presenting $\log_{10}BF < 3$ for the fine-mapping lead SNP in the association analysis. 313

Abbreviations

ABF	Approximate Bayes factor
AD	Atopic dermatitis
AMP	Antimicrobial peptide
APC	Antigen presenting cell
AS	Ankylosing spondylitis
ASE	Allelic specific expression
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
CD	Crohn's disease
ChIPm	(Chromatin immunoprecipitation)-mentation
ChIP-seq	Chromatin immunoprecipitation sequencing
CLE	Cutaneous lupus erythematosus
CNV	Copy number variation
DAR	Differentially accessible region
DGE	Differential gene expression
DHS	DNase I hypersensitive site
DMARDs	Disease-modifying anti-rheumatic drugs
DNase-seq	DNase I hypersensitive sites sequencing
DNMT	DNA methyl-transferase
eRNA	Enhancer RNA
FAIRE-seq	Formaldehyde-assisted isolation of regulatory elements sequencing
FANTOM5	Functional annotation of the mammalian genome
Fast-ATAC	Fast assay for transposase-accessible chromatin
FRiP	Fraction of reads in peaks
GWAS	Genome-wide association studies
HIV	Human immunodefficiency virus
HLA	Human leukocyte antigen
IBD	Inflammatory bowed disease
IDR	Irreproducibility discovery rate
IFN	Interferon
IL	Interleukin
KIR	Killer immunoglobulin-like receptor
KRT	Keratin
LCE	Late cornified envelop
LD	Linkage disequilibrium
lncRNA	Long non-coding RNA
miRNA	micro-RNA
MNase-seq	Micrococcal nuclease sequencing
MS	Multiple sclerosis
NBF	Nucleosome-bound fragment
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NFF	Nucleosome-free fragment
NGS	Next generation sequencing

NOD	Nucleotide-binding oligomerization domain
NSAID	Nonsteroidal anti-inflammatory drug
NSC	Normalised strand cross-correlation coefficient
Omni-ATAC	Omni- assay for transposase-accessible chromatin
OR	Odds ratio
PB	Peripheral blood
PBC	PCR bottle-necking coefficient
PBMC	Peripheral blood mononuclear cells
PCA	Principal component analysis
PI	Protein inhibitor
PP	Posterior probability
PPAR	Peroxisome proliferator-activated receptor
PRC	Polycomb repressor complex
PsA	Psoriatic arthritis
PTM	Post-translational modification
qPCR	quantitative polymerase chain reaction
RA	Rheumatoid arthritis
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
RSC	Relative strand cross-correlation coefficient
scRNA-seq	singe-cell RNA sequencing
SDS	Sodium dodecyl sulfate
SF	synovial fluid
SLE	Systemic lupus erythematosus
SpA	Spondyloarthritis
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TAD	Topological associating domain
TCR	T cell receptor
Th	T-helper
TLR	Toll-like receptor
TNF	Tumour necrosis factor
UCSC	University of California Santa Cruz
WGS	Whole genome sequencing
WHG	Wellcome Center for Human Genetics

Chapter 1

Introduction

Our growing knowledge of genetic associations with susceptibility to psoriasis and psoriatic arthritis (PsA) has not yet been matched by understanding the functional basis of these disease-specific associations and the translation to patients' benefit. To address this challenge, it is important to understand the regulatory genomic landscape within which disease-associated genetic variants may act. This thesis describes functional genomic approaches to establish genome-wide epigenetic and gene expression profiles for disease relevant tissues and blood-isolated immune cells in psoriasis and PsA, and explore their potential significance for disease pathogenesis and genetic variation. In this chapter, I begin by reviewing current knowledge of the pathophysiology of psoriasis and psoriatic arthritis, the role of genetic variation, and the challenge of functionally characterising genome-wide association studies (GWAS) in complex traits, including the different functional genomics approaches that can be applied.

1.1 Psoriasis and psoriatic arthritis

Psoriasis and PsA have been described are common complex disease entities that nonetheless share certain clinical features and genetic architecture. Psoriasis is a chronic inflammatory skin disease characterised by episodes of relapse and remittance, most commonly manifesting as well-demarcated erythematous plaques with silver scale and associated with increased risk of joint, eye and systemic disorders (Nestle *et al.* 2009). On the other hand, PsA is a seronegative

chronic inflammatory disease within the spondyloarthritis (SpA) family that usually develops after psoriasis skin manifestations (Moll *et al.* 1973; Coates *et al.* 2016a; Villanova *et al.* 2013). Both conditions present similarities and differences at the pathophysiological and genetic level.

1.1.1 Epidemiology and global impact

Psoriasis represents a serious global health problem that currently affects about 100 million people worldwide, including both children and adults with no sex bias (Organization 2016). Although the mean age of onset is 33 years, a bimodal distribution with psoriasis patients being classified as early-onset/ type I (with peaks between 16 and 22 years) or late-onset / type II (between 50-60 years) has also been described (Henseler and Christophers 1985; Perera *et al.* 2012). This classification based on the age of onset also correlates with distinctive clinical features including severity, relapse frequency and family history.

The risk of developing psoriasis shows ethnic differences with a lower prevalence among adult African, African American and Asian populations (between 0.4 and 0.7%) compared to American and Canadian (4.6 and 4.7 %, respectively) (Jacobson *et al.* 2011). In the UK the prevalence of psoriasis ranges between 2 and 3%, affecting approximately 1.8 million people (Perera *et al.* 2012). For PsA, incidence in the general population varies between 0.04 and 1.2%, with peak age of onset between the 35 and 45 years of age. The estimation of psoriasis patients with concominant PsA ranges from 10 to 30%, with arthritis onset occurring approximately ten years after the onset of the skin disease (Greb *et al.* 2016). This clearly evidences the strong association between both diseases (Gelfand *et al.* 2005; Reich *et al.* 2015; Perera *et al.* 2012).

Several severe comorbidities are associated with psoriasis and PsA, with comparatively greater prevalence in PsA. For example, intraocular inflammation (uveitis) affects 8% of PsA patients compared to only 2% of psoriasis patients

(Husted *et al.* 2011; Oliveira *et al.* 2015). Other comorbidities include inflammatory bowel disease (IBD), cardiovascular disease (CVD), type 2 diabetes (T2D) and metabolic syndrome (Gelfand *et al.* 2006; Shapiro *et al.* 2007; Cohen *et al.* 2008). Psoriasis and PsA have also been associated with an increased prevalence of depression and suicidal ideation, and they represent a significant burden for the economy due to treatment costs and associated morbidity (Sampogna *et al.* 2012). In fact, treatment and management-associated costs per psoriasis patient in UK during 2015 accounted for £4,000 to £14,000, before and after requirements of biological therapy, respectively, and were larger in PsA (Burgos-Pol and Martínez-Sesmero 2016).

1.1.2 Psoriasis and inflammatory dermatoses

The skin is the biggest organ in the human body and constitutes an effective barrier between the environment and the internal organs. The most external layer, the epidermis, plays an important role in innate and adaptive immunity and its alterations can lead to development of inflammatory skin conditions, such as psoriasis or atopic dermatitis (Johnson-Huang *et al.* 2009; Proksch *et al.* 2008). Lesions in psoriasis are very heterogeneous in type (pustular and non- pustular), location and severity, which complicates its clinical classification (Perera *et al.* 2012). As a result, several phenotypes including chronic plaque psoriasis (psoriasis vulgaris), guttate psoriasis, pustular psoriasis, erythrodermic psoriasis and nail psoriasis have been defined (Marrakchi *et al.* 2011).

1.1.3 PsA and spondyloarthropathies

PsA belongs to the SpA family, which includes diseases such as ankylosing spondylitis (AS), reactive arthritis, arthritis associated with idiopathic inflammatory bowel disease (IBD) and undifferentiated SpA (Goldman and Schafer 2011). All these SpA subtypes are characterised by structural damage

(bone formation and erosion) as well as inflammation of joints and extra-articular sites such as eyes, gut and skin. Broadly, SpA has been classified into axial and peripheral based on the affected joints (spine/sacroilicac or peripheral) and the presence of extra-articular features (Rudwaleit *et al.* 2009).

Major histocompatibility (MHC) class I molecules are encoded by the human leukocyte antigen (HLA) A, B and C genes are responsible for presenting intracellular peptides (self or from infectious agents) to CD8 $^{+}$ T cells. *HLA-B*27* is the strongest genetic association for the SpA family. Studies in human families and rat models with *HLA-B*27* positive status have shown manifestation of different SpA features, such as psoriasis and IBD, within a single family or individual (Hammer *et al.* 1990; Said-Nahal *et al.* 2001). Based on common pathophysiological foundations, some studies have supported the concept of SpA as a single disease that presents heterogeneous phenotypic manifestations (Said-Nahal *et al.* 2001; Baeten *et al.* 2013). Interestingly, the axial and peripheral classification of SpA may be supported by true immunopathological differences between the two, with the main producers of IL-17 being neutrophils or mast cells in facet joints and synovial membranes, respectively (Porcher *et al.* 2005; Appel *et al.* 2011; Vandooren *et al.* 2004). Nevertheless, additional research will be required to further dissect whether different cellular and molecular processes are driving axial and peripheral SpA manifestations, contributing to a more refined classification and understanding of SpA.

As a phenotype, PsA can be further subdivided in five clinical groups as per the Moll and Wright criteria: distal, destructive, symmetric, asymmetric and spinal (Moll *et al.* 1973). These subclasses mainly differ in the location, number and distribution of the affected joints and have been later modified to also include dactylitis (diffuse swelling of a digit). Dactylitis is a distinctive feature of PsA that occurs in 16–49% of patients with PsA, often early in disease as the inaugural symptom, and that has been established as a marker for disease

severity (Reich 2012). Importantly, the phenotypic heterogeneity of SpA, and also within PsA, impairs the design and achievement of meaningful outcomes from clinical studies. This may obscure findings and conclusions in pathophysiological and clinical studies and needs to be considered when interpreting results.

1.2 Pathophysiology of psoriasis and psoriatic arthritis

1.2.1 Clinical presentation and diagnosis

Approximately 90% of all psoriasis cases are psoriasis vulgaris, which manifests with well demarcated plaques, erythema and scaling. Plaque formation is the result of thickening (acanthosis) and vascularisation of the epidermis and can vary in size and distribution, with the most common locations being the elbows, knees and scalp (Perera *et al.* 2012; Griffiths and Barker 2007). The second most common clinical presentation is guttate psoriasis (10% of all cases) characterised by acute onset of small droplike papules usually in the trunk and proximal extremities (Vence *et al.* 2015). Psoriasis vulgaris and guttate are generally not life-threatening forms of the disease, but they pose a significant impact on the patient's wellbeing.

Early and late onset psoriasis (type I and type II) differ in clinical presentation. Type I psoriasis patients commonly present with guttate lesions followed very often by bacterial infection, particularly *Streptococcus* throat infection, and have a stronger family history (50% with affected parents) with a high prevalence of *HLA-C*06:02* (85.4% of the cases) (Telfer *et al.* 1992). In contrast, in type II psoriasis only 14.6% of the cases are positive for *HLA-C*06:02*, and it most commonly manifests as spontaneous chronic plaques (psoriasis vulgaris) (Perera *et al.* 2012).

For PsA, symmetric/polyarticular PsA constitutes the most common manifestation (more than 50% of the cases) followed by asymmetric/oligoarticular PsA (around 30%), which exclusively affects four or fewer distal interphalangeal or phalangeal joints and may become polyarticular as disease progressess (Reich *et al.* 2009; McGonagle *et al.* 2011). Skin psoriatic lesions precede joint inflammation in approximately 60 to 70% of the cases (Gladman *et al.* 2005; McGonagle *et al.* 2011). In particular, nail pitting as well as scalp and intergluteal skin lesions constitute a predictive biomarker for development of joint inflammation (Moll *et al.* 1973; Griffiths and Barker 2007; McGonagle *et al.* 2011).

The diagnosis of psoriasis and PsA is primarily based on clinical assessment of the patient's symptoms due to the lack of molecular biomarkers at early stages of the disease (Villanova *et al.* 2013). The evaluation of skin lesion severity poses an additional challenge, and different measures have been implemented for criteria unification. The Psoriasis Area and Severity Index (PASI) is the most widely used quantitative rating score of skin lesion severity in research and clinical trials (Fredriksson and Pettersson 1978; Finlay 2005).

PASI	Description
Body location	Head and neck, upper limbs, trunk and lower limbs
Clinical signs	Redness, thickness and scaling
Severity scale	Absent, mild, moderate, severe or very severe
Affected area (%)	0, 1-9, 10-29, 30-49, 50-69, 70-89 or 90-100

Table 1.1: Variables and scoring used in the Psoriasis Area and Severity Index (PASI). For each of the locations the test quantifies the percentage of affected area and the severity of those three clinical signs (redness, thickness and scaling). The percentage of affected area is scored in a scale 1 to 6 (1=1-9%;2=10-29%; 3=30-49%; 4=50-69%; 5=70-89%; 6=90-100%) and the severity of the three clinical signs in a scale from 0 to 4 (from none to maximum). A combined score for each of the body regions is calculated and the final PASI score is the addition of each of those scores for each body region. PASI ranges from 0 (no disease) to 72 (maximal disease severity).

PASI quantifies the lesional burden by body part based on the affected area and the severity of erythema, induration and scale at each location (Table 1.1).

Disease is considered mild for PASI scores below 7 and is classified as moderate-to-severe for PASI scores between 7 to 12, depending on the study (Finlay 2005; Schmitt and Wozel 2005; Langewouters *et al.* 2008). To diagnose PsA, modified Moll and Wright criteria known as Classification Criteria for Psoriatic Arthritis (CASPAR) are most widely used (Taylor *et al.* 2006). A positive diagnosis based on CASPAR requires the presence of inflammatory arthritis, enthesitis, and/or spondylitis and three points from a list of associated elements. In terms of disease activity and treatment efficacy, the PsA Response Criteria (PsARC) is the preferred measure (Mease 2011; Clegg *et al.* 1996). PsARC considers the number of tender joints (TJC) and swollen joints (SJC) over 68 and 66, respectively, as well as patient and physician global assessment of the individual's general health based on a short questionnaire (Table 1.2).

PsARC	Description
TJC68	Number of tender joints over 68
SJC66	Number of swollen joints over 66
Patient's global health assessment	Evaluation of the patient's health by the patient (scale 0 to 5)
Physician global health assessment	Evaluation of the of the patient's by the physician (scale 0 to 5)

Table 1.2: Variables and scoring used in the Psoriatic Arthritis Response Criteria (PsARC). The patient's global health assessment by the patient and the physician is scored using a 5-point Likert scale, where 0 corresponds to very good, no symptoms and 5 corresponds to very poor and severe symptoms. When used to evaluate overall improvement after 12 weeks of treatment, improvement in at least two of the four variables evaluated (one of which must be TJC or SJC score) with no worsening of any criteria is required.

1.2.2 Aetiology of psoriasis and PsA

Psoriasis and PsA are chronic inflammatory diseases characterised by a dysregulated immune response as the result of genetic predisposition and exposure to environmental cues (Figure1.1). The origins of both pathologies,

as well as the connection between skin and joint inflammation, still remain uncertain.

Environmental factors and disease

A variety of exposures are proposed as risk factors for the development and worsening of psoriasis and PsA, with controversy about which are really accountable for disease onset or relapse. A wide range of drugs including anti-depressants, anti-hypertensives and anti-cytokine therapies (e.g β-blockers and IFN-α, respectively) have been associated with initiation, exacerbation and worsening of psoriasis (Kim *et al.* 2010). Bacterial and viral infections are associated with triggering and exacerbation of psoriasis and PsA. Onset of guttate psoriasis is associated with throat infection of group C *Streptococcus* and infection with humman immunodefficiency virus (HIV). In PsA statistical association with antibody production against *Streptococcus pyogenes*, *Yersinia enterocolitica*, *Chlamydophila psittaci* and HIV has also been reported (Gudjonsson and Karason 2003; Valdimarsson *et al.* 2009; Diluvio *et al.* 2006; Thrastardottir and Love 2018). Moreover, recent studies have also observed perturbation in the composition of the gut and skin microbiota in both pathologies (Eppinga and Konstantinov 2014; Yan *et al.* 2017).

Physical trauma and mechanical stress can also trigger the appearance of skin lesions and digit joint inflammation (Weiss *et al.* 2002; Nestle *et al.* 2009). Increased risk of PsA onset amongst psoriasis patients was indeed associated with lifting cumulative heavy loads as well as with several types of injuries and infections that require treatment with antibiotics (Eder *et al.* 2011). Smoking has been the behavioral factor most confidently associated with psoriasis, particularly with palmoplantar pustulosis (Armstrong *et al.* 2014). Psoriasis has also shown association with obesity, alcohol dependency, vitamin

D deficiency and stress, but evidence still remains controversial (Meglio *et al.* 2014).

Interestingly, epidemiological data suggests a steady increase in psoriasis and PsA prevalence over the last 30 years, particularly in older age groups (Springate *et al.* 2017; Organization 2016). This trend may be the result of the increase in frequency of various environmental risk factors over the same period of time (for example prevalence of obesity and beta blockers in patients with myocardial infarction), rather than a consequence of the improvement in diagnosis and access to medical care (Icen *et al.* 2009). Altogether this reinforces the role of environmental factors in the risk of developing psoriasis and PsA.

Histopathological alterations in skin and joints

The epidermis is the most external compartment of the skin, comprising approximately 90% keratinocytes and is organised in a layer-like structure that self-renews in a spatial and time-dependent manner (Wikramanayake *et al.* 2014). Keratinocyte differentiation is associated with changes in morphology, replication ability and keratin composition of the intracellular matrix. In the context of psoriasis, impaired epidermis cell renewal leads to histological alterations and lesion development. Importantly, keratinocytes up-regulate their proliferation rate (hyperplasia) causing aberrant cell differentiation (parakeratosis), thickening of the epidermis and scale formation (Ruchusatsawat *et al.* 2011). Concomitantly, the skin lesion undergoes hypervascularisation driven by up-regulated expression of angiogenic factors and activation of the endothelium that contributes to immune cell infiltration and inflammation (Perera *et al.* 2012).

In PsA, the affected joint shows a wide range of histological changes, one of the most common being arthritis caused by the swelling and inflammation of the joints (Haddad and Chandran 2013; Schett *et al.* 2011). As a result,

alterations in bone remodelling lead to osteolysis, bone resorption and erosion at the affected joints (Mensah *et al.* 2008). Moreover, 35% of PsA patients also undergo inflammation of the connective tissue at the insertion of tendons or ligaments, a phenomenon known as enthesitis (McGonagle *et al.* 2011; Polachek *et al.* 2017). The inflammatory environment at the entheses favours bone spurs formation along the insertion sites, similar to RA, causing structural debilitation of the joints (Benjamin and McGonagle 2009; Finzel *et al.* 2014). Interestingly, enthesitis has been identified as the main mechanism driving dactylitis, which is importantly characterised by tenosynovitis of the flexor tendons that are constrained by accessory pulleys (sites of compressive and tensile biophysical stress). The accessory pulleys behave as functional entheses and high-resolution magnetic resonance imaging has shown inflammation of this structure in early dactylitis (McGonagle *et al.* 2019).

Dysregulation of the innate and adaptive immune response

The dysregulated immune response in psoriasis and PsA is the result of the interaction between innate and adaptive immune cells through feedback loops and a complex cytokine milieu (Figure 1.1). Interferon (IFN)-α, β and γ are innate immune cytokines involved in disease initiation and are produced mainly by circulating plasmacytoid dendritic cells (pDCs), T lymphocytes and natural killer (NK) cells in the lesional skin (Johnson-Huang *et al.* 2009; Perera *et al.* 2012; Hijnen *et al.* 2013). Increased mRNA levels for IFN-α and β have been detected in skin plaques and shown to contribute to lymphocyte recruitment and maintenance of DC activation (Schmid and Itin 1994). Tumour necrosis factor (TNF)-α is another key cytokine involved in the dysregulated innate immune response observed in psoriasis and PsA. TNF-α is produced by innate immune cells, including mDCs, activated keratinocytes, monocytes/macrophages, NK cells and also cells from the adaptive immune system, including the T helper

(Th)/CD4$^+$ activated cell subsets Th-1 and Th-17 present at skin lesions and inflamed joints (Perera *et al.* 2012; Lizzul *et al.* 2005). TNF-α causes down-stream activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), a master transcriptional regulator which induces expression of pro-inflammatory cytokines, antiapoptotic genes and genes involved in maintenance of chronic inflammation (Lizzul *et al.* 2005; Johansen *et al.* 2004). Moreover, TNF-α has a prominent role in bone turnover and bone remodeling (mainly enhancing bone resorption), key histopathological features of PsA (Mensah *et al.* 2008).

Interleukin-23 (IL-23) and interleukine-17 (IL-17) constitute a link between the innate and adaptive immunity as well as a key loop for the perpetuation of the psoriasis and PsA inflammatory response. IL-23 is an innate immune cytokine mainly produced by myeloid DC (mDCs) and, to a lesser extent, by lesional skin-resident macrophages and psoriatic keratinocytes (Lee *et al.* 2004; Li *et al.* 2018). IL-23 binds to the IL-23 receptor (IL23R) and is highly expressed by the lesion-resident DCs, and T cells, and also by circulating CD4$^+$ from psoriasis patients (Tonel *et al.* 2010). In psoriasis, IL-23 mediates the pathogenic loop between activated keratinocytes and T cells. Binding of IL-23 to the IL-23R receptor expressed by memory CD4$^+$ and CD8$^+$ T cells leads to their differentiation into Th-17 and Tc-17 cells, both characterised by producing high levels of the IL-17 cytokine that activate NF-κB, AP-1 and MAPK signalling, amongst others (McGeachy *et al.* 2009). IL-17 signalling maintains the Th-17 immune mediated response through recruitment and activation of neutrophils, induction of pro-inflammatory cytokines, including interleukin-1β (IL-1β) and interleukin-6 (IL-6), and sustained keratinocytes activation (Doyle *et al.* 2012).

More recently, interleukin 22 (IL-22) has gained relevance as mediator of dysregulated crosstalk between the innate and adaptive immune response. IL-22 levels are increased in the skin lesions and plasma of psoriatic patients and is mainly produced by a subset of CD4$^+$ cells known as Th-22 (Wolk *et al.* 2006).

IL-22 contributes to keratinocytes hyperproliferation and their production of antimicrobial peptides (AMPs) (Eyerich *et al.* 2009).

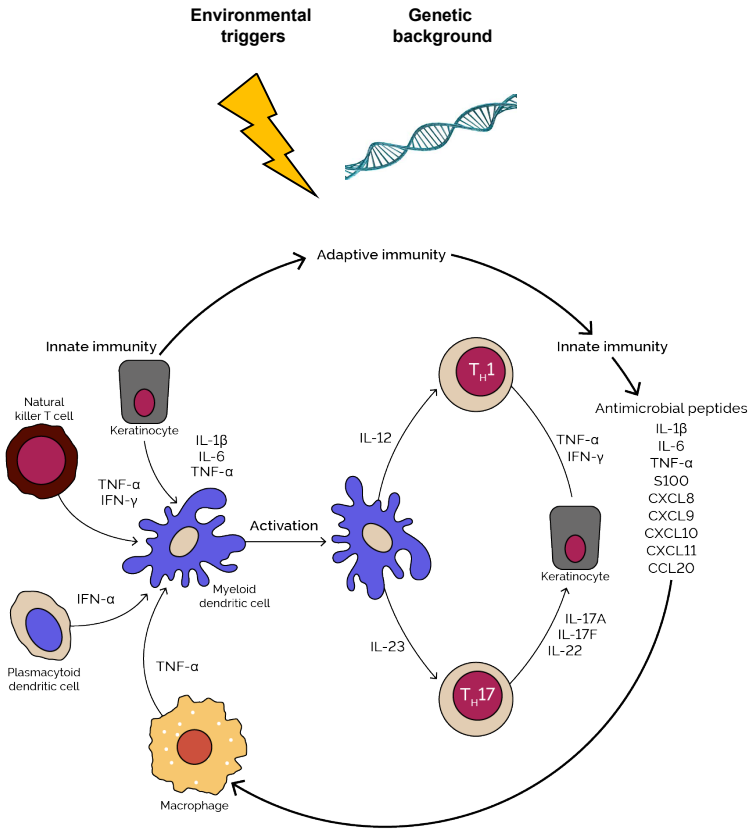


Figure 1.1: Environmental triggers and genetic predisposition leading to psoriasis and PsA (adapted from Nestle *et al.* 2009). The main cell types, cytokines and chemokines involved in the dysregulated innate and adaptive immune response found in these conditions are shown.

1.2.3 Cell types involved in psoriasis and PsA pathogenesis

Psoriasis and PsA are complex dynamic pathophysiological processes, and the understanding of the relative importance of different cell types at different disease stages still remains challenging.

Keratinocytes. In psoriasis, keratinocytes represent a bridge between the innate and adaptive immune response. Keratinocytes have the ability to act as immune sentinels through MHC-class II antigen presentation to $CD4^+$ and also T cytotoxic(Tc)/$CD8^+$ cells through MHC-class I (Black and Ardern-Jones

2007). Upon damage by environmental triggers, keratinocytes release cationic AMP, such as LL-37, and self-DNA/RNA that form a complex which acts as an antigen for skin-resident DCs activation and initiation of the inflammatory response (Lande *et al.* 2007). Pro-inflammatory cytokines such as IL-17A or IL-22 also induce keratinocyte proliferation and in turn production of cytokines, including IL-1β, IL-6 and TNF-α, and chemokines (e.g CXCL1, CXCL2, CXCL5, CXCL8 and CCL20) leading to recruitment of neutrophils and T cell to the site of inflammation (Feldmeyer *et al.* 2007; Arend *et al.* 2008; Nestle *et al.* 2009; Nestle *et al.* 2005). Keratinocytes also release vein endothelial growth factor (VEGF), a pro-angiogenic factor that activates endothelial cells and leads to pathogenic angiogenesis (Xia *et al.* 2003). The relevance of keratinocytes in the dysregulated immune response in psoriasis is further reinforced by the genetic association with variants located in keratinocyte-specific genes from the late cornified envelope (LCE) family (Tsoi *et al.* 2012).

Dendritic cells. mDCs and pDCs are important innate immune cells in initiation of the psoriasis dysregulated immune response through antigen presentation and T cell activation (Mahil *et al.* 2016). pDCs are circulating professional antigen presentation cells (APCs) that infiltrate into the uninvolved and lesional dermis and undergo activation by recognition of self-DNA and LL-37 complex through the Toll-like receptor (TLR)7/9 (Nestle *et al.* 2005; Lande *et al.* 2007). This activation prompts the activation and clonal expansion of antigen-specific $CD8^+$ T cells. In contrast, quiescent mDCs are epidermal resident cells that undergo maturation in presence of the IFN-α and β secreted by pDCs, expanding up to 30-fold in lesional skin (Zaba *et al.* 2007). Activated mDCs mediate the Th-1 and Th-17 response as well as perpetuation of keratinocytes activation through IL-23 , IL-12 and TNF-α production (Lee *et al.* 2004).

T cells. T lymphocytes have been considered the most relevant cell types in the initiation and maintenance of psoriasis and PsA. Skin-resident memory T

cells have been demonstrated to have a key role in psoriatic lesion development in a mouse model (Boyman *et al.* 2004). In human case reports, bone marrow transplantation has been shown to have the ability to both, terminate or initiate psoriasis in recipients (Gardembas-Pain *et al.* 1990; Eedy *et al.* 1990). *In vivo* studies demonstrated that transition to psoriatic lesions following engrafted human pre-lesional skin in immune-deficient mice was only dependent on injection of autologous activated $CD4^+$ and not $CD8^+$ cells (Wrone-Smith and Nickoloff 1996). Nevertheless, $CD8^+$ have shown preferential migration into the lesional epidermis where clonal populations have been isolated (Wrone-Smith and Nickoloff 1996; Chang *et al.* 1994). In psoriasis and PsA, IL-23 together with other cytokines, including IL-1β and IL-6, induce activation and differentiation of memory $CD4^+$ and $CD8^+$ into effector pathogenic Th-17 and Tc17 cells producing IL-17 (Weaver *et al.* 2007). IL-17^+ $CD8^+$ and $CD4^+$ cells have been found in lesional and uninvolved psoriatic skin and are enriched in PsA synovial fluid when compared to peripheral blood, showing correlation with markers of inflammation and structural changes in the joint (Menon *et al.* 2014; Ortega *et al.* 2009; Lowes *et al.* 2008; Pène *et al.* 2008).

Monocytes and macrophages. Resident macrophages in the healthy dermis undergo a 3-fold increase in psoriatic skin lesions and contribute to disease development through TNF-α production (Perera *et al.* 2012; Mahil *et al.* 2016). Similarly, a mouse model for chronic psoriasiform skin inflammation has demonstrated macrophage migration into affected skin and how production of TNF-α contributes to maintenance of skin lesions (Stratis *et al.* 2006; Wang *et al.* 2006). Initial studies showed greater phagocytic and bactericidal activity in monocytes from psoriasis patients compared to those from healthy individuals (M *et al.* 1979). Additionally, increased circulating intermediate-like monocytes ($CD14^{high}$ $CD16^{high}$) and monocyte aggregation was also observed in psoriasis patients, resulting in enhanced platelet activation and angiogenesis (Golden *et*

al. 2015). In PsA synovial membranes, the levels of monocytes/macrophage secreted metalloproteinases responsible for bone erosion upon differentiation into osteoclasts have been found to be similar to those in RA joints (Hitchon *et al.* 2002). Moreover, decreased number of macrophages in PsA synovial tissue have been observed following therapy and proposed as a biomarker to assess treatment response (Cañete *et al.* 2010).

Natural killer cells. NK cells are lymphoid-derived innate immune cells identified as $CD3^{-}$ $CD56^{+}$. The majority of circulating NK cells (90%) are $CD56^{dim}$ and show strong cytotoxicity (Mandal and Viswanathan 2015). In contrast, $CD56^{bright}$ commonly infiltrate into second lymph organs and other tissues, where they are activated by DCs and produce immunoregulatory cytokines such as IFN-γ, CCL5, CCL3 and GM-CSF that contribute to the shaping and regulation of other immune cells and promote Th-1 expansion and activation of the adaptive immune response (Martin-Fontecha *et al.* 2004; Ferlazzo *et al.* 2004). In psoriasis, a significant increase of cells expressing NK markers have been found in lesional compared to uninvolved skin (Cameron and Kirby 2003; Ottaviani *et al.* 2006). In a cohort including RA, PsA and AS patients, expansion of NK $CD3^{-}$ $CD56^{bright}$ cells was observed in inflamed joints (Dalbeth and Callan 2002). Moreover, NK cells in RA have shown to trigger osteoclastogenesis and bone destruction *in vitro* and also in mice (Soderstrom *et al.* 2010). Additionally, the NK receptors from the killer immunoglobulin-like receptor (KIR) family KIR2DL1 and KIR2DS1 (inhibitory and activatory, respectively) recognise HLA-Cw*06:02, strongly associated with psoriasis and PsA (Tobin *et al.* 2011). In fact, gene based studies have shown genetic variability in *KIR2DS1* associated with psoriasis and PsA susceptibility, as well as with AS and RA (Luszczek *et al.* 2004; Williams *et al.* 2005; Carter *et al.* 2007; Yen *et al.* 2001)

Neutrophils. Neutrophils are implicated in disease initiation through their ability to form neutrophil extracellular traps (NET) that contain host DNA and

LL-37 (Hu *et al.* 2016). Evidence of increased NET formation in peripheral blood and lesional skin of psoriasis patients has been found and seems to contribute to pDC and $CD4^+$ T cell activation (Hu *et al.* 2016). Neutrophils have also been identified as one of the main sources of IL-17 in skin lesions and to release a wide range of proteases, some of which induce keratinocyte proliferation (Lin *et al.* 2011; Mahil *et al.* 2016).

B cells. B cells are mainly involved in the humoral adaptive immune response through antibody production. However, they also act as APCs, regulate $CD4^+$ activation and differentiation into Th effector cells by providing coestimulatory signals and actively secrete cytokines (Bouaziz *et al.* 2007; Constant *et al.* 1995; Harris *et al.* 2000; Linton *et al.* 2003). The role of B cells in the pathophysiology of psoriasis and PsA remains unclear since both are negative for auto-antigens. Recent studies on the imiquimod-induced psoriasis mouse model have demonstrated a more severe inflammation in $CD19^{-/-}$ knock-out mice lacking a regulatory B cell subset producing IL-10 (Yanaba *et al.* 2013; Alrefai *et al.* 2016). Furthermore, different B cell subsets have been found in PBMCs from psoriasis patients and in lesional skin, and correlation with disease severity has been identified for some clinical subtypes (Lu *et al.* 2016). Moreover, enrichment of psoriasis GWAS risk SNPs in B cells regulatory elements (Farh *et al.* 2015; Patrick *et al.* 2018).

1.2.4 Therapeutic intervention

Psoriasis and PsA are currently incurable diseases, with available treatments focused on alleviating symptoms. For instance, topical therapies are advocated in cases of mild-to-moderate psoriasis, including emollients and short-term corticosteroids (Menter *et al.* 2009). Other treatments may be used in combination with corticosteroids, such as ultraviolet (UV) light therapy and vitamin D analogues, directed to inhibit T-cell and keratinocyte proliferation

and also induce keratinocyte differentiation (Rizova and Corroller 2001). In the case of PsA, for patients presenting with swelling of two or fewer joints, nonsteroidal anti-inflammatory drugs (NSAID) and intra-articular injection of glucocorticosteroids, together with joint aspiration, are used to reduce pain and inflammation (Coates *et al.* 2016a).

Treatment of most forms of PsA and moderate-to-severe psoriasis requires the use of systemic therapies. More severe forms of PsA require disease-modifying antirheumatic drugs (DMARDs) including antagonist of folic acid methotrexate and phosphodiesterase 4 inhibitor Apremilast, which act as immunosuppressors of activated T cells and cytokine production, respectively (Schmitt and Rosumeck 2014; Gossec *et al.* 2016; Keating 2017; Polachek *et al.* 2017). Biologic systemic agents represent the most specific treatment option for severe psoriasis and PsA notably TNF-α inhibitors (TNFi). Three TNFi have been approved for the treatment of psoriasis: etanercept, infliximab and adalimumab (Mahil *et al.* 2016). In addition, certolizumab pegol and golimumab are often used in the management of PsA (Coates *et al.* 2016b). However, side effects of TNFi such as increased risk of infection or reactivation of latent infections have been identified (Gottlieb *et al.* 2003). Moreover, 20 to 50% of patients fail to respond to the first TNFi administrated, requiring switching to an alternative TNFi (Abramson and Khattri 2016). New biologic therapies have been developed to target other key cytokines, such as IL-12, IL-23 (ustekinumab) or IL-17 (secukinumab and ixekizumab), which represent a substantial advance in treating patients failing to respond to TNFi (Mahil *et al.* 2016; Coates *et al.* 2016b).

1.3 Genetics of psoriasis and psoriatic arthritis

1.3.1 Heritability

The risk of developing psoriasis and PsA is not only influenced by environmental conditions but also by the genetic background of each individual. In fact, approximately 40% of patients with psoriasis or PsA have a family history in first degree relatives (Gladman *et al.* 1986). Twins studies in patients with psoriasis (only cutaneous lesions) have demonstrated that concordance of disease is greater in monozygotic (33-55%) compared to dizygotic twins (13-21%), giving a heritability estimate between 50 to 80% in populations of European descendants (Farber *et al.* 1974; Duffy *et al.* 1993; Grjibovski *et al.* 2007; Pedersen *et al.* 2008; Lonnberg *et al.* 2013). Moreover, psoriasis prevalence is also greater amongst first degree relatives compared to the general population, ranging between 4-19% (Myers *et al.* 2005; Chandran *et al.* 2009).

Interestingly, the only twins study conducted to date in PsA did not find differences in concordance between monozygotic and dizygotic twins, which could be due to lack of power in the study (Pedersen *et al.* 2008). Nevertheless, heritability estimates in PsA are between 80 to 100% and the recurrence rate in first-degree relatives has been shown to be greater for PsA (40-47-fold) compared to psoriasis (8-fold) (Myers *et al.* 2005; Chandran *et al.* 2009; Karason *et al.* 2009). These observations may highlight differences in the heritability between the two phenotypes and a stronger genetic contribution in PsA compared to psoriasis.

1.3.2 Genome-wide association studies

Initially, linkage studies were used to dissect the genetic architecture of psoriasis and PsA. A susceptibility loci in chr16p was identified by linkage analysis in families with PsA, suggesting a role of imprinting and

paternal transmission (Karason *et al.* 2003). A study of psoriasis in family pedigrees presenting an autosomal dominant condition revealed nine psoriasis susceptibility loci (PSORS1-9) with PSORS1 showing the strongest genetic association (Allen and Barker 2003). Nevertheless, the inability of independent studies to reproduce these results for regions other than PSOR1, 2 and 4 highlighted the limitations of this approach to understand the genetics of complex diseases (Capon 2017).

Advances in the ability to catalogue genome-wide common single base-pair changes known as single nucleotide polymorphisms (SNPs) led to the implementation of genome-wide association studies (GWAS). GWAS are focused on identifying disease-associated common SNPs (with minor allele frequency (MAF)≥1 to 5%), showing differences in allele frequency between patients and controls (Ku *et al.* 2010). GWAS are thus based on the hypothesis that complex diseases are caused by the interaction of multiple common variants, with modest effect size, with OR between 1.2 and 2 (Schork *et al.* 2009; Cui *et al.* 2010). The genotyped SNPs in GWAS are used as a proxy for the disease causative variant, which can be non-genotyped SNPs or copy number variants (CNVs) (Hirschhorn 2005; Ku *et al.* 2010).

The first psoriasis GWAS published in 2007 by Cargill *et al.* has been followed by other studies with larger sample sizes and meta-analysis across different cohorts (Table 1.3). The vast majority of these GWAS have combined psoriasis (only cutaneous lesions) and PsA patients in the cases group. Up to date, psoriasis GWAS studies have identified a total of 63 independent associations at genome-wide significance (p-value<$5x10^{-8}$) in European population, which only account for 28% of the overall estimated psoriasis and PsA heritability (Tsoi *et al.* 2017). The majority of studies have been performed in Caucasian European or North American cohorts, but increasing numbers of GWAS in large Chinese cohorts are also being published, increasing the number of independent

associated loci to 70 (Zhang *et al.* 2009; Sun *et al.* 2010; Yin *et al.* 2015). Early GWAS with moderate power confirmed association with loci overlapping the PSOR1 and PSOR2 genomic regions (Cargill *et al.* 2007; Strange *et al.* 2010).

The informativeness of GWAS was significantly enhanced with the use of the Immunochip genotyping chip, which covers 186 immune relevant loci identified in previous GWAS across different inflammatory diseases at a greater genotyping density (Tsoi *et al.* 2012). The psoriasis Immunochip study uncovered 15 new associations, including *CARD14* at the PSOR4, and also performed a meta-analysis incorporating the largest available psoriasis cohorts at the time (Tsoi *et al.* 2012). This meta-analysis has since been further expanded, yielding sixteen additional associations and reinforcing the importance of NFκB and cytotoxicity pathways in disease pathophysiology (Tsoi *et al.* 2015a; Tsoi *et al.* 2017). Meta-analysis of GWAS across Caucasian and Chinese populations revealed four new associations as well as population-specific effect or allelic heterogeneity in eleven loci (including MHC-I genes), demonstrating the value of a trans-ethnic approach to further understand the heterogeneous genetic susceptibility to psoriasis in different populations (Yin *et al.* 2015).

Non MHC genome-wide associations in psoriasis and PsA

A limited number of PsA GWAS have been conducted,being the best powered the Immunochip study performed by Bowes and colleagues (Liu *et al.* 2008; Hüffmeier *et al.* 2010; Ellinghaus *et al.* 2012; Bowes *et al.* 2015a). These studies identified a total of thirteen PsA associated loci at genome-wide significance (p-value<5x10−8). The Immunochip PsA GWAS also unveiled PsA-specific associations at the *IL23R*, 5q31 and *PTPN22* loci, which did not show significant associations (not even at nominal p-value<0.05) in psoriasis patients (Bowes *et al.* 2015a; Bowes *et al.* 2015b).

Following the Immunochip PsA GWAS, Stuart and colleagues published a larger study than the one conducted in 2010, comparing associations between PsA patients and those psoriasis patients with skin symptoms that had not developed PsA for at least ten years (Table 1.3). This study revealed nine regions associated with PsA (*IFNLR1*,*IL23R*, *REL*,*IFIH1*,*TNIP1*, *IFNLR1*, *IL12B*, *TRAF3IP2*, *NFKBIA* and *TYK2*) and ten with cutaneous psoriasis (*TNFRSF9*, *LCE3C/B*, *IL13*, *TNIP1*, *IL12B*, *TRAF3IP2*, *TNFAIP3*, *IL23A*, *NFKBIA* and *NOS2*) at genome-wide significance.

Amongst the psoriasis (combined cutaneous psoriasis and PsA together) GWAS variants reported by Stuart and colleagues, those nearby *TNFRSF9* and *LCE3A* showed stronger association with cutaeous psoriasis than PsA whereas variants at the *IL23R* and *TNFAIP3* loci presented stronger association with PsA compared to cutaneous psoriasis. Moreover,the PsA significantly associated variants at *IL23R* (previously reported by Bowes *et al.*, 2015) and *TNFAIP3* loci were independent from previously identified psoriasis variants at the same loci.

Interestingly, PsA-specific signals previously identified by less powered studies failed to show significant differences in association between cutaneous psoriasis and PsA in the Stuart and colleagues analysis (including the 5q31 loci reported by the Immunochip PsA GWAS), and only the *PTPN22* locus reached nominal stronger association in PsA (Stuart *et al.* 2015). The latest phenotype-specific association analysis published by Patrick and colleagues in 2018 has increased the number of genome-wide association with PsA and cutaneous psoriasis to thirteen and fifteen, respectively, with all of these loci previously identified as psoriasis associated (Patrick *et al.* 2018).

Differences in the MHC associations between psoriasis and PsA

Regarding MHC associations with psoriasis, *HLA-Cw*06:02* has been consistently identified as the most significantly associated locus with the greatest

effect size in psoriasis and also in independent analysis of cutaneous psoriasis and PsA compared to controls (Okada *et al.* 2014; Bowes *et al.* 2015a; Stuart *et al.* 2015). Step-wise conditional analysis has revealed additional HLA-I and HLA-II associations, including *HLA-C*12:03*, HLA-B amino acid positions 67 and 9, HLA-A amino acid position 95, and HLA-DQA1 amino acid position 53, with similar results found in the independent analysis of cutaneous psoriasis and PsA compared to controls (Okada *et al.* 2014; Bowes *et al.* 2015a).

Dissecting the *HLA-Cw*06:02* association in psoriasis and PsA has remained challenging. *HLA-Cw*06:02* has been demonstrated to increase risk of PsA when compared to controls. However, in the comparison between cutaneous psoriasis and PsA *HLA-Cw*06:02* appeared as more strongly associated with cutaneous psoriasis and showed a protective effect for PsA within psoriasis (Eder *et al.* 2012; Stuart *et al.* 2015). Since *HLA-Cw*06:02* is also associated with earlier age of onset in psoriasis, accounting for age of onset in the comparison of cutaneous psoriasis vs PsA *HLA-Cw*06:02* did not show protective association in PsA (Bowes *et al.* 2017). Instead the most significant MHC association when comparing both phenotypes was HLA-B aminoacid 97 (asparagine, mostly found in HLA-B*27), which differentiated PsA from cutaneous psoriasis (Bowes *et al.* 2017).

Table 1.3: Main GWAS in psoriasis and PsA. Summary table describing the most relevant psoriasis and PsA GWAS. Information regarding sample size, patients phenotypes and the main reported associations in each study is included. The Stuart *et al.* 2010, Stuart *et al.* 2015 and Patrick *et al.* 2018 studies included stratified association analysis of cutanous psoriasis (patients with skin symptoms that had not developped PsA for at least ten years) and PsA independently. WA=white American; Eur=European; ⋆ Meta-analysis performed.

Study	Etnicity	Sample size (Cases/Controls)	Phenotype	Main associations (putative genes)
(Cargill *et al.* 2007)	WA	1,446/1,432	Psoriasis(+PsA)	HLA-C (PSOR1) and *IL-12B*
(Nair *et al.* 2009)	Eur	1,409/1,436	Psoriasis(+PsA)	*IL-23A*, *IL23R*, *IL-12B*, *TNIP1*, *TNFIP3*, *IL-4* and *IL-13*
(Stuart *et al.* 2010)	WA, Eur	1,831/2,546	Cutaneous psoriasis, PsA	*NOS2*, *FBXL19*,*PSMA6-NFKBIA*
(Ellinghaus *et al.* 2010)	German	472/1,146	Psoriasis (skin only)	*TRAF3IP2*
(Strange *et al.* 2010)	Eur	2,622/5,667	Psoriasis(+PsA)	*LCE3D* (PSOR2), *IL28RA*, *REL*, *IFIH1*, *ERAP1*, *TYK2* and *HLA-C/ERAP1* epistasia
(Zhang *et al.* 2008)	Chinese	1,139/1,132	Psoriasis (type I)	*LCE* gene family and *IL-12B*
(Sun *et al.* 2010)	Chinese	8,312/12,919	Psoriasis(+PsA)	*ERAP1*, *PTTG1*, *CSMD1*, *GJB2*, *SERPINB8* , *ZNF816A*

(Tsoi *et al.* 2012)*	WA, Eur	10,588/22,806	Psoriasis(+PsA)	*CARD14* (PSOR4), *RUNX3*, *B3GNT2*, *ELMO1*, *STAT3*
(Tsoi *et al.* 2015a)*	WA, Eur	15,000/27,000	Psoriasis(+PsA)	1q31.1, 5p13.1, *PLCL2*, *NFKBIZ*, *CAMK2G*
(Bowes *et al.* 2015a)	British, Irish, Australians	1,962/8,923	PsA	5q31 PsA-specific
(Stuart *et al.* 2015)	WA and Eur	1,430/1,417	Cutaneous psoriasis, PsA	PsA-specific secondary signals (main text), new psoriasis-specific (*CDKAL1* and *CAMK2G*), stronger psoriasis *LCE* association
(Yin *et al.* 2015)	WA, Eur, Asian	15,369/19,517	Psoriasis(+PsA)	*LOC144817*, *COG6*, *RUNX1* and *TP63*; signals with ethnic heterogeneity
(Tsoi *et al.* 2017)*	WA, Eur	19,032/39,498	Psoriasis(+PsA)	*CHUK*, *IKBKE*, *FASLG*, *KLRK1*, *PTEN*
(Patrick *et al.* 2018)	WA, Eur	11,024/16,336	Cutaneous psoriasis, PsA	Thirteen and fifteen loci associated with PsA and cutaneous psoriasis, respectively

Overall, GWAS have demonstrated shared and distinct genetic architectures in MHC and non-MHC loci between cutaneous psoriasis and PsA, adding a layer of complexity in the understanding of the similarities and differences bewteen the two pathologies.

1.3.3 Relevance of non-coding variants in disease susceptibility

Approximately 88% of all GWAS associations map within non-coding regions (Welter *et al.* 2013). Psoriasis exome association studies in Chinese and Caucasian populations have increased the number of coding variants with putative effects on the protein structure (Tang *et al.* 2014; Zuo *et al.* 2015; Dand *et al.* 2017). These studies have confirmed some previously identified missense associations in *CARD14* and *ERAP1*, revealed new common coding variants at these previously associated loci and identified rare protective missense changes, for example in the *TYK2* gene (Tang *et al.* 2014; Dand *et al.* 2017). Nevertheless, results from extensive exome studies suggest that non-synonymous SNPs have a limited contribution to the overall genetic risk of psoriasis compared to non-coding variants (Tang *et al.* 2014).

Efforts have been made to elucidate the association of non-coding variants to disease by their ability to regulate gene expression in a cell and context specific manner (Fairfax *et al.* 2012; Fairfax *et al.* 2012; Naranbhai *et al.* 2015; Nica *et al.* 2011). Susceptibility variants can be located in different regulatory elements, including enhancers, silencers, promoters and the 5' and 3' untranslated regions (UTRs) (Ward and Kellis 2012). Non-coding GWAS variants can alter the expression of target genes through different mechanisms including changes in chromatin accessibility, histone modifications, protein binding such as transcription factors (TFs), DNA methylation and binding of non-coding RNA molecules (Knight 2014) (see 1.4.2).

Identification of the target genes regulated by non-coding variants poses a challenge in the field of functional genetics. This limitation can be partially addressed by conducting expression quantitative trait loci (eQTL) analysis, which identifies genome-wide statistical associations between gene transcript levels and SNPs in *cis* (<1Mb) or *trans* to the gene. For instance, in T2D this approach revealed a *cis*-eQTL involving the TF *KLF4* and a haplotype of non-coding GWAS SNPs located 14kb upstream (Small *et al.* 2011). Moreover, this haplotype also showed association with genes in *trans*, highlighting indirect downstream targets regulated by KLF4. Nonetheless, eQTL mapping alone only provides statistical suggestion of transcriptomic regulation, and additional functional assays, such as chromatin conformation and genome editing, are required to demonstrate causality (Edwards *et al.* 2013).

1.3.4 The role of GWAS in highlighting immune-relevant cell types and pathways

GWAS represent a biologically unbiased approach to shed light on pathophysiologically relevant cell types and molecular pathways associated with disease. The functional relevance and cell-type specificity of GWAS variants can be interrogated by overlapping them with epigenetic features mapped in cell lines or primary cells isolated from healthy controls (Farh *et al.* 2015). In psoriasis and PsA, enrichment of associated variants for regulatory elements has been found in several cell types (e.g Th-1 and Th-17 cells) and the majority of GWAS risk loci have been linked to genes from a limited number of pathways, as detailed below (Tsoi *et al.* 2017; Capon 2017). Most of these genes are reported based on proximity to the associated variants and fail to take account of long-range gene regulation, which represents a limitation when interpreting GWAS results. Additional criteria has also been used to link non-coding GWAS variants to a target gene, including LD with a deleterious variant, direct functional

characterisation of the regulatory element and/or genetic variant or association of gene expression with the genotype of the GWAS lead SNP (Capon *et al.* 2008; Tsoi *et al.* 2012; Tsoi *et al.* 2017; Meglio *et al.* 2011).

Systematic comparison of the genetic architecture across different diseases revealed that loci associated with psoriasis and PsA are shared in the same or opposite directions, with AS, Crohn's disease (CD), multiple sclerosis (MS), RA or type 1 diabetes (T1D) (Tsoi *et al.* 2012). This has also supported the use of therapeutics such as anti- TNF, IL-23 and IL-17 antibodies across a number of immune-related phenotypes (Visscher *et al.* 2017).

Antigen presentation

In psoriasis HLA-Cw*0602 represents the strongest GWAS association, also shared with other diseases such as hepatitis C, primary sclerosing cholangitis and Grave´s disease (Blais *et al.* 2011). No differences at the transcript level have been identified for HLA-Cw*0602 when comparing psoriasis patients versus controls, suggesting alterations in antigen presentation (Hundhausen *et al.* 2012). The relevance of antigen presentation in psoriasis and PsA has been reinforced by the GWAS association of the endoplasmic reticulum aminopeptidase 1 *ERAP1* gene, involved in the trimming of peptide antigens. Moreover, GWAS identified that *ERAP1* was associated with psoriasis and PsA only in individuals carrying one copy of the rs10484554 *HLA-C* risk allele (Strange *et al.* 2010). Such epistatic phenomenon, whereby association of one gene is dependent on the presence of another, has also been reported between *HLA-B*27* and *ERAP1* in AS (Evans *et al.* 2011; Cortes *et al.* 2015). Interestingly, the *ERAP1* haplotype associated with increased risk of SpA increases *ERAP1* expression and also alters splicing, resulting in an ERAP1 protein isoform with increased activity in monocyte-derived DCs and lymphoblastoid B cell lines (Costantino *et al.* 2015; Hanson *et al.* 2018).

Skin barrier

GWAS have highlighted keratinocyte-specific genes such as the *LCE* gene cluster and genes with a key role in skin biology such as *CARD-14*. Further studies in the *PSORS2* region have revealed that its association with disease is driven by the deletion of *LCE3B* and *LCE3C* (*LCE3C_LCE3B_del*), which also shows an epistatic interaction with the *HLA-Cw*0602* locus (Cid *et al.* 2009). Lack of *LCE3B* and *LCE3C* expression in psoriasis patients has been hypothesised to impair the repair following skin disruption, potentially facilitating microorganism infection and triggering a dysregulated immune response (Bergboer *et al.* 2011). Similarly to the *LCE* gene cluster, common and rare pathogenic mutations of the PSOR4 gene *CARD-14* lead to increased activation of NF-κB and overexpression of psoriasis pathophysiological relevant genes, such as *IL-6* and *TNFA*, in keratinocyte cell lines (Jordan *et al.* 2012).

NF-κB and TNF pathways

The NF-κB pathway is involved in the regulation of the innate and adaptive immune response and dysregulation of NF-κB contributes to the development of many chronic inflammatory diseases (Liu *et al.* 2017). In fact, elevated levels of NF-κB are present in psoriatic lesions compared to uninvolved and normal skin (Lizzul *et al.* 2005).

Several psoriasis and PsA GWAS loci have been mapped to gene members of the NF-κB and TNF signalling pathways including *TNIP1*, *TNFAIP3*, *NFKBIA*, *REL*, *TRAF3IP2*, *CHUK*, *IKBKE* and *FASLG* (Nair *et al.* 2008; Ellinghaus *et al.* 2010; Hüffmeier *et al.* 2010; Wang *et al.* 2008; Idel *et al.* 2003; Bowes *et al.* 2012; Tsoi *et al.* 2017). For example, a haplotype including missense mutations and intronic variants in *TRAF3IP2* has been reported to drive psoriasis and PsA association by reducing its affinity for TRAF interacting proteins and concomitantly altering NF-κB activation and the IL-17/IL-23 axis (Hüffmeier *et*

al. 2010; Ellinghaus *et al.* 2010). In addition, exome-sequencing studies have identified variants with predicted influence on protein structure and function at *TNFSF15*, a TNF ligand family protein which regulates NF-κB and MAP kinases activation in endothelial cells (Dand *et al.* 2017; Wang *et al.* 2014).

Type I IFN and innate host defense

Members of the type-I IFN signalling pathway have been associated with psoriasis and PsA, highlighting genes involved in host response to viruses and bacteria. Exome-sequencing and GWAS have identified two independent protective missense mutations predicted to impair the catalytic activity of the Janus kinases (JAK) protein member TYK2, and thus the initiation of the IFN-I downstream inflammatory cascade in psoriasis and PsA (Strange *et al.* 2010; Tsoi *et al.* 2012; Dand *et al.* 2017). A JAK inhibitor approved for RA is currently undergoing clinical trials in psoriasis and PsA, alongside with development of more specific JAK inhibitors and drugs targeting the upstream members of the type I IFN pathway, such as *TLR7* and *TLR9* (Yogo *et al.* 2016; Baker and D. 2017).

IL-17/IL-23 axis

Together with the TNF pathway, the IL-17/IL-23 axis is the most common target of biological therapeutics. In fact, some studies have reported greater efficacy of individual IL-17A or IL-23 blockade compared to TNF inhibition in the treatment of psoriasis and PsA (Griffiths *et al.* 2015; Blauvelt *et al.* 2017). The cytokine IL-23 is formed of two subunits: IL-23A/p19 and IL-12B/p40. Transcriptional studies have shown increased levels of p40 and p19 in psoriatic lesional skin and a role of both subunits in abnormal keratinocyte differentiation (Lee *et al.* 2004; Zhu *et al.* 2012). Psoriasis and PsA GWAS have reported associations with *IL23R*, including a protective two SNP haplotype shared with

CD (Nair *et al.* 2008; Strange *et al.* 2010; Tsoi *et al.* 2012). GWAS have also implicated *IL-23A* and *IL-12* in psoriasis and PsA (Cargill *et al.* 2007; Strange *et al.* 2010; Tsoi *et al.* 2012; Patrick *et al.* 2018). Interestingly, an *IL23R* signal secondary to the one reported by Tsoi *et al.* 2012 has been specifically associated with PsA (Bowes *et al.* 2015a). Regarding the genetics of the Th-17 pathway, its relevance is partly explained through the cross-talk with the IL-23 response, which mediates Th-17 cell differentiation and activation. Additionally, TFs regulating Th-17 polarisation, such as *IRF4* and *STAT3*, have also been implicated through GWAS with psoriasis and PsA pathophysiology (Tsoi *et al.* 2012; Huber *et al.* 2008; Harris *et al.* 2007).

Genome-wide pathway enrichment analysis and intergenic regions

New approaches using genetic association data have shed light on relevant biological processes by conducting genome-wide pathway analysis. In psoriasis, this has revealed association of novel processes, such as retinol metabolism, transport of inorganic ions and aminoacids and post-translational protein modifications (PTMs) (Aterido *et al.* 2016).

As previously mentioned, the majority of the non-coding GWAS associations are located in intergenic regions and often lack functional characterisation. These variants tend to be assigned to the nearest gene but may reside in intergenic regions distant from any gene, such as chr1p36.23, chr2p15 and chr9q31.2 in psoriasis. One of the most interesting regions is chr1p36.23, shared with UC and proximal to a number of gene candidates including *RERE*, *SLC45A1*, *ERRFI1* and *TNFRSF9* (Tsoi *et al.* 2012). Unpublished capture-HiC data using the immortalised keratinocyte cell line HaCaT have revealed interactions of SNPs in this locus with the promoter of *ERRFI1* gene, an inhibitor of epidermal growth factor receptor signalling required for normal keratinocyte proliferation (Ray-Jones *et al.* 2017).

1.3.5 Psoriasis and PsA: the same or distinct disease entities

It still remains unclear whether psoriasis and PsA are distinct entities or manifestations of a single disease since commonalities but also differences are found at the pathophysiological and genetic level, as previously reviewed. The fact the PsA prevalence amongst psoriasis patients is greater than the expected by chance and that skin lesions precede joint affection in 90% of the cases have supported the fact that PsA may be an extracutaneous manifestation of psoriasis. One of the links between skin and joint inflammation has been speculated to be biomechanical. The skin and the enthesis share associated tissue-specific factors since both structures integrate two different tissues, epidermis and dermis or tendon and ligaments, respectively, and they undergo high stress concentration at the tissue interface usually followed by micro-inflammation (McGonagle *et al.* 2007). Studying correlation between skin and joint symptoms has yielded opposing results. Some studies have shown lack of correlation between skin and joint lesions thus supporting independence of the two diseases, whereas others have demonstrated correlation within a subpopulation with concomitant onset of skin and joint inflammation suggesting a shared pathophysiology under particular clinical setting (Jones *et al.* 1994; Elkayam *et al.* 2000).

Regarding the genetic predisposition, the greater heritability of PsA compared to psoriasis have supported some independence between the two entities. Likewise, despite cutaneous psoriasis and PsA sharing a large number of GWAS risk loci, differences in the strength of association have also been identified for some of them, importantly the PsA association with the *HLA-B*27* allele and the stronger association of *HLA-Cw*0602* with cutaneous psoriasis (as previously detailed in 1.3.2). This heterogeneity could indicate that some of the associations are involved in disease susceptibility whereas other may have a role in modifying disease expression, which together with the influence of environmental factors, shape incidence and disease severity (Ciocon and Kimball 2007).

In terms of pathophysiology, the role of TNF-α and the IL-23/IL-17 axis has been demonstrated to be important for the establishment and perpetuation of skin and joint inflammation. Transcriptomic studies in skin and synovial membrane from the same patients have demonstrated a stronger IL-17 signature in the skin, consistent with the greater efficacy of anti-IL17 biologic therapies in the treatment of skin lesions compared to joints symptoms (Belasco *et al.* 2015; Furue *et al.* 2018). The role of T cells, particularly mCD8$^+$ expansion in epidermis and synovial fluid, has been well documented in psoriasis and PsA, respectively (Wrone-Smith and Nickoloff 1996; Costello *et al.* 1999). Nevertheless, it remains unclear whether differences in T cells biological subsets and T cell migration exist between skin and joints. Two studies revealed biologically distinct T cell populations in the two compartments based on the higher levels of expression of the cutaneous lymphocyte-associated antigen (CLA) in skin T cells compared to the joints (Pitzalis *et al.* 1996; Jones *et al.* 1997). Such finding could be the result of a homogenous activated T cell population undergoing distinct differentiation to travel to specific tissue sites. Moreover, T cell clonality studies have revealed polyclonal and oligoclonal T cell populations in skin and joints, highlighting common antigens driving T cell activation across different patients (Tassiulas *et al.* 1999; Borgato *et al.* 2002). However, when comparing clonality between skin and joints of the same individual only in three of them shared clones were found across tissues, inconclusively supporting the existence of a homogenous T cell population driving psoriasis and PsA simultaneously.

Altogether, epidemiological, pathophysiological and genetic data have shown evidence for psoriasis and PsA to be consider as a common disease, but they have also revealed heterogeneity between the two conditions. Additional research in the genetics, cellular and molecular components contributing to both pathophysiological processes will be required to further understand the overlap in the aetiology of cutaneous and articular disease.

1.3.6 Limitations of GWAS

GWAS have made a great contribution to our understanding of the genetic basis of complex diseases. However, this approach has a number of limitations to consider. One is the challenge of fine-mapping causal variants due to linkage disequilibrium (LD) in the genome, where the causal variant could be potentially any of the highly correlated and in high LD with the GWAS lead SNP. This can be addressed in part by dense genotyping, statistical fine-mapping methods and the incorporation of epigenetic data.

Another concern is the missing heritability when considering the estimated heritability from twin and family studies (Ku *et al.* 2010; Yang *et al.* 2010). Since complex traits are influenced by polygenic effects, where the genetic contribution is driven by multiple variants with small effect size, larger experimental cohorts have led to the discovery of new genome-wide significant associations (Visscher *et al.* 2017; Yang *et al.* 2010). To overcome the need of increasingly larger cohorts, the calculation of genome-wide polygenic risk scores making use of existing GWAS data has started to be implemented in the study of complex diseases. New methodologies for the calculation of polygenic risk scores leverage the aggregation of GWAS variants under the genome-wide significant threshold to predict the genetic liability of disease on the basis of an individual's genotype and identify subgroups of the population at high risk (Khera *et al.* 2018). A validated genome-wide polygenic risk score across five common diseases has recently shown successful results in coronary artery disease, identifying 8% of the population at greater than three-fold increased risk (Khera *et al.* 2018). In psoriasis, a polygenic risk score based on 200 genetic markers has been developed to predict the risk of PsA amongst psoriasis patients, showing comparable accuracy (0.82) to that of polygenic risk scores used to discriminate IBD subtypes (Patrick *et al.* 2018).

Other sources of unexplained heritability can be rare variants and structural variation, such as copy number variants (CNVs), small (<1Kb) insertions/deletions (indels) and inversions, poorly tagged by common SNPs (Wray 2005; Glessner *et al.* 2009; Marshall *et al.* 2017; Visscher *et al.* 2017). Interrogation of these sources of variation in complex immune diseases has partly been overcome by improved genotyping arrays like Immunochip, which incorporates SNPs with MAF<1%, together with long reads whole genome sequencing (WGS) technologies (Cortes and Brown 2011; Visscher *et al.* 2017). Exome studies have also shown contribution of coding and intronic rare variants (MAF<5%) in the genetic architecture of complex traits, such as height or psoriasis (Marouli *et al.* 2017; Dand *et al.* 2017). Nevertheless, studies in autoimmune diseases have revealed that the impact of rare coding-variants in the missing heritability can be considered negligible (Hunt *et al.* 2013). Lastly, the missing heritability may also be the consequence of the overestimated heritability due to assumption of additive genetic effects instead of epistatic interaction between the different associated loci (Zuk *et al.* 2012).

1.4 Functional interpretation of GWAS in complex diseases

1.4.1 Overcoming GWAS limitations

A number of strategies can be implemented to overcome the aforementioned limitations of GWAS and identify the true causal variant(s) amongst the large number of SNPs in high LD with the GWAS lead variant located at the same haplotype block (Edwards *et al.* 2013). Refinement of the number of candidate causal variants for each GWAS locus is required prior to further functional validation of their putative pathogenic effect through time and cost consuming molecular and cellular assays, and *in vivo* models. Statistical fine-

mapping can sometimes be used to narrow down the number of most likely causal variants within each GWAS LD block. The integration of statistical fine-mapping with cell type and context specific epigenetic data, including chromatin accessibility, histone modifications and DNA methylation, can help to determine the chromatin state where the fine-mapped variants are located and their potential role in regulating gene expression (Petronis 2010). Additionally, the incorporation of gene expression, eQTL analysis and chromatin interaction data can establish a relationship between non-coding variants and putative gene targets (Calderon *et al.* 2018).

1.4.2 Understanding the epigenetic landscape in complex diseases

Epigenetic effects are heritable changes in the phenotype and/or gene expression that do not involve changes in the DNA sequence (Feil and Fraga 2012). Epigenetic modifications include additions of chemical groups to histone proteins which serve as scaffold for the DNA compaction into chromatin, as well as changes in DNA methylation and regulatory functions of non-coding RNAs (Figure 1.2). Environmental and intrinsic factors can trigger changes in the epigenome that result in dysregulation of gene expression and function. Likewise, ther genetic background can increase the predisposition to epigenetic changes. Several studies have demonstrated differences in response to environmental factors by different mice breeds as well as greater differences in the epigenetic landscape between human dizygotic twins when compared to monozygotic (Pogribny and Tryndyak 2009; Kaminsky *et al.* 2009). Importantly, disease-associated GWAS variants have consistently shown enrichment for DNA regulatory elements, characterised by the combination of epigenetic marks, including accessible chromatin, histone modifications and DNA methylation

(Trynka and Raychaudhuri 2013a; Trynka and Raychaudhuri 2013b; Gusev *et al.* 2014).

The plasticity of the epigenetic landscape is required for cell differentiation and identity, being particularly important for adaptation and response to infections by the immune system (Yosef and Regev 2016). The role of cell-type specificity in regulating chromatin accessibility and gene expression was demonstrated in QTL studies, where 50 to 90% the genetic variants are cell type and stimulus dependent (Dimas *et al.* 2009; Nica *et al.* 2011; Fairfax *et al.* 2012; Fairfax *et al.* 2014; Raj *et al.* 2014; Naranbhai *et al.* 2015; Kasela *et al.* 2017). Recent methodological advances based on low-cell-input techniques coupled to next generation sequencing (NGS) have made possible the personalised study and understanding of the epigenome (Buenrostro *et al.* 2013; Schmidl *et al.* 2015; Oudelaar *et al.* 2017). Furthermore, cell-to-cell epigenomic heterogeneity is also being explored with single-cell methods, which may help to elucidate the impact of genetic variability in the regulation of gene expression and disease mechanisms (Buenrostro *et al.* 2015; Cusanovich *et al.* 2015; Rotem *et al.* 2015; Nagano *et al.* 2013; Smallwood *et al.* 2014).

1.4.3 The chromatin landscape

In the cell nucleus, DNA is compacted into highly organised chromatin. The nucleosome is the basic repeating unit of chromatin formed by a 147bp segment of DNA wrapped around an octamere core of histone proteins regularly spaced by 10bp of linker DNA (Luger *et al.* 1997). In general, highly compacted DNA will remain more inaccessible for the assembly of the transcriptional machinery, consequently preventing gene expression (Figure 1.2). Chromatin accessibility can be altered by PTM of the histones that affects their affinity for DNA within the nulcleosome as well as their interaction across nucleosomes in the vicinity (Polach *et al.* 2000; Pepenella *et al.* 2014). Additionally, chromatin structure

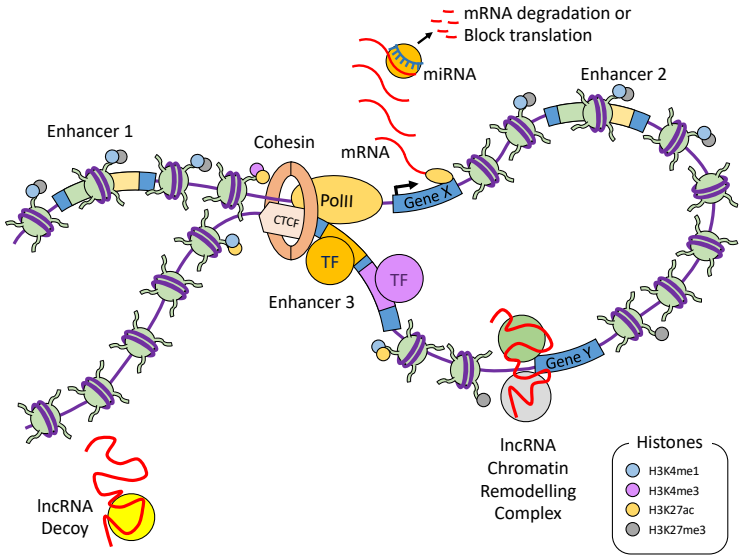


Figure 1.2: The epigenetic landscape and its role in regulation of gene expression. Long range regulation of gene expression is mediated by changes in chromatin conformation responsible for bringing promoters into proximity with distal regulatory elements and recruited TF complexes. CTCF and the cohesin complex mediate chromatin looping by serving as a scaffold protein in the maintenance of the 3D genomic structure. Transcription can also be modified by lncRNA recruiting chromatin remodelling complexes or by the down-stream regulation of transcription through alteration of mRNA stability. Similarly, miRNA can regulate the levels of mRNA by either preventing translation or leading transcript degradation.

can also be influenced by adenosin triphosphate (ATP)-remodelling complexes that facilitate sliding of individual nucleosomes, increasing temporary chromatin accessibility at particular sites (Cosma *et al.* 1999). From the biochemical point of view, the signature of chromatin accessibility, histone modifications, transcription factor occupancy and DNA methylation has been used to identify *cis*-regulatory elements such as promoters, enhancers, silencers, insulators and locus control regions, and define the cell type specific chromatin landscape (Boyle *et al.* 2012; Kundaje *et al.* 2015).

Methods to ascertain chromatin accessibility

Accessible chromatin represents a very robust marker for histone modifications, early replication regions, transcription start sites (TSS) and TF binding sites (ENCODE 2007). The informativeness of chromatin accessibility

to understand gene regulation has driven the development of several high-throughput techniques for accurately tagging these regions. Amongst those techniques, the "gold standard" is DNase I hypersensitive site sequencing (DNase-seq), which uses the non-specific double strand endonuclease DNase I to preferentially cut in nucleosome-free regions known as DNase hypersensitive sites (DHSs). In this approach, isolation of the chromatin-free DNA is followed by further enzymatic digestion and DNA library preparation prior to NGS (John *et al.* 2013). DNase-seq also provides high quality information regarding TF binding sites, generating footprints that can identify precise TF binding sites (Hesselberth *et al.* 2009; Boyle *et al.* 2010).

Another method to interrogate the accessible genome is formaldehyde-assisted isolation of regulatory elements (FAIRE-seq), which uses formaldehyde cross-linking, sonication and phenol-chloroform extraction to remove the DNA-protein complexes and only sequence nucleosome-depleted regions (Giresi *et al.* 2006). Both methods have enabled ENCODE to map regulatory elements in several cell lines, primary cells and tissues, revealing that 76.6% of all non-coding GWAS SNPs together with those in complete LD are located within broadly accessible chromatin (ENCODE 2007; Buck *et al.* 2014; Gaulton *et al.* 2010; Maurano *et al.* 2012). Indirect measurement of chromatin accessibility has been performed using micrococcal nuclease-sequencing (MNase-seq), which retains nucleosome-bound material for downstream sequencing, providing a qualitative and quantitative comprehensive map for nucleosome positioning and also TF occupancy (Axel 1975; Ponts *et al.* 2010). A high number of cells (10 millions or more) required by these assays to achieve good quality output limits their application to particular biological and clinical samples.

Recently, a groundbreaking assay for transposase-accessible chromatin using sequencing (ATAC-seq) provided a milestone in the ability to characterise the genomic regulatory landscape (Buenrostro *et al.* 2013). ATAC-seq is

based on an engineered hyperactive transposase enzyme, known as Tn5, that preferentially accesses and tags nucleosome-free and inter-nucleosomal DNA, using a lower number of cells and shorter processing times compared to DNase-seq (Gradman *et al.* 2008; Adey *et al.* 2010). This makes ATAC-seq a very versatile technique to interrogate the chromatin landscape in a clinical set-up, where sample availability and time-efficiency are key factors (Scharer *et al.* 2016; Qu *et al.* 2015; Qu *et al.* 2017).

The role of histone modifications and TF occupancy in the chromatin landscape

Identifying the combination of histone modifications and binding of TFs is essential to characterise regulatory regions of the genome. Histone modifications take place in the NH_2 terminal tail that protrudes from the nucleosome, with the most common modifications being acetylation, phosphorylation and methylation. The co-localisation of different histone marks modulates the affinity for DNA-binding proteins and the interactions with neighboring nucleosomes in varied manners, contributing to the overall chromatin accessibility landscape of the nucleus (Figure 1.2) (Jenuwein and Allis 2001; Bannister and Kouzarides 2011).

The combination of histone modifications can be used to broadly divide chromatin into condensed non-transcribed heterochromatin and accessible transcriptionally active euchromatin. Further studies have identified facultative and constitutive heterochromatin, which distinguishes spatially and temporally regulated genes from those permanently silenced, respectively. Facultative heterochromatin is enriched for H3K27me3 and the polycomb repressor complexes (PRCs), whilst constitutive heterochromatin is marked by H3K9me3 (Hansen *et al.* 2008; Bannister *et al.* 2001).

Several types of chromatin corresponding to different regulatory elements have also been defined. Enhancers and promoters, regardless of their activation state, are tagged by high levels of H3K4me1 or H3K4me3, respectively, and both features co-localise with H3K4me2 modifications (Heintzman *et al.* 2007; Hon *et al.* 2009). H3K9ac is specifically enriched at active promoters whereas H3K27ac generally designates activation at both promoters and enhancers (Hon *et al.* 2009; Creyghton *et al.* 2010). Conversely, H3K27me3 together with the heterochromatin mark H3K9me3 indicates gene repression (Hansen *et al.* 2008; Bannister *et al.* 2001; Pan *et al.* 2007). Interestingly, GWAS variants for different complex diseases have been demonstrated to be relatively enriched for some of those modifications, importantly H3K4me3, H3K9ac, H3K79me2, H3K4me1 and H3K36me3 (Ernst *et al.* 2011; Trynka and Raychaudhuri 2013a). Functional understanding and interpretation of histone mark co-localisation remains challenging and incorporation of additional epigenetic information is usually required. Together with histone modifications, TFs also play a role in nucleosome positioning as well as in acting as boundary elements to separate chromatin states (Vierstra *et al.* 2014; Zhang *et al.* 2009; Bell and Felsenfeld -G. 2000).

Chromatin immunoprecipitation sequencing (ChIP-seq) can precisely locate histone modifications and TF binding in the genome. This technique assays protein-DNA binding using antibodies that specifically recognise histone modifications or TF after DNA-protein cross-linking and sonication. Following immunoprecipitation of the desired DNA-protein complexes, the cross-linking is reversed and the proteins digested prior to DNA library preparation and sequencing (Solomon *et al.* 1988; Barski *et al.* 2007; Johnson *et al.* 2007). ChIP-seq has been used to analyse a wide range of histone modifications and TF binding in different cell lines, primary cells and tissues, providing a valuable resource (ENCODE 2007; Bernstein *et al.* 2010; Adams *et al.* 2012).

Similarly to the first generation of chromatin accessibility techniques, chromatin immunoprecipitation sequencing (ChIP-seq) requires at least between 5 to 10 million cells per experiment, restricting its application to the availability of biological material. In order to overcome this limitation, modified protocols have been developed, of which ChIPmentation (ChIPm) is the simplest and most cost-effective method, requiring only 10,000 and 100,000 cells to assay histone modifications or TF binding, respectively (Schmidl *et al.* 2015). ChIPm uses the Tn5 transposase, accelerating library preparation and increasing the sensitivity of the results.

DNA methylation

DNA methylation involves the transfer of a methyl group to the 5' carbon of a cytosine that precedes a guanine nucleotide (CpG sites) by a group of enzymes known as DNA methyl-transferase (DNMTs). CpG islands (long stretches of C and G nucleotides) are found along the genome and their methylation is generally associated with repression of gene expression (Herman and Baylin 2003). The pathogenicity of changes in the methylome has been studied in a range of diseases including RA, systemic lupus erythematosus (SLE) and psoriasis (Lei *et al.* 2009; Liu *et al.* 2013; Zhang *et al.* 2010). For example, regulation of TNF-α production upon inflammatory stimuli involves a complex network of DNMTs that alter the methylation signature at this locus (Sullivan *et al.* 2007).

Chromatin interactions and gene expression

The functional understanding of non-coding variants has benefited from eQTL studies. Nevertheless, eQTLs only provide indirect evidence of the effect of a SNP on regulating expression of a particular gene. Since enhancers may not control the expression of the closest gene, functional interpretation of GWAS variants requires genome-wide mapping of physical chromatin interactions

(Smemo *et al.* 2014). Chromatin is organised into topologically associating domains (TADs) of several hundred kb insulated from other TADs by the binding of CTCF protein, amongst others (Nora *et al.* 2017). Chromatin loops between promoters and regulatory elements mostly take place within the same TAD and are highly cell- and context-specific (Smith *et al.* 2016). Hence, interrogation of chromatin interactions provides evidence for physical contact between enhancers and gene promoters coordinating assembly of the transcriptional machinery and consequently regulating expression (Figure 1.2). For example, the obesity risk non-coding variants located within the *FTO* gene physically loop to and regulate the expression of *IRX3*, located 1Mb downstream (Smemo *et al.* 2014).

A wide range of genome-wide and high-throughput methods to investigate the 3D chromatin conformation have been developed. Of particular interest is Capture-C, that simultaneously scales up the number of interactions investigated at high resolution and minimises the number of cells required (Davies *et al.* 2017; Oudelaar *et al.* 2017). Other techniques such as promoter capture HiC have yielded comprehensive immune-specific maps of promoter-enhancer interactions in seventeen human primary hematopoietic cell types (Javierre *et al.* 2016).

1.4.4 Transcriptional profiles in disease

The role of environmental and genetic factors in altering gene expression in complex diseases has been investigated through extensive comparison of case-control transcriptional profiles. The informativeness of this approach is conditional on studying the disease-relevant tissue, which often remains challenging due to lack of pathophysiological understanding of disease mechanisms or difficulties in accessing the tissue. In immune-mediated diseases, peripheral blood mononuclear cells (PBMCs) differential gene expression analysis between patients and controls has enabled identification of relevant molecular pathways in a number of chronic immune diseases, including psoriasis

and PsA (further detailed in Chapter 4 and 5, respectively) (Miao *et al.* 2013; Junta *et al.* 2009; Baechler *et al.* 2003; Assassi *et al.* 2010; Batliwalla *et al.* 2005). The growing evidence supporting cell type and context-specificity has also prioritised the use of disease-specific affected tissue over PBMCs, when possible, including skin biopsies in psoriasis, synovial-isolated macrophages in RA and B cells and monocytes in SLE (Katschke *et al.* 2001; Dozmorov *et al.* 2015; Jabbari *et al.* 2012).

Likewise, the extensive overlap of GWAS variants with non-coding regions potentially dysregulating gene expression has highlighted the importance of performing context-specific eQTL studies. In this respect, consortia such as the Genotype Tissue Expression (GTEx) have generated publicly accessible comprehensive tissue-specific eQTL studies that have greatly contributed to the functional understanding of GWAS risk alleles in many complex diseases (Lonsdale *et al.* 2013; Fagny *et al.* 2017).

Long non-coding RNAs and enhancer RNAs

In addition to protein coding mRNAs, non-coding RNAs have been demonstrated to have a role in regulation of gene expression. One category of non-coding RNAs are the long non-coding RNAs (lncRNAs), transcripts between 200bp and 100Kb long that undergo splicing, 5' capping and 3' poly-adenylation (Derrien *et al.* 2012). lncRNAs can positively and negatively regulate transcription through different mechanisms including guidance of chromatin modifiers such as DNMTs and PRCs to specific loci, alteration of mRNA stability, translational control, and acting as a decoy for other non-coding RNAs and regulatory proteins (Figure 1.2) (Pandey *et al.* 2008; Faghihi *et al.* 2008; Gong and Maquat 2011; Carrieri *et al.* 2012; Kino *et al.* 2010).

Amongst the characterised lncRNAs, many have been demonstrated to play a role in the regulation of the innate and adaptive immune response, for

example in T cell activation and host-pathogen interactions (Pang *et al.* 2009; Rossetto and Pari 2012). Differential case-control gene expression analyses have underscored the contribution of lncRNAs in several chronic inflammatory conditions, including RA, SLE and psoriatic skin (Müller *et al.* 2014; Shi *et al.* 2014; Li *et al.* 2014; Ahn *et al.* 2016).

A particularly relevant type of lncRNAs are the enhancer RNAs (eRNAs), shorter molecules compared to the canonical lncRNAs (approximately 346 nucleotides) that do not undergo splicing or poly-adenylation (Forrest *et al.* 2014). Although traditionally chromatin segmentation maps have defined enhancers as DNA regions with particular epigenetic characteristics, later studies have shown their ability to be bi-directionally transcribed into eRNAs molecules (De Santa *et al.* 2010; Kim *et al.* 2010). Importantly, the transcriptional activity of enhancers has been demonstrated to be an excellent proxy in identifying functionally active regulatory regions, which have been successfully validated by reporter assays (Forrest *et al.* 2014; Andersen *et al.* 2014).

Another class of non-coding RNAs are micro-RNAs (miRNAs), between 21 to 24 nucleotides long (Lee *et al.* 2002). Under particular conditions, expression of genes containing complementary sequences to miRNAs can be negatively regulated through assembly of the miRNA-induced silencing complex followed by mRNA degradation, mRNA destabilisation or translational repression (Figure 1.2) (Ameres *et al.* 2010; Braun *et al.* 2011; Petersen *et al.* 2006; Lewis *et al.* 2005; Friedman *et al.* 2008).

Methods to assay gene expression

RNA sequencing (RNA-seq) involves reverse-transcription of the extracted RNA into cDNA and PCR amplification preserving relative abundance of each transcript, followed by library preparation and NGS (Mortazavi *et al.* 2008). Systematic comparison has shown superior dynamic range of detection for RNA-

seq compared to micro-arrays, particularly for low abundance transcripts (Zhao *et al.* 2014). Furthermore, RNA-seq allows the capture of additional information, including the identification of new exons, alternative splicing events and allele-specific expression (ASE). Quantification of ASE through RNA-seq has provided direct evidence for local/*cis*-eQTLs driven by allele-specific mechanisms in up to 88% of the genes with an associated *cis*-eQTL (Yan *et al.* 2002; Pickrell *et al.* 2010).

Sequencing of the 5'-end of RNA molecules using methods such as cap analysis of gene expression (CAGE) has been used to quantify eRNAs by the functional annotation of the mammalian genome 5' (FANTOM5) consortium, contributing to a better definition of enhancers and their spatial and temporal specificity in hundreds of human primary cells and tissues (Forrest *et al.* 2014; Andersson *et al.* 2014). Lastly, development of single-cell RNA-seq (scRNA-seq) has enabled the identification of cell sub-populations within a tissue in an unbiased way (Tang *et al.* 2009; Tang *et al.* 2010).

1.4.5 Transcriptional regulation in complex diseases

Non-coding GWAS variants can exert pathogenic effects by affecting one or many of the previously described mechanisms responsible for the fine regulation of gene expression in homeostatic conditions. For example, intronic SNPs can influence mRNA splicing through exon skipping, resulting in truncated but functional proteins. For instance, exon skipping caused by an intronic risk allele at the TNF Receptor Superfamily Member 1A (*TNFRSF1A*) associated with increased MS risk results in a soluble isoform of the TNFRS1A protein with TNF antagonistic function, which is consistent with the negative effects of anti-TNF agents in the treatment of this disease (Gregory *et al.* 2012). On the other hand, non-coding variants at enhancers, silencers and promoters can dysregulate gene expression by altering affinity at TF binding sites, histone modifications and

chromatin accessibility (Figure 1.2). For instance, in thyroid autoimmunity the risk allele of an intronic SNP in the thyroid stimulating hormone receptor (TSHR) gene reduces its expression in IFN-α stimulated thyroid cells by increasing the affinity for a TF that recruites histonce deacetylases (HDACs) to the locus (Stefan *et al.* 2014). Alterations in TF binding sites can also affect looping and long-range chromatin interactions between enhancers and promoters (Figure 1.2). For instance, in prostate cancer this phenomenon causes up-regulated expression of the oncogene *SOX9* due to increased enhancer activity and enhancer-promoter interaction (Zhang *et al.* 2012).

Alternatively, non-coding SNPs can regulate gene expression by creating a new promoter-like element, as in α-thalassaemia, where this phenomenon leads to dysregulated downstream activation of all α-like globin genes in erythroid cells (Gobbi *et al.* 2006). Genetic variants at eRNAs can also affect regulation of gene expression as it has been demonstrated in the nuclear receptor for anti-diabetic drugs PPARγ in mice (Soccio *et al.* 2015). Lastly, non-coding variants placed in UTRs and intergenic regions can affect binding of miRNAs and lncRNA to the target genes, as in the CD associated variant at the 3'UTR of the gene immunity related GTPase M (*IRGM*), which reduces binding of miR-196 (Brest *et al.* 2011). In psoriasis, some specific SNPs located at 3' UTR of genes such as *IL-23*, *TRAF3IP2* or *SOCS1* have been hypothesised to disrupt or create *de novo* miRNA binding sites, but no experimental evidence has been provided yet (Pivarcsi *et al.* 2014).

1.4.6 The use of fine-mapping to prioritise functional causal variants

The aim of fine-mapping is to reduce the size of GWAS genomic intervals and yield a minimal set of SNPs containing the causal variant that will explain most of the association for that particular locus (Spain and Barrett 2015). Fine-

mapping studies require extensive genotyping to meet the assumption that the putative causal variant will be likely interrogated in the analysis. This can be achieved by WGS, dense genotyping arrays and *in silico* imputation using publicly available data. The use of the Immunochip array across most of the immune-mediated inflammatory diseases has increased the genotyping density at previously associated immune-relevant loci (Trynka *et al.* 2011). Similarly, imputation methods using WGS reference panels, such as HapMap and 1000 Genomes Project, have offered genome-wide coverage for SNPs and CNVs with MAF >1% across different ancestry groups (Abecasis *et al.* 2012). More recently, the UK10K project has improved the quality of imputation specifically for rare variants with MAF between 0.01 and 0.5% (Chou *et al.* 2016).

Bayesian statistical analysis is the method of choice over the frequentist approach (based on p-value calculations) to increase the resolution of the GWAS associations. Bayesian fine-mapping quantifies the evidence of association for each of the genotyped or imputed SNPs as Bayes Factor (BF). BFs are later used to calculate posterior probabilities (PP) which represent the probability of each SNP to drive a particular association (Wakefield 2007). Since including only the most significant fine-mapped SNP would miss the causal variant in 97.6% of loci, the Bayesian strategies report a credible set of SNPs which includes variants capturing 95 or 99% of the cumulative PP in each loci (Bunt *et al.* 2015). This strategy has shown further refinement and reduction of false-positives compared to inclusion of all SNPs in high LD (e.g $r^2 \geq 0.8$) with the lead variant(Bunt *et al.* 2015). Furthermore, inclusion of functional data from publicly available sources as priors of the approximate Bayesian model has demonstrated a reduction of the number of SNPs in the credible set and also increased the proportion of successfully fine-mapped loci (Bunt *et al.* 2015; Kichaev and Pasaniuc 2015). The integration of fine-mapping data generated with the Bayesian probabilistic identification of causal SNPs (PICS) method and

a map of genomic regulatory elements, revealed that approximately 60% of the top fine-mapped SNPs overlapped enhancer elements (importantly, stimulus-specific) and were very close but not within TF binding sites (Farh *et al.* 2015).

1.4.7 Generation, integration and interpretation of genomic data

The generation of multi-omics clinical datasets has a number of challenges related to experimental design, availability of suitable patients, logistics for sample processing and limited cohort size. Experimental design is influenced by logistics, costs and accessibility to the affected tissue, plus the ability to generate paired datasets from the same patients and most appropriate tissues. Sample size is another limitation when working in a clinical setting and represents an important challenge when studying complex diseases. Identifying genome-wide significant differences in epigenetic features or gene expression between patients and controls is influenced by the sample size, causing many effects to be missed. Small cohorts can still be used as hypothesis generating studies, requiring replication in larger ones (Hackshaw 2008).

The evolution of different omics methods enable the generation of datasets at a high-throughput scale, also poses challenges for data interpretation and integration. The relationship between different omics datasets and genetic variability is required to improve the understanding of pathophysiological mechanisms, as well as the role of genetic variability in the context of disease. The large-scale epigenetic and expression data generated by International consortia such as ENCODE, Blueprint, the Roadmap Epigenomics Project, GTEx or FANTOM include comprehensive website resources for browsing and downloading (ENCODE 2007; Lonsdale *et al.* 2013; Forrest *et al.* 2014; Adams *et al.* 2012). Moreover, integration of these publicly available epigenetic datasets has led to the assembly of cell-type specific chromatin states maps to functionally

annotate each position of the genome based on concurrence of several epigenetic marks (Ernst and Kellis 2010; Ernst *et al.* 2011; Hoffman *et al.* 2013; Kundaje *et al.* 2015). Resources such as University of California Santa Cruz (UCSC) genome browser, allows an integrated visualisation of these publicly available epigenetic and functional resources together with in-house generated data (Kent *et al.* 2002).

In addition to data integration, the other main bottleneck encountered by functional genomics is determining the clinical relevance of GWAS SNPs, eQTLs, differentially expressed genes or differences in epigenetic modifications. This can be addressed by performing enrichment analysis, which tests for statistically significant over-representation of particular annotation terms (e.g ontologies, signalling pathways or functional elements). For instance, pathway enrichment analysis uses functional units containing related genes defined by prior knowledge, such as The Kyoto Encyclopedia of Genes and Genomes (KEGG) and the REACTOME (Kanehisa and Goto 2000; Fabregat *et al.* 2018). Tools such as eXploring Genomic Relations (XGR) integrates a wide range of ontologies and up to date publicly available functional data to perform different types of annotation and enrichment analysis (Fang *et al.* 2016).

Altogether, technical cost-effective advances in multi-omic technologies together with development and improvement of appropriate tools for data integration will improve the prioritisation of putative GWAS causal variants and their functional characterisation.

1.4.8 Aims and objectives

The aim of this thesis is to investigate the epigenetic regulatory landscape of psoriasis and PsA to identify disease and cell-type specific changes in putative regulatory regions and gene expression, with the longer-term goal of advancing understanding of the pathophysiology of these diseases and informing interpretation of genetic associations arising from GWAS.

Specific objectives are:

1. **To establish ATAC protocols appropriate for cells, tissues and clinical samples of interest, with optimisation of the required methodologies and analytical tools** (Chapter 3). Establishment of the methodological and analytical tools to study chromatin accessibility in clinical samples was required at the start of the project. This work will investigate different aspects of ATAC protocols, with a focus on:

    - quality control measurements, peak filtering, differential analysis, transposition time and approaches to reduce mitochondrial DNA contamination.
    - application of ATAC to skin biopsies
    - assessment of the impact of cryopreservation and fixation methods in the study of chromatin accessibility

2. **To define the differential epigenetic landscape and gene expression profile in psoriasis patients vs healthy controls**. Identifying differentially accessible or histone modified regions together with changes in gene expression related to disease may lead to a better understanding of disease pathophysiology (Chapter 4). This work aims to investigate:

    - the differences in chromatin accessibility and H3K27ac modifications between psoriasis patients and healthy controls in four primary immune cell types isolated from peripheral blood.
    - the differences in gene expression between patients and controls in the same four cell types.
    - the differences in gene expression between lesional and uninvolved psoriasis epidermis isolated from skin biopsies

- the fine-mapping of psoriasis GWAS loci and the integration with psoriasis epigenetic and gene expression profiles to prioritise putative causal variants.

3. **To identify differences in chromatin accessibility, gene and protein expression of immune relevant cells between synovial fluid and peripheral blood of PsA patients** (Chapter 5). Understanding epigenetic, transcriptional and proteomic differences between the two tissues in a cell-type specific manner may shed light on relevant molecular processes contributing through particular cell subsets to inflammation in the PsA synovium. This chapter will aim to:

   - investigate differences in chromatin accessibility and gene expression between synovial fluid and peripheral blood of PsA patients in a number of immune-relevant cell types.
   - to further explore particular cell types of interest using single-cell transcriptomics.
   - to integrate differences in chromatin accessibility, gene expression and protein expression for relevant cell types.
   - to perform fine-mapping of PsA GWAS loci and integrate it with PsA-specific epigenetic and expression datasets to prioritise putative causal variants with particular relevance in the synovium.

Chapter 2

Material and Methods

2.1 Ethical approval and recruitment of study participants

A summary of the samples and cohorts recruited for this thesis is in Table 2.1.

2.1.1 Psoriasis patient recruitment

Patient blood samples and skin biopsies were collected in collaboration with Professor Graham Ogg at the Weatherall Institute of Molecular Medicine, University of Oxford, and the Dermatology Department research nurses at the Churchill Hospital, Oxford University Hospitals NHS Trust. This was under approval from the Oxfordshire Research Ethics Committee (REC 14/SC/0106 and REC 14/NW/1153). After written informed consent, up to 60mL of blood from eligible psoriasis patients were collected into 10mL anticoagulant ethylenediaminetetraacetic acid (EDTA)-containing blood tubes (Vacutainer System, Becton Dickson) (Table 2.1). Detailed clinical information describing the recruited psoriasis cohorts is included in Chapter 4 Table 4.2.

Eligibility criteria. Psoriasis patients were eligible for recruitment when aged 18 years or older, previously or newly diagnosed as having plaque psoriasis (fulfilling the Psoriasis Area and Severity Index (PASI) classification, Table 1.1) and in a flare (active disease state). Recruited patients were required to have

moderate to severe disease (PASI>5), not to have taken antibiotics in the two weeks before sampling and to be naïve for biological therapy.

2.1.2 PsA patient recruitment

Sample recruitment was performed as part of the Immune Function in Inflammatory Arthritis (IFIA) study established in 2006 (REC/06/Q1606/139) in collaboration with Dr Hussein Al-Mossawi at the Botnar Research Centre and research nurses at the Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Trust. Following informed written consent, approximately 30mL of both blood and synovial fluid aspirate (variable upon disease severity) were collected into 10mL anticoagulant sodium heparin coated tubes (Vacutainer System, Becton Dickson) (Table 2.1). Further details about the cohort and collected clinical information can be found in Chapter 5 Table 5.1.

Eligibility criteria. Eligible patients were aged 18 years or older, previously or newly diagnosed with PsA as per the PsA Response Criteria (PsARC) (Table 1.2) that includes a physician global assessment questionnaire, with concomitant psoriasis and in a flare. Patients had to present an oligoarticular phenotype (four or fewer joints affected), not having taken antibiotics in the two weeks before sampling and be naïve for biological therapy and preferably for any other treatment.

2.1.3 Healthy volunteer recruitment

Recruitment of healthy volunteers was conducted as part of the Genetic Diversity and Gene Expression in White Blood Cells study with approval from the Oxford Research Ethics Committee (REC 06/Q1605/55). Up to 80mL of blood were collected into 10mL anticoagulant EDTA-containing blood tubes. Healthy individuals recruited in the study were required to be 18 years old or older without family history of psoriasis, PsA, RA or SpA and not having

suffered from an infection in the two weeks prior to sample recruitment (Table 2.1). Detailed information describing the recruited matched healthy control volunteers for psoriasis patients is included in Chapter 4 Table 4.3.

Chapter	Samples/cohorts	Description	Assay
Chapter 3	Healthy controls (n=3)	$CD14^+$ and $CD4^+$ from PB (fresh, frozen and fixed)	ATAC-seq
	Skin biopsies (n=6)	Healthy and psoriatic skin	ATAC (ATAC-seq and Fast-ATAC)
Chapter 4 (Tables 4.2 and 4.3)	Cohort 1A and 1B (n=18)	Psoriasis and healthy control cells isolated from PB	ATAC, ChIPm and RNA-seq
	Skin biopsies (n=3)	Paired lesional and uninvolved	RNA-seq
Chapter 5 (Tables 5.1 and 5.2)	PsA samples (n=6)	Cells isolated from SF and PB	ATAC, qPCR, scRNA-seq and mass-cytometry

Table 2.1: Summary table of samples/cohorts recruited in this thesis for generation of various datasets. For each of the chapters, the table specifies the recruited samples or name of the cohort (when applicable), a short description and the assays conducted. Additional tables with further details of the patient and healthy control cohorts (see in brackets) are included within the relevant result chapters. PB=peripheral blood; SF=synovial fluid

2.2 Sample processing

Blood, synovial fluid and skin biopsies were processed straight after recruitment, following the appropriate protocols.

2.2.1 PBMCs and synovial fluid cells isolation

PBMCs were isolated from blood samples through density gradient separation using Ficoll-Paque (GE Healthcare) with centrifugation at 500g for 30 minutes at room temperature with minimum acceleration and no braking.

Total synovial fluid mononuclear cells (SFMCs) were isolated by centrifugation at 500g for 5 min at room temperature in absence of density gradient. Samples were placed on ice, washed twice in ice cold Hank's balanced salt solution (HBSS) without calcium or magnesium (Thermo Fisher Scientific) and resuspended in phosphate saline buffer (PBS, Gibco) supplemented with 0.5% fetal calf serum (FCS, Invitrogen) and 2mM EDTA (Sigma), prior to separation of the different cell types. Cell numbers and viability were determined by manual counting using a haemocytometer with trypan blue (Sigma) for viability assessment.

2.2.2 Skin biopsies processing and adherence assays

Keratinocytes enrichment from skin biopsies was performed as described in Gutowska-Owsiak and colleagues (Gutowska-Owsiak *et al.* 2012). Skin biopsies (approximately 4mm) were washed with PBS, cut in 1mm width strips and incubated in 2U/mL of dispase II (Sigma) overnight at 4°C. Following incubation, the epidermis was separated from the dermis. For RNA extraction, the epidermis was snap-frozen in liquid nitrogen. For chromatin accessibility assay, the epidermis was further digested in trypsin (Invitrogen) at 37°C for 5 min to obtain a cell suspension that was filtered through a 70µm nylon strainer (BD) and washed with PBS. In some instances cells were manually counted and aliquoted in PBS prior to chromatin accessibility assay. In others, cells from each of the biopsies were resuspended in 100µL of KGM-2 BulletKit (Lonza) supplemented with 0.06mM Ca^{2+} (low Ca^{2+}) and cultured in a collagen IV coated 96-well plate at 37°C 5% CO_2 for 3 hours. After culture, cells in 96-well plates were washed twice with 200µL of PBS and kept at 37°C for downstream chromatin accessibility processing.

2.2.3 Fixation, cryopreservation and cell culture

In Chapter 3, cells (50,000) were fixed using dithio-bis(succinimidyl propionate) (DSP) as described in Attar and colleagues and stored at 4°C for 24h (Attar *et al.* 2018).

Liquid nitrogen storage of 40-50x10^6 PBMCs was carried out using a modified version of the protocol described by Kent (Kent 2009), where cells were pre-conditioned in RPMI 1640 complete medium (Lonza) supplemented with 2 mM L-glutamine (Sigma), 100U penicillin and streptomycin 100µg/mL (Sigma) and 50% FCS for 30 minutes and, afterwards, diluted 1 in 2 in complete RPMI 1640 (supplemented as previously described) with 20% dimethyl sulfoxide (DMSO, Sigma). PBMCs followed slow cryopreservation at -80°C at -1°C per minute and then transferred and stored for a minimum of two weeks in liquid nitrogen. PBMCs were thawed quickly in a 37°C water bath, resuspended in supplemented complete RPMI 1640 with 10% FCS at a density of 10^6 cells/mL (Fairfax *et al.* 2014) and rested for 30 min at 37°C, 5% CO_2 in 25mL non-adherent polypropylene cell culture flasks (Greiner) followed by filtering through a 40µm cell strainer to obtain a homogenous cell suspension for FACS separation.

Cryopreserved normal human epidermal keratinocytes (NHEK, Lonza) in passage three were recovered and cultured at a cell density of 5x10^6 cells/mL in a 75 mL adherent cell culture flask (Greiner) in EpiLife basal medium (Gibco), following manufacturer's instructions. After recovery, NHEKs were trypsinised at room temperature for 8 minutes and trypsin was inactivated with EpiLife 10% FCS. Cells were centrifuged at 180g for 10 min at room temperature and then manually counted with trypan blue for viability staining. NHEKs (16,000 cells) were seeded in a 96-well plate in 100uL of medium and cultured for 2 days at 37°C, 5% CO_2 to reach 90-100% confluence (approximately 50,000 cells) before performing any ATAC protocol on the plate (further detailed in Chapter 3). When used for Omni-ATAC, NHEKs after trypsinisation were processed through Ficoll

density gradient (as previously explained for PBMCs isolation) to remove dead cells as recommended by Corces and colleagues (Corces *et al.* 2017).

2.2.4 Primary cell isolation using magnetic-activated cell sorting

Primary cell subpopulations in Chapter 4 (Table 2.1) were separated using magnetic-activated cell sorting (MACS, Miltenyi) following the manufacturer's instructions. Consecutive positive selection was performed using Miltenyi beads for $CD19^+$, $CD8^+$, $CD14^+$ monocytes and $CD4^+$ cells (catalogue numbers 130-050-201, 130-045-101, 130-045-201 and 130-050-301, respectively) and AutoMACS Pro (Miltenyi) followed by a manual cell count with trypan blue.

MACS separation was chosen over fluorescence-associated cell sorting (FACS) due to time and logistic constraints during sample processing. Moreover MACS positive selection was preferred over depletion as it allowed step-wise isolation of the four cell populations of interest with an expected purity of approximately 99%. Nevertheless, using positive selection with antibodies targeting CD4 and CD8 can cause T cell activation to some extent as both surface markers are co-receptors involved in the recognition by the T cell receptor (TCR) of antigens presented through MHC-I or MHC-II molecules, respectively. This pre-activation is of particular relevance when performing downstream cell culture and assays.

2.2.5 Primary cell isolation using fluorescence-activated cell sorting

Isolation of cell subpopulations from PBMCs and SFMCs (Chapter 3 and 5 in Table 2.1) was performed by FACS. PBMCs and SFMCs were resuspended in 1mM EDTA PBS (FACS buffer) at $10x10^6$ cells/mL, stained with the appropriate antibody cocktail (Table 2.2) for 30 min at 4°C, washed with FACS buffer and

centrifuged at 500g for 5 min at 4°C. For the cell separation in the Chapter 3 samples (Table 2.1), a modified FACS buffer supplemented with 3mM EDTA, 2% FCS and 25mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Invitrogen) was used to avoid cell clumping after cryopreservation and short recovery in culture (as detailed previously). After removing the supernatant from the washing step, cells were resuspended in FACS buffer prior to separation.

From the healthy control samples of Chapter 3 (Table 2.1), $CD14^+$ monocytes and $CD3^+$ $CD14^-$ $CD4^+$ T cells were isolated using the SONY SH800 cell sorter. For the PsA samples (Chapter 5 in Table 2.1 and Table 5.2), separation of $CD19^+$ B cells, memory T cells ($CD3^+$ $CD14^-$ $CD4^+$ $CD45RA^-$ and $CD3^+$ $CD14^-$ $CD8^+$ $CD45RA^-$), $CD14^+$ monocytes and $CD3^-$ $CD56^+$ NK from PBMCs and SFMCs was performed using FACS Aria (BD) cell sorter. Sorted cells were collected in 1.5mL tubes containing PBS 1% FCS when used for ATAC-seq or only PBS when processed for scRNA-seq to avoid potential RNAse contamination. OneComp eBeads (eBioscience) were used for compensation of fluorescence spill over.

Surface marker	Fluorochrome PsA/Control	Manufacturer PsA/Control	Clone PsA/Control	Dilution PsA/Control
Viability	eFluor780	-	eBioscience	1:500/1:250
CD3	FITC/AF700	SK7/UCHT1	BioLegend	1:50/1:50
CD4	APC	RPA-T4/RPA-T4	BioLegend	1:50/1:50
CD8a	PE	RPA-T8	BioLegend	1:100/-
CD45RA	BV421	HI100	BioLegend	1:25/-
CD19	PerCP-Cy5.5	SJ25C1	BioLegend	1:50/-
CD14	Pe-Cy7/FITC	M5E2/TUK4	BioLegend /Miltenyi	1:50/1:100
CD56	BV510	HCD56	Biolegend	1:25/-

Table 2.2: Antibody panel used for FACS separation of primary cell populations in Chapter 3 controls and Chapter 5 PsA samples. Details regarding target molecule, fluorochrome, clone, supplier and dilution used for PBMCs and SFMCs staining are provided for each surface marker in the panel. For cell separation from Chapter 3, control PBMCs staining was only performed for CD3, CD4, CD14 and viability markers.

2.3 Experimental protocols

2.3.1 ATAC - Chromatin Accessibility

Three different versions of the ATAC-seq protocol were progressively used in this thesis, following their publication, for assessment of chromatin accessibility in different primary cells, including $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells, $CD19^+$ B cells and $CD3^-$ $CD56^+$ NK cells, as well as in epidermal keratinocytes isolated from skin biopsies. ATAC (used in this thesis to refer to any of the three protocols in subsequent chapters) interrogates accessible chromatin using the transposase enzyme Tn5 to simultaneously access nucleosome-free and inter-nucleosomal DNA and tag both ends of the fragments with Illumina adapters. Fast-ATAC and Omni-ATAC were two subsequent protocols following the ATAC-seq protocol from Buenrostro and colleagues (Buenrostro *et al.* 2013), aiming to reduce the amount of mitochondrial DNA in the sequencing libraries and improve the signal-to-noise ratio of the original protocol. When using MACS separation, primary cells were manually counted, as specified above, and resuspended in PBS 1% FCS.

ATAC-seq

ATAC-seq was used to generate data from NHEKs, skin biopsies and healthy volunteers (Chapter 3 in Table 2.1) as well as cohort 1A primary immune cells isolated from blood of psoriasis and control samples (Chapter 4 in Table 2.1 and Tables 4.2 and 4.3). ATAC-seq was performed using an estimated number of 50,000 cells as described in Buenrostro and colleagues with minor modifications (Buenrostro *et al.* 2013). Cells lysis was carried out for 10 min, the isolated nuclei were transposed for 40 min at 37°C using the Nextera Tn5 transposase (Illumina) and DNA was purified with the PCR MinElute kit (Qiagen), following the manufacturer's instructions. When using DSP fixed cells, two washes with

50µL of PBS were performed to remove any remaining fixative prior to the ATAC-seq protocol. After the transposition reaction, the Tn5 enzyme was inactivated with 500 mM EDTA for 30 min at 70°C followed by de-crosslinking using 50 mM dithiothreitol (DTT) for 30 min at 37°C and DNA column purification, as previously detailed. All transposed samples were simultaneously amplified and singled indexed for 11 PCR cycles using modified Nextera primers from Buenrostro *et al.*,2013, after appropriate assessment of the approximate required number of qPCR cycles. The resulting DNA libraries were purified using the MinElute PCR purification kit (Qiagen) and a 1.8X (v/v) of Agencourt AMPure XP Magnetic Beads (Beckman Coulter) to remove adapters excess and primer dimers.

Additional modifications of the protocol were implemented when processing keratinocytes isolated from skin biopsies and NHEKs in 96-well plates (Bao *et al.* 2015) (as later detailed in Chapter 3).

Fast-ATAC

An improved ATAC-seq protocol was published in Nature Methods by Corces and colleagues called Fast-ATAC (Corces *et al.* 2016). Optimised for hematopoietic cells, the protocol combined cell lysis and transposition into a single step. In this thesis, Fast-ATAC was performed in skin biopsies (Chapter 3 in Table 2.1), cohort 1B primary immune cells isolated from blood of psoriasis and control samples (Chapter 4 in Table 2.1 and Tables 4.2 and 4.3) and primary immune cells isolated from blood and synovial fluid of PsA patients (Chapter 5 in Table 2.1 and Table 5.1). Fast-ATAC was conducted as described by Corces and colleagues with minor modifications. Following the advice from the authors, approximately 20,000 cells (MACS or FACS sorted) were washed with 200µL of PBS, centrifuged at 500g for 5 min at 4°C and incubated in the lysis/transposition buffer containing digitonin (Roche) for 30 min at 37°C and agitation at 400rpm.

Following transposition DNA was prepared and purified, as per the previously detailed ATAC-seq protocol, except 13 cycles of PCR amplification were used after appropriate cell cycle determination in a pilot set of samples.

Omni-ATAC

Omni-ATAC was performed in 50,000 viable NHEKs in suspension (Chapter 3), as described in the latest publication from Corces and colleagues (Corces *et al.* 2017). Transposed DNA was simultaneously amplified and indexed, as detailed in the ATAC-seq standard protocol, for 8 PCR cycles and purified using MinElute PCR purification columns (Qiagen) only.

Quality control and sequencing

Indexed and amplified ATAC samples were assessed for tagmentation profile on an Agilent 2200 or 4200 Tapestation with the D1000 high sensitivity DNA tape (Agilent) as part of the quality control. Quantification of the library concentration was performed by qPCR using the Kapa assay from Roche, following the manufacturer's instructions. Pools of 12 to 16 libraries were sequenced on up to 3 lanes of the HiSeq4000 Illumina platform aiming for at least 30 million paired-end reads after filtering. NGS was performed at the Oxford Genomics Centre, at the Wellcome Centre for Human Genetics (WHG) for this and all other work involving NGS in this thesis.

2.3.2 Chromatin immunoprecipitation with sequencing library preparation by Tn5 transposase

For chromatin immunoprecipitation (ChIP) a low cell input protocol known as ChIPmentation (ChIPm) was used (Schmidl *et al.* 2015). This protocol combines DNA cross-linking and chromatin sonication and immunoprecipitation, as per standard ChIP-seq, with sequencing library preparation by tagmentation

using Tn5 transposase. ChIPm introduces sequencing-compatible adapters by performing the transposition reaction directly on bead-bound chromatin, prior to de-crosslinking, preventing overtagmentation of the DNA and thus reducing required input material. Overall ChIPm represents a more time- and cost-effective protocol compared to ChIP-seq or other low-input ChIP-seq protocols.

The H3K27ac histone mark (active enhancer and promoter marker) was assayed in four cells types ($CD14^+$ monocytes, total $CD4^+$, total $CD8^+$ and $CD19^+$) from samples of cohort 1B (Chapter 4 in Table 2.1 and Tables 4.2 and 4.3). For each sample and cell type, 600,000 MACS sorted cells, as described in 2.2, were fixed with 1% formaldehyde (Sigma) and snap frozen in dry ice and ethanol prior to storage at -80°C. Fixed cells were thawed, resuspended in SDS lysis buffer and sonicated for 8 min using Covaris M220 (Covaris) with a duty factor of 5%. After sonication, chromatin was split into 6 aliquots (100,000 cells per aliquot), snap frozen and stored at -80°C. Aliquots as needed were thawed on ice and then processed downstream for ChIPm as described by Schmild and colleagues for immunoprecipitation, tagmentation of bed-bound chromatin, de-crosslinking and DNA amplification (Schmidl *et al.* 2015). Immunoprecipitation was carried out with 1µg of the Diagenode antibody (C15410196). For each sample, an aliquot of chromatin was processed in parallel without incubation with the anti-H3K27ac antibody (control input) to account for bias in enrichment of more accessible genomic regions that may be sonicated and tagmented more efficiently and therefore overrepresented in the antibody pull-down. Tagmentation of the control input was performed using 1ng of DNA. Amplification by qPCR was carried out in each of the samples and control inputs to determine the appropriate number of full cycles required to reach one-third of the final fluorescence to minimise the presence of PCR replicates upon NGS. Libraries were then amplified for the number of determined cycles minus one and simultaneously dual indexed using the primers optimised by Buenrostro

and colleagues (Buenrostro *et al.* 2015). A pool of 64 libraries (including control input samples) were sequenced over a number of lanes in the HiSeq4000 Illumina platform aiming for 25 million paired-end reads

2.3.3 RNA extraction and gene expression quantification

RNA extraction

Following MACS isolation of the different cell types 2 to $3x10^6$ cells were resuspended in 350µL of RNAProtect or RLT buffer (Qiagen) supplemented with 0.1% of beta-mercaptoethanol (BM, Sigma) and snap frozen in dry ice before storage at -80°C. Cells isolated from Cohort 1A psoriasis and control samples (Chapter 4 in Table 2.1 and Tables 4.2 and 4.3) were preserved in RNAProtect, which stops any biochemical reaction and transcriptional activity whilst maintaining cell integrity. At early stages of the project, the time frame to process the acquired samples was uncertain and RNAProtect was chosen as the most appropriate strategy to preserve cells for future RNA extraction to guarantee high quality in case storage exceeded 6 months. In the psoriasis and control samples from Cohort 1B (Chapter 4 in Table 2.1 and Tables 4.2 and 4.3) and in the PsA samples (Chapter 5 in Table 2.1 and Table 5.2), cells were resuspended in 0.1% BM supplemented RLT buffer, which lysis cells and prevents RNA degradation. When starting from RNAProtect preserved material, cells were centrifuged at 300g for 10 min at room temperature, the supernatant were removed and the pellets were resuspended in 350µL of RLT 0.1% BM buffer. All cell lysates were homogenised using the QIAshredder (Qiagen) prior to RNA extraction.

Total RNA was extracted using the AllPrep DNA/mRNA/microRNA Universal kit (Qiagen) following the manufacturer's instructions. RNA extractions were performed in batches of 12 samples, including all cell types from each individual processed and balanced numbers of psoriasis and

control samples, to minimise batch effect correlating with phenotype. Basic quantification was performed with NanoDrop (Thermo Scientific) before storage at -80°C.

RNA-seq

RNA-seq quality control (QC), quantification and library preparation were carried out by Oxford Genomics Centre followed by NGS in two independent batches of samples, each including Cohort 1A or Cohort 1B, respectively (Chapter 4 in Table 2.1 and Tables 4.2 and 4.3). Processing of samples in two batches was due to logistics of patient recruitment in the project. RNA quality control and quantification were assessed with the Bioanalyzer (Agilent). Samples were depleted from ribosomal RNA using Ribo-Zero rRNA Removal kit (Illumina) prior to cDNA synthesis and library preparation using TruSeq Stranded Total RNA (Illumina). This method preserves non-polyadenylated transcripts including nascent pre-mRNA (unspliced) and functionally relevant lncRNAs. For each of the cohorts, all libraries were pooled together and sequenced over several lanes of HiSeq4000 aiming a depth of approximately 25 million paired-end reads per sample to maintain an appropriate level of sensitivity for subsequent expression analysis.

Gene expression quantification by qPCR array

Expression of immune-relevant genes for a number of the PsA cohort samples (Chapter 5 in Table 2.1 and Table 5.2) was profiled using the RT2 Profiler PCR Array (PAHS-3803Z, Qiagen) by UCB collaborators. The qPCR array consisted of a 384-well plate which includes SYBR Green-optimised primers assays to test expression for 370 key genes involved in immune response during autoimmunity and inflammation, as well as appropriate house-keeping genes for normalisation of expression. This PCR array allows simultaneous amplification

of different gene-specific products under the same PCR conditions. In brief, RNA was extracted, as detailed previously, from $CD14^+$ monocytes, memory $CD4^+$ and $CD8^+$ cells, followed by reverse-transcription for cDNA synthesis using RT Primer Mix (Qiagen). Synthesised cDNA was mixed with the ready-to-use PCR mastermix and equal volumes were aliquoted to each well of the plate as specified in the manufacturer´s instructions. Real-time PCR cycling program was run and abundance of the amplified product quantified by measuring SYBR Green fluorescence.

Single-cell RNA-seq

scRNA-seq data was generated using 10X Genomics technology Chromium single cell 3' expression library preparation kit (PN-120267) by the Oxford Genomics Centre at the WHG for a number of the PsA cohort samples (Chapter 5 in Table 2.1 and Table 5.2). Briefly, PBMCs and SFMCs were made into a cell suspension. Approximately 3,000 cells from the PBMCs or SFMCs suspensions were partitioned into single-cell gel beads in emulsion (GEMs) using the 10X Chromium controller system. Reverse-transcription for cDNA synthesis was performed within the GEMs, which included a 16bp 10x barcode, a 10bp unique molecular identifier (UMI) and poly-dT primers. The cDNA was released from the GEMs, followed by PCR amplification, enzymatic fragmentation and size selection (to optimise the insert size before library preparation). Afterward, appropriate sequencing Illumina indexes were incorporated into the samples through TrueSeq library preparation using end repaire and A-tail ligation. Sequencing was performed using PE HiSeq4000 with 26bp for read 1 (barcodes) and 98bp for read 2 at a depth of approximately 50,000 reads per cell, following standard 10X library sequencing requirements.

2.3.4 DNA extraction and rs4672405 genotyping

DNA isolation was performed using the AllPrep DNA/mRNA/microRNA Universal kit (Qiagen) following the manufacturer's instructions. Quantification was performed using NanoDrop (Thermo Scientific) and samples were kept at -80°C for long term storage. The extracted DNA was amplified by PCR using forward (5'-CACTGTGGAGGGAGGAACAA-3') and reverse (5'-CGTGTTGGCCAGGATAGTCT-3') primers annealing up and down stream of the SNP rs4672505, respectively. An aliquot of the sample was run on a 1% agarose gel to check for amplification of a 390bp PCR fragment. The remaining was purified using MinElute PCR purification kit, quantified by dsDNA Qubit kit (Invitrogen) according to the manufacturer's instructions and prepared for Sanger sequencing using the Mix2Seq kit and service (Eurofins). The forward and reverse sequences were analysed with BioEdit software (`http://www.mbio.ncsu.edu/BioEdit/bioedit.html`).

2.3.5 Mass cytometry using cytometry by time of flight (CyTOF)

CyTOF represents an improved version of flow cytometry that allows a single-cell high-dimensional phenotypic and functional analysis. CyTOF interrogates up to 42 intracellular and surface markers by using antibodies coupled to metal isotopes instead of fluorescence molecules. Replacement of fluorescence molecules by metal isotopes allows increasing the number of functional markers simultanously assayed and minimises the spillover of signal, two of the main limitations in flow cytometry (Yao *et al.* 2014).

Mass cytometry using CyTOF was performed by Dr Nicole Yager in collaboration with UCB in all the PsA samples detailed in Chapter 5 Table 2.1 and Table 5.2, and in additional four PsA patients not processed for any

of the other techniques. Briefly, an aliquot of peripheral blood and synovial fluid were fixed for 5 min with 1.6% paraformaldehyde (PFA) within 30 min of venipuncture/aspiration, respectively. These samples were defined as time 0h. In addition, another aliquot of peripheral blood and synovial fluid were incubated at 37°C for 6h in the presence of the protein transport inhibitors 1X BD GolgiStop (BD) and 1X BD GolgiPlug (BD), containing monesin and brefeldin A, respectively. Treatment with monensin and brefeldin A prevents the extracellular transport of cytokines from the cells and allowed measuring the intrinsic cytokine production rate in basal conditions. After 5h 45 mins the samples were treated with cisplatin to facilitate discrimination of dead cells, and then fixed 5 min with 1.6% PFA. These samples were defined as time 6h. After fixation of time 0h or 6h samples, red blood cells were lysed and cell suspensions were washed with PBS and stained with antibodies against the cell surface markers (Table 2.3). The samples were then permeabilised and labelled with antibodies for intracellular functional markers (Table 2.3) and data was acquired on a CyTOF Helios instrument.

Markers from the CyTOF antibody panel
CD248, CD19, GP38, FAP, CD8a, IL8, CD16, CD25, CD123, IL-17F, IL-17A, IL-10, CD11c, CD14, IL6, IFN-γ, CD-11b CD45RO, CD56, HLA-DR, IL-13, CD117, CD4, IL4, IL-2, TNFα, IL-21, FceR, CD3, CD161, IP-10, GM-CSF, CD45, MCP-1, osteopontin

Table 2.3: Molecules targeted by the mass-cytometry antibody panel for peripheral blood and synovial fluid samples. The molecules targeted by the isotope-conjugated antibodies of the CyTOF panel are listed and include intra-cellular and surface markers.

2.4 Computational and statistical analysis

2.4.1 ATAC data analysis

ATAC-seq, Fast-ATAC and Omni-ATAC data were analysed using an in house pipeline. The pipeline performs single sample data processing and also builds a combined master list for each of the conditions of interest for chromatin accessibility characterisation and differential analysis.

Next generation sequencing data processing

NGS data for each sample was trimmed for low quality base pairs and Nextera adapter sequences using cutadapt (Martin 2011) before general QC assessment using fastqc (Andrews 2010). Trimmed reads were aligned to the reference genome (build hg19) using bowtie2 (Langmead and Salzberg 2012) and the following parameters were used, -k 4 -X 2000 -I 38 ––mm -1, consistent with other publications (Buenrostro *et al.* 2013; Corces *et al.* 2016). Samtools (Li *et al.* 2009) was used to remove PCR duplicate reads marked with Picard Tools (`http://broadinstitute.github.io/picard/`) as well as low MAPQ (<30), non-uniquely mapping and non-properly paired reads. The resulting bam file was additionally filtered to remove mitochondrial DNA and reads were adjusted by +4bp on the plus strand and by −5bp on the minus strand to represent the center of the transposition binding event. Pileup tracks (bigWig files) representing the number of reads per bp position were generated using bedtools genomeCoverageBed (Quinlan and Hall 2010) and the UCSC genome browser bedGraphToBigWig tool (Kent *et al.* 2010). For visualisation purposes, normalised bigWig files were generated from normalised bedgraph files with bedtools genomecov and the library size factor estimated by the differential analysis algorithm.

Peak calling, filtering and sample quality assessment

Peak calling was performed using MACS2 callpeak (Zhang *et al.* 2008) and the parameters ––nomodel ––shift -100 ––extsize 200 ––p 0.1 ––keep-dup all ––call-summits. Peaks overlapping blacklisted features from the ENCODE project (`https://www.encodeproject.org/annotations`) were removed. The ––shift and ––extsize parameters were set according to the recommendations of MACS2 for DHS and following other ATAC-seq publications (Buenrostro *et al.* 2015; Corces *et al.* 2016). The p-value cut off for filtering called peaks was determined for each cell type using Irreproducibility Discovery Rate (IDR) analysis (as further detailed in Chapter 3). For this, the filtered bam file of each sample was partitioned into two equal size files (pseudoreplicates). Peak calling was performed in each pseudoreplicate, followed by filtering for a range of p-values (from 0 to 10^{-45}) and IDR analysis done for the resulting pairs of filtered peak sets. For each of the filtering p-values, the percentage of peaks sharing IDR rank between the two pseudoreplicates was determined, and the optimal p-value filter identified. When more than one summit was identified in a peak, the median of the summits was used. For all peaks, summits were extended ±250bp to create a non-overlapping homogenous 500bp peak list for each sample (Buenrostro *et al.* 2015; Corces *et al.* 2016).

Sample quality was determined by the fold enrichment of ATAC-seq signal across all the hg19 TSSs annotated by Ensembl (175,114 features), computed as in (Buenrostro *et al.* 2015) using a script provided by Dr Silvia Salatino. In brief, transposition events were calculated in 1bp bins ±1,000bp surrounding all TSSs and normalised to the mean value of background reads (signal from -1,000 to -800). For overall library quality assessment all ATAC fragments were considered. When assessing chromatin structure within or across the TSS, fragments of <150bp or 260-340bp were used, respectively (Scharer *et al.* 2016). Fraction of reads in peaks (FRiP) was calculated for samples in Chapter 3 as

the overlap between the peak list filtered for FDR<0.01 and all ATAC fragments using bedtools intersect with the parameter −−f 0.1.

Combined peak master list and differential analysis

To perform differential open chromatin analysis a non-overlapping 500bp peak master list including all samples for a particular contrast was built upon removal of peaks in the X and Y chromosomes. Each master list was built by union of all the peaks present in at least 30% of the samples included, regardless of the subgroup to which they belonged (e.g patients or controls, synovial fluid or peripheral blood). Reads overlapping each of the peaks in the master list were retrieved for each sample using the HTSeq-count algorithm (Anders *et al.* 2015). Principal component analysis (PCA) was performed on all counts normalised with the vsd function from the DESeq2 v1.20 R package (Love *et al.* 2014).

For differential chromatin accessibility analysis, an empirical 80% confidence cut-off was calculated and used to pre-filter peaks with high noise that could be confounding the analysis (detailed in Chapter 3). Differential analysis was performed using DESeq2 with a paired design (in Chapter 3 and Chapter 5) or including batch as a covariate in the model (in Chapter 4). Peaks were annotated with proximal genes (≤5Kb) using xGR2nGenes function from the XGR R package (Fang *et al.* 2016). These gene lists were intersected with differentially expressed genes (DEGs) from RNA-seq/scRNA-seq analysis or genes reported by psoriasis and PsA GWAS.

2.4.2 ChIPm data analysis

Next generation sequencing data processing

ChIPm NGS data from samples and inputs were processed similarly to ATAC-seq (Section 2.3.1) for trimming, mapping and filtering with minor modifications. Specifically, the MAPQ30 score for filtering reads was lowered

to 10. For visualisation, bedgraph files with noise subtracted using the control input were generated using MACS2 bdgcmp -m subtract followed by conversion to bigWig with bedGraphToBigWig tools.

Peak calling, filtering and sample quality

Peak calling for each ChIPm sample was performed accounting for background signal using the control input samples with MACS2 callpeak −−bw 200 −−p 0.1 −−keep-dup all −−call-summits. In this case the average library fragment size (−−bw) was used by MACS2 to first empirically find the model that best represents the precise protein-DNA interactions and calculate the appropriate −−shift parameter. For ChIPm PCA, filtering and downstream analysis peak homogenisation was performed similarly to Section 2.4.1 to build a combined master list for all samples and cell types from cohort 1B (Chapter 4 Table 4.2 and 4.3).

Sample quality was determined by a combination of measurements. For library complexity the non-redundant fraction and PCR bottleneck coefficients (PBC1 and PBC2) were calculated following ENCODE guidelines (`https://www.encodeproject.org/chip-seq/histone/`) from unfiltered bam files. PBC1 corresponds to number of genomic locations where exactly only one read maps uniquely divided by the number of genomic locations to which more than one read maps uniquely. PBC2 is defined as the number of genomic locations where only one read maps uniquely divided by the number of genomic locations where exactly two reads map uniquely. Enrichment of the ChIPm signal was evaluated based on the normalised strand cross-correlation coefficient (NSC) and relative strand cross-correlation coefficient (RSC), calculated with SPP using bam files filtered for low MAPQ30, duplicated and non-properly paired reads.

Combined peak master list and differential analysis

DiffBind (default parameters unless specified) was used to build a peak master list and perform differential H3K27ac analysis between psoriasis patients and healthy controls for each cell type. DiffBind used the unfiltered sample peak files generated by MACS2 and the filtered bam files (from samples and control inputs) to generate a master list including high quality reproducible peaks present in at least 50% of the samples (modification from default parameters) and retrieve counts of the reads mapping at the location of each peak.

2.4.3 Gene expression analysis

qPCR analysis

Pre-processing of the qPCR data up to calculation of fold changes for each gene was conducted by UCB collaborators. When comparing synovial fluid and peripheral blood isolated cells from patients, fold changes were calculated for each patient as the ratio between the 2^{-dCt} in each of the tissues. Therefore, in a particular cell type three fold changes (one per individual) were provided for each gene. When comparing expression in cells isolated from peripheral blood of PsA patients versus healthy controls, the fold change for each gene in each patient was calculated as the ratio between the average 2^{-dCt} in the three controls and the 2^{-dCt} of a particular patient. In order to determine the significance of the calculated fold changes, one sample t-test was performed with the null hypothesis being $H_0 : \mu = 1$ (no change) and the alternative hypothesis $H_1 : \mu \neq 1$.

Bulk RNA-seq analysis

NGS RNA-seq data processing was performed using an in-house pipeline developed by Dr Katie Burnham. Ribo-depleted RNA-seq data was mapped against the reference genome (build hg19) using the aligner STAR (Dobin *et*

al. 2013). Mapping allowed multiple alignments and only retained those with the best score and a mis-match percentage lower than 0.04%. Duplicates were marked and removed using Picard Tools. Gene counts were retrieved using HTSeq-count and the gencode hg19 annotation file comprising 2,840,278 gene entities, including lnc-RNAs. Differential gene expression analysis was performed with DESeq2 on genes with five or more reads in at least eight samples (smallest group size corresponding to the psoriasis patients samples). Independent filtering of genes with low expression levels, outlier removal using Cook's distance and moderation of $\log_2$ fold changes were enabled when using DESeq2. Differentially expressed transcripts were identified based on p-values calculated using Wald test and False discovery rate (FDR) multiple testing correction, with a threhold of FDR<0.05. Batch effect was included as a covariate in the contrast between psoriasis patients and healthy controls. This effect related to the RNA extraction, library preparation and sequencing of cohort 1A and cohort 1B samples from the psoriasis study (including healthy controls). Lnc-RNAs were annotated using the list provided by gencode.v19 `https://www.gencodegenes.org/releases/19.html`. A paired design of the psoriasis skin DGE analysis was taken into account by the DESeq2 model.

Single-cell RNA-seq analysis

Raw Illumina sequencing data from the 10X Genomics technology Chromium single cell 3' expression libraries generated in bulk PBMCs and SFMCs from three PsA patients (see Chapter 5) were first processed using Cell Ranger v2.2 software provided by 10X Genomics technology `https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/`. Illumina sequencing base call files (BCLs) were demultiplexed and converted into fastq files using cellranger mkfastq. For each of the samples, mapping of the fastq files to the compatible human transcriptome reference (GRCh38-1.2.0)

and retrieval of counts for each transcript included in the reference genome were performed with cellranger count using default parameters.

The count matrix files were processed downstream using the package scanpy in python 3.6 (Wolf *et al.* 2018). Each of the PBMCs and SFMCs individual count matrices were downsampled to 3,000 cells, when necessary (scanpy.subsample). Those cells expressing less than 200 genes were filtered out and genes expressed in less than 20 cells were also discarded (scanpy.filter_cells and (scanpy.filter_genes). Cells showing expression of *CD14* and *LYZ* greater than 0 were isolated as $CD14^+$ monocytes. Additional filtering was conducted on the $CD14^+$ to remove cells presenting more than 15% of mitochondrial reads, a number of expressed genes greater than 3,000 and a maximum of 15,000 UMI. Gene expression was normalised per cell by total counts over all the genes, UMI and percentage of mitochondria were regressed out and counts were transformed to natural log. Imputation of missing gene expression was conducted using the Markov Affinity-based Graph Imputation of Cells (MAGIC) (scanpy.magic) (Dijk *et al.* 2018) followed by data scaling (zero centered) and PCA. Batch effect was corrected using batch balanced k nearest neighbours (BBKNN) algorithm incorporating the calculated PCs and using patient ID as key for batch discrimination (scanpy.bbnnn). Cluster identification was performed using the Leiden algorithm with default resolution 1 (scanpy.leiden) (Traag *et al.* 2018) and manual merge of some of the clusters. Dimensional reduction and visualisation was conducted using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) algorithm (scanpy.umap). To identify marker genes of a particular cluster, differential analysis between cells in that cluster and the rest was conducted using scanpy.rank_genes_groups function with p-values calculation using Wilcoxon-signed rank test and FDR using Benjamini-Hochberg test. The same function was used to perform genome-wide DGE analysis between synovial fluid vs peripheral blood $CD14^+$ monocytes.

2.4.4 Genomic region annotation, enrichment analysis and pathway visualisation

Genomic region annotation with chromatin states was performed by overlapping peaks with the appropriate cell type chromatin segmentation map from Epigenome Roadmap (`https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html`) using bedtools intersect and a minimum overlap of 20% of the peak with a nominated state.

Genomic feature enrichment analysis for XGR built-in data (including FANTOM enhancers, histone marks, TFBS and chromatin segmentation maps) was performed using binomial test from the xGRviaGenomicAnno function (Fang *et al.* 2016). When annotating regions (ATAC or ChIPm peaks) the summit of each peak were used to increase the specificity of the analysis. Pathway enrichment analysis was conducted for the built-in KEGG, Reactome MsigdbC2BIOCARTA and MsigdbC2CPall pathways with the xEnricherGenesAdv XGR functionality using hypergeometric test. Input data for the enrichment analysis were proximal genes to differentially accessible ATAC peaks or the differentially expressed genes. The background was defined as all the ATAC peaks from the consensus list tested in differential chromatin accessibility analysis or all the gencode hg19 genes interrogated in the differential gene expression analysis. Enrichment analysis for the Gene Ontology biological processes (GOBP) dataset was conducted using XGR to explore the functional relevance of the scRNA-seq synovial fluid and peripheral blood $CD14^{+}$ monocytes subpopulations. The input were all the up-regulated genes (FDR<0.01 and fold change>1.5) unique for that particular cluster and the background all the detected genes after data filtering. For the MC-4 cluster, GOPBP enrichment analysis was also conducted using all the significant differentially expressed genes (FDR<0.01) identified when comparing to all the other clusters combined.

LD-based enrichment analysis of psoriasis and PsA GWAS LD blocks for differentially accessible regions (DARs) was conducted with the XGR function xLDenricher. This function allows enrichment of the psoriasis and PsA GWAS LD blocks for an annotation feature, in this case differentially accessible ATAC peaks, to be calculated by comparing this overlap (observed) to the overlap between differential ATAC regions and a null distribution of LD blocks generated via sampling from all common SNPs (background). The sampling strategy to generate this null distribution of LD blocks respects the minor allele frequency (MAF) and the distance of the lead SNP to the nearest gene in each of the GWAS LD blocks. 20,000 permutations were conducted. Similarly, enrichment analysis for PsA GWAS fine-mapped SNPs and differentially accessible regions or consensus ATAC peaks was conducted using 1,000 backgrounds of SNPs generated by SNPsnap (`https://data.broadinstitute.org/mpg/snpsnap/match_snps.html`). These backgrounds included SNPs that preserved the same LD structure, allele frequencies and proximity to genes as the fine-mapped SNPs. The fine-mapped SNPs and the matched backgrounds were intersected using an in-house R script with (1) differentially accessible regions (DARs) or consensus ATAC peaks and (2) a window of 10Kb around the DARs or the consensus ATAC peaks for each of the cell types. Empirical p-value was computed to determine the significance of the overlap between fine-mapped SNPs and each of the features using binomial test.

Visualisation of the signalling-pathway enrichment from the RNA-seq results was performed using an R function part of the Atlas and Analysis of systems-biology-led pathways website resource, developed by Dr. Hai Fang (manuscript in preparation). The manually curated KEGG pathways expanded with genes for each gene family were coloured based on the fold change from the corresponding RNA-seq differential analysis and highlighted in bold when passing the FDR threshold for significance.

2.4.5 Statistical fine-mapping

Fine mapping of the psoriasis and PsA GWAS signals was performed using a Bayesian approach, aiming to overcome the incomplete coverage of genotyping arrays and the hundreds of associations per locus due to the LD structure of the genome. Fine-mapping was conducted using two different strategies due to availability of summary statistics or genotyping data from the psoriasis and PsA Immunochip GWAS studies, respectively. Both strategies include the same main steps of imputation and calculation of posterior probability (PP) and were implemented in collaboration with Dr Adrián Cortés.

Psoriasis fine-mapping using Immunochip summary statistics

Fine mapping was performed for 26 of the risk loci reported by the psoriasis Immunochip GWAS study from Tsoi and colleagues, for which comprehensive summary statistics were only publicly available for the GAPC cohort (2,997 cases and 9,183 controls) at the time of this analysis (Tsoi *et al.* 2012). The summary statistics file included the p-value, OR, z-score and standard error (SE) calculated for each of the genotyped SNPs using a logarithmic regression model and correcting for ten principal components. The statistic z-scores from the genotyped SNPs were used in the direct imputation of summary statistics (DIST) method to impute the z-scores for allele 1 (reference allele) of the missing SNPs based on the correlation in linkage disequilibrium (r^2) from the 1000 Genome Project Version 3 (Lee *et al.* 2013). Imputation was performed genome-wide for all autosomes and the results were filtered based on the quality of the imputation (>0.8). Calculation of the ABF were performed using Wakefield approximation for a 2Mb window around the GWAS lead SNP of each locus of interest. This approximation was applied under the priors of (1) normally distributed OR with mean and variance (σ^2); (2) the greater the variance the bigger the obtained size effects will be; (3) mean is 0 and variance is fixed to 0.2 (accepted variance

for GWAS studies). In this approach ABF was calculated using effect size (β) and standard error (SE). For the imputed SNPs, SE is calculated as a function of the allele frequencies and β is determined using the z-scores of each of the interrogated SNPs as: β=z-score*SE. It is important to note that step-wise conditional analysis is not performed when using summary statistics since new association analysis cannot be conducted in absence of genotyping data. PP for the SNPs in a particular window (signal) were calculated and ranked to set the threshold of the 90 % credible set of SNPs.

PsA fine-mapping using Immunochip genotyping data

Fine-mapping was performed for a number of non-MHC PsA Immunochip GWAS susceptibility loci using the UK cohort (1,103 patients and 8,900 controls) from Bowes and colleagues PsA Immunochip (Bowes *et al.* 2015a). Access to the data post-quality control was kindly provided by Dr Anne Barton (University of Manchester). PCA was performed using only pruned SNPs with flashpca (Abraham and Inouye 2014) and the first ten calculated PCs were used as covariates in the association analysis to correct for population stratification.

For each of the fine-mapped loci a 2Mb window around the lead SNP was defined and SNPs extracted from the data using PLINK 1.9 (Chang *et al.* 2015). Phasing of the genotype data was performed with SHAPEIT (Delaneau *et al.* 2012) and used to impute missing genotypes with IMPUTE2 (Howie *et al.* 2009) and the 1000 Genomes Project Version 3 as the reference panel (October 2015 release). SNPs for which imputation was not successful in at least 70% of the samples (info-score<0.7) were filtered out using QCtool. The association and conditional analysis was conducted using a Bayesian additive model implemented with SNPTEST and including the previously calculated PCs as covariates (Burton 2007). Bayes Factors (BF) were calculated for the lead signal and step-wise conditional analysis was performed if $\log_{10}BF$ >3. Credible sets of

SNPs containing the variants likely to explain 50% and 90% of that association were identified for each of the signals in the locus, along with their corresponding posterior probabilities (PP) as further detailed in Bunt and colleagues (Bunt *et al.* 2015).

2.4.6 Mass cytometry data analysis

Mass cytometry analysis was performed by Dr Nicole Yager. Cell populations in synovial fluid and peripheral blood at 0h and 6h of BFA incubation were identified using manual annotation bsed on expression of surface markers $CD14^+$ $CD3^-$. For the $CD14^+$ monocyte population identified within each tissue at 0h and 6h, the percentage of TNF-α, MCP-1, IP-10 and osteopontin positive staining cells were calculated as the difference between 6h and 0h. Significance of the differences in the percentage of positive staining cells for each cytokine between synovial fluid and peripheral blood isolated $CD14^+$ monocytes was determined using Wilcoxon signed-rank test.

Chapter 3

Establishing laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples

3.1 Introduction

3.1.1 Principle of ATAC-seq and compatibility with clinical samples

Several techniques including DNase-seq, FAIRE-seq and MNase-seq have been used during the last few decades to map the accessible genome in different cell lines and some abundant sources of primary cells (reviewed in Chapter 1). All these techniques require a large number of cells as input material, making them unsuitable for use in a clinical setting. The publication of ATAC-seq represented a revolution in the field to interrogate chromatin accessibility. ATAC-seq uses a hyperactive modification of the bacterial transposase Tn5 to perform simultaneous fragmentation and insertion of synthetic oligonucleotides (adapters) into native chromatin from 50,000 cells and also at single-cell resolution (Buenrostro *et al.* 2013; Buenrostro *et al.* 2015). The Tn5 reaction incorporates, in a non-strand specific manner, adapters containing the complementary sequences to the i5-R1 and i7-R2 elements required for Illumina NGS. ATAC-seq provided a fast two-step protocol, not requiring cross-linking, enzyme titration or sonication, that was able to profile

nucleosome-free DNA (fragments ≤150bp) and DNA spanning nucleosomes (fragments >150bp). ATAC-seq data can also be used to identify TF foot-printing as well as nucleosome positioning in the genome. This technique opened a new avenue to interrogate the chromatin landscape in clinical samples with limited input material as well as in rare cell populations, with a shorter preparation and results turn-over time.

3.1.2 ATAC-seq limitations and advances in optimisation

Despite these advantages, early data from ATAC-seq revealed two major limitations, namely, a high percentage of mitochrondrial DNA tagged by the Tn5 enzyme, and insufficient sensitivity to detect all the accessible regions, partly due to high background noise (Corces *et al.* 2016; Sos *et al.* 2016). An optimised version of the protocol specifically for haematopoietic cells, named Fast-ATAC, replaced the NP-40 detergent used in ATAC-seq with digitonin. This prevents solubilisation of the mitochondrial membrane, and performs lysis and transposition in a single step. This modification efficiently reduced the percentage of mitochondrial reads to approximately 10% and increased the signal at annotated TSS, using only 5,000 cells as input material (Corces *et al.* 2016). Optimisation of the ATAC-seq protocol for keratinocytes was also published by Bao and colleagues, where they performed the two ATAC-seq steps directly on the 96-well plate containing adherent NHEKs using an increased concentration of Tn5 in the transposition reaction (Bao *et al.* 2015).

The third generation of ATAC (in this thesis ATAC will be used to refer to the technique regardless of the specific protocol), known as Omni-ATAC, was released in 2017 and offered a generic version of the protocol optimised to yield high quality data in any cell type and fresh or frozen tissue (Corces *et al.* 2017). The Omni-ATAC protocol consisted of lysis, wash and transposition steps. In addition to the NP-40 and digitonin, used by the previous ATAC protocols, Omni-

ATAC also included Tween-20 in the lysis buffer to improve cell permeabilisation. Comparison of Omni-ATAC with ATAC-seq and Fast-ATAC data demonstrated a higher variability in sample quality and sensitivity in the latter two protocols (Corces *et al.* 2017). Moreover, Omni-ATAC achieved greater signal-to-noise ratios by modifying the transposition buffer. Notably, this versatile protocol represented a particular improvement in keratinocytes with data demonstrating the inability of ATAC-seq and Fast-ATAC to yield good quality data in this cell type (Corces *et al.* 2017).

3.1.3 Challenges of ATAC data analysis

Although some guidelines for DHS data analysis were available, the release of the new ATAC methods also led to a need to adapt and develop additional tools and strategies for chromatin accessibility data analysis. In contrast to DNase-seq or FAIRE-seq data starting from high number of cells (minimum of $10x10^6$ cells), ATAC-seq and Fast-ATAC showed lower signal-to-noise ratios and higher variability across samples. This required appropriate implementation of quality control measurements in order to confidently identify good quality samples prior to downstream analysis. Regarding peak calling in ATAC, different algorithms have been applied, with MACS2 being preferred by the majority of studies, including ENCODE (`https://www.encodeproject.org/atac-seq/`)(Table 3.1). The criteria for filtering out poor quality peaks in ATAC is another critical aspect, particularly for the libraries at the lower end of the quality spectrum. False discovery rate (FDR) has been the most widely applied criterion except for ENCODE data, where technical replicates are generated and irreproducible discovery rate (IDR) analysis is used to identify robust peaks (`https://www.encodeproject.org/atac-seq/`) (Table 3.1).

Table 3.1: Summary table of ATAC methodology analysis for peak calling, filtering and differential analysis. NA indicates the study did not perform or detail that aspect of the analysis. Alasoo *et al.* 2018 was published in bioRxiv in 2017 and access to Thurner *et al.* 2018 work was also available earlier through collaboration in the WHG. Thus both publications were considered at the time of establishing the pipeline. * Access to ENCODE ATAC pipeline at `https://www.encodeproject.org/atac-seq/`.

Publication	Peak calling and filtering	Consensus peak list	Differential analysis
Corces *et al.* 2016	MACS2 (-nomodel,-shift -75, –extsize 150), peak summit extension +/−250bp, rank summits by p-value	Maximally significant non-overlapping peaks.	Quantile normalisation and unsupervised hierarchical clustering.
ENCODE*	MACS2 -nomodel, pairwise IDR analysis, filtering IDR<10%	IDR peaks: those consistent based on rank procedure across replicates or pseudoreplicates	NA
Thurner *et al.* 2018	MACS2 (-nomodel –q 0.01)	Merging all filtered called peaks from the different cell types.	*De novo*: DiffReps with fragment size 50bp.
Alasoo *et al.* 2018	MACS2 (-nomodel -shift -25 -extsize 50 –q 0.01	Union of peaks from all conditions present in at least three samples of the same condition.	Peak based: TMM normalisation and lima voom (FDR<0.1).

Qu *et al.* 2015	ZINBA PP>0.99 (Bayesian approach)	Merging of filtered peaks from each individual sample.	Quantile normalisation and peak based in house Pearson correlation method.
Rendeiro *et al.* 2016	MACS2 (-nomodel -extsize 147 -shift 73.5)	Merge of peaks from all samples in an iterative process including permutations	Peak based: quantile normalisation and Fisher exact text (FDR<0.05).
Scharer *et al.* 2016	HOMER (-style dnase)	Merge of all overlapping peaks between all samples using HOMER mergePeaks	Peak based: TMM normalisation and edgeR package (FDR<0.05).

The feasibility of generating ATAC data from low numbers of cells and clinical samples presents an opportunity to perform differential chromatin accessibility analysis between conditions, cell types or groups of patients and healthy control samples. The most common approach is a peak based strategy, which requires building a non-redundant and non-overlapping list of high quality peaks, counting the reads mapping to those locations and performing normalisation across samples before conducting differential analysis with microarray or RNA-seq based-methods. An alternative is known as the *de novo* approach, which has previously been implemented for ChIP data, and consists of using a sliding window to scan the genome and identify those regions showing read count differences between two groups of samples, avoiding peak calling bias (Shen *et al.* 2013).

3.1.4 The challenge of working with clinical samples

The opportunity to apply epigenetic assays to clinical samples has also highlighted a logistical problem. Often patient recruitment takes place at distant geographical locations or out of normal working hours. This requires the application of preservation methods that provide a snap shot of the *in vivo* cellular characteristics and avoid introducing confounders. The main methods to preserve cell structure and DNA integrity involve cryopreservation or DNA-protein fixative compounds such as formaldehyde. Regarding preservation of pure cell populations, a study in motor neurons demonstrated that slow-cooling using DMSO but not snap-freezing maintained intact cell nuclei and chromatin organisation and overall yielded comparable ATAC-seq data to those generated in fresh neurons (Milani *et al.* 2016). When working with mixed cell populations such as PBMCs, slow temperature cryopreservation with DMSO allows long term storage and also offers the flexibility of retrospective separation of distinct cell populations by FACS following thawing and recovery. However, in a mixed

population such as PBMCs some cell types are more sensitive to cryopreservation and that may lead to distinct alterations in the chromatin accessibility landscape and gene expression profile. In terms of fixatives, the Oxford Genomic Centre at the WHG had incorporated the use of dithio-bis(succinimidyl propionate) (DSP) to stabilise cell samples for single-cell transcriptomic applications demonstrating only moderate differences from fresh samples profiles (Attar *et al.* 2018). DSP is a reversible cross-linker of free amine groups that fixes proteins without damaging RNA and is compatible with microfluidics-based scRNA-seq systems, unlike formaldehyde fixation. DSP preservation does not require sample freezing after fixation and samples can undergo successful immuno-staining as well as FACS cell separation (Mueller *et al.* 2013).

3.1.5 Aims

The aim of this chapter is to establish the ATAC protocols in the laboratory, perform a thorough optimisation of the methodological and analytical tools required to study chromatin accessibility in clinical samples of interest, and to determine the suitability of relevant methods for sample preservation to overcome the inherent logistical constraints of working with clinical samples.

The specific aims of this chapter are:

1. To establish an in-house pipeline to analyse ATAC data including quality control measurements, peak filtering and a method for differential analysis in order to maximise the use of available samples.

2. To investigate the effect of transposition time on sample quality in the ATAC-seq protocol.

3. To validate the reduction of mitochondrial DNA and improvement of signal in Fast-ATAC protocol compared to ATAC-seq.

4. To test and optimise ATAC-seq protocols in order to adapt them to the mapping of the chromatin accessibility landscape in psoriasis skin biopsies and in cultured primary keratinocytes.

5. To determine the effect of cryopreservation and DSP fixation on ATAC-seq data quality and on the overall chromatin accessibility landscape of $CD14^+$ monocytes and total $CD4^+$ ($CD4^+$) cells.

3.2 Results

3.2.1 Establishment of an ATAC-seq data analysis pipeline

A robust ATAC-seq data analysis pipeline was required as this was a new methodology (Buenrostro *et al.* 2013) for which established pipelines were not available in the group at the time this work was undertaken. Such a pipeline requires appropriate quality control, peak calling and filtering, and a method for identification of differential chromatin accessibility between groups of samples.

The most appropriate criteria and parameters to implement were investigated using ATAC-seq data generated for paired $CD14^+$ monocytes and $CD4^+$ T cells from three healthy individuals (referred as CTL1-3). This data corresponds to the fresh samples generated to test the effect of cryopreservation and fixation in the chromatin landscape, further detailed in section 3.2.5 (Table 2.1) and was appropriate to establish quality control measurements and the differential analysis approach later implemented for the study of psoriasis and PsA chromatin accessibility landscape (Chapters 4 and 5).

Sample quality control

The variability in performance of ATAC experiments and protocols requires appropriate quality control of the samples before proceeding with downstream differential analysis. ATAC-seq fragment size distribution was analysed in each

library containing between 25 and 30 million reads. The observed fragment size distribution demonstrated nucleosome periodicity protecting the DNA during the transposition event (Figure 3.1A) and indicated chromatin integrity, one of the requirements for good quality ATAC libraries. All six libraries showed appropriate nucleosome periodicity (every ~200bp) up to 600bp, clearly distinguishing chromatin organisation into mono-, di- and tri-nucleosomes. Some variation in the relative intensity of nucleosome-free fragments (NFF) (≤147bp, approximately) compared to nucleosome-bound DNA was seen across samples. However, NFF were clearly distinguished in all of the samples, which is considered a compulsory feature for ATAC-seq libraries to pass quality control, according to ENCODE recommendations (`https://www.encodeproject.org/atac-seq/`)(Table 3.1).

Another quality control measurement investigated and also implemented in the pipeline was the enrichment of ATAC-seq signal over a random background of reads across all the TSSs identified for Ensembl genes (Figure 3.1B). This is a useful measure as nucleosome repositioning and an increase in chromatin accessibility occur at the TSS to allow TF binding and initiation of gene transcription. Fold-enrichment signals over the TSS ranged from 5 to 7 for the $CD4^+$ samples, and were much higher (17 to 20) in the $CD14^+$ samples. The lower sample quality of the $CD4^+$ compared to $CD14^+$ samples indicated by the TSS enrichment values were further evidenced by visualising the ATAC-seq read pile up at the promoters of the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and the NOP2 Nucleolar Protein (*NOP2*) genes, showing more background reads and lower signal for the $CD4^+$ samples (Figure 3.1C).

As part of the quality control assessment, the percentage of mitochondrial reads and the fraction of reads in peaks (FRiP) were also investigated (Table 3.2). FRiP score is an alternative to TSS enrichment for assessing the background

signal in different types of assays that are based on peak calling, including ChIP-seq.

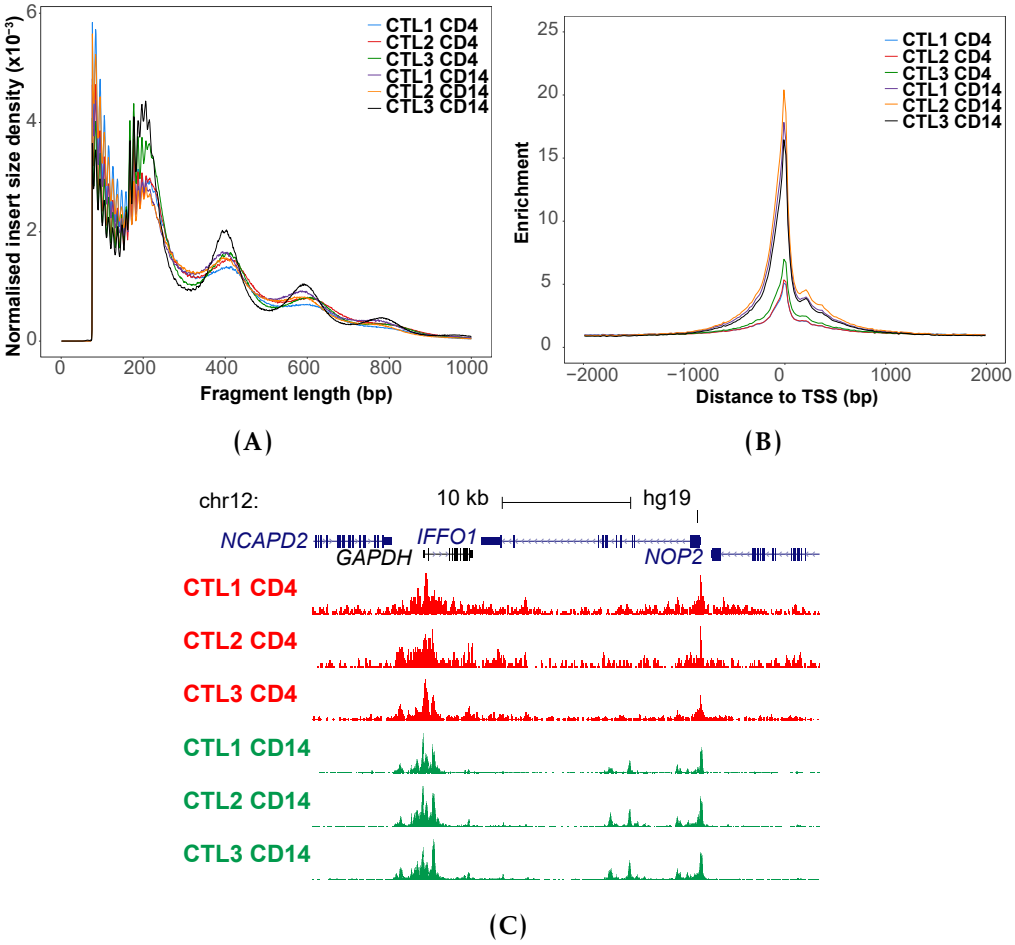


Figure 3.1: Measurements for quality control assessment in ATAC-seq samples. For each of the CD14$^+$ monocytes and CD4$^+$ samples used to establish the ATAC analysis pipeline quality control measures (A) density distribution of ATAC-seq fragment sizes, (B) enrichment of ATAC-seq fragments across the TSS of all Ensembl genes and (C) UCSC Genome Browser view illustrating the ATAC-seq normalised read density (y-axis) at the promoters of *GAPDH* and *NOP2* genes are shown. CD14$^+$ monocytes and CD4$^+$ tracks are colour-coded in green and red, respectively.

Positive correlation between the TSS fold-change enrichment and FRiP was observed and suggested that both are appropriate inter-dependent quality control measures to evaluate sample noise. For the six samples analysed here, mitochondrial content varied between (14.9-43.3%), was higher in CD14$^+$ than in CD4$^+$ cells and was not directly related to any of the other quality control measurements (Table 3.2). Therefore, mitochondrial reads in this range did

not appear to reflect sample quality and the main issue related to the need for deeper sequencing to achieve the desired number of non-mitochondrial reads for downstream analysis.

Sample	% mitochondrial reads	Fraction of reads in peaks
CTL1 CD4	14.9	9.8
CTL2 CD4	30.5	11.2
CTL3 CD4	28.8	11.6
CTL1 CD14	43.3	32.2
CTL2 CD14	36.8	57.0
CTL3 CD14	37.6	49.9

Table 3.2: ATAC-seq percentage of mitochondrial reads and fraction of reads in called peaks (FRiP). The percentage of mitochondrial reads was calculated over the total number of sequencing reads (before filtering). FRiP was calculated as the proportion of ATAC-seq fragments overlapping significant peaks with standard filtering for all the samples (FDR<0.01).

Both TSS and FRiP are appropriate signal-to-noise measures, being the recommended threshold values by ENCODE and Alasoo *et al.* 2018 (previously published in bioRxiv in 2017) 10-20% for FRiP and 6-10 for TSS enrichment. Importantly, ENCODE has recently prioritised the use of TSS over FRiP as a more stable measure to determine noise and therefore this was the chosen measure for this thesis. In summary from this analysis, all six samples showed appropriate ATAC-seq patterns of fragment size distribution, FRiP and TSS, with the exception of $CD4^+$ CTL1 and CTL2, being borderline for the 6 fold-enrichment ENCODE TSS recommended threshold.

Peak calling and filtering

Next, criteria for peak calling and filtering were investigated. Although different peak callers have been used to analyse ATAC-seq data, MACS2 has been the preferred algorithm by ENCODE and most publications at the time of this thesis (Table 3.1). MACS2 was initially developed for ChIP data, but has also been used for DHS and ATAC-seq disabling the model option and manually

setting the shift (–shift) and extension size (–extsize) parameters, which refer to the number of bp and direction for the reads to be shifted and the number of bp for them to be extended, respectively. Since the –extsize should correspond to the average fragment size, this was set to 200bp, the average fragment size calculated for the ATAC-seq libraries in this project. The –shift was set to -100, as it is recommended to be -1/2 of the fragment size when analysing chromatin accessibility data so that the summit of the peak is located at the Tn5 insertion site.

A systematic analysis of the effect of sequencing depth and the sample quality on peak calling was conducted to better understand the effect of both variables on the downstream analysis. For each of the six samples, random sub-sampling of reads was performed in every 5 million increment, ranging from 5 to 30 total million reads, followed by peak calling with arbitrary filtering for false discovery rate (FDR)<0.01. The number of called peaks passing filtering showed a steady increase with read depth (Figure 3.2A), beginning to plateau at approximately 25 million reads (Figure 3.2B). Moreover, lower number of peaks were detected in $CD4^+$ samples compared to $CD14^+$ monocytes when using standard FDR<0.01 filtering, highlighting the influence of sample quality on the total number of called peaks. Interestingly, sample quality as measured by FRiP (which relies on peak calling) showed very small changes with read depth and was stable from 15 million reads onwards for all six samples (Figure 3.2C), similarly to TSS (Figure 3.2D). This confirmed that measurement of sample quality using FRiP or TSS was not affected by the sequencing depth.

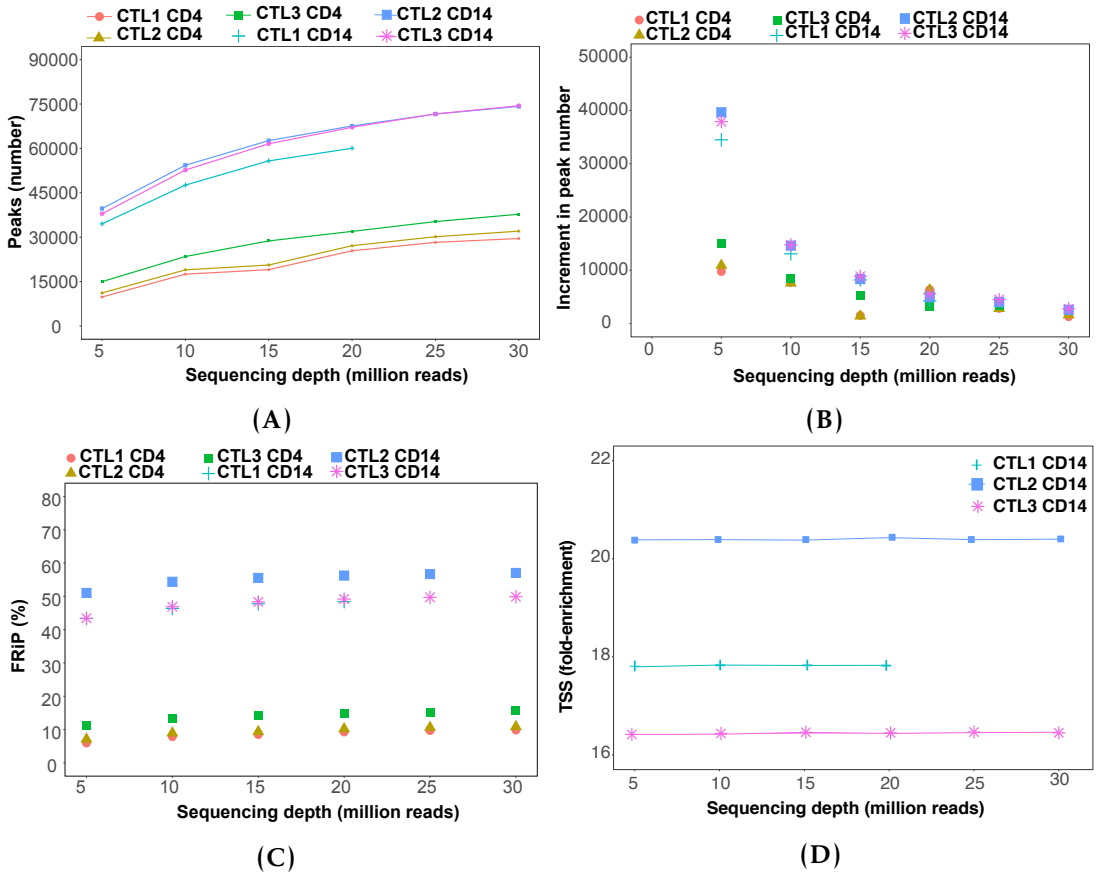


Figure 3.2: FRiP and peak calling at different sequencing depths in ATAC-seq libraries. For a series of sequencing depths (from 5 to 30 million reads after filtering) representation of (A) number of called peaks (standard filtering using FDR<0.01, (B) the increment on the number of called peaks and (C) FRiP and (D) TSS as a function of the sequencing coverage in $CD4^+$ and/or $CD14^+$ samples.

For peak filtering, an arbitrary FDR<0.01 in MACS2 is typically used (Table 3.1) but may not remove low quality peaks equally successfully in lower quality samples and does not take into account the reproducibility of the called peaks. Thus IDR was used to experimentally identify the most appropriate p-value threshold to filter the called peaks in each individual sample. Filtered reads from each sample were partitioned in half to create two pseudoreplicates, peaks were called in each pseudoreplicate and the percentage of peaks sharing IDR rank position when filtered at decreasing p-values was calculated (Figure 3.3A and B). This strategy was tested across a range of total read counts (as above) to determine the effect of sequencing depth on the suitability of this peak calling filtering approach. The optimal p-value giving the largest percentage of IDR

shared peaks between the two pseudoreplicates varied more erratically when the sequencing depth was lower than 10 million reads (Figure 3.3A and B), suggesting this analysis was not appropriate for lower read depths. This variation was more pronounced and extended in $CD4^+$ T cells, which had lower TSS values compared to $CD14^+$ monocytes (Figure 3.3A and B). The shape of the curves were also influenced by the sample quality. For appropriate sequencing depth (minimum of 15-20 million reads), the $CD14^+$ monocytes (TSS enrichment $\geq$10) samples presented a profile reaching a single maximum of shared IDR peaks for a particular filtering p-value (Figure 3.3B), which was -log10 p-value 8 in all three samples (data not shown). In contrast, the same analysis in $CD4^+$ samples reached two local maxima (Figure 3.3A).

Filtering the $CD4^+$ peaks at the p-value of the first local maximum (10^{-11}) reduced the percentage of peaks overlapping regions suggestive of noise (e.g repressed and quiescent), and increased the proportion of accessible regions located at active TSS when compared to using the list of significant peaks filtered based on FDR<0.01 (Figure 3.3C). In summary, this IDR analysis provided a systematic method to identify an optimum p-value with which to perform sample-specific filtering of technically reproducible peaks when the sequencing depth was over 10 million reads. These resulting filtered peaks will be used downstream to build the consensus list of ATAC peaks across all the samples and perform differential chromatin accessibility analysis.

Differential chromatin accessibility analysis

A peak-based approach using the number of reads overlapping the peaks included in the consensus list of ATAC peaks (CP_all) was implemented to perform differential analysis. One of the main limitations of the ATAC-seq and Fast-ATAC protocols (discussed in Sections 3.2.2 and 3.2.3) is the background signal. Therefore, an empirical cut-off was identified to minimise the impact

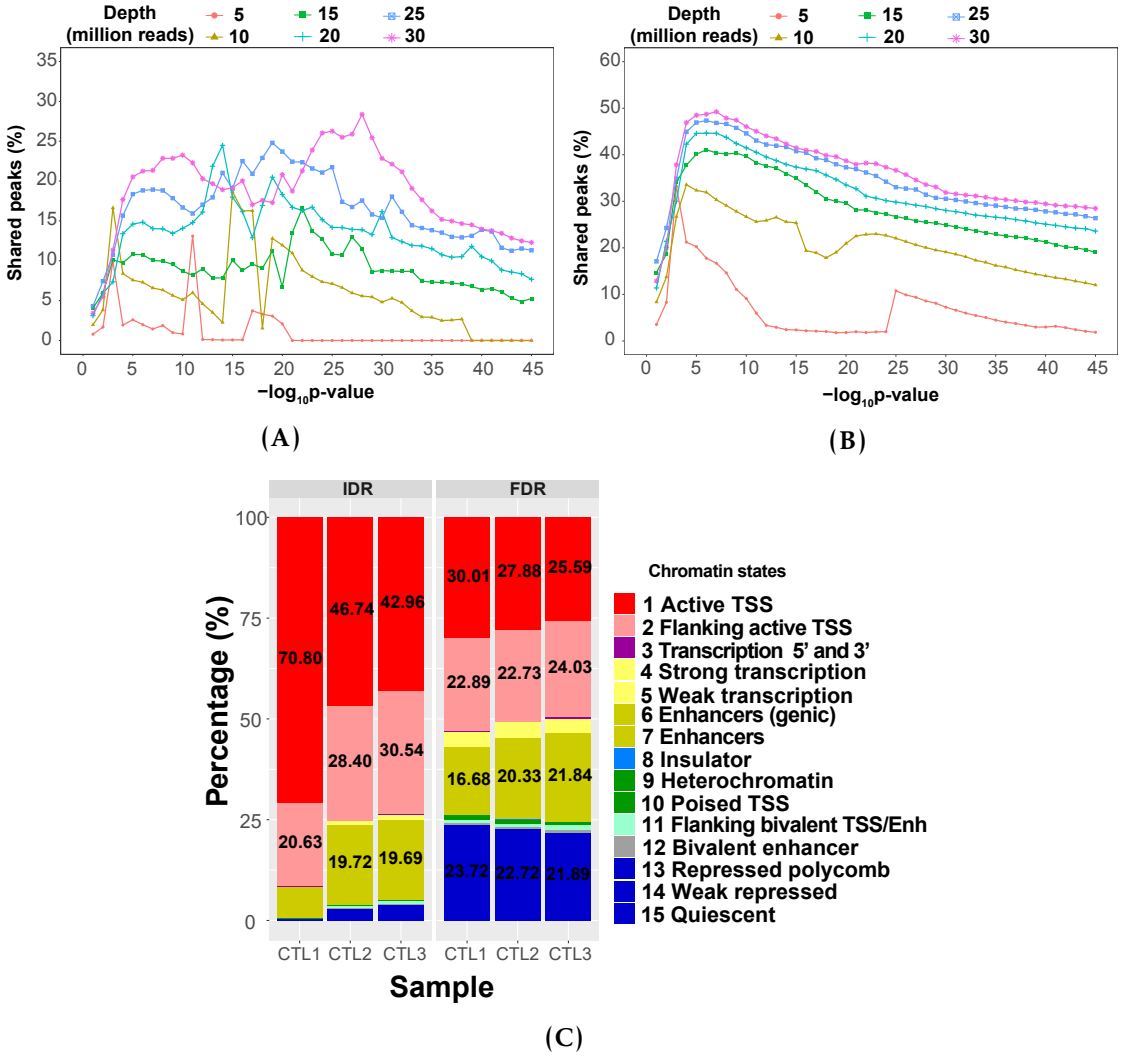


Figure 3.3: Peak calling filtering using IDR analysis in ATAC-seq samples. For each of the sequencing depths tested (from 5 to 30 million reads after filtering), an illustration of the percentage of peaks sharing IDR rank between the two pseudoreplicates is shown when using different p-value filtering thresholds in CTL2 (A) $CD4^+$ T cells and (B) $CD14^+$ monocytes, two representative samples differing in quality for this analysis. (C) Annotation (as percentage of total) of the CTL1, CTL2 and CTL3 $CD4^+$ ATAC-seq peaks filtered for FDR<0.01 or optimal p-value from the IDR analysis (p-value=10^{-11}) with the corresponding cell-type specific Roadmap Epigenomics Project chromatin segmentation map.

of background read counts on the peaks included for the differential analysis (Xinmin *et al.* 2005; Jonker *et al.* 2014). Moreover, due to lack of consensus in terms of normalisation and differential analysis methods in ATAC-seq (as reviewed in Table 3.1), two strategies were tested (Figure 3.4). The first involved quantile normalisation of the read count matrix and differential chromatin accessibility analysis using limma voom lineal model. The second strategy consisted of DESeq2, which performs internal normalisation calculating size factors using median of ratios and differential analysis based on negative binomial distribution.

The CP_all matrix includes all peaks called significant (present) in at least 30% of the six samples used in this section (Figure 3.4 Step 1). This consensus list of ATAC peaks can include what may be considered background counts, i.e. reads overlapping regions not considered significant in a particular sample (absent) but called as significant (and therefore present) in other samples. The number of reads overlapping peaks in samples where that peak was not called significant (absent) were used to generate a density distribution plot of the background reads (Figures 3.4 Step 2). From this plot, a sequence of twenty cut-offs was defined, with each value representing the number of counts shown by an increasing percentage of the total absent peaks (Figure B.1) and each cut-off (corresponding to a minimum number of mapping reads) was applied to filter out peaks from the CP_all raw count matrix (not normalised) with values lower than the cut-off in more than three samples (Figure 3.4 Step 3). A filter of three samples was chosen as it corresponds to the smallest number of biological replicates in this particular experimental design and ensures that peaks absent in one condition but not in the other are retained. The resulting reduced matrix of low-noise peaks for each cut-off value was then normalised using quantile or DESeq2 (normalisation using size factors as detailed in Love *et al.* 2014) and used to conduct differential analysis with limma voom or DESeq2, respectively.

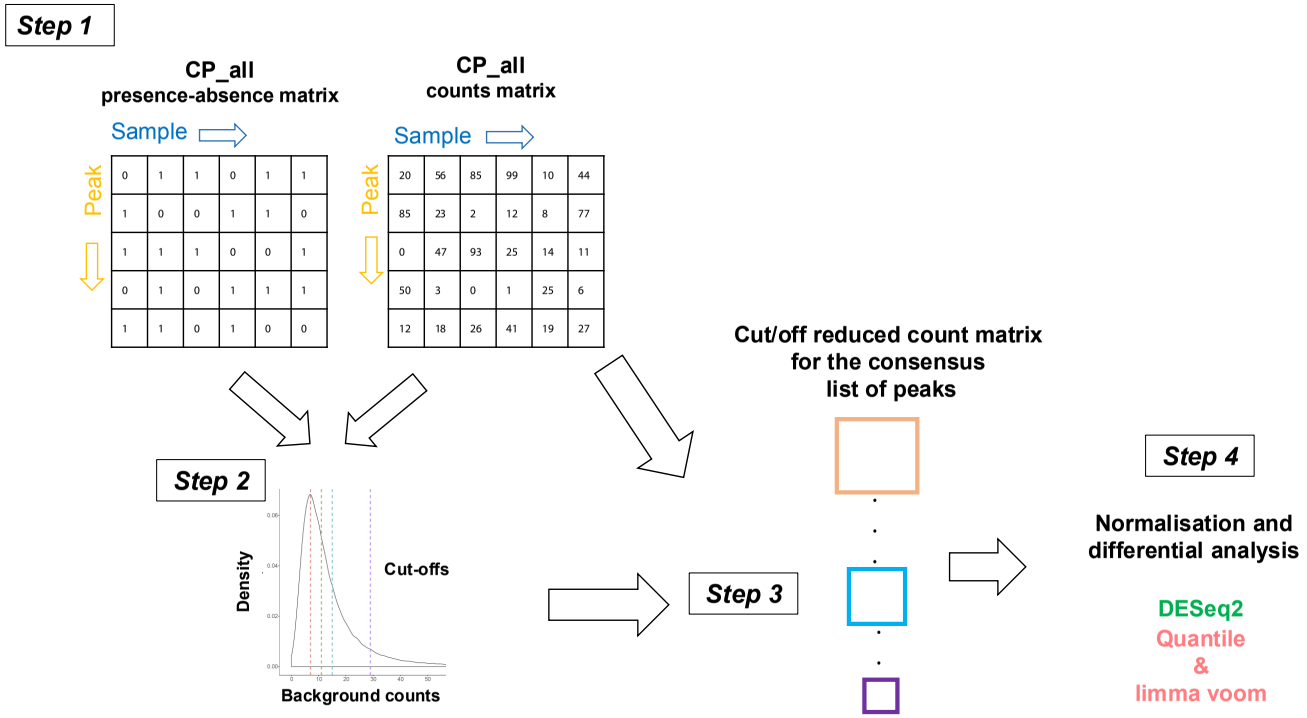


Figure 3.4: Work flow illustrating the strategy to account for ATAC background noise prior to differential analysis. A combined consensus list of ATAC peaks (CP_all) was built as explained in Chapter 2, including all the significant peaks that were shared by at least 30% of the samples included in the analysis. The peaks were further transformed to obtain non-overlapping 500bp homogenous entities. The data was then represented by two matrices (Step 1). The first is the significance matrix with each cell indicating the significance of a peak (in rows) in each sample (in columns) as presence (1, significant) or absence (0, non-significant). The second matrix is the count matrix storing the number of reads mapped to the peak (in rows) for each sample (in columns). A density distribution plot of the read counts from the peaks absent (0) in each sample was used to define a sequence of twenty cut-offs. These cut-offs correspond to the number of counts showed by a particular percentage of the absent peaks and can be considered background counts (Step 2). The defined cut-offs were used to filter out peaks from the CP_all and generate a series of reduced matrices (Step 3) that were tested for normalisation and differential chromatin accessibility analysis by two methods (quantile & limma voom or DESeq2) (Step 4).

The results from this were used to determine an appropriate threshold for filtering background reads, as well as to compare the two analytical methods.

Both normalisation methods performed appropriately across all filter values and samples, with the quantile normalisation showing slightly better consistency across the two groups (Figure 3.5A and B). Differential chromatin accessibility analysis using limma voom showed a greater number of significant (FDR<0.01 and fold change>1.5) differentially accessible regions (DARs) compared to DESeq2 across all filtering cut-offs. For both approaches, the number of DARs progressively decreased as the filtering cut-off increased from 75% onwards (Figure 3.5C).

This suggested a reduction in the number of false positive hits resulting from peaks with high background reads. However, increasingly stringent filtering would also remove true positives so an intermediate value of 80% was chosen here. In future analyses this value may vary depending on the noise inherent to the experiment.

The vast majority of the 19,855 DARs called as significant using the more conservative method (DESeq2) were recapitulated by limma voom at the same significance threshold (FDR<0.01 and fold change>1.5) (Figure 3.6A). FDR rank revealed that out of the first 19,855 limma voom DARs 18,768 matched the rank of those retrieved by DESeq2. Moreover, very significant positive correlation was found between fold changes for all the regions included in the 80% cut-off reduced matrix reported by the two methods (R=0.999, p-value=2.2^{-16}) (Figure 3.6B). Overall, this suggested that the differences in the number of significant DARs reported by limma voom and DESeq2 could partly be driven by differences in the models used by the two methods to estimate dispersion of counts.

Lastly, the significant DARs identified by DESeq2 and filtered for FDR<0.01 and fold change>1.5 were divided in those more accessible in $CD14^+$ monocytes (open in monocytes) or $CD4^+$ (open in $CD4^+$). Enrichment analysis for cell

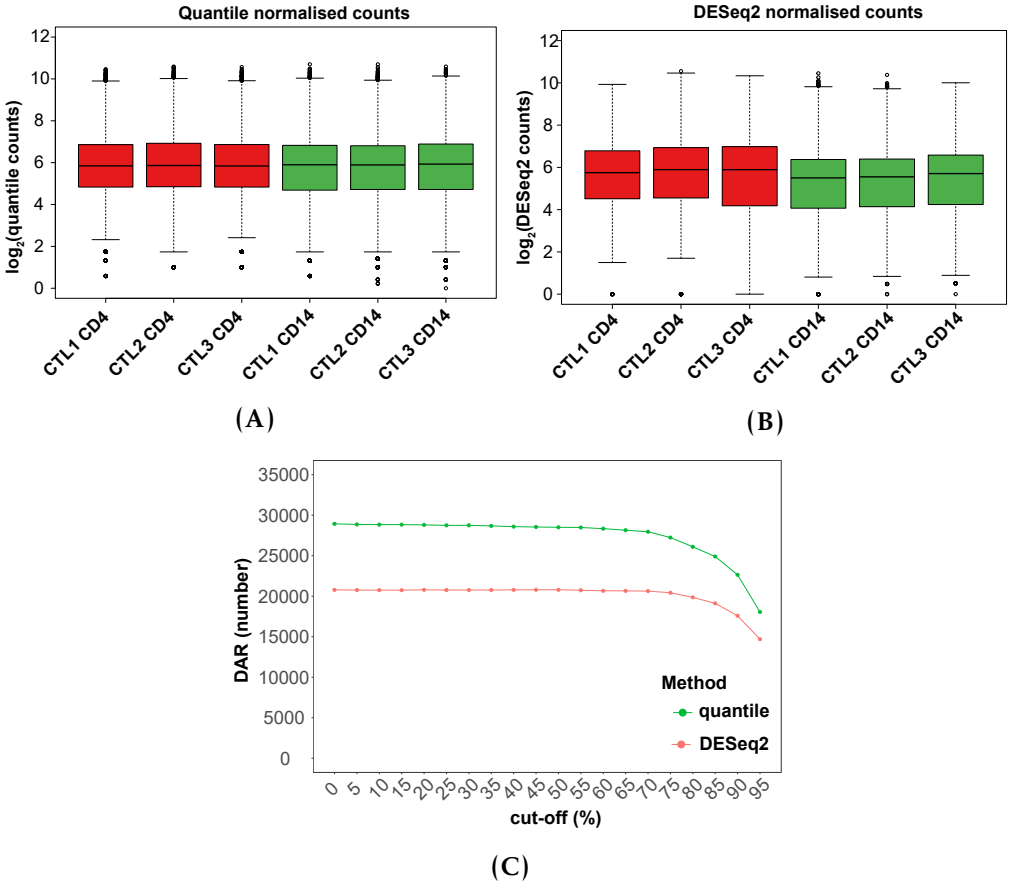


Figure 3.5: Normalisation and differential chromatin accessibility analysis for different cut-offs using quantile normalisation& limma voom and DESeq2. Boxplots representing the $\log_2$ read counts for each of the peaks from the unfiltered consensus list of peaks normalised by (A) quantile or (B) DESeq2 in the three $CD14^+$ monocytes and total $CD4^+$ healthy control paired samples. (C) Representation of the number of significant DARs (FDR<0.01 and fold change>1.5) detected in the differential analysis by limma voom (using quantile normalisation) or DESeq2 when using a sequence of empirical background noise cut-offs to filter the consensus list of peaks.

type-specific epigenetic features, including FANTOM5 eRNAs, histone marks and DHSs, was conducted in each of the two groups of DARs. The $CD4^+$ open DARs were enriched for T cell eRNAs, $CD4^+$ H3Kme1 and H3K27ac and $CD14^+$ DHSs (Figure 3.7). Conversely, the top enriched features for $CD14^+$ monocytes open DARs included eRNAs, H3K27ac and DHSs in monocytes. Overall, this enrichment analysis confirmed the ability of this differential analysis method to identify significant and robust DARs that highlight cell-type specific regulatory regions for $CD14^+$ monocytes and $CD4^+$ cells.

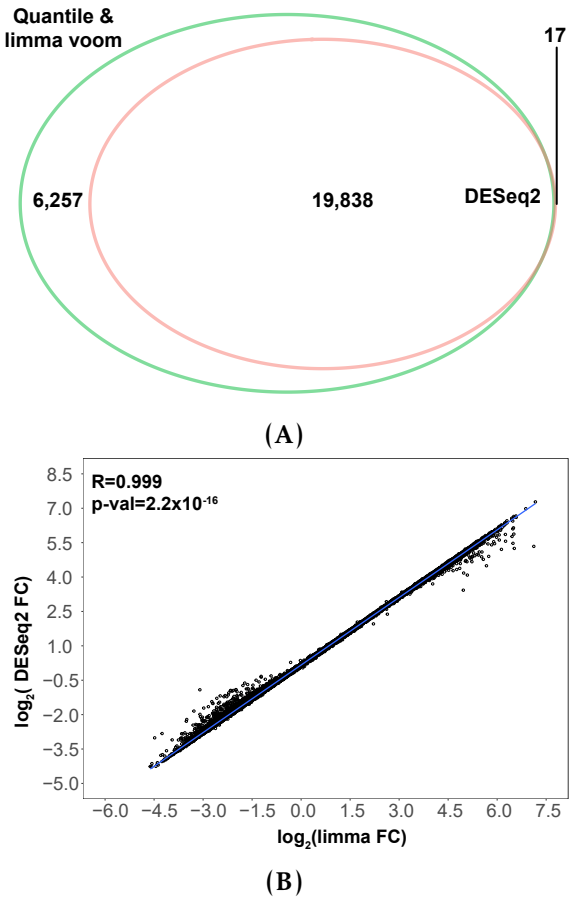


Figure 3.6: Comparison of the DARs identified by differential analysis using limma voom or DESeq for the consensus peak list filtered at the optimal 80% cut-off. (A) Venn diagram illustrating the common and distinct significant (FDR<0.01 and fold change>1.5) DARs identified by differential analysis in the filtered consenus list of ATAC peaksfor the 80% optimal cut-off using limma voom or DESeq2. (B) Representation of the correlation between limma voom and DESeq2 $\log_2$ fold changes (no FDR filtered) in each the peaks from the fconsensus list of peaks filtered at the 80% optimal cut-off. Pearson correlation coefficient (R) and significance (p-value) are indicated. Limma voom was applied to quantile normalised count data.

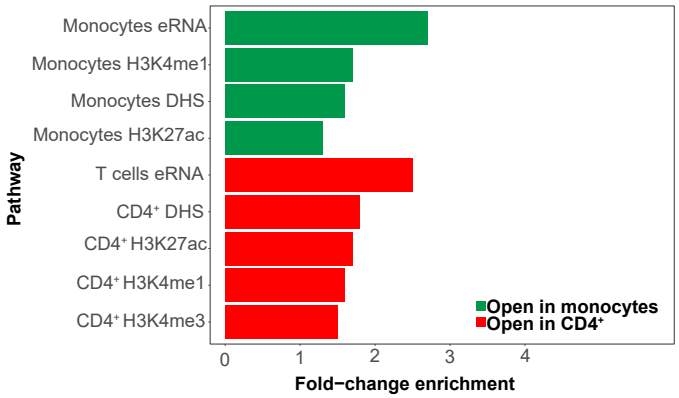


Figure 3.7: Enrichment analysis for the significant DARs identified by DESeq2 between CD14$^+$ monocytes and CD4$^+$ cells. Barplot representing the fold change for the top significantly enriched (FDR<0.01) FANTOM5 eRNAs, histone marks and DHSs from Blueprint. Enrichment analysis was performed separately for the significant (FDR<0.01 and fold change>1.5) DARs more accessible in CD14$^+$ monocytes (open in monocytes) or CD4$^+$ (open in CD4$^+$).

3.2.2 Assessment of ATAC-seq transposition times in relevant cell types

The duration of the transposition reaction was further investigated for the main immune cell types of interest for this thesis. ATAC-seq was performed for three different transposition times (20, 30 and 40 min) in CD14$^+$ monocytes, CD4$^+$, total CD8$^+$ (CD8$^+$) and CD19$^+$ cells in one healthy control sample included in cohort 1A from Chapter 4 (Tables 2.1 and 4.3). The impact of transposition times in a number of ATAC related readouts was explored. The duration of the transposition, the concentration of Tn5 and the number of cells can have an effect on the relative proportion of nucleosome-free vs nucleosome-bound (mono-nucleosomes and beyond) regions tagged by the adapters. Ideally, the transposition reaction should yield NFF (≤150bp), where TF and other proteins bind, without losing mono- and di-nucleosome fragments characteristic of the ATAC libraries. In this case, all three transposition times produced appropriate fragment size distributions which recapitulated the ATAC characteristic nucleosome periodicity pattern (Figure 3.8A). This was reflected

by the ratio of NFF vs mono-nucleosome bound fragments (mNBF) (151-250bp, the most abundant of the nucleosome-bound fraction) being around 1, which indicates at least similar abundance of both types of fragments. The variation of this ratio with the transposition times showed very moderate changes and was heterogenous across cell types (Figure 3.8B). For example, $CD8^+$ presented the greatest NFF/mNBF ratio for 20 min of transposition whereas the NFF/mNBF ratio reached a maximum at 40 min for the $CD14^+$ monocytes. Overall, the tested transposition times in this experiment showed neglectable impact on the NFF/mNBF ratio, which was lower than 7 (cut-off to designate ATAC libraries lacking appropriate fragment size distribution according to Alasoo and colleagues (Alasoo *et al.* 2018)) in all instances.

This variation in NFF/mNBF did not significantly impact the signal-to-noise ratio, with the largest TSS enrichment values corresponding to 30 min in three cell types (Figure 3.8C and D) and the major differences found across cell types. Differences between TSS fold-enrichments at 30 and 40 min were very modest and did not differ in more than four units in $CD8^+$ cells, where the largest variation was found. Before performing this formal comparison of transposition times, some sample recruitment had already been conducted using ATAC-seq with transposition for 40 min, as it was found to be the most appropriate condition based on the relative abundance of DNA fragment sizes profiles from the pre-sequencing library quality control. Although this analysis suggested 30 min as the best condition for the majority of the cell types tested in this single repetition, the differences between 30 min and 40 min were minor and therefore for consistency across the cohort 40 min was used for all other patient samples using ATAC-seq.

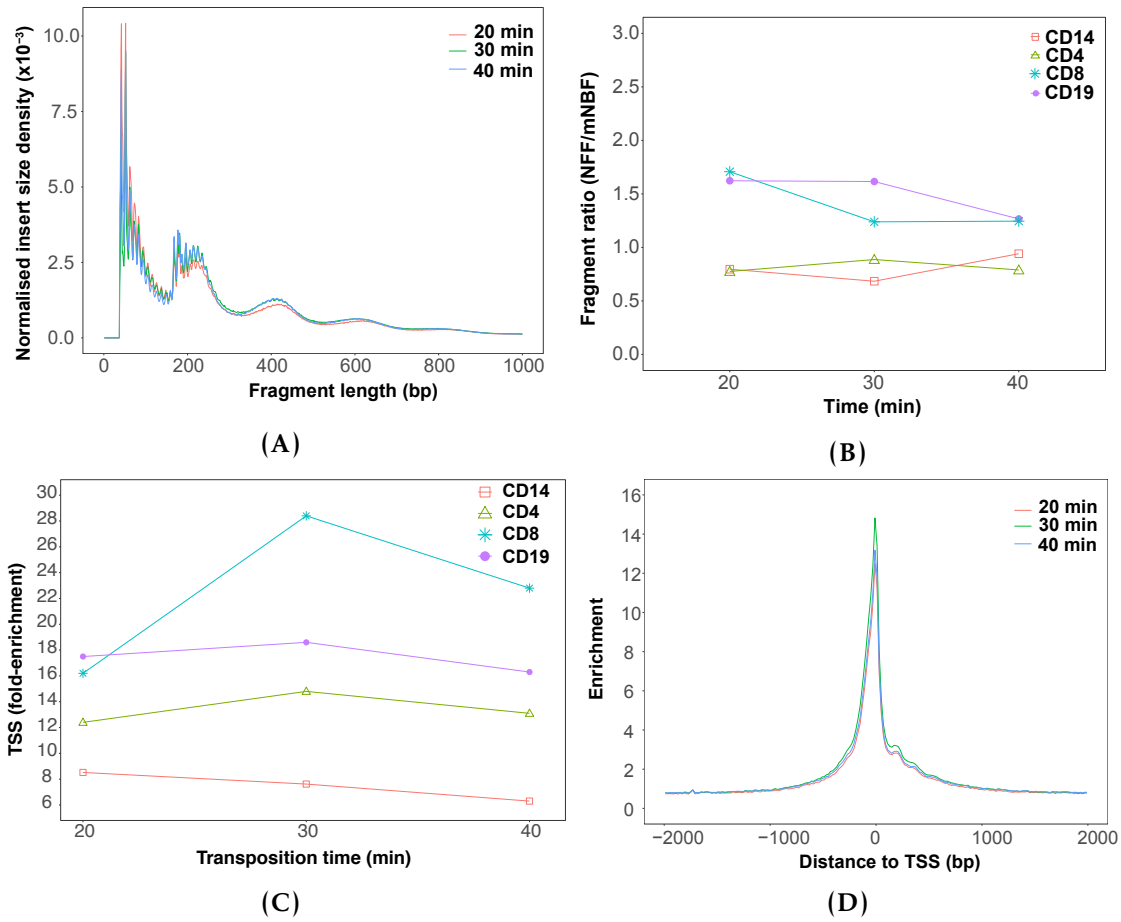
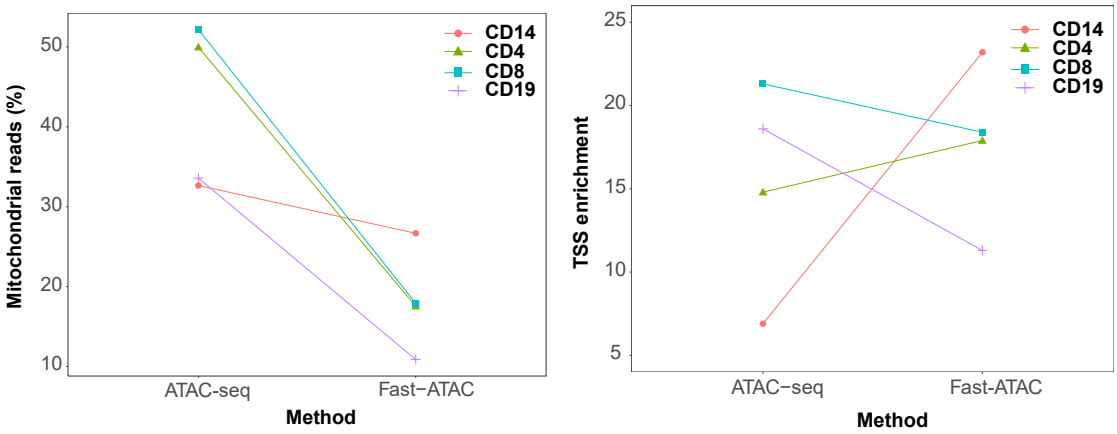


Figure 3.8: Assessment of the effect of transposition times on the ATAC-seq measurements. (A) Representative plot of the ATAC-seq fragment sizes density distribution following 20, 30 and 40 min of transposition in $CD8^+$ cells. Changes across different transposition times in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells of (B) the ratio between ATAC-seq nucleosome-free fragments (NFF) (fragments ≤150bp) and mono-nucleosome bound fragments (mNBF) (>151bp) and (C) TSS fold-enrichment. (D) $CD4^+$ cells enrichment of ATAC-seq fragments for different transposition times across the TSS of all Ensembl genes.

3.2.3 Comparison of ATAC-seq with Fast-ATAC protocol

An improved Fast-ATAC protocol from Corces and colleagues (Corces *et al.* 2016) was reported whilst the work in this thesis was being undertaken and was compared with the ATAC-seq protocol from Buenrostro and colleagues (Buenrostro *et al.* 2013). The aim was to confirm the two main reported advantages of Fast-ATAC, namely reduction of mitochondrial reads and signal enhancement, before implementing it as the replacement for the current ATAC-

seq protocol in use to process patient and control samples (Chapter 4 cohort 1A in Tables 2.1, 4.2 and 4.3). Fast-ATAC was conducted in one healthy volunteer sample included in cohort 1B from Chapter 4 (Tables 2.1 and 4.3). Fast-ATAC was specifically optimised for haematopoietic cells and used 30 min of transposition. Thus, for consistency, Fast-ATAC samples were here compared to the ATAC-seq library from Section 3.2.2 transposed for 30 min in the same four cell types.



(A) **(B)**

Figure 3.9: Differences in mitochondrial DNA abundance and TSS enrichment between ATAC-seq and Fast-ATAC protocols. Representation of changes in (A) percentage of mitochondrial reads and (B) TSS fold-enrichment between ATAC-seq and Fast-ATAC libraries for $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. Fast-ATAC protocol was specifically optimised by Corces and colleagues for hematopoietic cells and recommends 30 min of transposition (2 and Corces *et al.* 2016). Therefore, for consistency, Fast-ATAC samples were compared to the ATAC-seq libraries from Section 3.2.2 transposed for 30 min.

The percentage of mitochondrial reads in the Fast-ATAC libraries was lower for all cell types analysed compared to ATAC-seq (Figure 3.9A). $CD4^+$ and $CD8^+$ cells showed the largest reduction from 50% to less than 20% when using ATAC-seq and Fast-ATAC, respectively. With regard to the background signal, differing trends were observed across cell types (Figure 3.9B). Improvement in TSS enrichment was observed for $CD14^+$ monocytes and $CD4^+$ T cells whereas $CD8^+$ and $CD19^+$ cells showed a reduction when using Fast-ATAC compared to ATAC-seq in this single repetition. Overall, the large decrease in mitochondrial reads together with the reduced duration of the experimental protocol supported

the replacement of ATAC-seq by Fast-ATAC for future patients recruitment (see Chapter 4).

3.2.4 Limitations of ATAC-seq and Fast-ATAC to assess chromatin accessibility in keratinocytes

A particular aim of this thesis was to characterise the regulatory landscape in psoriatic keratinocytes, one of the most relevant cell types in this pathophysiology. In order to assess the feasibility of using the ATAC-seq protocol from Buenrostro and colleagues (Buenrostro *et al.* 2013) (referred to as ATAC 1 in this subsection), epidermis from a psoriatic lesional skin biopsy was isolated, digested with trypsin and filtered through a cell strainer to ensure a single-cell suspension, as detailed in Chapter 2. Cell suspensions obtained from skin biopsies using trypsinisation of the epidermal layer are enriched in keratinocytes, constituting approximately 90% of the cells (Haftek *et al.* 1986). Approximately 50,000 cells from the suspension were counted and ATAC 1 was performed for two different transposition times (30 and 40 min).

Pre-sequencing library quality control to assess the relative abundance of different DNA fragment sizes recapitulated the characteristic ATAC nucleosome pattern every ~200bp resulting from transposition of nucleosome-free and nucleosome-bound DNA in both samples (Figures 3.10A and B.2). Following sequencing, calculation of TSS fold enrichment values for the two ibraries demonstrated to be below the acceptable cut-off threshold of 6, with slightly better signal (3.5 fold-enrichment) for the 30 min transposition time (Figure 3.10B). Fragment size distribution using NGS data from the same libraries was consistent with the pre-sequencing assessment, showing approximately equal abundance of the NFF and mono-nucleosome fragments for both transposition times (Figure 3.10C). Approximately similar abundance of NFF and NBF can also be found in good quality libraries with TSS fold-enrichment above 6 (as seen

previously). However, in this case, the extremely low TSS enrichments could suggest that a significant proportion of the NFF in these libraries are free-DNA released by apoptotic cells as a result of the trypsinisation step. Moreover, poor lysis of the keratinocytes may also result in inefficient access of the transposase into the nuclei compartment. This would cause under-transposition and greater NBF abundance when compared to NFF, which in this case may not be revealed as a result of large amounts of NFF being generated from tagmented free-DNA.

Following the ATAC-seq protocol, a modified ATAC-seq version for keratinocytes by Bao and colleagues (named ATAC 2 in this subsection) and the Fast-ATAC protocols were published (Bao *et al.* 2015; Corces *et al.* 2016). Interestingly, Bao's protocol was performed directly on NHEKs adhered to the cell culture plate, avoiding the trypsinisation step that could compromise cell viability. This protocol reduced the percentage of NP-40 (weaker lysis) but increased the amount of Tn5 enzyme per cell, which may address potential under-transposition due to insufficient Tn5 availability. In line with Bao and colleagues approach (avoiding trypsinisation), the three ATAC protocols (ATAC 1 30 min, ATAC 2 and Fast-ATAC using C1 conditions, Table 3.3) were tested directly on keratinocytes isolated from skin biopsies and/or NHEKs cultured on 96-well plates to avoid the trypsinisation step and the subsequent impact of dead cells in data quality. For this purpose, the cell suspension of epidermal keratinocytes obtained from skin biopsies were cultured for 3h in a 96-well plate following a washing step to remove remaining non-viable cells. This procedure is known as adherence assay and allows the isolation of viable undifferentiated keratinocytes. Similarly, 50,000 NHEKs were cultured per well and used as control to test the performance of the different protocols in keratinocyte cells.

For the three tested protocols, the NHEKs library size distribution of sequenced fragments showed presence of NFF but almost absent nucleosome pattern, particularly for the ATAC 2 protocol (Figure 3.11A). Moreover, the TSS

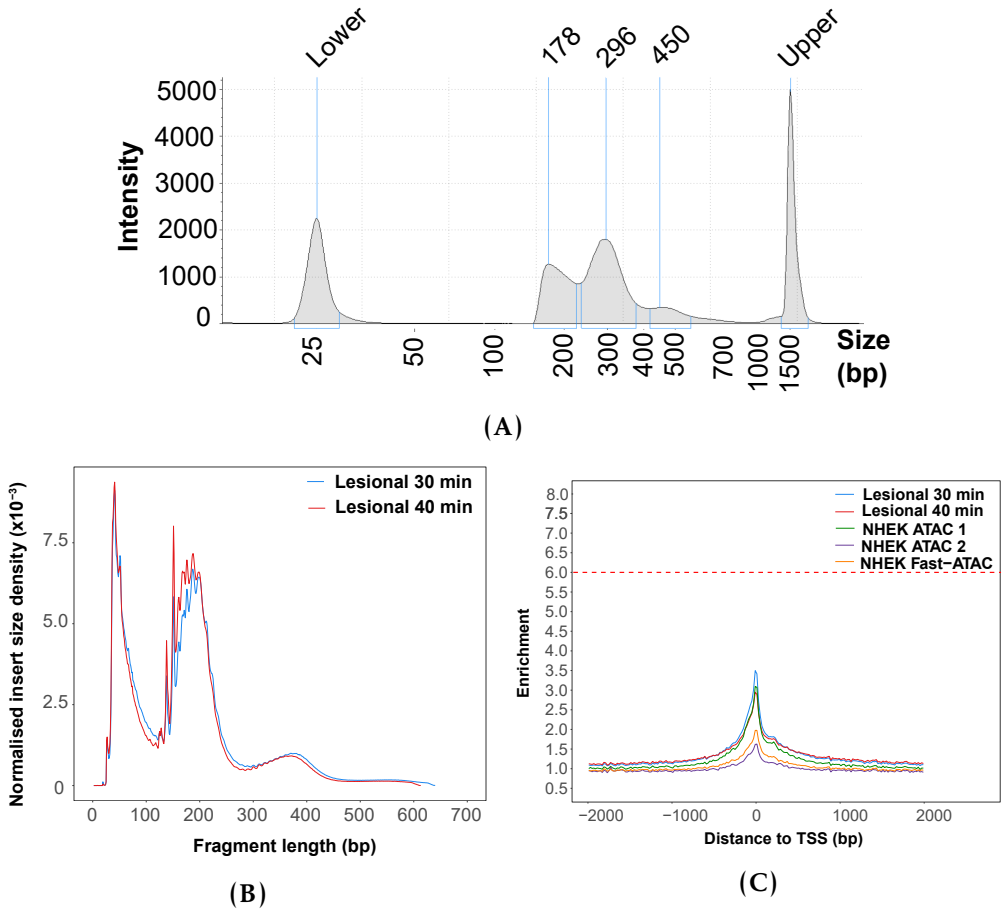


Figure 3.10: Quality control assessment of different ATAC protocols in keratinocytes from a psoriatic lesion and NHEKs. (A) Pre-sequencing quality control showing relative abundance of different DNA fragment sizes in the ATAC libraries generated in 50,000 cells from a suspension of keratinocytes isolated from lesional skin of one psoriasis patient. Cells from the same suspension were used to test two transposition times (30 and 40 min) and the profile for the 30 min library is shown here. The corresponding pre-sequencing profile for the sample transposed for 40 min is included in Figure B.2. (B) Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from the ATAC 1 psoriasis lesional keratinocytes libraries, previously mentioned in (A), and NHEK libraries generated by performing ATAC 1, ATAC 2 and Fast-ATAC protocols directly on adherent cells cultured in 96-well plates. (C) Density distribution of sequenced fragments for ATAC 1 libraries in (A) and Figure B.2.

Protocol	Lysis and transposition	Key parameters
ATAC 1 (Buenrostro *et al.* 2013)	Two steps	0.1% NP-40 and 2.5µL Tn5
ATAC 2 (Bao *et al.* 2015)	Two steps	0.05% NP-40 and 5µL Tn5
Fast-ATAC (Corces *et al.* 2016)	One step	C1*: 0.01% digitonin, 2.5µL Tn5 C2: 0.01% digitonin, 0.5µL Tn5 C3: 0.025% digitonin, 0.5µL Tn5 C4: 0.025% digitonin, 2.5 µL Tn5

Table 3.3: Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies. Transposition times for all protocols was 30 min. * corresponds to the original Fast-ATAC conditions from Corces *et al.* 2016.

enrichment was under the recommended cut-off of 6 in skin biopsy keratinocytes and NHEKs libraries from the three protocols (Figures 3.10C and B.3). Potential poor lysis was further addressed through additional modifications of the Fast-ATAC protocol by increasing the concentration of the detergent digitonin (from 0.01% in the published Fast-ATAC protocol to 0.025%) in combination with standard or lower Tn5 amounts (2.5 and 0.5 µL, respectively)(Table 3.3 C2 to 4). Library quality control prior to sequencing to assess the relative abundance of different DNA fragment sizes failed to show the characteristic nucleosome pattern profile expected in ATAC for any of the three tested conditions, and therefore NGS was not performed (Figure B.4A, B and C).

At the end of the experimental work of this thesis, a new protocol called Omni-ATAC was published (Corces *et al.* 2017). Omni-ATAC was a protocol suitable for every cell type, in contrast to ATAC 1 and Fast-ATAC, which were optimised for hematopoietic cells (Buenrostro *et al.* 2013; Corces *et al.* 2016). Performance of this protocol in 50,000 viable NHEKs in suspension yielded the expected fragment size distribution for sequenced fragments, with the greatest abundance for NFF followed by mono and di-nucleosome fragments (Figure

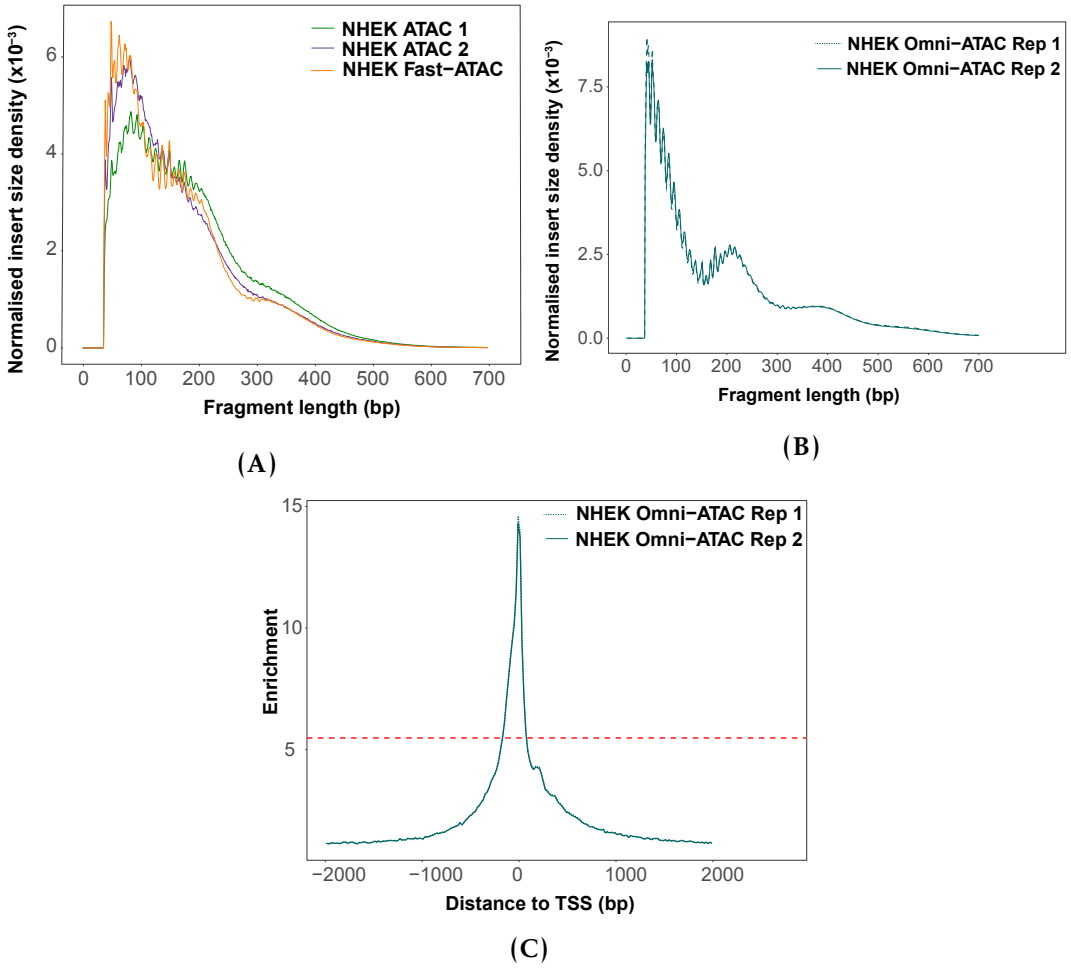


Figure 3.11: Quality control assessment of Fast-ATAC and Omni-ATAC in cultured NHEK. Representation of the fragment sizes density distribution in NHEKs libraries generated using (A) ATAC 1, ATAC 2, Fast-ATAC or (B) Omni-ATAC protocols. (C) Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from two Omni-ATAC technical replicates. In (A) the results for one of the two replicates is shown.

3.10B). Moreover, high TSS enrichment values (approximately 15 fold change) were observed for the two replicates (Figure 3.10C). When performing overlap between the Omni-ATAC sample peaks filtered based on low stringency p-value (p-value=0.01) and the ENCODE DHSs from 125 cell types, the highest percentage of overlap (approximately 55%) was observed for NHEKs (Figure 3.12A). In contrast, the same analysis using peaks from the NHEKs libraries generated with ATAC 1, ATAC 2 and Fast-ATAC protocols only showed 20% or less overlap of each sample called peaks with the ENCODE DHS data of the same cell type, supporting the higher quality and specificity of Omni-ATAC accessible regions in keratinocytes, even at a very low p-value filtering. The differences in the quality of ATAC signal between Omni-ATAC and the prior ATAC protocols was clearly observed at the keratin (*KRT*) genes (main components of keratinocyte cytoskeleton) in chr17, where Omni-ATAC showed the lowest background noise, the highest signal intensity and the greatest number of high quality peaks (Figure 3.12B).

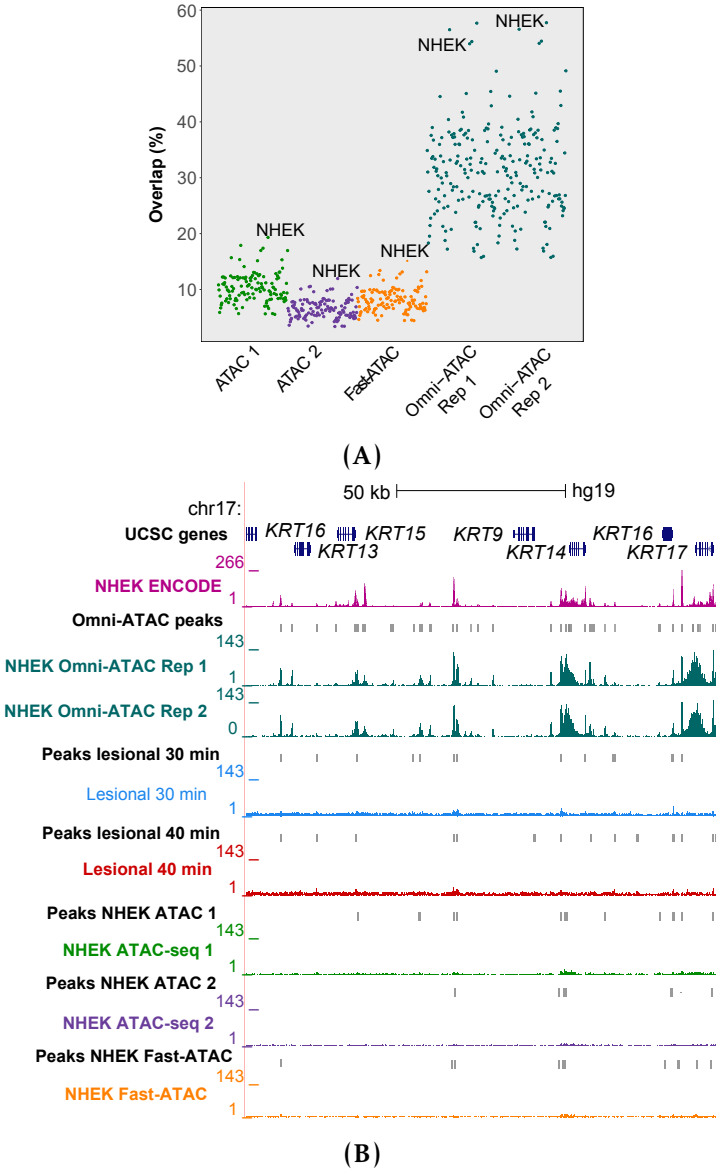


Figure 3.12: Comparison of four different ATAC protocols applied to psoriasis and/or healthy keratinocytes and NHEKs cells with published ENCODE DNase-seq data. (A) Enrichment (% of overlap) of observed peaks called in each ATAC sample (using low stringent p-value) with open DHS chromatin regions in 125 ENCODE cell types. (B) UCSC Genome Browser track showing ATAC read density (y-axis) at the chr17 keratin (KRT) family gene locus (x-axis) for different ATAC protocols (all subsampled to 30 million total reads).

3.2.5 Effect of cryopreservation and fixation in the chromatin landscape of immune primary cells

Experimental design and sample description

As previously introduced, research using clinical samples represents a logistical challenge as immediate processing of freshly acquired cells may not be feasible. In the context of this thesis, two different possible approaches involving cryopreservation and fixation were of interest to test the preservation of chromatin accessibility and a collaborative project was established with High-Throughput Genomics core at the WHG. The first approach was the cryopreservation of PBMCs in liquid nitrogen using DMSO followed by thawing, recovery and FACS isolation of the cell population of interest (Figure 3.13). Secondly, the performance of an optimised protocol developed by High-Throughput Genomics using DSP in scRNA-seq (Attar *et al.* 2018) was investigated as a short term preservation method for FACS-isolated relevant cell types.

In order to investigate the performance of these two strategies, blood from three healthy volunteers (Chapter 2 in Table 2.1) matched for sex and age (three females, mean age 25.6 years old) was processed on different days to simulate the experimental design when using patient samples (experimental design summarised in Figure 3.13). PBMCs were prepared from 90mL blood using a Ficoll gradient and $CD14^+$ monocytes and $CD4^+$ T cells were isolated by FACS, as detailed in Chapter 2. ATAC-seq was performed on 50,000 $CD14^+$ monocytes and $CD4^+$ T cells, either freshly isolated or after fixation with DPS, stored at 4°C for 24h and then processed for ATAC-seq (Figure 3.13 Day 1, ATAC-seq fresh and ATAC-seq fixed, respectively).

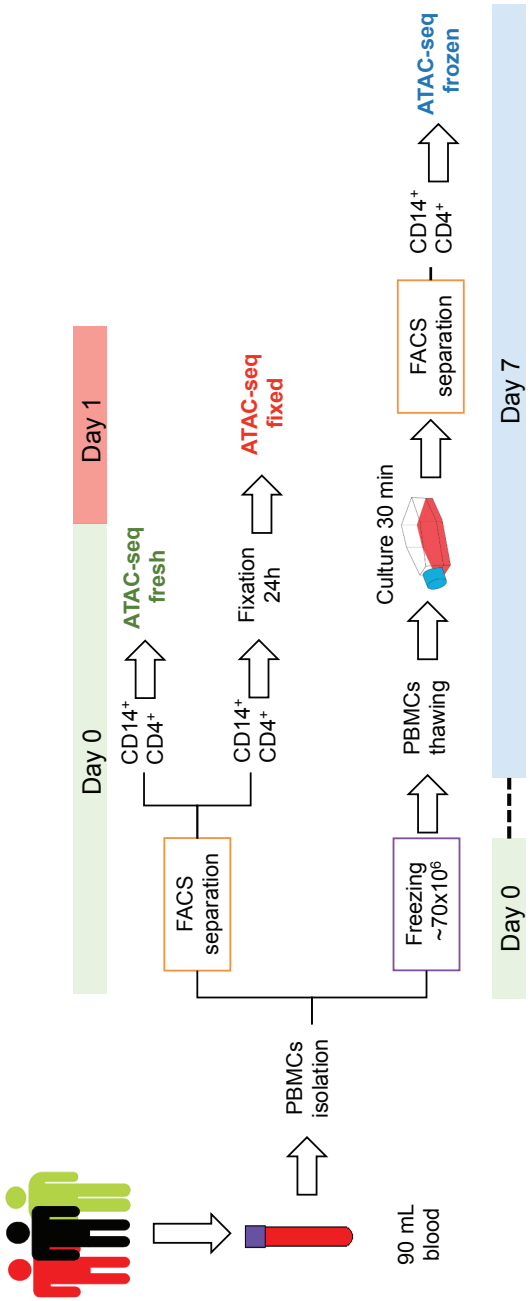


Figure 3.13: Experimental design to assess the impact of cryopreservation and fixation in the chromatin accessibility of immune primary cells. Three healthy control individuals were recruited on different days, PBMCs were isolated and freshly isolated $CD14^+$ monocytes or $CD4^+$ T cells processed for ATAC-seq immediately (ATAC-seq fresh), after fixation with DPS (ATAC-seq fixed) or after cryopreservation of PBMCs (ATAC-seq frozen).

To investigate the effect of cryopreservation and cell recovery in the chromatin landscape of primary immune cells, $70x10^6$ million PBMCs were cryopreserved with DMSO on the day of collection and stored in liquid nitrogen followed by thawing and recovery in culture for approximately 30 min (as detailed in Chapter 2). After recovery, $CD14^+$ monocytes and $CD4^+$ T cells were isolated from the PBMCs by FACS and then ATAC-seq was performed (Figure 3.13 Day 7, ATAC-seq frozen). Altogether, for each volunteer, three matched ATAC-seq libraries were generated in two different cell populations: ATAC-seq fresh, ATAC-seq fixed and ATAC-seq frozen.

Assessment of the chromatin structure preservation in the different conditions

All samples from each of the two cell types had more than 15 million reads, which has previously been shown as the minimum for successful ATAC-seq analysis and peak calling (Figure 3.14). The median number of reads across the fresh, frozen and fixed were more similar for $CD14^+$ monocytes (58.6, 64.2 and 39.6 million reads, respectively) than in the $CD4^+$ samples, where the frozen and fixed presented lower median total reads compared to the controls (43.8, 32.9 and 28.8 million reads respectively) (Figure 3.14A and B).

The ATAC-seq signal-to-noise ratios across the TSS showed a similar median for the fresh and fixed $CD14^+$ monocytes libraries (17.4 and 16.5 fold-enrichment, respectively) and was higher for the frozen samples (26.3 fold-enrichment)(Table A.1). The median TSS enrichments in the frozen and fixed $CD4^+$ samples were considerably higher (16.1 and 14.3 fold-enrichment, respectively) than the fresh samples (5.6), which were borderline for the ENCODE recommended threshold. For one volunteer (CTL1) the fixed samples of both cell types showed considerably lower TSS enrichment (2.5 and 7.9, respectively) compared to the other fixed samples (Table A.1). As expected, the

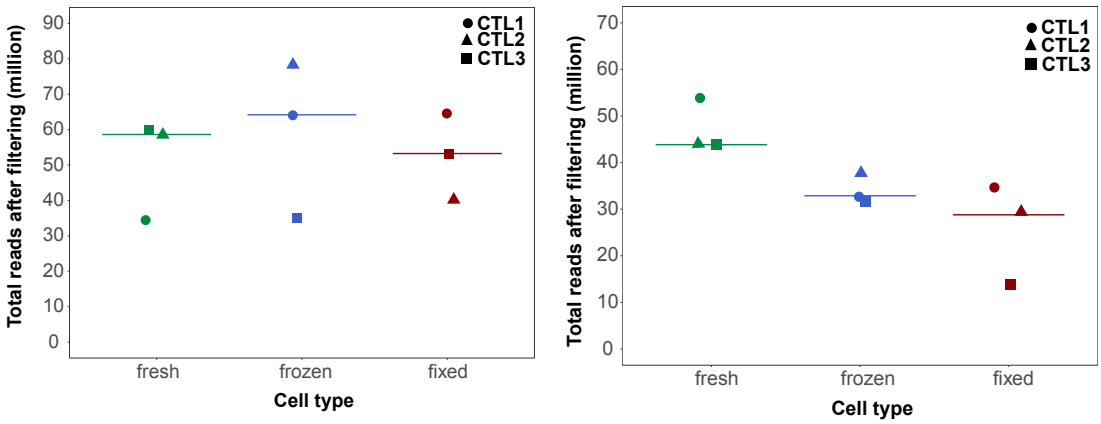


Figure 3.14: Total number of ATAC-seq reads for the fresh, frozen and fixed CD14$^+$ monocytes and CD4$^+$ samples for three volunteers (CTL1-3). Representation of million reads after filtering for the fresh, fixed and frozen ATAC-seq libraries in (A) CD14$^+$ monocytes and (B) CD4$^+$ cells.

fragment size distribution profiles of the fresh and frozen samples for both cell types were very similar, showing appropriate relative abundance of NFF and NBF corresponding to mono-, di-, tri- and tetra-nucleosomes (Figure 3.15A and B green and blue lines). Conversely, fixed samples showed differences in the fragment size distribution profiles, where for both cell types a fraction of NFF was absent in CTL1 and lower than NBF in CTL2 and CTL3 (Figure 3.15 red lines).

Chromatin structure across and within the TSS was then investigated following Scharer and colleagues approach (Scharer *et al.* 2016). The nucleosome-free fragments (<150bp) from all the samples showed a single peak of enrichment at the nucleosome-depleted TSS position, with CD4$^+$ fresh samples presenting the lowest relative enrichment (Figure 3.16A and C).

The pattern of enrichment of di-nucleosome fragments (ranging between 260 and 340bp) demonstrated in the majority of the samples periodicity in the TSS surroundings, with two peaks of enrichment mapping at the characteristic up-stream and down-stream positioned nucleosomes (Figure 3.16B and D). Although fixation in CD14$^+$ and CD4$^+$ CTL1 samples drastically reduced the

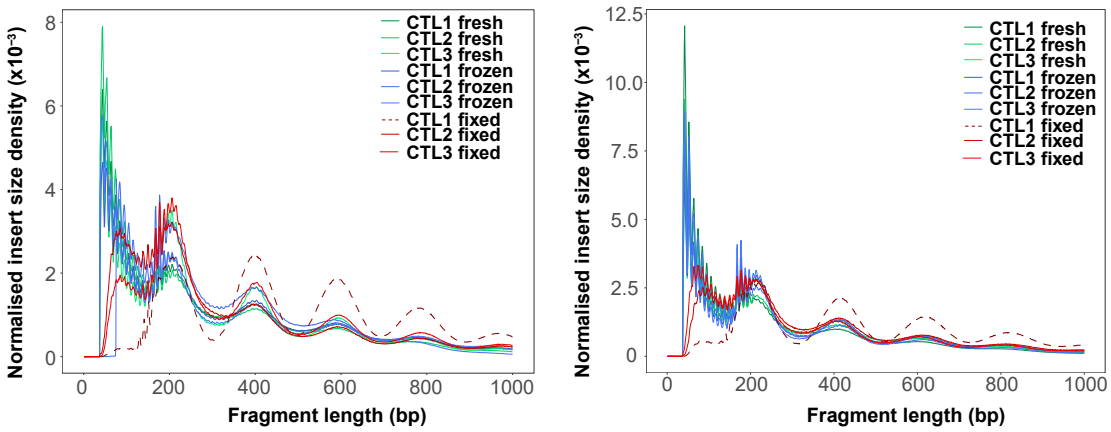


Figure 3.15: Fragment size density distribution for ATAC-seq fresh, fixed and frozen in CD14$^+$ monocytes and CD4$^+$ cells. The distribution of ATAC-seq fragment lengths are illustrated for (A) CD14$^+$ monocytes and (B) CD4$^+$ cells and colour-coded by condition (fresh=green, frozen=blue and fixed=red).

abundance of NFF, the ATAC-seq signal for the remaining ones was clearly enriched at TSS (Figure 3.16A and C). Interestingly, despite fixed CD14$^+$ CTL1 having increased density of NBF (in this case di-, tri- and tetra-nucleosomes) loss of chromatin structure around the TSS suggested that nucleosomes may have been displaced from their original position, likely due to the fixative not cross-linking proteins to DNA, and therefore explaining this lack of enrichment (Figure 3.16B). CD4$^+$ fresh samples showed comparatively low enrichment at the TSS for NFF and di-nucleosome fragments, consistent with the overall borderline signal-to-noise ratio of the samples (Figure 3.16B and D and Table A.1).

Annotation of the significant peaks identified in each sample (filtered using the IDR strategy as explained in 3.2.1) revealed that the highest proportion of ATAC-seq peaks localised to promoters, introns and intergenic regions (Figure 3.17A and B), consistent with previous studies (Buenrostro *et al.* 2013; Scharer *et al.* 2016). Interestingly, the ATAC-seq libraries showing higher percentage of peaks at promoter regions corresponded to fixed samples, in which chromatin structure was incompletely preserved, and to fresh CD4$^+$ T cells, with borderline signal-to-noise ratios based on TSS enrichment analysis. The higher percentage

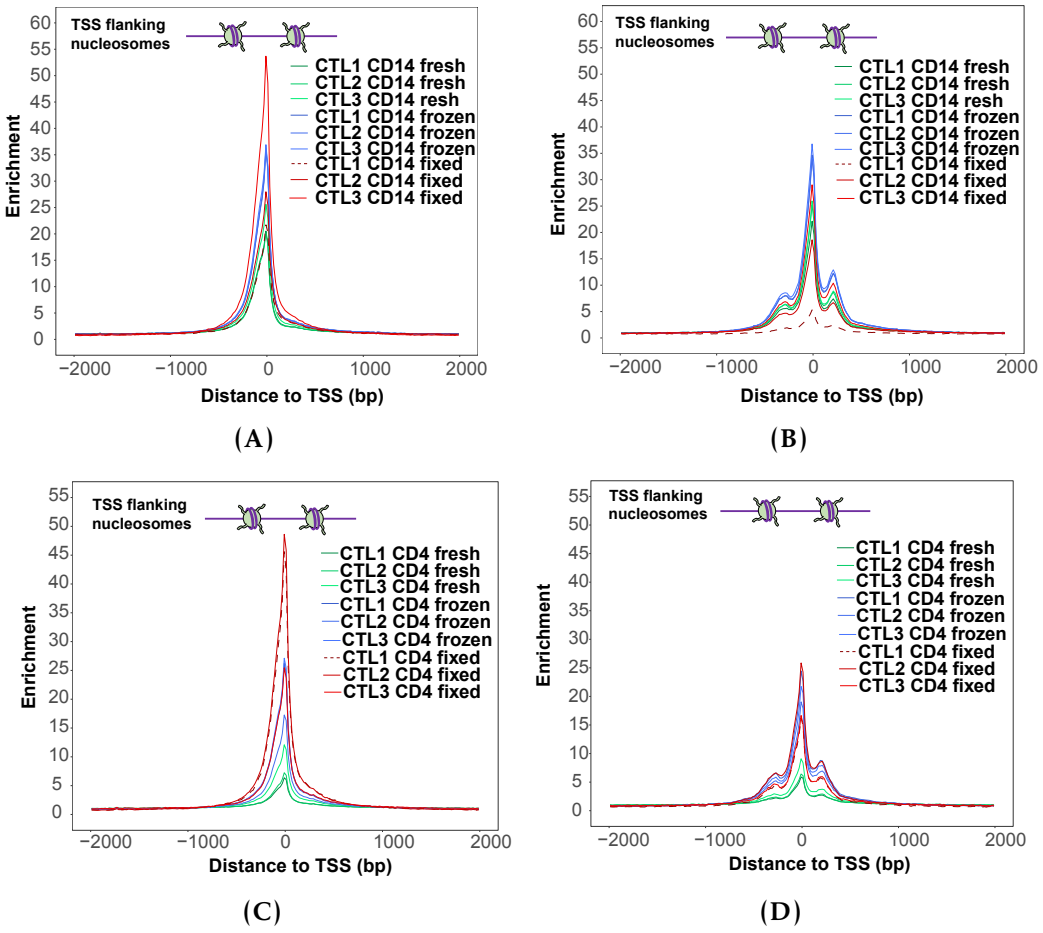
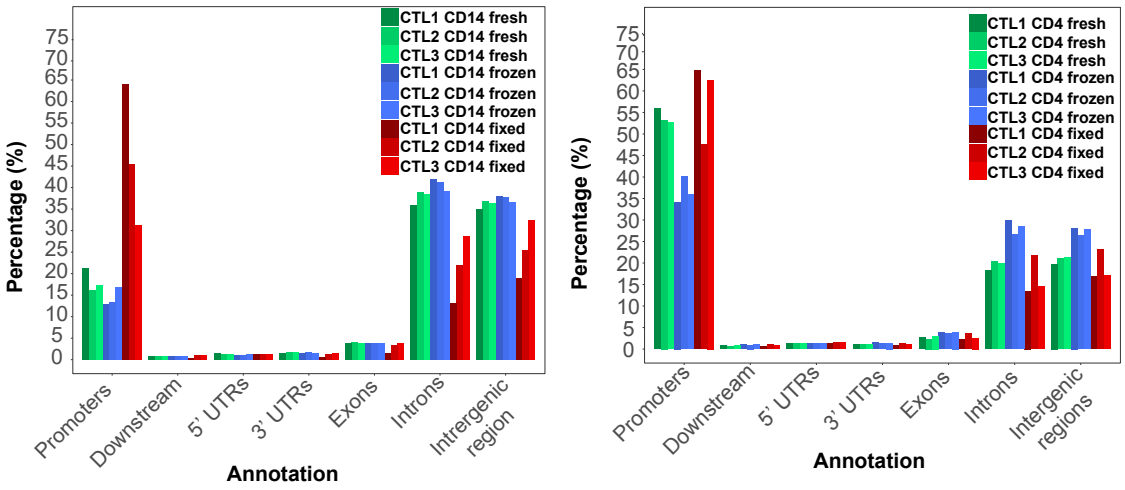


Figure 3.16: ATAC-seq enrichment of nucleosome-free and di-nucleosome fragments at the TSS and surroundings in CD14$^+$ monocytes and CD4$^+$ samples for the three conditions. For each of the samples, nucleosome-free fragments (<150bp) or di-nucleosome (between 260 and 340bp) fragments were selected *in silico*. Enrichment at +/−1Kb across all the Ensembl annotated TSS is shown for (A) CD14$^+$ monocytes or (C) CD4$^+$ T cells nucleosome-free fragments and for (B) CD14$^+$ monocytes and (D) CD4$^+$ T cells di-nucleosome fragments. The two nucleosomes flanking the TSS are depicted at the top of each graph.

of peaks annotated as promoters in in low quality samples revealed reduced ability of these samples to capture genomic features with more moderate chromatin accessibility and bias for the the location of significant robust peaks at the most stable and distinctive open chromatin sites in the genome.

Overall, this analysis demonstrated that the chromatin structure from DSP fixed ATAC-seq libraries was not fully preserved in either of the two cell types, as they all showed a distinct pattern of fragment size distribution in comparison to fresh and frozen samples, as well as loss of nucleosome positioning across the TSS only in CTL1 fixed samples.



(A) **(B)**

Figure 3.17: Genomic features annotation for the ATAC-seq peaks called in each of the fresh, frozen and fixed samples from CD14$^+$ monocytes and CD4$^+$. Overlap was performed between the genomic features and the list of (A) CD14$^+$ monocytes and (B) CD4$^+$ peaks, filtered for the corresponding optimal p-value based on the pseudoreplicates IDR analysis in each sample and condition (fresh=green, frozen=blue and fixed=red).

Lastly, principal component analysis (PCA) was conducted separately in CD14$^+$ monocytes and CD4$^+$ T cells using the normalised counts retrieved at each of the ATAC peaks from the consensus list built including samples from the three conditions (CP_CD14_all_cond and CP_CD4_all_cond, respectively), with the exception of fixed CTL1 (considerably different from the other two fixed samples). Plotting the first two principal components (PCs) showed sample

clustering based on condition in both cell types and demonstrated that the largest source of variability correlated with the way samples were processed (Figure 3.18).

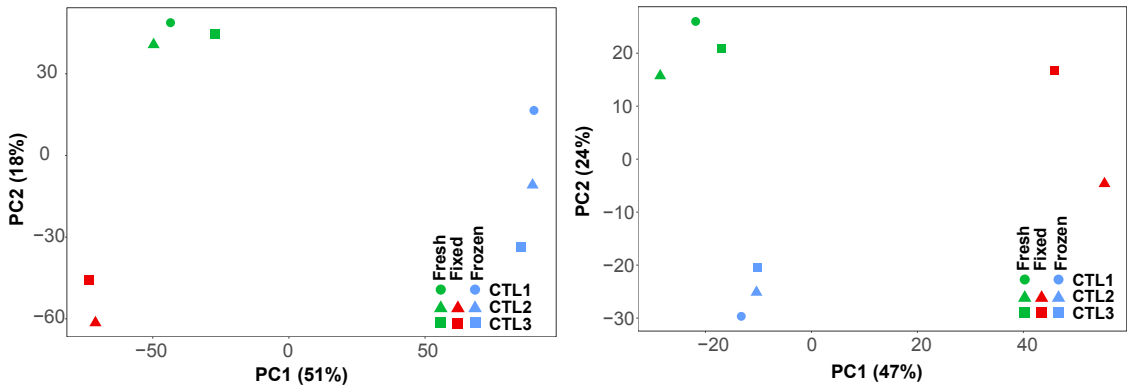
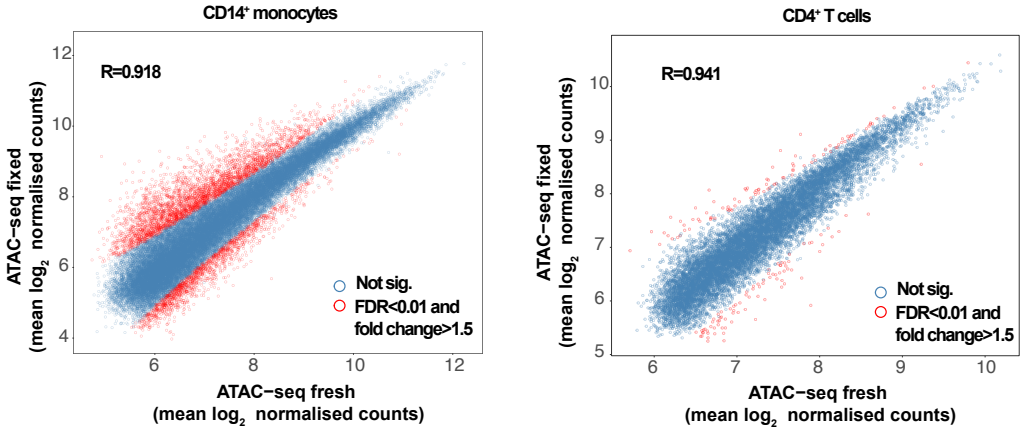


Figure 3.18: PCA based on the ATAC-seq chromatin accessibility landscape in fresh, fixed and frozen samples. PCA was performed using the normalised counts across the consensus list of ATAC peaks from the combined fresh, fixed and frozen samples (CP_CD14_all_cond and CP_CD4_all_cond) in (A) $CD14^+$ monocytes or (B) $CD4^+$ cells from the same three healthy individuals. The first two PCs (x-axis and y-axis, respectively) of all the regions included in each of the consensus peak list are plotted. Each point represents a sample, where shape codes for individual (CTL1, CTL2, CTL3) and colour means condition (fresh, fixed, frozen). The proportion of variation explained by each principal component is indicated in brackets.

Differential chromatin accessibility analysis between fresh and frozen samples

In order to determine the effect of the freezing and recovery process in the chromatin landscape of $CD14^+$ monocytes and $CD4^+$ T cells, genome-wide differences between ATAC-seq fresh (biological reference) and frozen were investigated. Since the use of DSP as a fixative was already found to impact preservation of the chromatin structure, differential analysis between fresh and fixed samples was not conducted as the results may be confounded. Comparison between ATAC-seq fresh and ATAC-seq frozen within each cell types was performed using the normalised read counts for the list of ATAC peaks included in the consensus list. Overall, ATAC-seq normalised counts showed

high correlation between fresh and frozen samples in the two cell types, with the lowest correlation (R=0.918) found in CD14$^+$ monocytes (Figure 3.19A, B).



(A) (B)

Figure 3.19: Comparison of the $\log_2$ normalised ATAC-seq counts at the consensus consensus list of peaks in fresh, fixed and frozen conditions. Each plot shows the comparison of ATAC-seq $\log_2$ mean normalised counts from the CP_CD14_all_cond or CP_CD4_all_cond filtered for background noise (80% empirical cut-off) between fresh and frozen (A) CD14$^+$ monocytes and (B) CD4$^+$ T cells. Pearson correlation coefficient (R) is indicated.

The differential chromatin accessibility analysis between ATAC-seq fresh and frozen samples using DEseq2 revealed a total of 5,123 and 177 significant DARs (FDR<0.01 and fold change>1.5) and a larger proportion of regions (12.6 and 1.3%) changing chromatin accessibility between fresh and frozen samples in CD14$^+$ monocytes compared to CD4$^+$ T cells, respectively. Nevertheless, the percentage of sites identified as DARs in relation to the total number of investigated regions for both cell types suggested only moderate changes in the chromatin accessibility landscape between fresh and frozen conditions.

In order to identify biological processes that may altered by the changes in chromatin accessibility between conditions, enrichment analysis was conducted using as input the genes proximal (≤5Kb) to the significant DARs (FDR<0.01 and fold change>1.5) identified in each cell type. Significantly enriched Gene Ontology (GO) biological processes (FDR<0.01) were only identified for proximal genes to DARs in CD14$^+$ (Figure 3.20). This included changes in

Cell type	Total regions in consensus list	DARs (percentage)
$CD14^+$ monocytes	40,600	5,123 (12.6%)
$CD4^+$	13,025	177 (1.3%)

Table 3.4: Summary results from the differential chromatin accessibility analysis comparing ATAC-seq frozen or fixed chromatin landscape to the reference ATAC-seq fresh. The number of regions in the consensus list (CP_CD14_all_cond or CP_CD4_all_cond) following filtering for background counts using 80% cut-off and the number of significant DARs (FDR<0.01 and fold change>1.5) are shown. In brackets, the percentage of DARs over the total number of regions included in the differential analysis is indicated.

chromatin accessibility of regions in proximity to genes involved in cell shape and cell adhesion, lipopolysaccharide-mediated signalling, inflammation and signal transduction, amongst others, suggesting that $CD14^+$ monocytes may undergo some activation as a result of the cryopreservation and subsequent recovery in incubation.

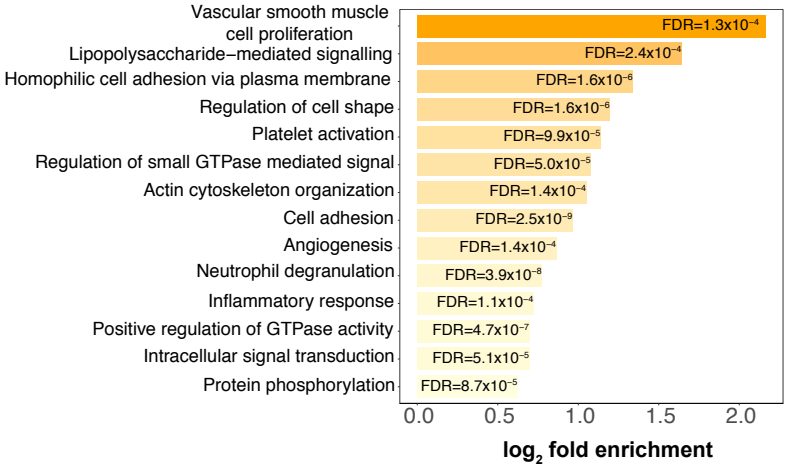


Figure 3.20: Top fourteen gene Ontology biological processes enriched for DARs between ATAC-seq fresh and ATAC-seq frozen samples in $CD14^+$ monocytes. The barplot shows the top fourteen most significant (FDR<0.01) Gene Ontology (GO) biological processes enriched for genes proximal to DARs identified between fresh and frozen $CD14^+$ monocytes. The GO terms are ordered based $\log_2$ fold change enrichment and the exact significance (FDR) for each of them is also indicated.

For example, three of the DARs identified when comparing ATAC-seq fresh to ATAC-seq frozen $CD14^+$ were found at the promoter, an exon and the 3'UTR of the *TNFSF14* gene and showed greater accessibility in frozen compared to fresh

samples (Figure 3.21. TNFSF14 is the ligand for a receptor from the TNF-receptor superfamily and is involved in T cell activation, induction of apoptosis and also in bone destruction mediated by monocytes and synovial cells interactions in RA.

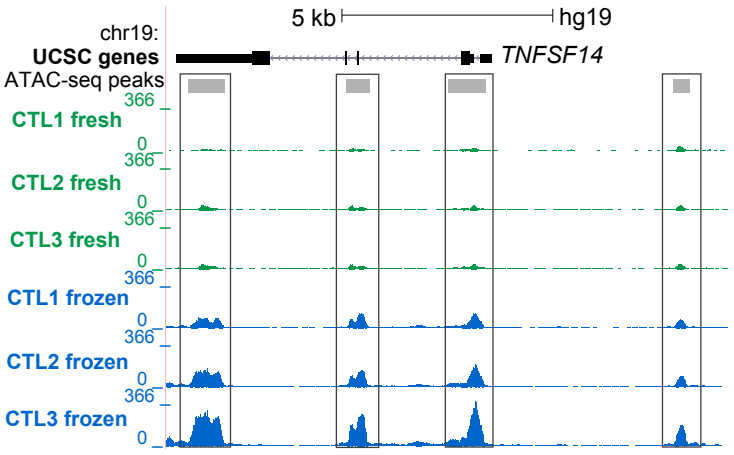


Figure 3.21: Differential chromatin accessibility at the *TNFSF14* gene between ATAC-seq fresh and ATAC-seq frozen in CD14$^+$ monocytes. UCSC Genome Browser view illustrating the normalised read density (y-axis) at four significant (FDR<0.01 and fold change>1.5) DARs (x-axis) within and upstream the *TNFSF14* gene in CD14$^+$ monocytes. The four DARs were more accessible in ATAC-seq frozen when compared to ATAC-seq fresh. Tracks are colour-coded by condition (green=fresh and blue=frozen).

3.3 Discussion

The aim of this chapter was to establish a data analysis pipeline for ATAC and compare various experimental protocols, as this was the first time this technique was used in the research group. A particular focus was to consider appropriate methodologies for clinical studies, where sample availability and quality may be severely limiting. A number of alternative protocols, metrics and algorithms described in early ATAC reports were evaluated in the pilot experiments presented in this chapter. This enabled the establishment of a pipeline and approach to be implemented for investigation of the psoriasis and PsA chromatin landscape (Chapters 4 and 5).

3.3.1 ATAC: methodological aspects and pipeline establishment

At the time of the first ATAC-seq publication (Buenrostro *et al.* 2013), well established protocols for complete processing and data analysis of ATAC were lacking. Since then, several publications have implemented ATAC-seq and modifications of this protocol together with a wide range of data analysis strategies to answer different biological questions (Table 3.1) as well as alternative protocols such as THS-seq to assess chromatin accessibility in low number of cells (Sos *et al.* 2016).

The data and analysis presented in this chapter has confirmed some limitations of the ATAC-seq and Fast-ATAC protocols. Quality assessment and variability across samples was difficult to detect through pre-sequencing library quality control based on relative abundance of the different DNA fragment sizes. In fact, successful pre-sequencing profiles of the ATAC fragment size distribution recapitulating nucleosome patterns would still lead to libraries with high background noise when visualising read density in the UCSC Genome Browser. This required the identification and establishment of appropriate data analysis and quality control measures beyond pre-sequencing library quality assessment.

In this chapter different quality metrics were explored, including TSS enrichment and FRiP. Both correlated well with the overall differences in sample quality from the ATAC-seq libraries used as an exemplar here. Importantly, TSS and FRiP were shown to be independent of sequencing depth, and therefore can be applied in low depth sequenced samples when performing optimisation or preliminary quality control before increasing the coverage, as also recently shown in other studies (Corces *et al.* 2017). Similarly to TSS, FRiP proved to be informative in evaluating signal-to-noise ratios; however it relies on peak calling and thus is more likely to be biased by the filtering strategy. In agreement with these findings, enrichment of ATAC signal across Ensembl annotated TSS is now

recommended by ENCODE as the preferred means of assessing overall sample quality, and was implemented as the metric to evaluate signal-to-noise in our pipeline.

The variability in quality of ATAC-seq libraries was also addressed at the peak calling level in this chapter, with the implementation of a peak filtering strategy that for each sample could identify good quality and reproducible peaks using IDR analysis between pseudoreplicates. This approach was demonstrated to reduce repetitive and non meaningful regions that could be confounders for downstream analysis. In terms of sequencing depth, analysis in this chapter showed that approximately 15 to 20 million reads after filtering were the minimum required to identify an appropriate proportion of accessible regions (peaks) as well as to obtain meaningful results in the peak filtering based on pseudoreplicate IDR analysis. These observations have also been confirmed by Qu and colleagues (Qu *et al.* 2015), where IDR analysis at different sequencing depths was used to evaluate consistency across replicates but not implemented for peak filtering.

Establishment of appropriate measurements for post-sequencing library quality control allowed formal testing, beyond the conditions from Buenrostro's publication, for the effect of transposition times, one of the variants that can affect the quality of ATAC libraries in a cell-type specific manner. At the start of the project transposition for 40 min appeared the most appropriate for all cell types according to pre-sequencing library quality control (relative abundance of DNA library fragment sizes). Longer transposition times are known to reduce the length of the yielded fragments and to increases, to some extent, the abundance of NFF (Raurell-Vila *et al.* 2018). Assessment of three different transposition times for the ATAC-seq protocol showed minor and heterogenous effects on the NFF/mNBF ratio and the signal enrichment across the TSS for the different cell types. For example longer transposition showed a slight increase of NFF

abundance in $CD14^+$ monocytes but reduction in $CD8^+$ cell. Altogether, these results from a single replicate for each transposition time suggested that changes in the duration of the tagmentation reaction between 30 and 40 min do not have considerable impact on the ATAC-seq fragment size distribution and the signal-to-noise ratios, with the largest variability being cell type dependent.

The use of the improved Fast-ATAC protocol addressed some of the limitations identified by the ATAC-seq data generated within our group. Using only one replicate, Fast-ATAC showed reduction in the percentage of mitochondrial reads in all cell types when compared to ATAC-seq. Retrospective analysis comparing ATAC-seq to Fast-ATAC-seq using all samples described in this thesis confirmed that Fast-ATAC significantly reduced the percentage of mitochondrial reads in the four studied cell types (unpaired Wilcoxon signed-ranked test p-values<$4.1x10^{-05}$) (Figure B.5A).

Regarding signal-to-noise in the libraries, a trend for improved TSS enrichments when using Fast-ATAC was only found in $CD14^+$ monocytes and $CD4^+$ cells in the preliminary analysis using one replicate per cell type. Conversely, the retrospective analysis including all the libraries generated in this thesis only showed statistically significant increased TSS fole-enrichment in $CD4^+$ cells (unpaired Wilcoxon signed-ranked test p-value=0.024) and was not a significant improvement for any of the remaining three haematopoietic cell types (p-value>0.05 Figure B.5B). In fact, Corces and colleagues Fast-ATAC publication only showed data demonstrating improved TSS by Fast-ATAC for $CD4^+$ T cells (Corces *et al.* 2016). In contrast, the Omni-ATAC protocol publication included a comprehensive comparison of the three ATAC protocols across a large number of cell types showing that Fast-ATAC did not improve TSS fold-enrichment when compared to ATAC-seq in a number of the haematopoietic cells, for example $CD19^+$ cells, in line with the results presented in this chapter (Corces *et al.* 2017).

3.3.2 The challenges of performing differential chromatin accessibility analysis

Studying chromatin accessibility in clinical samples first requires the definition of a consensus list of accessible regions for which no accepted method has yet been agreed. In this work, a consensus list of ATAC peaks containing all the peaks identified in at least 30% of the samples included in analysis has been chosen. This represents an unbiased approach to include peaks that can vary across individuals (regardless of biological subgroup) but still be differentially accessible across conditions. Other publications have preferred building condition-specific consensus lists of peaks that are later merged or simply include all the significant peaks called in all the analysed samples (Alasoo *et al.* 2018; Thurner *et al.* 2018).

When used for differential analysis, an additional filtering step has been implemented to account for background noise in peaks included in the consensus list for being significant (present) in at least 30% of all the samples but did not pass quality filtering (therefore considered absent) in a number of them. In terms of the algorithm to perform normalisation and differential chromatin accessibility analysis, no consensus has been reached in the literature. The majority of the studies reviewed at the time of implementing differential analysis were peak-based and relied on RNA-seq or microarray algorithms such as EdgeR, limma or DESEq2 (Table 3.1). The analysis here, revealed DESeq2 as a more stringent method compared to quantile normalisation & limma voom. Limma has been reported to be affected by low quality samples and that may also explain the increase in differential hits observed when compared to DESEq2 (Alasoo *et al.* 2018). For both methods, the implementation of the additional filtering cut-off to control for high number of background reads has shown a reduction in the number of significant differentially accessible regions. Given the difficulties of recruiting large numbers of suitable clinical samples, DEseq2 in combination

with the additional filtering step was chosen at the time of the study as an appropriately stringent method to account for variability in sample quality and to control, to some extent, for potential false positive hits. As specific tools for ATAC analysis are developed, further comparison of the different outputs will be of interest in future work.

3.3.3 Studying the chromatin landscape from psoriasis skin biopsies

At the time of writing, only RNA-seq studies have been performed in keratinocytes from psoriasis skin biopsies. The relevance of keratinocytes in psoriasis pathophysiology and the ability to sample this tissue represented a great opportunity to investigate the chromatin accessibility landscape at the main site of inflammation using ATAC. The low signal-to-noise ratios observed in ATAC-seq libraries generated from lesional keratinocytes in suspension was hypothesised to be the result of low cell viability together with under-transposition of the nuclei due to insufficient Tn5 and/or poor cell lysis. The poor performance of ATAC-seq (both, Buenrostro *et al.* 2013 and Bao *et al.* 2015 versions) and Fast-ATAC protocols on NHEKs and keratinocytes isolated from skin biopsies on a 96-well plate, without prior cell dettachment using trypsinisation, suggested that the main limitation was intrinsic to the cell type and not substantially driven by compromised cell viability in these particular cases. Importantly, ATAC-seq on cultured NHEKs using Bao *et al.* protocol (ATAC 2) not only failed to reproduce the successful results presented in the publication but also did not show improvement in the quality of the library by increasing the Tn5 concentration, reinforcing the lysis step as the main limitation.

The characteristic insoluble protein structure (mainly formed by keratins) synthesised by keratinocytes as differentiation progresses (cornification) and eventually replacing the plasma membrane is likely to difficult cell

permeabilisation, lysis and efficient transposition. In fact, inefficient lysis by the Buenrostro *et al.* 2013 ATAC-seq version of the protocol has been demonstrated to cause poor quality libraries in breast tissue biopsies (Fujiwara *et al.* 2019). The lysis buffer from the latest Omni-ATAC protocol combines three non-ionic detergents (NP-40, digitonin and Tween-20) with an overall greater detergent concentration (0.21% vs 0.1 and 0.01% in ATAC-seq and Fast-ATAC, respectively) likely resulting in a more efficient cell lysis (Corces *et al.* 2017). The Omni-ATAC protocol demonstrated a considerable improvement in performance compared to ATAC-seq and Fast-ATAC in NHEKs, consistent with data presented by Corces *et al.* 2017 in the same cell type. The required two-fold increase in detergent concentration used by the Omni-ATAC protocol compared to ATAC-seq (0.1 vs 0.021%) is consistent with the failure of the 0.025% digitonin concentration tested with the Fast-ATAC protocol C4 to improve cell lysis. Furthermore, the Omni-ATAC publication also showed quality measures for ATAC-seq and Fast-ATAC data in NHEKs, which further supports the poor quality of my results when using those protocols.

Altogether, the successful results of Omni-ATAC in NHEKs encourage testing its performance in keratinocytes isolated from skin biopsies and open an avenue to ultimately explore the chromatin landscape in lesional and uninvolved skin biopsies from psoriasis patients. Unfortunately, this aim was not compatible with the time frame of this thesis and could be conducted as future work for the project.

3.3.4 Characterisation of the effect of preservative techniques in the chromatin landscape

The use of clinical samples sometimes involves logistical limitations that require sample preservation. At the time of starting this thesis, the Oxford Genomics Centre at the WHG had implemented the use of DSP as a compatible

fixative for microfluidics-based scRNA-seq methods (Attar *et al.* 2018). Moreover, at the time of the experiment ATAC-seq has not yet been proven to be compatible with traditional fixatives such as formaldehyde. It was therefore of interest to test the ability of this fixative to maintain the chromatin structure prior to performance of ATAC-seq, which would allow simulatanous preservation of gene expression and chromatin accessibility in the samples. Despite of DSP showing appropriate preservation of cell and RNA integrity with minimal differences in scRNA-seq when compared to fresh material (Attar *et al.* 2018), my data demonstrated abnormal chromatin structure in the the fixed ATAC-seq libraries which showed altered fragment size distribution profiles with partial or total loss of NFF when compared to the fresh and frozen libraries. The performance of DSP was particularly poor in CTL1 samples, which despite showing relatively high abundance of di-nucleosome fragments failed to reproduce enrichment at the position of the TSS flanking nucleosomes, suggesting nucleosome displacement or unspecific tagmentation. Although the cross-link of free amine groups upon DSP fixation allows to preserve tissue and RNA integrity (Mueller *et al.* 2013; Attar *et al.* 2018), the data here presented suggests that this is not sufficient to maintain the chromatin structure of cells likely due to lack of cross-linking between DNA and proteins.

Cryopreservation of PBMCs had historically been used but formal assessment of the effect of this process on the chromatin landscape of different cell types had not yet been conducted. In this chapter cells were recovered in cell culture after thawing and subsequently sorted using FACS, as it has been done historically in the group prior to downstream analysis. As expected, the chromatin structure of the frozen ATAC-seq samples was appropriate and similar to the fresh libraries. PCA demonstrated that global differences in chromatin accessibility due to condition were greater than the differences between individuals, as fresh and frozen samples could be clearly separated in

both cell types. Differential chromatin accessibility analysis revealed a moderate number of DARs between frozen and fresh samples in both cell types, with a greater percentage in $CD14^+$ monocytes compared to $CD4^+$ T cells. This suggested that $CD14^+$ monocytes were more sensitive to the cryopreservation and recovery process than T cells, which was further supported by the enrichment of $CD14^+$ monocytes DARs in biological processes characteristic of cell activation, including cell adhesion and shape, LPS response and inflammation. Cryopreservation of PBMCs using DMSO has been demonstrated to alter transcription of a number of genes involved in stress responses, immune activation and cell death (Yang *et al.* 2016b), in line with the results presented here. Importantly, cryopreservation of monocytes and immature DCs have shown a skewed cytokine profile compared to fresh counterparts (Meijerink *et al.* 2011). The impact of cryopreservation and thawing on the chromatin landscape may have been minimised by avoiding the recovery step in culture and proceeding directly into PBMC staining with FACS antibodies; however additional data would be required to confirm this hypothesis.

It is important to note that the results presented here is limited by the small sample size and borderline quality of some of the libraries, particularly $CD4^+$ fresh samples, likely due to the intrinsic inconsistency issues of the ATAC-seq protocol. This may have impacted to some extent the power to identify DARs between fresh and frozen $CD4^+$ ATAC-seq samples. Nevertheless, this data still provides some useful information regarding the effect of using DSP in sorted cell populations and cryopreservation of PBMCs, which should be considered when planning the experimental design. Following these results and for the purpose of this thesis, fresh isolated cells were used in the experiments presented in Chapters 4 and 5, since the sample size was limited and the main question was to assess the chromatin landscape as close as possible to *in vivo* disease conditions.

Fresh sample processing may not always be feasible, for example in studies aiming to maximise the use of existing biobank cryopreserved samples to achieve large sample sizes or in longitudinal studies looking at the same individual at different points in times (e.g pre- and post- disease or treatment). In such instances, studying chromatin accessibility in cells isolated from cryopreserved PBMCs represents a suitable alternative, where acknowledging that differences may be found when comparing to fresh processed samples can be useful when trying to validate these results. Importantly, the Buenrostro *et al.* 2013 ATAC-seq protocol with some modification has been successfully used in formaldehyde fixed samples and it may represent a better alternative to preserve freshly purifed cell populations to study the epigenetic landscape of clinical samples retrospectively (Corces *et al.* 2017; Chen *et al.* 2016).

3.3.5 Limitations

Several limitations have been described throughout this chapter. Many of the results have been generated in a small number of samples, with low number of replicates, limiting power to identify differences and, in some instances, lacking statistical evidence. Larger sample sizes would enable more robust conclusions about the effect of cryopreservation and DSP fixative in the chromatin accessibility landscape of peripheral blood and a more precise characterisation of alterations that may be of interest when investigating particular biological processes. Similarly, the comparison of ATAC-seq vs Fast-ATAC, regarding effect on the percentage of mitochondrial reads and TSS fold enrichment, was first conducted in one replicate of each protocol (Section 3.2.3) and a comparison in a larger cohort with statistical evidence for the results could only performed at the end of the thesis. The trends observed in the pilot comparison for the effect of the Fast-ATAC protocol in the mitochondrial reads and the signal-to-noise ratios were found to be a good indicator further confirmed

when using the entire thesis cohort sample in retrospective. Interestingly, the results here presented did not agree with one of the claims made by Cores *et al.* 2016 publication, in which this new protocol was supposed to improve the signal in all the hematopoietic cell types. This highlights the importance of testing published protocols in the relevant experimental set up, as discrepancies can be identified.

3.4 Conclusions

The work described here has compared commonly used strategies from ATAC publications in order to establish an appropriate method to perform chromatin accessibility analysis in the context of psoriasis and PsA, maximising the use of samples in a clinical setting whilst accounting for some quality limitations and the resources and expertise in the group available at the time. A robust pipeline was established and conclusions drawn regarding optimised protocols and conditions. As ATAC has become a more commonly used technique, new methods and updates have been introduced both in the experimental protocols and analysis, which means that the results presented throughout this thesis could be revisited in the future with further optimised analytical methods. Moreover, the implementation of Omni-ATAC for future sample recruitment is likely to further improve the quality and confidence in reported differentially open regions.

Chapter 4

Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis

4.1 Introduction

4.1.1 The systemic and skin-specific manifestations of psoriasis

In psoriasis, skin lesions represent the main manifestation of the dysregulated innate immune response triggered by the interaction between genetic and environmental factors (reviewed in Chapter 1). In addition to keratinocytes, circulating immune cells, such as T cells or DCs, are actively recruited to the site of inflammation contributing to disease initiation and progression (Johnson-Huang *et al.* 2009). A number of studies have identified systemic components of psoriasis, including an increase of circulating Th-17, Th-1 and Th-22 cells and an impaired inhibitory function of circulating Tregs (Kagami *et al.* 2010; Sugiyama *et al.* 2005). Autologous activated T cells isolated from patients blood have the ability to induce skin lesions in xenotransplantation models of psoriasis upon injection (Wrone-Smith and Nickoloff 1996; Nickoloff and Wrone-Smith 1999). Moreover, psoriasis patients also have increased risk for PsA as well as for other co-morbidities, such as CVD (Ibrahim *et al.* 2009; Shapiro *et al.* 2007). Increased levels of cytokines including TNF-α, IFN-γ, IL-17A, IL-6, CXCL8 and IL-2, amongst others, have also been reported in psoriasis

serum in a large case control study (Sahmatova *et al.* 2017; Bai *et al.* 2018). In addition to this, in patients with mild-to-moderate psoriasis sub-clinical foci of inflammation at skin, liver, joints, tendons and aorta have been identified when compared to healthy control individuals (Mehta *et al.* 20119). Overall, these findings reinforce the systemic component of psoriasis and highlight the importance of investigating relevant circulating immune cells to better understand disease pathophysiology.

4.1.2 The personalised epigenome in disease

The technical revolution in the epigenetic field has opened avenues to profile the epigenome in clinical samples. Particularly, ATAC-seq and ChIPm have enabled the interrogation of chromatin accessibility, histone modifications and TF binding using only a few thousand cells (Buenrostro *et al.* 2013; Schmidl *et al.* 2015). This has allowed mapping the regulatory landscapes of a wide range of cell types and tissues from clinical samples and has provided disease-specific understanding of the molecular location and status of *cis*-regulatory elements.

ATAC-seq has been used to identify inter- and intra- individual differences and pathological changes in chromatin accessibility (Qu *et al.* 2015). For example, differential analysis in B cells isolated from SLE patients and healthy controls has revealed changes in chromatin accessibility near genes involved in B cell activation and enrichment for TF binding sites potentially regulating pathogenic processes (Scharer *et al.* 2016). Similarly, a study in age-related macular degeneration (AMD) has identified the retinal epithelium as the main tissue driving disease onset through a global loss of chromatin accessibility (Wang *et al.* 2018).

In addition to chromatin accessibility, the characterisation of histone modifications provides further functional information to understand the cell-type specificity of the regulatory landscape. For example, in chronic lymphocytic

leukaemia ChIPm has been used to identify B cell subtype-specific epigenome signatures based on the interrogation of several histone marks (Rendeiro *et al.* 2016). As GWAS SNPs are mostly located in intergenic regions that may act as regulatory elements, assessing the active enhancer mark H3K27ac can be of particular interest here to identify active enhancers in a cell type and context specific manner. However, disease specific data to use in this type of analysis is currently unavailable for most of the complex diseases. A relevant example of enhancer profiling through H3K27ac assay has been conducted in juvenile idiopathic arthritis, where a disease specific H3K27ac super-enhancer (those spanning up to 50Kb) signature has been identified in synovial fluid mCD4$^+$ cells (Peeters *et al.* 2015). Concordantly, inhibitors of histone de-acetylases (HDACs) are being investigated as potential therapeutic agents for RA and SLE, amongst others (Hsieh *et al.* 2014; Shu *et al.* 2017).

4.1.3 Transcriptional profiles in psoriasis

Trancriptomics in psoriatic skin

Characterisation of transcriptional profiles in complex diseases has been performed to better understand disease pathophysiology and assess the role of genetic variability in regulating gene expression. In psoriasis, the majority of transcriptional studies have been performed in inflamed skin (lesional) using pre-lesional (uninvolved) skin, adjacent to the lesion, as the best internal control (Table 4.1). Other studies have also incorporated healthy control skin biopsies to ascertain the extent of dysregulation of the transcriptomic profile prior to lesion development (Table 4.1). Of note, discrepancies in the transcriptional similarities between normal and uninvolved skin have been identified, likely due to different filtering criteria for magnitude of effect (Keermann *et al.* 2015; Tsoi *et al.* 2015b).

The latest transcriptomic studies in psoriasis using RNA-seq have demonstrated greater sensitivity as well as the ability to identify non-coding RNA

species, such as lncRNAs, in an unbiased way (Jabbari *et al.* 2012; Li *et al.* 2014; Tsoi *et al.* 2015b). LncRNAs expression also have shown to have a role in psoriasis pathophysiology, with approximately 1,000 species differentially expressed between lesional and uninvolved skin (Tsoi *et al.* 2015b). However, comparison of protein abundance and differential gene expression in psoriatic skin has revealed that only 5% of the dysregulated transcripts present a similar trend at the protein level (Swindell *et al.* 2015). The majority of those transcriptional studies have been performed in full-thickness skin biopsies containing a mix of tissues from the epidermis, dermis, and basal layer containing keratinocytes, muscle and adipose tissue (Table 4.1). Lately, studies in psoriatic cultured keratinocytes (from lesional and uninvolved biopsies) and epidermis from split-thickness skin grafts have identified differences in gene expression and functional pathway enrichment compared to the studies based on full-thickness skin biopsies (Swindell *et al.* 2017; Tervaniemi *et al.* 2016). These results highlighted the effect of cell type heterogeneity in dissecting the dyregulation of biological processes contributing to skin inflammation in psoriasis.

Table 4.1: Summary table of the most comprehensive transcriptional studies in psoriasis skin and blood. SB= full-thickness skin biopsy; EpB=epidermal biopsy; CK=cultured keratinocytes; C=control; L= psoriatic lesional skin; U=psoriatic uninvolved skin.

Author and year	Sample type and size	Technology	Description
(Jabbari *et al.* 2012)	SB (L=3, U=3)	RNA-seq and microarray	Technology discrepancies
(Li *et al.* 2014)	SB (L=92, C=82)	RNA-seq and microarray	Technology discrepancies and lncRNAs targets co-regulation
(Keermann *et al.* 2015)	SB (L=12, U=12, C=12)	RNA-seq	Dormant psoriasis signature and *IL36* expression in psoriasis skin
(Tsoi *et al.* 2015b)	SB (L=97, U=29, C=90)	RNA-seq	Psoriatic skin-specific new lncRNAs
(Swindell *et al.* 2015)	SB (L=14, U=14)	RNA-seq and mass-spectometry	209 co-regulated mRNA-proteins
(Swindell *et al.* 2017)	CK (L=4, U=4, C=4)	RNA-seq	Decreased differentiation gene signature in lesional skin
(Tervaniemi *et al.* 2016)	EpB (L=6, U=6, C=9)	5'-end RNA-seq	NOD-like and inflammasome pathways
(Coda *et al.* 2012)	PBMCs (PS=6, C=5) and SB (L=5, U=5)	Microarray	Partial overlap between PBMCs and skin DEGs
(Lee *et al.* 2009)	PBMCs (PS=5, C=8)	Microarray	202 DEGs, circulating gene expression signature
(Mesko *et al.* 2010)	PBMCs(PS=15, IBD=12,RA=12, C=18)	TaqMan customised array (96 genes)	6 psoriasis-specific DEGs

(Palau *et al.* 2013)	Activated CD4+$^{+}$ and CD8^{+} (PS=17, C=7)	Microarray	42 DEGs in T cell activation (*SPATS2L* and *KLF6*)
(Jung *et al.* 2004)	IL-10 stimulated PBMCs and CD14^{+} (C=5), IL-10 therapy PBMCs (PS=4)	Microarray	High correspondence between *in vitro* and *in vivo* IL-10 driven DEGs

Transcriptomics in circulating immune cells

A limited number of transcriptional studies comparing circulating immune cells between psoriasis patients and healthy controls have been conducted. The majority of these studies have investigated changes in gene expression between psoriasis and healthy controls in mixed PBMC populations using microarray technologies (Table 4.1).

A study conducted by Coda and colleagues explored the overlap between differentially expressed genes (DEGs) in PBMCs (psoriasis vs controls) and skin (lesional vs uninvolved skin biopsies) (Coda *et al.* 2012). The results revealed a limited overlap with more than 50% of the genes in common presenting opposite directions of modulation between these two tissues. At the cell-type specific level, some studies have performed *in vitro* culture and stimulation of T cells and monocytes (Palau *et al.* 2013; Jung *et al.* 2004). For instance, Palau and colleagues found forty-two DEGs enriched for cytokine and IFN (α, β and γ) signalling pathways when comparing activated $CD4^+$ and $CD8^+$ T cell from psoriasis patients and healthy controls.

4.1.4 Chromatin accessibility, gene expression and genetic variability

As described in Chapter 1, accessible chromatin is more likely to be bound by TFs and other co-regulatory proteins, and so can be used as a proxy to tag genomic loci involved in regulation of gene expression and to infer the putative functional relevance of GWAS SNPs. The orchestration of cell-type specific changes in the chromatin landscape and gene expression is pivotal for an appropriate immune response (Goodnow *et al.* 2005). For example, integration of ATAC-seq data and gene expression in pancreatic islets has revealed chromatin accessibility to be a better predictor for gene activation in

α- compared to β cells, which could be explained by the heterogeneity within each cell population or cell type intrinsic differences in gene regulation. In AMD clinical samples, integration of ATAC and gene expression found moderate correlation between the two in retina and pigmented epithelium retina (Wang *et al.* 2018). In the context of genetic variability, the relationship between chromatin accessibility and gene expression in homeostasis and stimulated conditions has been addressed by integrating eQTL and chromatin accessibility QTLs (ca-QTLs). For example, enhancer priming events have been described in human iPS derived macrophages, where the same genetic variants leads to changes in chromatin accessibility in the naïve state prior to changes in gene expression upon stimulation (Alasoo *et al.* 2018).

4.1.5 Fine-mapping using summary statistics

Generation of cell-type specific epigenetic maps can be used to inform statistical fine-mapping in the effort to identify putative causal SNPs for further investigate through functional follow-up studies. Integration of Bayesian fine-mapping for twenty-one complex immune diseases demonstrated the greatest enrichment of fine-mapped causal variants in immune cell enhancer elements, particularly from activated conditions (Farh *et al.* 2015). In this study, psoriasis PICS showed the most significant enrichment for Th-1, Th-2 and Th-17 subsets. Furthermore, exhaustive fine-mapping using a customised genotyping array has been conducted for eight psoriasis GWAS loci using a frequentist approach which measure the association of each SNP through p-values, finding additional signals at the MHC and *IL12B* regions and a number of non-coding variants overlapping chromatin segmentation maps NHEKS enhancers (Das *et al.* 2014).

Traditional Bayesian fine-mapping requires GWAS genotyping data to perform genotype phasing and imputation prior to association analysis and calculation of posterior probabilities (PP) and a definition of credible sets of

SNPs. Restricted access to GWAS genotyping data, commonly due to ethical reasons, can be a limitation when performing this type of analysis. A number of methods have been developed to use the summary statistics form the association analysis of genotyped SNPs to then impute the statistics of additional SNPs not typed in the array (Pasaniuc and Price 2017). These methods represent an advantage since they are computationally more cost-effective and summary statistics for GWAS studies are widely available. One of the most widely-used methods is direct imputation of summary statistics for unmeasured SNPs (DIST) (Lee *et al.* 2013), yielding successful results, for example, in schizophrenia (Edwards *et al.* 2015).

4.2 Aims

The aim of this chapter is to determine chromatin accessibility, histone modification and gene expression differences between psoriasis patients and controls in four circulating immune cell types ($CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells and $CD19^+$ B cells) and to complement this with analysis of differential gene expression in lesional and uninvolved epidermis isolated from psoriatic skin biopsies. The long term goal is to identify disease and cell-type specific changes in putative regulatory regions and integrate them with observed differences in gene expression to improve the understanding of systemic and skin inflammatory features of psoriasis and prioritise putative causal GWAS variants.

The specific aims for this chapter are:

1. To identify differences in chromatin accessibility and the H3K27ac active enhancer mark modifications between psoriasis patients and healthy controls in immune cells isolated from peripheral blood.

2. To determine changes in genes expression between psoriasis patients and healthy controls in immune cells isolated from peripheral blood.

3. To identify differentially expressed genes between lesional and uninvolved epidermis isolated from psoriatic skin biopsies.

4. To compare the differentially expressed genes between patients and controls in circulating immune cells with those found when contrasting lesional and uninvolved epidermis from patients.

5. To conduct fine-mapping analysis for a number of psoriasis GWAS loci using summary statistics.

6. To integrate the fine-mapped credible set of SNPs with disease and cell-type specific epigenetic maps, gene expression profiles and publicly available data to narrow down the putative causal variants at GWAS risk loci.

4.3 Results

4.3.1 Psoriasis and healthy controls: cohort description and datasets

Peripheral blood samples were collected from a cohort of psoriasis patients and healthy individuals in order to isolate four relevant immune cells types ($CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$) and perform ATAC-seq, RNA-seq and ChIPm analyses. Additionally, epidermis from paired uninvolved and lesional skin biopsies collected from three psoriasis patients were processed for RNA-seq.

A total of eight psoriasis patients, six males and two females (Table 4.2) were recruited following eligibility criteria detailed in Chapter 2. The mean age of the cohort was 55 years old and the mean disease duration 331.5 months. All patients showed active skin disease and none of them had reported joint involvement at the time of sample collection. Mean disease severity quantified using the PASI score (reviewed in Chapter 1) was 10. Currently, there is no consensus on

PASI thresholds to define mild and moderate-to-severe disease, with a review study suggesting to consider chronic plaque psoriasis as moderate when PASI ranges between 7 to 12, and severe for PASI>12 (Schmitt and Wozel 2005). On the other hand, NICE and other studies defined psoriasis as severe based on PASI≥10 (Woolacott *et al.* 2006; Finlay 2005). In this cohort, six out of eight patients had PASI≥10, and so were categorised as having severe psoriasis and two had PASI<7 showing a milder phenotype. All patients were naïve for biologic therapies. PS2319 was currently on methotrexate therapy and the remaining patients have only been treated occasionally with topical steroids or UVB therapy. Interestingly, PS2014 showed the most severe PASI score (17) and was a non-responder to methotrexate in the past.

Table 4.2: Description and metadata of the psoriasis patients cohort. For each of the individuals information relating to sex, age, disease duration (in months), PASI score, nail involvement and family history has been recorded. Patients are divided into cohort 1A and cohort 1B based on the timing (batch) of ATAC and RNA-seq processing and type of ATAC-seq protocol applied. PASI score is detailed in Table 1.1. Available datasets from peripheral blood isolated cells (ATAC, ChIPm, RNA-seq) and skin biopsies (skin RNA-seq) are indicated for each sample. The skin RNA-seq samples include lesional and uninvolved paired-skin biopsies from each of the three individuals.

Sample ID	Sex	Age (years)	Disease duration	PASI	Nails affected	Family history	ATAC	ChIPm	RNA-seq	Skin RNA-seq
Cohort 1A										
PS1011	Male	55	420	11	Yes	No	Yes	No	Yes	Yes
PS2014	Female	65	588	17	No	No	Yes	No	Yes	No
PS2015	Male	56	384	5	Yes	No	Yes	No	Yes	Yes
PS2016	Male	40	180	10	No	No	Yes	No	Yes	Yes
Cohort 1B										
PS2000	Male	61	156	10	No	Yes	Yes	Yes	Yes	No
PS2001	Male	56	432	10	Yes	No	Yes	Yes	Yes	No
PS2314	Male	42	120	6.5	Yes	No	Yes	Yes	Yes	No
PS2319	Female	64	372	10.2	No	Yes	Yes	Yes	Yes	No
Mean (±SD)	–	55 (±9.4)	331.5 (±163.3)	10 (±3.5)	–	–	–	–	–	–

Patients PS1011, PS2015, PS2001 and PS2314 presented nail pitting, which has been defined as one of the markers for increased risk of developing joint involvement and PsA (Moll *et al.* 1973; Griffiths and Barker 2007; McGonagle *et al.* 2011). A family history of psoriasis was reported by PS2000 and PS2319. In addition to the psoriasis samples, peripheral blood was collected from ten sex and age-matched healthy individuals (Table 4.3).

Sample ID	Sex	Age (years)	ATAC	ChIPm	RNA-seq
Cohort 1A					
CTL1	Male	36	Yes	No	Yes
CTL2	Male	53	Yes	No	Yes
CTL3	Male	34	Yes	No	Yes
CTL4	Female	46	Yes	No	Yes
CTL5	Male	42	Yes	No	Yes
Cohort 1B					
CTL6	Male	31	Yes	Yes	Yes
CTL7	Male	57	Yes	Yes	Yes
CTL8	Female	50	Yes	Yes	Yes
CTL9	Male	50	Yes	Yes	Yes
CTL10	Male	67	Yes	Yes	Yes
Mean (±SD)	–	46.6 11.2			

Table 4.3: Description of the healthy control cohort. Controls are divided in cohort 1A and cohort 1B based on the timing (batch) of ATAC and RNA-seq processing and type of ATAC protocol, similarly to the psoriasis patients samples. For each of the samples, availability of ATAC, ChIPm and RNA-seq generated from peripheral blood isolated cells are indicated.

For cohort 1A (Tables 4.2 and 4.3) ATAC data was generated using the ATAC-seq protocol from Buenrostro *et al.* 2013, and this was later replaced by the Fast-ATAC method from Corces and colleagues (Corces *et al.* 2016) in cohort 1B, due to the improvements of this protocol explained in Chapter 3. Additionally, samples from cohort 1B were also assayed for differences in H3K27ac patterns using ChIPm. For three psoriasis patients (PS2014, PS2015 and PS2016) paired biopsies from lesional and uninvolved skin were collected and the epidermal sheets were isolated to perform RNA-seq (Table 4.2). This

was a pilot study aiming to refine the previous RNA-seq studies performed in full-thickness skin biopsies, with a more heterogeneous cell type composition compared to epidermis, which could not be expanded due to time constraints.

4.3.2 Investigation of psoriasis-specific changes in the enhancer mark H3K27ac in peripheral blood immune cell populations

Data processing and quality control

A total of 32 ChIPm libraries from four patients and four controls in four immune cell types were sequenced, and reads filtered as detailed in Chapter 4. After filtering, the total number of reads ranged between 46.9 and 60.5 million, compliant with the 40 million total reads recommended by ENCODE (Figure 4.1A). Based on ENCODE recommended non-redundant fraction and PCR bottlenecking coefficients PBC1 and PBC2 (see Section 2.4.2), most of the libraries passed these filtering criteria (Table A.3). Only $CD8^+$ CTL7 and the $CD19^+$ PS2000 and PS2314 libraries failed the recommended complexity according to non-redundant fraction (NRF) values and showed more severe PCR bottlenecking (based on NRF<0.5 and PBC1<0.5, Table A.3), consistent with a greater number of duplicates (>50% of the total sequenced reads) and lower number of reads after filtering when compared to the other libraries (Figure 4.1A). ChIPm libraries had appropriate signal-to-noise ratios based on the ENCODE standards (Landt *et al.* 2012) for the NSC and RSC coefficients (above 1.05 and 0.8, respectively) calculated using cross-correlation analysis (Figure 4.1).

PCA per individual cell type identified $CD8^+$ CTL7 and PS2319 and $CD19^+$ PS2000 and PS2314 libraries as outliers when compared to the rest of the samples from the same cell type (data not shown). Due to the small sample size and the potential effect of borderline quality control samples and outliers in the differential analysis, $CD8^+$ CTL7 and PS2319 and $CD19^+$ PS2000 and

PS2314 libraries were removed from the downstream analysis. Additionally, PCA using a combined list of consensus H3K27ac called peaks in the 29 samples, which included patients and controls in all four cell types (excluding the aforementioned low quality samples), confirmed appropriate quality of this data to identify cell-type specific differences in the enhancer landscape (Figure 4.2A).

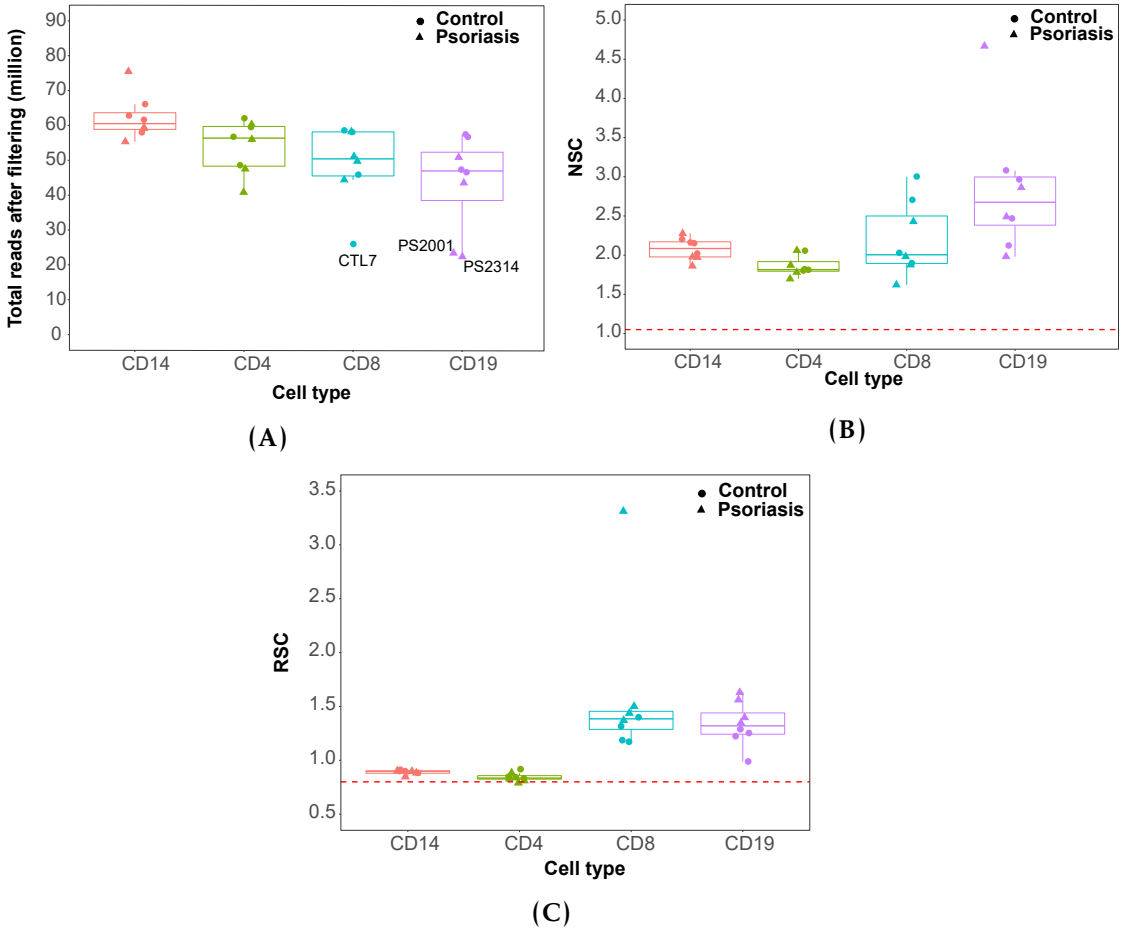


Figure 4.1: Quality control evaluation of the H3K27ac ChIPm libraries in immune cells isolated from psoriasis and control samples. For each of the cell types boxplots representing (A) million of reads after filtering, (B) normalised strand cross-correlation coefficient (NSC) and (C) relative strand cross-correlation coefficient (RSC). NSC and RSC are measures of signal enrichment independent of peak calling, where 1 and 0 indicate no enrichment, respectively. In (B) and (C) the dashed red line indicates the ENCODE threshold for low enrichment (NSC<1.05 and RSC<0.8). For each point, colour codes for cell type and shape for phenotype (psoriasis or control).

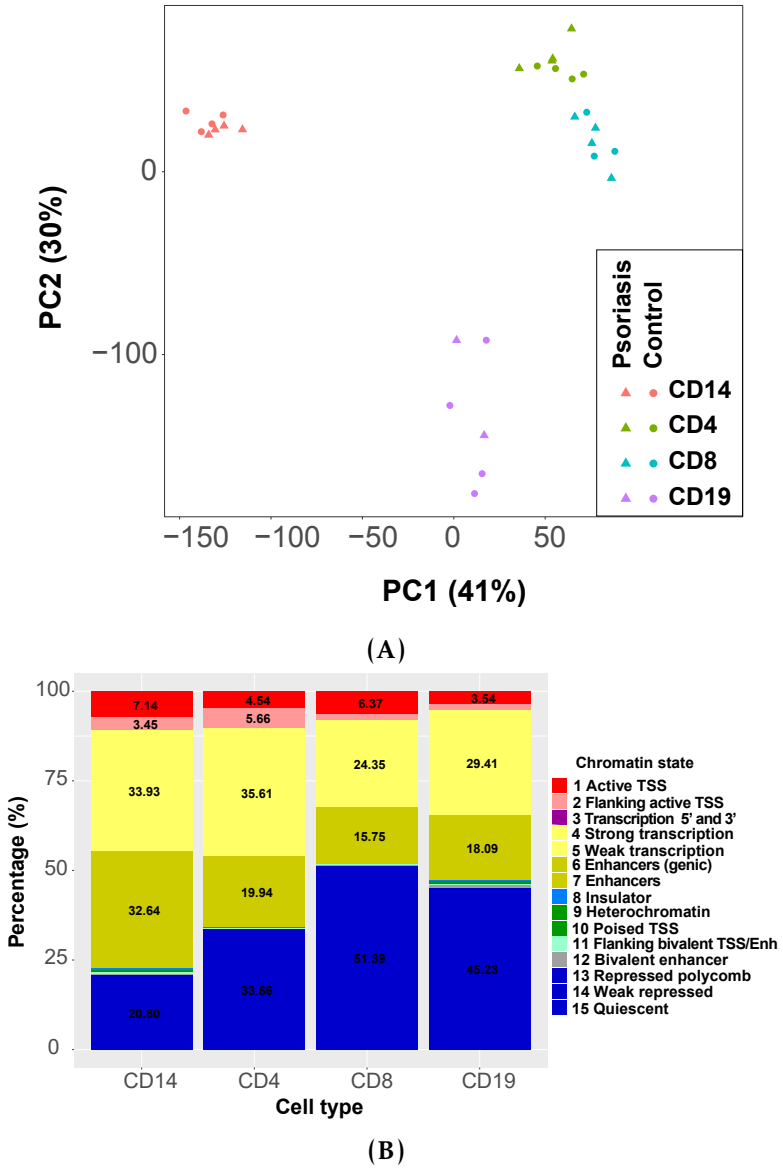


Figure 4.2: PCA and chromatin annotation states of the consensus list of H3K27ac called peaks in four immune primary cell types from psoriasis and healthy control samples. (A) PCA was performed using the normalised counts across a consensus master list of the combined H3K27ac enriched regions in psoriasis patients and healthy control samples across $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. (B) Annotation of the H3K27ac list of consensus enriched sites built by DiffBind for each cell was performed using the appropriate cell-type specific Roadmap Epigenomics Project chromatin segmentation maps. Results are expressed as the percentage of regions annotated with a particular chromatin state over the total number of H3K27ac enriched sites in each individual cell type master list.

H3K27ac differential analysis

An exploratory analysis was conducted to find differences in H3K27ac patterns between patients and controls using DiffBind in each cell type. DiffBind assembled a consensus list of H3K27ac peaks used to perform differential analysis (as explained in Chapter 2 and Table A.2). Annotation with chromatin states of each cell type consensus list of H3K27ac peaks showed a high percentage of sites annotated as repressed and quiescent (Figure 4.2B), ranging from 20.8% in $CD14^+$ monocytes to 51.39% in $CD8^+$ cells. Such sites are less likely to be relevant since H3K27ac is a histone modification mainly enriched at enhancer elements. Therefore, the differential analysis for H3K27ac modifications between psoriasis and healthy control samples in each cell type was restricted to those H3K27ac peaks annotated as enhancers or transcribed (Table A.2). $CD14^+$ monocytes had the greatest number of differentially modified enhancers (8 significant sites), followed by $CD4^+$ (4) and $CD8^+$ (1) (Table 4.4).

Cell type	Differential H3K27ac genome-wide analysis	Differential H3K27ac enhancers analysis
$CD14^+$	15	8
$CD4^+$	0	4
$CD8^+$	8	1
$CD19^+$	12	0

Table 4.4: Summary results from the differential H3K27ac analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. Differential analysis was performed genome-wide in all the DiffBind H3K27ac consensus sites or only at the consensus sites annotated as enhancers (according to the chromatin segmentation map from Roadmap Epigenomics Project). Genome-wide differential significant sites in $CD14^+$ monocytes and $CD8^+$ also contain the sites identified in the enhancer restricted analysis. Significant differentially H3K27ac modified regions were determined using FDR<0.05 and no fold change threshold.

When comparing patients vs controls, several H3K27ac differentially modified regions were proximal to genes relevant in chronic inflammation. For example, a differential H3K27ac region in $CD14^+$ monocytes was located between

the *SLC15A2* and *ILDR2* genes (Figure 4.3). *ILDR2* has recently been identified as relevant for negative regulation of T cell response in RA (Hecht *et al.* 2018). This region showed lower H3K27ac levels in psoriasis patients compared to controls and was annotated as an enhancer by the Roadmap Epigenomics Project chromatin segmentation map. Additionally this site was overlapping a DHS and H3K4me1 (enhancer mark) modification and showed a CTCF-binding site in K562 cells.

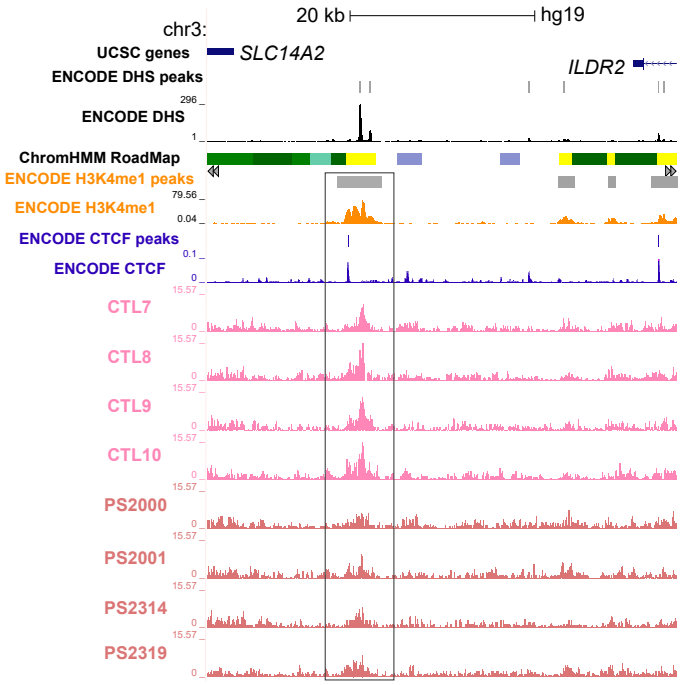


Figure 4.3: Differential H3K27ac modification at a putative intergenic enhancer region in circulating CD14$^+$ monocytes between psoriasis patients and healthy controls. UCSC Genome Browser view illustrating the normalised H3K27ac fold-enrichment (y-axis) at an intergenic differentially modified region located between *SLC14A2* and *ILDR2* genes (x-axis) in CD14$^+$ monocytes (lower H3K27ac enrichment in psoriasis patients compared to healthy controls). CD14$^+$ monocytes publicly available epigenetic data from ENCODE (including DHS, H3K4me1 and CTCF ChIP-seq) and the Roadmap Epigenomics Project chromatin segmentation track are also shown. Differential H3K27ac modified regions were considered significant based on FDR<0.05 and no fold change cut-off. H3K27ac tracks are colour-coded by condition: control(CTL)=pink and psoriasis (PS)=sienna.

In addition to the restricted analysis at enhancers and transcribed regions, genome-wide contrast between psoriasis and control samples revealed additional H3K27ac differential regions in CD14$^+$ monocytes, CD8$^+$ and CD19$^+$ cells, which

also included those already identified in the restricted enhancer analysis (Table 4.4). Comparison of both approaches revealed that restricting the analysis to enhancer and transcribed annotated regions based on chromatin segmentation maps did not significantly increase the number of observed differentially modified H3K27ac sites when compared to the genome-wide analysis in any of the four cell types. Altogether, the results presented in this pilot cohort did not show relevant global epigenetic changes in H3K27ac sites between psoriasis patients and controls for these cell types and sample size.

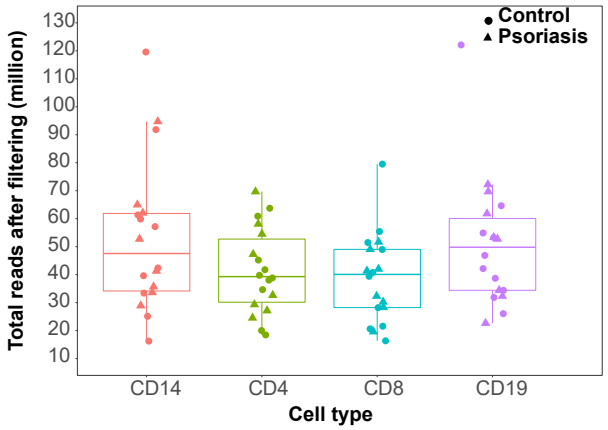
4.3.3 Interrogation of chromatin accessibility changes in psoriasis peripheral blood immune cells

In order to interrogate genome-wide changes in chromatin accessibility between psoriasis patients and healthy controls, ATAC (ATAC-seq or Fast-ATAC) was performed in the same four cell types in eight patients and ten controls (Tables 4.2 and 4.3) yielding a total of 72 libraries.

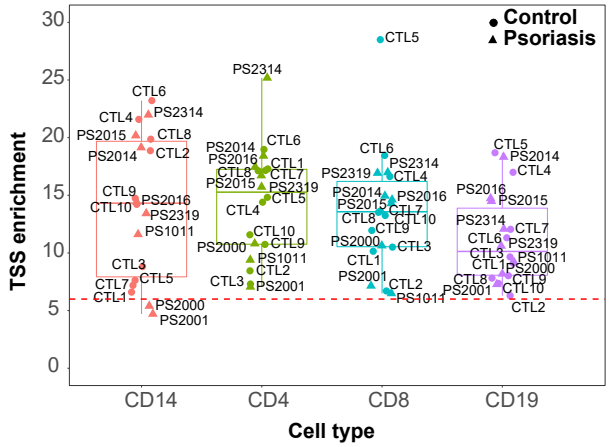
Data processing and quality control

The median total reads after filtering for all the ATAC libraries ranged between 39.2 and 49.8 million, above the 15 million reads determined as the minimum depth for downstream analysis (detailed in Chapter 3) (Figure 4.4A). All samples showed the required characteristic ATAC-seq fragment size distribution recapitulating nucleosome periodicity (not shown) and differences in the percentage of mitochondrial reads between ATAC-seq and Fast-ATAC libraries were also found (detailed in 3 and Figure B.6A). Analysis of ATAC signal enrichment across gene TSSs revealed that most samples had enrichment over 6 (Figure 4.4B). $CD14^+$ PS2000 and PS2001 monocytes showed TSS<6 and were removed from downstream analysis together with $CD19^+$ CTL2, which TSS was borderline.

Regarding the number of called peaks, most samples showed between 10,000 and 65,000 peaks after IDR filtering (Figure B.6B) and the majority of differences in numbers were intrinsic to cell type and signal-to-noise variability across libraries (Figure 4.4C). For example, CD19^{+} CTL2 showed lower number of filtered peaks compared to other samples with similar sequencing depth but larger signal-to-noise enrichment, supporting removal of this sample from downstream analysis.

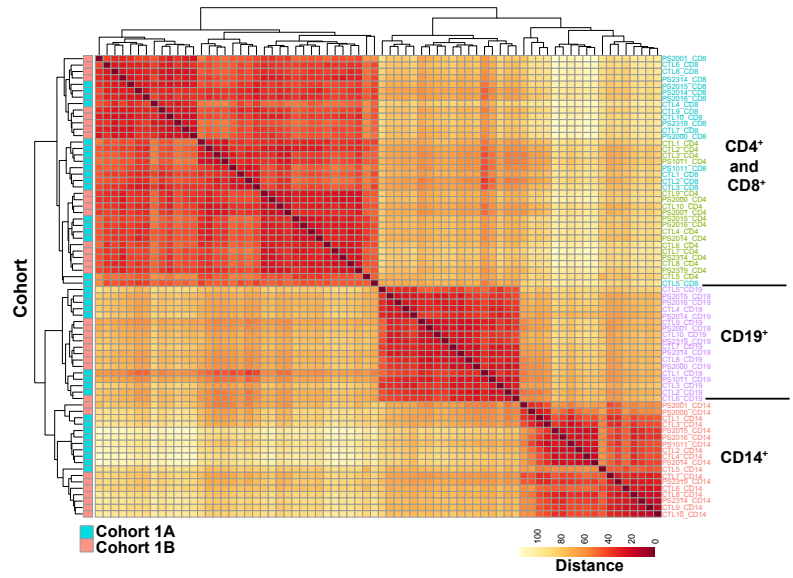


(A)

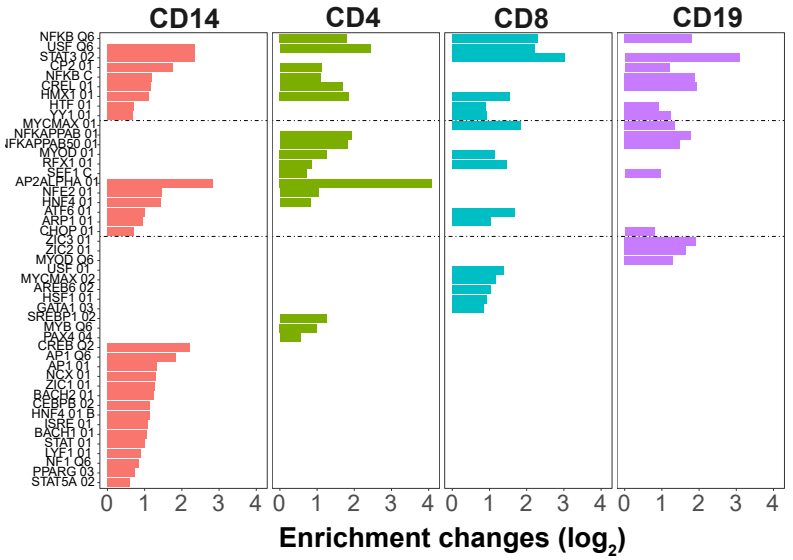


(B)

Figure 4.4: Quality control assessment of the ATAC libraries generated from circulating immune cells in psoriasis and control samples. For each of the cell types and samples, boxplots showing (A) million of reads after filtering and (B) values for fold-enrichment of ATAC fragments across the Ensembl annotated TSS. In (B) the dashed red line indicates the recommended ENCODE threshold for TSS enrichment values. For each point, colour codes for cell type and shape for phenotype (psoriasis or control). In (B) sample IDs are included.



(A)



(B)

Figure 4.5: Clustered heatmap and conserved TF binding sites enrichment analysis in the consensus list of called ATAC peaks in CD14$^+$ monocytes, CD4$^+$, CD8$^+$ and CD19$^+$ cells from the patients and controls cohort. (A) Distance matrix and hierarchical clustering for the 72 samples was performed based on the normalised read counts retrieved for each sample at the regions included in a combined consensus list of called ATAC peaks across all four cell types (CP_all). Clusters have been additionally annotated using cohort identity. (B) Enrichment analysis for the conserved TF binding sites was performed for each of the consensus list of ATAC peaks per cell type used for downstream differential analysis between psoriasis and control individuals (named here as CP_CD14, CP_CD4, CP_CD8, and CP_CD19). Enrichment was tested for 258 human conserved TF binding sites identified by Transfac using position-weight matrices based on experimental results in the scientific literature. Significant enrichment using FDR<0.01.

A heatmap illustrating sample distance using read counts retrieved at the combined consensus list of called ATAC peaks across all the samples (named here as CP_all) showed successful separation of the samples according to cell type into three main clusters corresponding to $CD14^+$ monocytes, $CD19^+$ and $CD4^+/CD8^+$ T cells (Figure 4.5A). Within each cell type cluster, samples did not separate based on disease condition, which could suggest an absence of major global differences in the chromatin accessibility landscape between psoriasis patients and healthy controls, but some grouping by batch was observed (Figure 4.5A).

PCA per individual cell type confirmed batch effect between cohort 1A and cohort 1B ATAC samples, likely driven by the use of ATAC-seq and Fast-ATAC protocols in each batch, respectively (representative example in $CD8^+$ cells Figure B.7A). This analysis also revealed CTL5 as an outlier sample in cohort 1A for all the cell types and was also removed from the differential analysis.

Differential chromatin accessibility analysis

To perform differential chromatin accessibility analysis between patients and controls, a consensus list of called ATAC peaks (accessible chromatin regions) was built for each of the four cell types (named here as CP_CD14, CP_CD4, CP_CD8, and CP_CD19) (detailed in Chapter 2). To determine the functional relevance of this consensus list of peaks used in the downstream comparison between patients and controls, enrichment analysis for conserved TF binding sites was conducted. All four cell types consensus list of peaks (CP) showed significant enrichment (FDR<0.01) for a number of TF binding sites (Figure 4.5B). For example, enrichment for conserved NFκB binding motifs was identified across the four cell types consensus list of ATAC peaks. Conserved binding motifs for TF involved in T cell biology, such as AREB6 (ZEB1), ATF6 and heat-shock transcription factor HSF1 were enriched in the CP_CD8 (Guan *et al.* 2018; Yamazaki *et al.* 2009; Gandhapudi and Murapa 2013). Additionally,

peaks from each cell type consensus list also showed enrichment for relevant cell type *cis*-eQTLs. For example, eQTLs from unstimulated monocytes were the top significantly enriched (FDR<0.01) in the CP_CD14 (fold-enrichment 5.1 from Fairfax *et al.* 2014 data) from a number of immune cell eQTLs. These observations reinforced the functional relevance of the accessible chromatin regions further investigated in psoriasis and healthy control individuals.

The differential chromatin accessibility analysis between patients and controls was performed on the ATAC normalised read counts of each cell type consensus list using DESeq2 and including the ATAC protocol as a covariate to account for the aforementioned batch effect. Genome-wide differential chromatin accessibility analysis in $CD8^+$ cells revealed 55 significant (FDR<0.05) differentially accessible regions (DARs) between psoriasis patients and healthy controls (Table 4.5), of which 17 showed FDR<0.01. In contrast, $CD14^+$ monocytes, $CD4^+$ and $CD19^+$ cells only showed one or no DARs at this level of significance.

Cell type	Number of DARs FDR<0.05
$CD14^+$	1
$CD4^+$	0
$CD8^+$	55
$CD19^+$	1

Table 4.5: Summary results from the differential chromatin accessibility analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. The number of DARs refers to those statistically significant when using a cut-off for background reads of 80% (see Chapter 3) and a FDR<0.05. No threshold for the fold change was applied.

Annotation of the 55 $CD8^+$ DARs using cell-type specific Roadmap Epigenomics Project chromatin segmentation maps revealed the potential for some of those regions to be involved in regulation of gene expression, including 26 (48.1%) enhancers, 7 (12.9%) active promoters and 6 (11.1%) transcribed

regions. The functional relevance of DARs in terms of regulation of gene expression was further investigated by integration of the CD8$^+$ T cell eRNA data from the FANTOM5 project. Of the CD8$^+$ DARs, 8 overlapped significantly expressed eRNAs in the same cell type. These included a region at the TSS of the *TNFSF11* (or *RANKL*) gene and another upstream of the *IL7R* promoter, both more accessible in psoriasis patients compared to the healthy controls (Figure 4.6A and B). The two DARs also overlapped chromatin harbouring H3K4me3, a histone mark indicating an active promoter, and H3K27ac consistent with the transcription of those regions as eRNAs.

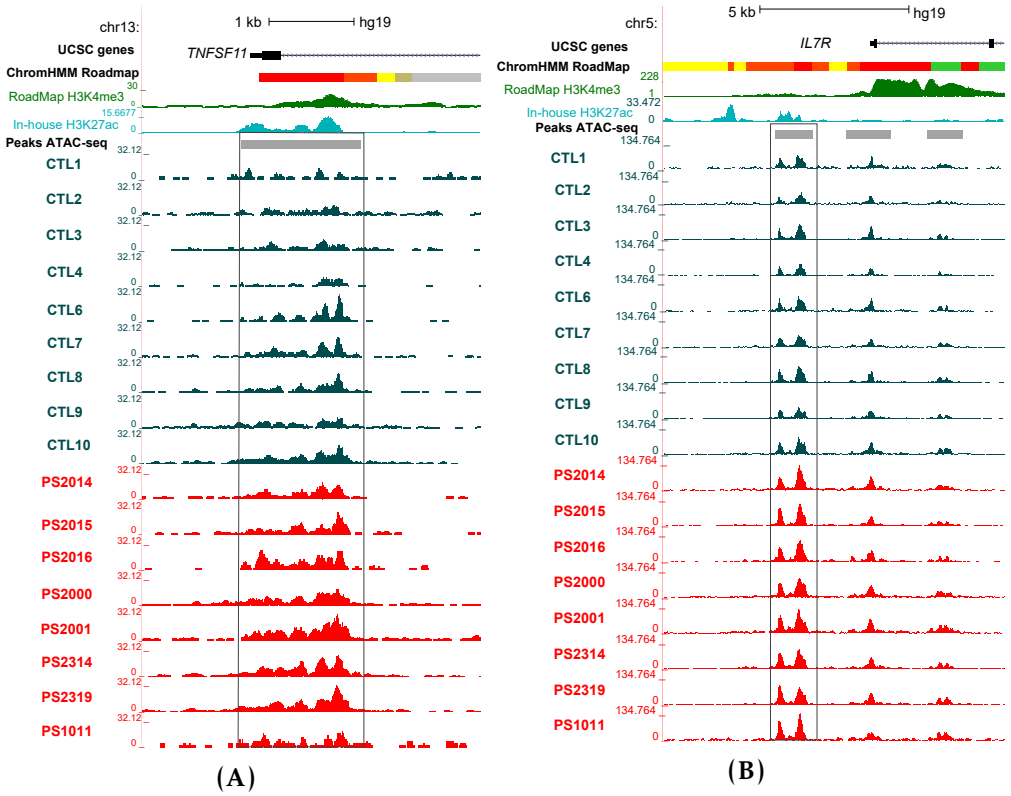


Figure 4.6: Epigenetic landscape at two ATAC differential accessible regions between patients and controls in CD8$^+$ cells. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) in DARs located at (A) the promoter of *TNFSF11* gene and (B) up-stream the *IL7R* gene (x-axis). Both DARs were more open in CD8$^+$ cells from psoriasis compared to controls. Tracks are colour-coded by condition: control (CTL)=dark turquoise and psoriasis (PS)=red. The Roadmap Epigenomics Project chromatin segmentation map and H3K4me3 for CD8$^+$ cells are also shown, together with a representative track from the in-house ChIPm H3K27ac in this cell type. All DARs were significant based on FDR<0.05 and no fold change cut-off.

Other interesting $CD8^+$ DARs were found nearby *MAP3K7CL* and *NFKB1* (encoding for the p105/p50 subunits) genes, but they did not co-localised with regions annotated as enhancers or transcribed into eRNAs.

Integration of H3K27ac ChIPm and ATAC-seq chromatin accessibility profiles

Integration of the observed H3K27ac ChIPm and ATAC differential analysis between patients and controls only showed overlap for one region in $CD8^+$ cells. Namely, an intron of the D-tyrosyl-tRNA deacylase 1 (*DTD1*) gene showed lower levels of H3K27ac and chromatin accessibility in psoriasis patients when compared to healthy controls (Figure 4.7). This differential region was annotated as an active enhancer by the Roadmap Epigenomics Project segmentation map and did not interact with the promoter of any gene in $CD8^+$ cells according to Hi-C and promoter Hi-C data (Javierre *et al.* 2016). Interestingly, SNPs within this region are reported to be eQTL for *DTD1* in whole blood (`https://gtexportal.org/home/eqtls`).

4.3.4 Differential gene expression analysis in circulating immune cells in psoriasis

Data processing and quality control

In addition to characterising the chromatin accessibility landscape, gene expression profiles in psoriasis and healthy individuals were analysed by RNA-seq in the same four primary circulating immune cell types (Table 4.2). All 72 samples showed percentages of RNA-seq reads mapping to a unique location in the genome above the recommended 70-80% (Figure B.8A). After filtering, all samples had at least 20 million reads (as required by ENCODE standards) mapping to a comprehensive list of Ensembl features, including protein coding genes and lncRNAs (Figure B.8B). In all four cell types, greater mapping rates and total reads mapping to Ensembl features were observed for cohort 1B samples

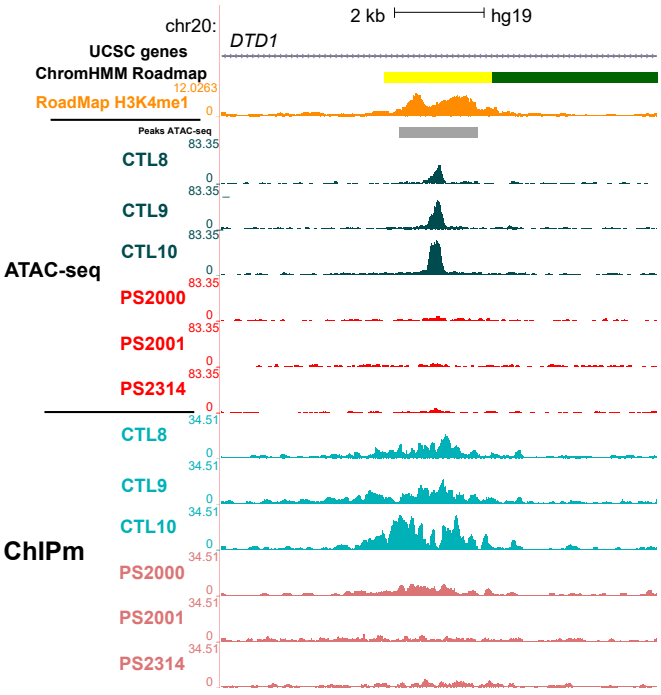


Figure 4.7: Epigenetic landscape at *DTD1* locus showing differential H3K27ac and chromatin accessibility between psoriasis patients and controls in CD8$^+$ cells. UCSC Genome Browser view illustrating the normalised ATAC read density and H3K27ac normalised fold-enrichment (y-axis). Tracks are colour-coded by condition and assay: control (CTL)=dark and light turquoise and psoriasis (PS)=light and dark red, for ATAC and ChIPm respectively. The Roadmap Epigenomics Project chromatin segmentation map and H3K4me1 for CD8$^+$ cells are also shown.

when compared to cohort 1A. These differences were attributed to the library preparation and sequencing of each cohort in two batches.

PCA using normalised reads showed that most of variability was driven by cell type differences, with three main clusters corresponding to CD14$^+$ monocytes, CD4$^+$ and CD8$^+$ T cells, and CD19$^+$ cells (Figure 4.8). Within each cell type cluster, samples were further grouped by cohort (1A and 1B) but not by condition (psoriasis and control) (Figure B.10). In fact, PC4 (explaining 3% of the total variance) from the PCA correlated with batch and clearly separated the 72 samples into cohort 1A and 1B (Figure B.7B).

mRNA and lncRNA differential expression analysis

Differential gene expression analysis between eight psoriasis patients and ten healthy controls was performed using DESeq2 and including cohort identity

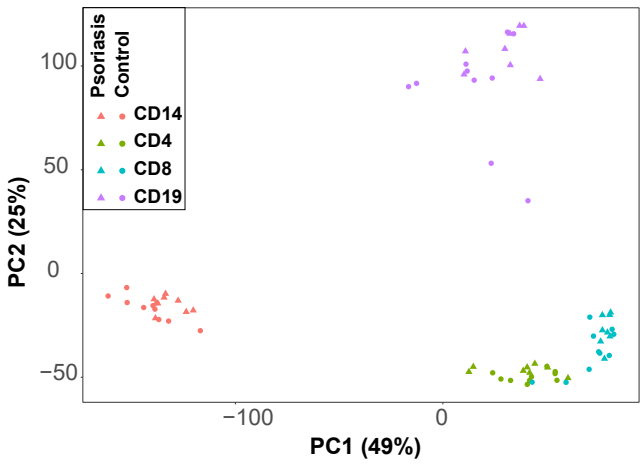


Figure 4.8: PCA illustrating the sample variability based on the gene expression profiles for all 72 samples. The first and second PCs (x-axis and y-axis, respectively), where each point represents a sample, colour coding for cell type and the shape for cohort (batch). The proportion of variation explained by each principal component is indicated.

as a covariate to account for the aforementioned batch effect. For each cell type, differentially expressed mRNAs were identified using two significance thresholds: FDR <0.05 and FDR <0.01 (Table 4.6). $CD14^+$ monocytes and $CD8^+$ T cells showed the largest number of differentially expressed mRNAs between psoriasis patients and controls with the highest magnitude of fold change (Figure 4.9).

$CD14^+$ monocytes and $CD4^+$ T cells presented similar numbers of genes up-regulated and down-regulated in psoriasis patients when compared to the healthy controls. In contrast, for $CD8^+$ and $CD19^+$ cells a larger number of modulated genes were down-regulated in patients compared to controls. In $CD14^+$ monocytes, the most significant up-regulated genes (fold change>1.5) included *SDC2*, *CD83* and *MIR22G* (Figure 4.9A). In $CD8^+$, *SOCS3* and *CXCR4* were amongst the top differentially expressed genes with fold change>1.5 (Figure 4.9C), whereas in $CD4^+$ cells, the G protein-coupled receptor *GPR18* showed the most significant down-regulation with a fold change=1.4. Overall, the dysregulation of gene expression in psoriasis patients compared to controls in all four analysed cell types only was demonstrated to be moderate.

Cell type	mRNA FDR<0.05/0.01	lncRNA FDR<0.05/0.01	Up-regulated FDR<0.05/0.01	Down-regulated FDR<0.05/0.01
$CD14^+$	671/229	28/8	331/112	368/125
$CD4^+$	108/40	12/4	56/20	64/24
$CD8^+$	651/175	31/5	269/67	418/113
$CD19^+$	167/71	6/2	29/13	144/60

Table 4.6: Summary results from the differential gene expression analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. The number of statistically differentially expressed mRNAs and lncRNAs are listed for two FDR threshold (FDR<0.05 and FDR<0.01). No threshold for the fold change was applied in this analysis. For each of the FDR threholds the number of up- and down-regulated genes is included.

None of the genes coding for well-known psoriasis drug targets, such as TNF-α and IL-17 or key cytokines including IL-6 and IL-36, were up-regulated in any of the four cells types in psoriasis patients compared to healthy controls. $CD8^+$ showed the largest number of significant DEGs (FDR<0.05) overlapping putative psoriasis GWAS genes from the NHGRI-EBI catalog (`https://www.ebi.ac.uk/gwas`) (Table 4.7, 7 hits), followed by $CD14^+$ monocytes and $CD4^+$ cells (3 hits each). Some of the GWAS genes were found to be differentially expressed in more than one cell type, including *NFKBIA*, *TNFAIP3*, *B3GNT2* and *NFKBIZ* (Table 4.7). Enrichment of psoriasis GWAS genes amongst the DEGs in patients compared to controls was statistically significant only for $CD4^+$ and $CD8^+$ T cells (Fisher exact test, p-values $2.6\text{x}10^{-3}$ and $9.36\text{x}10^{-4}$, respectively).

The role of lncRNAs in psoriasis circulating immune cells

DEGs between psoriasis patients and controls also included lncRNAs in all four cell types. $CD8^+$ and $CD14^+$ monocytes showed the largest number of dysregulated lncRNAs between psoriasis patients and controls (Table 4.6). However, $CD4^+$ T cells presented the largest proportion of lncRNAs over the total number of DEGs (11% for an FDR<0.05).

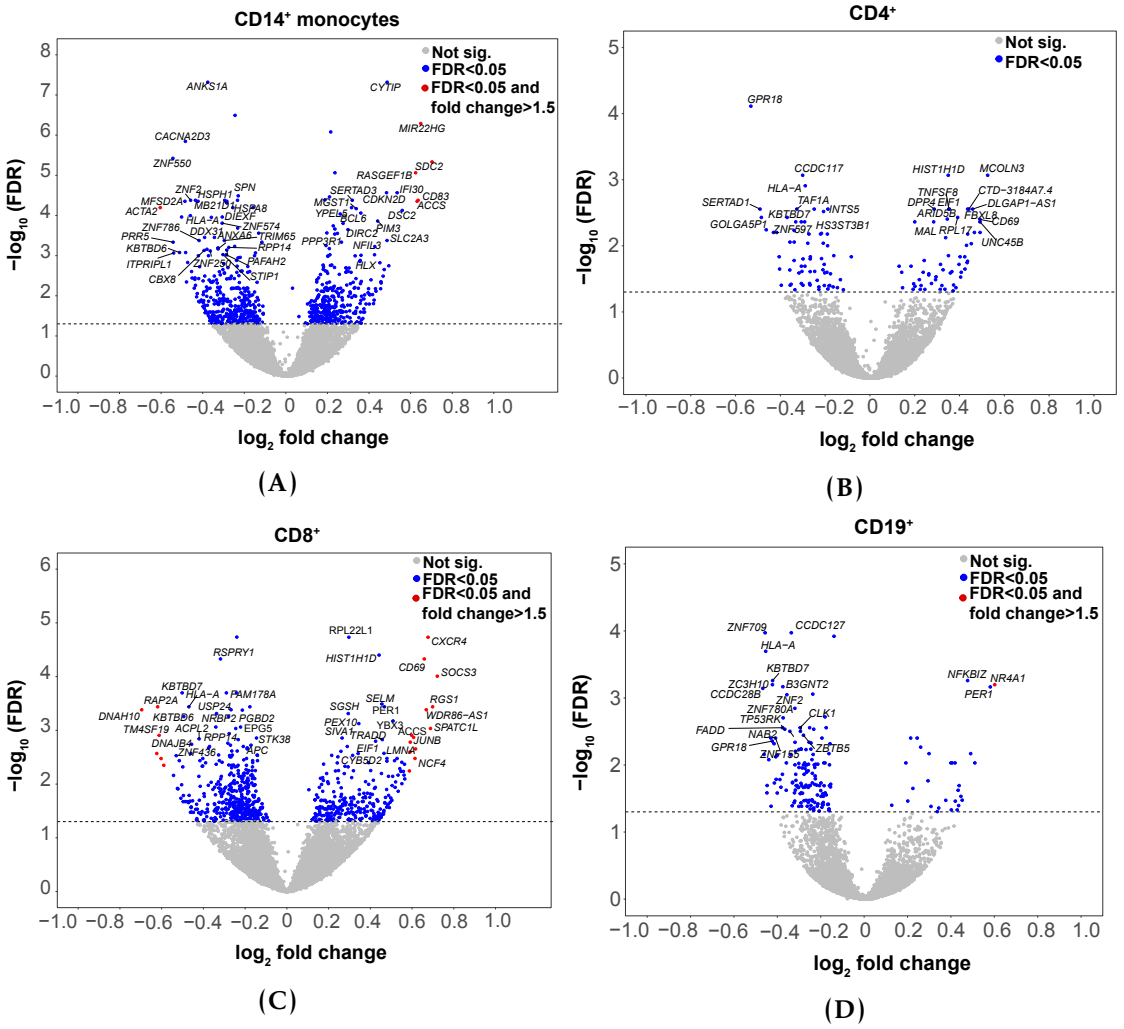


Figure 4.9: Magnitude and significance of gene expression changes between psoriasis patients and healthy controls in four immune cell types. Volcano plots illustrating the differential gene expression analysis results for (A) $CD14^+$ monocytes, (B) $CD4^+$, (C) $CD8^+$ and (D) $CD19^+$ cells. The log_2fold change represents change in expression for each gene in the psoriasis group relative to the healthy controls. Significant DEGs (FDR<0.05) in blue for fold change<1.5 and red for fold change >1.5. The volcano plots include mRNAs and lncRNAs species.

Cell type	Number of GWAS overlaps	Up-regulated genes	Down-regulated genes
$CD14^+$	3	*NFKBIA*	*IL23A*, *FASLG*
$CD4^+$	3	*TNFAIP3*, *NFKBIZ*	*FASLG*
$CD8^+$	7	*TNFAIP3*, *NFKBIA*, *ETS1*, *SOCS1*,*NFKBIZ*	*B3GNT2*, *FASLG*
$CD19^+$	2	*NFKBIZ*	*B3GNT2*

Table 4.7: Overlap between putative psoriasis GWAS genes and the reported significantly DEGs in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. DEGs list based on FDR<0.05.

The majority of differentially expressed lncRNAs (FDR<0.05) had an experimentally validated functional interacting partner based on the database NPInter (Table A.4), which retrieves published functional interactions between non-coding RNAs and biomolecules (proteins, RNAs and DNAs) (Hao *et al.* 2016). Some of these lncRNAs were previously found to be differentially expressed in PBMCs when comparing PsA patients vs healthy controls in a study conducted by Dolcino and colleagues (Dolcino *et al.* 2018) (Table A.4). Examples of cell type-specific differentially expressed lncRNAs between psoriasis patients and healthy controls included *DYNLL1-AS1* (or *NRAV*), *HOTAIRM1* and *NEAT1* in $CD14^+$ monocytes, *DLGAP1-AS1* in $CD4^+$, *UBXN8* and *MIR146A* in $CD8^+$ and *LINC00324* in $CD19^+$. Only *RP11-218M22.1* was dysregulated between psoriasis and healthy controls in all four cell types.

Relevant cell type specific lncRNAs included the down-regulation of *DYNLL1-AS1* (fold change=0.78) in $CD14^+$ monocytes from psoriasis patients. *DYNLL1-AS1* has been shown to affect histone modifications of critical IFN-stimulated genes (ISGs), including *IFITM3* and *MxA*, in human alveolal epithelial cells leading to their transcriptional down-regulation (Ouyang *et al.* 2014). Despite down-regulation of *DYNLL1-AS1*, up-regulation of *IFITM3* and *MxA* was not observed in psoriasis $CD14^+$ monocytes in this data. Conversely, *HOTAIRM1* up-regulation (fold change=1.23) in $CD14^+$ monocytes from

psoriasis patients (Figure 4.10A) was accompained by down-regulation of its NPInter experimentally validated target *UPF1* (fold change=0.87), a gene coding for a RNA helicase and ATPase protein (Hao *et al.* 2016) (Figure 4.10B). Lastly, *NEAT1* expression was up-regulated (fold change=1.27) in psoriasis patients compared to controls in $CD14^+$ monocytes and has also been reported to be up-regulated in SLE $CD14^+$ monocytes when compared to healthy individuals (Zhang *et al.* 2016).

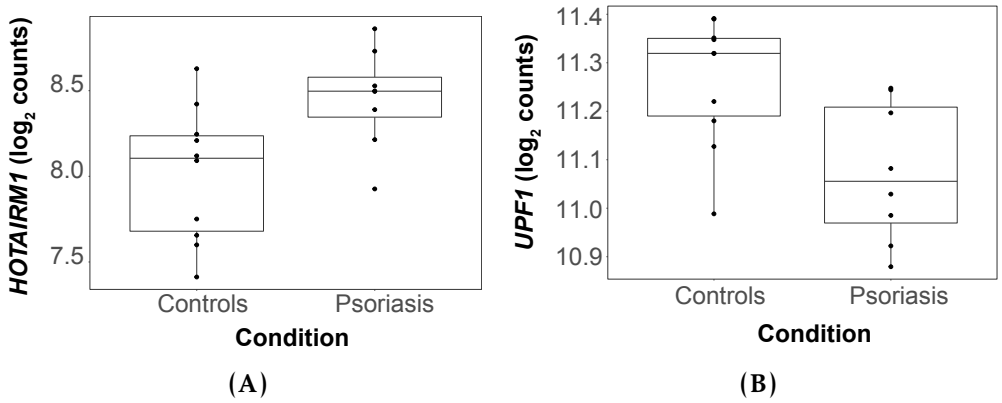


Figure 4.10: RNA-seq expression levels of the lncRNA *HOTAIRM1* and its experimentally validated target *UPF1* in psoriasis and healthy controls $CD14^+$ monocytes. Expression is illustrated as the $\log_2$ of the normalised read counts mapping to (A) the lncRNA *HOTAIRM1* and (B) *UPF1*, which has been experimentally identified as one of the genes regulated by this lncRNA according to NPInter database.

Although the RNA-seq library preparation used in this thesis was not specific to capture species of 20nt or less, some miRs were identified in the analysis. The miR *MIR146A* was the most relevant non-coding RNA dysregulated in $CD8^+$ T cells. *MIR146A* showed down-regulation (fold change=0.73) in $CD8^+$ cells from psoriasis patients when compared to controls and has been shown previously to have a role in negative regulation of the innate immunity, inflammatory response and antiviral pathways (Taganov *et al.* 2006).

Pathway enrichment analysis

To investigate the biological role of the differences in gene expression between psoriasis patients and healthy controls, pathway enrichment analysis

was performed for each cell type using DEG with FDR<0.05 and no fold change cut-off. This analysis revealed biologically relevant pathways significantly enriched (FDR<0.01) for DEGs in $CD14^+$ monocytes and $CD8^+$ cells (Table 4.8 and A.5). In $CD19^+$ cells, only one pathway ("generic transcription", p-value=$8.2x10^{-12}$, fold change=6.26) showed significant enrichment, whereas in $CD4^+$ cells no enrichment was seen for any pathway.

MAPK signalling and IL-12 mediated signalling pathways were enriched in both, $CD14^+$ monocytes and $CD8^+$ cells (Table 4.8). MAPK signalling enrichment was contributed by a number of MAPK genes that were differentially expressed between psoriasis patients and healthy controls (Figure 4.11). Two of those MAPK genes, *MAP3K4* and *MAPK14* (identified in a family-based AS GWAS study by Costantino *et al.* 2014), were down-regulated in psoriasis compared to controls in $CD14^+$ monocytes and $CD8^+$ T cells (Figure 4.11A and B). MAP3K4 is a member of the MAPKKK family, whose expression is down-regulated in LPS-stimulated PBMCs from CD patients leading to a relative immune deficiency in TLR-mediated cytokine production.

Moreover, differential gene expression analysis of members of the dual-specificity phosphatases (DUSP) family, involved in fine-tuning of the immune response (Qian *et al.* 2009), also contributed to the enrichment of the MAPK pathway in both cell types. *DUSP10* was down-regulated (fold change=0.83) in psoriasis $CD14^+$ monocytes (Figure 4.11C) and the knock-out of this gene in mice has been shown to enhanced inflammation (Qian *et al.* 2009). Conversely, *DUSP4* was up-regulated (fold change=1.47) in psoriasis $CD8^+$ cells compared to healthy controls (Figure 4.11D) and has been demonstrated to have a pro-inflammatory role in a sepsis mice model (Cornell *et al.* 2010).

Amongst members of the enriched IL-12 signalling pathway, *STAT4* and *STAT5A* were down-regulated (fold change 0.82 and 0.90, respectively) in $CD14^+$ monocytes from psoriasis patients. Neither *STAT4* or *STAT5A* were differentially

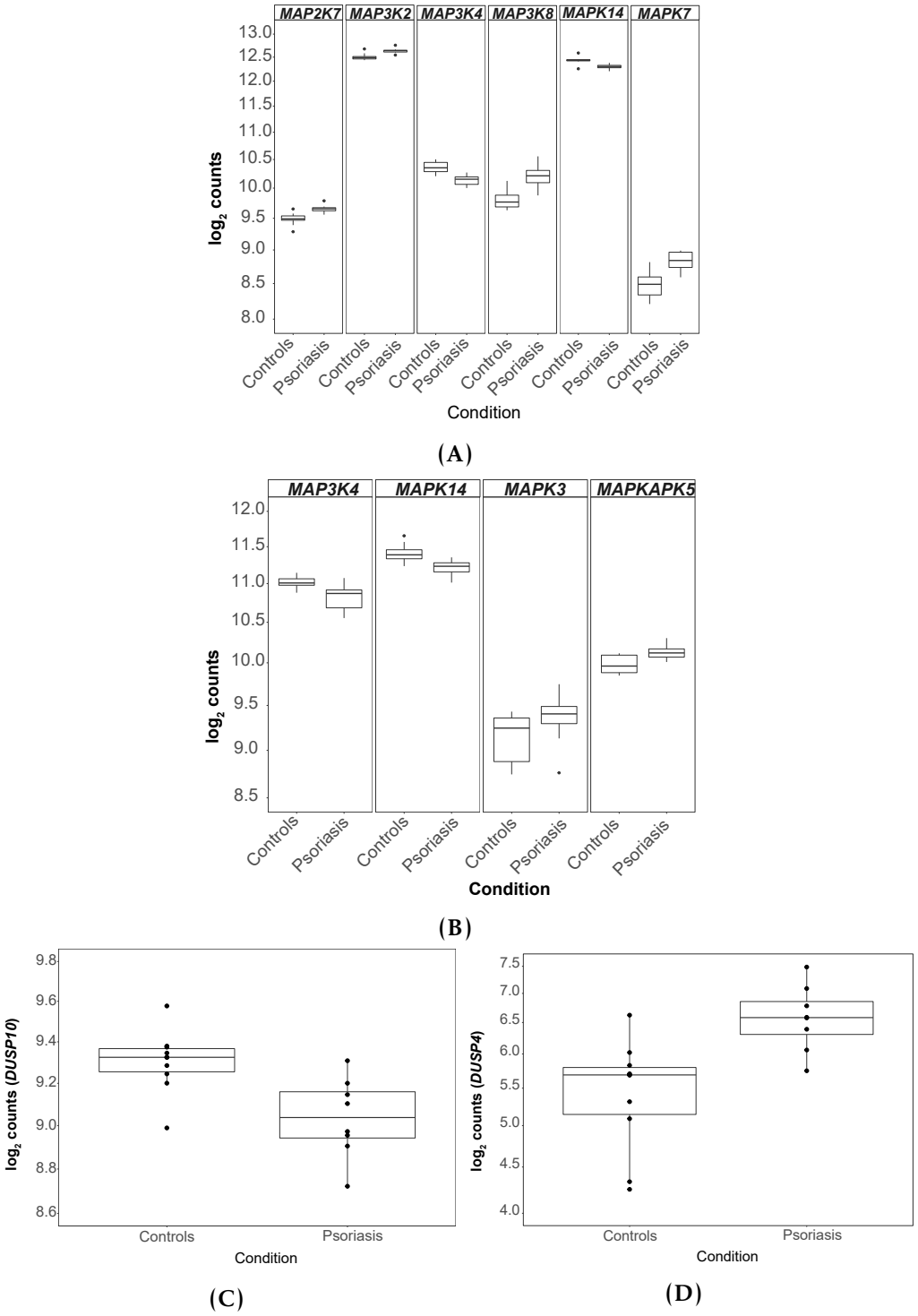


Figure 4.11: Differential expression between psoriasis and healthy controls of *MAPK* and *DUSP* genes contributing to MAPK signalling enrichment in CD14$^+$ monocytes and CD8$^+$ T cells. Expression for MAPK genes in psoriasis patients and controls is illustrated as the $\log_2$ of the normalised read counts, including (A) *MAP2K7*, *MAP3K2*, *MAP3K4*, *MAP3K8*, *MAPK14* and *MAPK7* in CD14$^+$ monocytes and (B) *MAP3K4*, *MAPK14*, *MAPK3* and *MAPKAPK5* in CD8$^+$ T cells. Similarly, expression as $\log_2$ of the normalised read counts is also illustrated for the DUSP gene members (C) *DUSP10* in CD14$^+$ monocytes and (D) *DUSP4* in CD8$^+$ T cells, in patients and control samples.

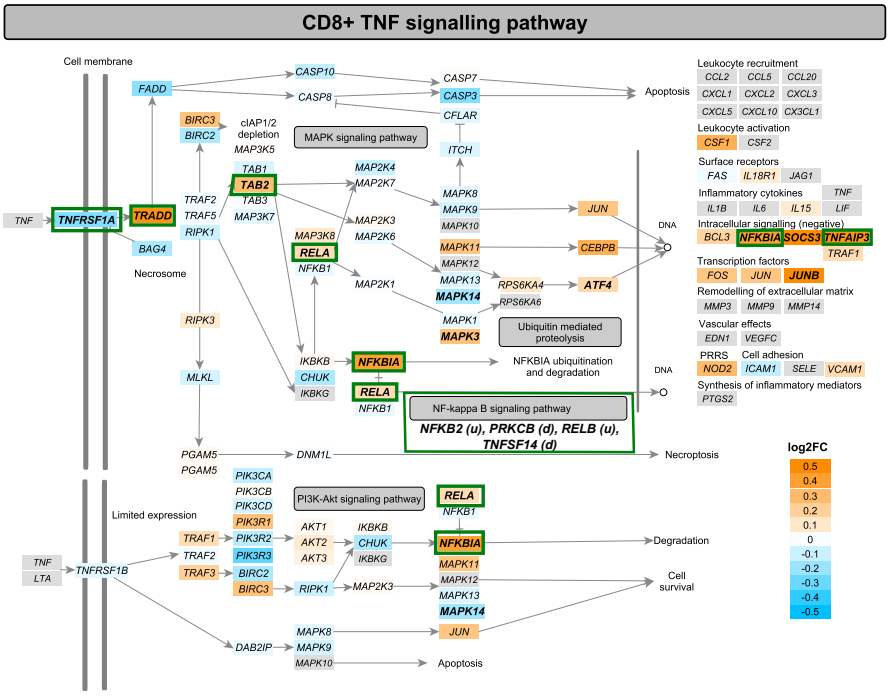
Cell type	Pathways	FDR	Fold change
$CD14^+$	MAPK signalling	$3.4x10^{-5}$	2.83
	IL-12 mediated signaling events	$6.2x10^{-6}$	4.82
	Th-1 and Th-2 cell differentiation	$2.4x10^{-4}$	3.34
	Th-17 cell differentiation	$2.4x10^{-4}$	3.21
	TCR signalling	$5.3x10^{-5}$	3.18
	Platelet-derived growth factor (PDGF-β)	$1.8x10^{-3}$	2.59
	Forkhead box O (FoxO) signalling	$3.6x10^{-3}$	2.38
$CD8^+$	Osteoclast differentiation	$7.2x10^{-5}$	3.45
	MAPK signalling	$1.5x10^{-3}$	2.33
	TNF signalling	$1.5x10^{-3}$	2.93
	IL-12 mediated signalling events	$5.5x10^{-4}$	3.86
	NF-κB signalling	$2.2x10^{-3}$	2.95
	Chemokine signalling	$2.3x10^{-3}$	2.43

Table 4.8: Pathways enriched for DEGs between psoriasis patients and healthy controls in $CD14^+$ monocytes and $CD8^+$ cells. The enrichment analysis was conducted using significant DEGs (FDR<0.05) and no fold change threshold. Significant enriched pathways (FDR<0.01) had a minimum of ten members overlapping DEGs in the corresponding cell type.

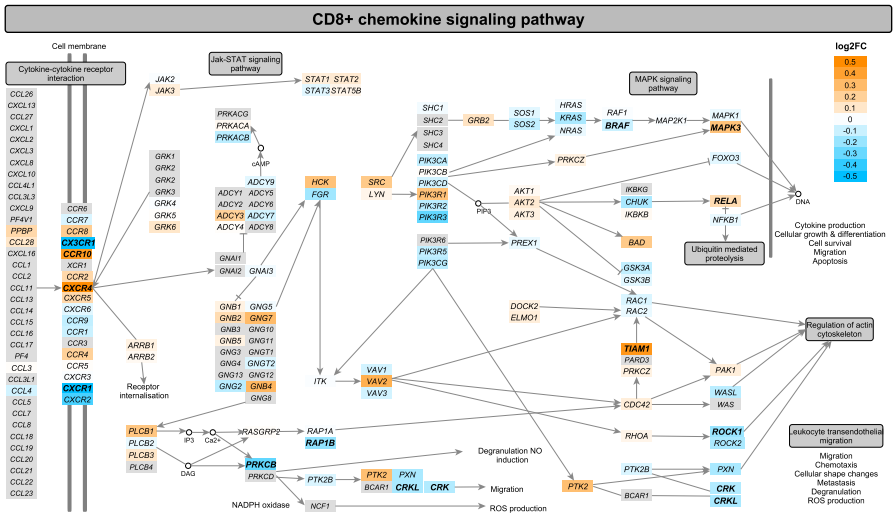
expressed in $CD8^+$ cells. In contrast, *IFNG* expression in psoriasis patients was lower than in healthy controls $CD8^+$ cells (fold change=0.66), showing no differences in $CD14^+$ monocytes. Also, *IL2RA* only demonstrated up-regulation (fold change=1.36) in $CD8^+$ T cells and this may have an effect in the formation of the receptor IL2-Rα and consequently on IL-2 signalling, which is involved in effector and Treg cell differentiation (Malek and Immunity 2010).

Pathways only enriched for $CD14^+$ monocytes DEGs involved platelet-derived growth factor (PDGF-β) signalling, and unexpectedly T cell related pathways such as TCR signalling and Th17 cell differentiation (Table 4.8). Dysregulated genes involved in the PDGF-β signalling included the down-regulation (fold change=0.84) in psoriasis samples of the *SLA*, for which a knock-out mice model has shown impaired IL-12 and TNF-α production and failure of T cell stimulation by GM-CSF treated bone marrow-derived DCs (Liontos *et al.* 2011).

Notably, a number of very relevant inflammatory pathways were enriched only for $CD8^+$ DEGs between psoriasis patients and controls. These included TNF, NF-κB and chemokine signalling (Table 4.8), with some of the DEGs contributing to more than one of the pathways. Enrichment of these three pathways resulted from combined up- and down-regulation of activators and inhibitors of the immune response. This included unexpected up-regulation in psoriasis $CD8^+$ T of the anti-inflammatory genes NF-κB inhibitor A (*NFKBIA*) and the TNF-α induced protein 3 (*TNFAIP3*) (fold change 1.32 and 1.36, respectively), both involved in TNF and NF-κB signalling (Figure 4.12A (green box) and B, in bold). As previously mentioned, *NFKBIA* and *TNFAIP3* were also up-regulated in psoriasis $CD14^+$ monocytes and $CD4^+$ cells, respectively (Table 4.7). Pro-inflammatory genes that showed down-regulation in psoriasis when compared to healthy controls included the activating transcription factor 2 (*ATF2*) (Figure 4.12A, in bold) and the protein kinase C beta (*PRKCB*) (Figure 4.12A (green box) and B, in bold). In contrast, *ATF4*, JunB proto-oncogene (*JUNB*) coding for one of the subunits of the TF AP-1 and three of the NF-κB subunits including *RELA* (p65), *RELB* and *NFKB2* (p52) were amonsgt the pro-inflammatory genes up-regulated in psoriasis compared to controls (Figure 4.12A). Furthermore, enrichment of the chemokine signalling pathway involved up-regulation in psoriasis $CD8^+$ T cells of *CCR10* (fold change=1.34), receptor for the chemotactic skin-associated chemokine CCL27, and *CXCR4* (fold change=1.59), receptor for the chemokine CXCL12, highly expressed in skin (Zgraggen *et al.* 2014) (Figure 4.12B), which would facilitate T cell migration into the affected epidermis. In contrast, *CXCR1* and *CX3CR1* , receptors of the pro-inflammatory molecules IL-8 and CX3CL1 respectively, were down-regulated in psoriais compared to healthy controls. Notably, none of the chemokines showed differential expression in this analysis (Figure 4.12B, not in bold).



(A)



(B)

Figure 4.12: Mapping of the DEGs identified in CD8$^+$ cells between psoriasis patients and healthy controls onto the TNF-α and the chemokine signalling pathways. The (A) TNF-α and (B) chemokine pathways were sourced from KEGG, manually curated in a way that all member genes are maximised visually and then automatically color-coded by the $\log_2$fold change expression between psoriasis patients and healthy controls CD8$^+$ cells isolated from PB. Significant DEGs (FDR<0.05) are highlighted in bold. In (A), members of the TNF-α pathway shared with the NF-κB are highlighted with a green box. Additional members of the NF-κB pathway differentially regulated in CD8$^+$ cells have also been indicated in brackets. Enrichment for (A) and (B) was identified by using only the CD8$^+$ DEGs (FDR<0.05).

These data highlight the dysregulation in $CD14^+$ monocytes and $CD8^+$ T cells of relevant pathways in psoriasis pathophysiology resulting from combined up- and down-regulation of pro- and anti-inflammatory genes which makes difficult the interpretation of the overall effect of those interactions in the inflammatory response.

4.3.5 RNA-seq in epidermis from psoriasis patients

Data processing and quality control

All three paired uninvolved-lesional samples (Table 4.2) had a mapping rate greater than 80% (Figure B.9A) and showed between 29.5 and 33.2 million reads mapping to Ensembl genes after filtering (Figure B.9B). The mapping rate and final number of reads mapping to genes was greater in the lesional samples compared to the controls (Figure B.9). PCA using normalised number of reads mapping to genes showed substantial variation between the lesional and uninvolved samples (Figure 4.13), PC1 37% of the variance) together with biological variability across individuals (PC2, 30% of the variance), for which the subsequent paired design in the differential gene expression analysis accounted.

Summary of differential gene expression between lesional and uninvolved epidermis

Differential gene expression analysis revealed a total of 1,227 (FDR<0.05) or 702 (FDR<0.01) genes differentially expressed between uninvolved and lesional epidermis skin biopsies, including mRNAs and lncRNAs (Table 4.9). Amongst the 1,227 DEGs, a similar proportion of up- and down-regulated genes (559 and 629, respectively) were identified between lesional and uninvolved epidermis (Figure 4.14) and 46 were annotated as lncRNAs (Table 4.9). The magnitude of changes in gene expression between lesional and uninvolved skin were notably

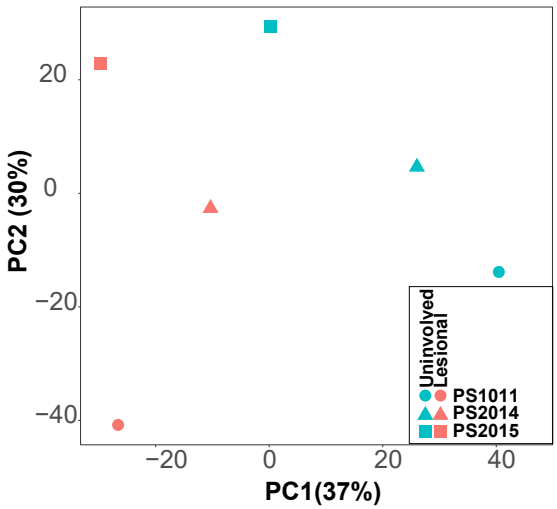


Figure 4.13: PCA for RNA-seq in the uninvolved and lesional epidermis from psoriasis patients. First and second component of the PCA performed on the normalised number of reads mapping to the Ensmbl list of mRNAs and lncRNAs detected in lesional and uninvolved epidermis isolated from psoriasis patient skin biopsies. Colour corresponds to condition (lesional or uninvolved) and shape refers to the patient ID.

larger when compared to the changes in expression in peripheral blood, with 874 out of 1,227 genes showing fold change>1.5.

FDR threshold	mRNA	lncRNA	Overlap with GWAS genes
0.05	1181	46	up(*IFIH1*, *NOS2*, *STAT3*, *LCE3D*), down(*TNFAIP3*)
0.01	677	25	*NOS2*, *STAT3*,*TNFAIP3*,*LCE3D*

Table 4.9: Summary of the differential gene expression analysis between uninvolved and lesional psoriatic epidermal biopsies. Number of differentially expressed mRNAs and lncRNAs are reported for two thresholds of significance (FDR<0.05 and FDR<0.01). The DEGs overlapping putative psoriasis GWAS genes and the directionality in the change of expression are listed.

Amongst DEGs between the uninvolved and lesional skin, five genes (FDR<0.05) overlapped with putative GWAS genes (Table 4.9). *IFIH1* (fold change=1.47), *NOS2* (fold change=1.74), *LCE3D* (fold change=1.94) and *STAT3* (fold change=1.54) were up-regulated in lesional epidermis, whereas *TNFAIP3* was down-regulated (fold change=0.47). All these genes but *IFIH1* showed significant differential expression at the more stringet FDR<0.01 threshold.

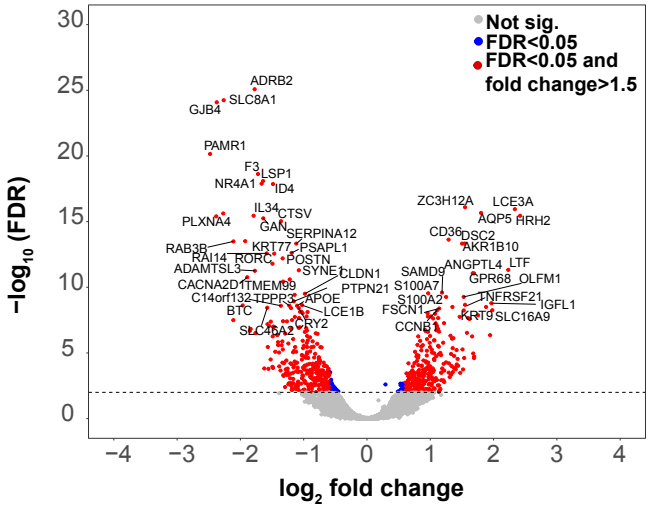


Figure 4.14: Magnitude and significance of the gene expression changes between matched lesional and uninvolved epidermal biopsies from three psoriasis patients. The volcano plot represents for each gene the significance ($-\log_{10}FDR$) of the $\log_2$fold change in expression for that gene in lesional skin group vs uninvolved skin. Significant DEGs (FDR<0.05) in blue for fold change<1.5 and red for fold change>1.5. The volcano plot includes mRNAs and lncRNAs species.

Comparison with other skin transcriptomic studies

As detailed in Chapter 2, in this thesis the epidermal layer was isolated from the rest of the biopsy in contrast to most published transcriptomic studies using full-thickness biopsies to compare lesional and uninvolved skin from psoriasis patients. This allowed to obtain purer sample composition with the majority of cells being epidermal keratinocytes and a smaller proportion infiltrated immune cells. During the course of this project a study was published by Tervaniemi and colleagues characterising the transcriptional profiles epidermis from lesional and uninvolved psoriatic skin (Tervaniemi *et al.* 2016). In order to explore the similarities across the two studies, a comparison of DEGs identified between lesional and uninvolved matched samples was conducted.

The Tervaniemi study identified 2,589 DEGs (filtering criteria fold change<0.75 or fold change>1.5 and FDR<0.05), a larger number of DEGs compared to the number found here, and reported a greater number of up-regulated than down-regulated hits (Figure 4.15 bottom panel). A total of 359

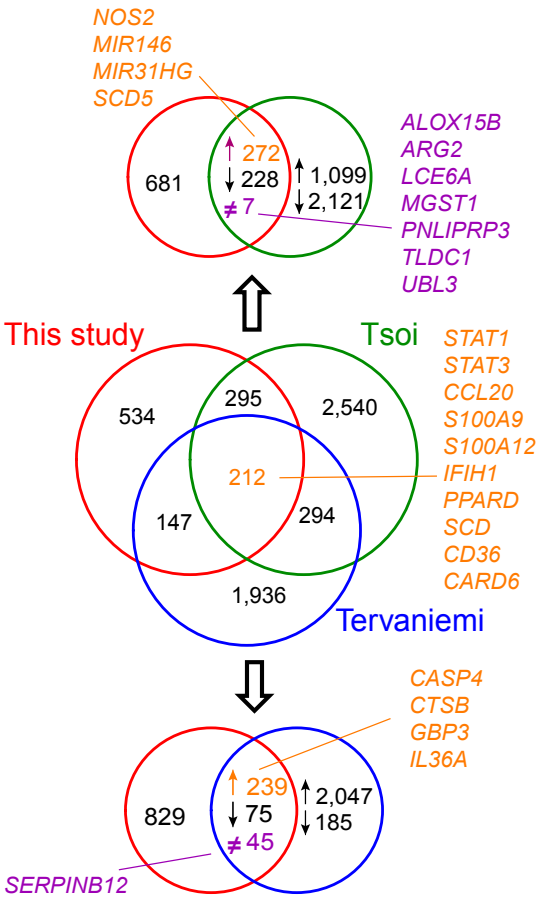


Figure 4.15: Overlap of the significant differentially expressed genes between lesional and uninvolved epidermal sheets, split epidermis and full-thickness skin biopsies. The central venn diagram illustrated the DEGs between this study (epidermal sheets), Tervaniemi *et al.* 2016 (split epidermis) and Tsoi *et al.* 2015a (full-thickness) lesional and uninvolved skin biopsies. Overlap is considered regardless the direction of the change. Two additional venn diagrams provide more detail about the total overlap and directionality in the change of gene expression between data rom this study and the Tsoi *et al.* 2015a (top) or Tervaniemi *et al.* (bottom) publications. Some of overlapping genes are listed.

out of the 1,227 DEGs (29.25%) identified here were shared with the Tervaniemi results, of which 239 and 75 were up- and down-regulated, respectively (Figure 4.15 bottom panel). Some examples of this overlap included up-regulation of *STAT1*, genes from the *S100* family (e.g *S100A9* and *S100A12*) and genes nearby psoriasis GWAS loci such as *STAT3* and *IFIH1*. Notably, 45 genes were differentially expressed in both studies but showed opposite direction. For example *SERPINB2* was down-regulated in this study and up-regulated in the Tervaniemi results.

Additionally, these results were also contrasted with one of the most recent comprehensive RNA-seq studies comparing lesional and uninvolved full-thickness skin biopsies from psoriasis patients (Tsoi *et al.* 2015b). Out of the 3,725 DEGs reported by the Tsoi analysis, 507 genes were shared between the two studies (41% of the in-house DEGs), of which 24 corresponded to dysregulated lncRNAs. Amongst the 507 commonly dysregulated genes, 272 were up-regulated, 228 down-regulated and only 7 showed opposite direction of change (Figure 4.15 top panel).

Simultanous overlap across the three studies identified only 212 DEGs. Despite having a larger sample size, the Tsoi study did not capture the majority of the DEGs from this dataset or the Tervaniemi one (Figure 4.15 middle panel).

Dysregulated lncRNAs in the psoriatic lesional skin

In addition to protein coding genes, 46 lncRNAs were significantly (FDR<0.05) differentially expressed between uninvolved and lesional skin and 37 of them had an interacting partner experimetally validated according to NPInter database (Hao *et al.* 2016). An interesting example was *H19*, which was significantly down-regulated (fold-change=0.43) in lesional epidermis and has been described to directly bind miR-130b-3p, which down-regulates by acting as a sponge the expression of desmoglein 1 (*DSG1*), a gene promoting keratinocyte

differentiation (Li *et al.* 2017). Nevertheless, *DSG1* itself did not show differential expression between lesional and uninvolved skin.

Four miRNAs (*MIR146A*, *MIR22HG*, *MIR31HG* and *MIR205HG*) were also captured with the standard library preparation for mRNAs and lncRNAs. The relevance of miR-146a has been already noted in the differential analysis from circulating immune cells (Section 3.3.4). In lesional skin *MIR146A* was up-regulated (fold change=2.0) when compared to uninvolved skin, consistent with other studies in skin (Lerman *et al.* 2011; Tsoi *et al.* 2015b).

Another relevant finding was the up-regulation of *MIR31HG* (fold change=1.84) in lesional skin previously reported by the Tsoi study (Tsoi *et al.* 2015b). Additionally, in a study in head and neck carcinoma, *MIR31HG* expression was identified to target *HIF-1A*, inducing its up-regulation by an unknown mechanisms (Wang *et al.* 2018). In this data, *HIF-1A* was up-regulated (fold change=1.81) in lesional skin compared to uninvolved, but only a trend for positive correlation was observed between normalised counts of this gene and the putative regulator *MIR31HG* (Figure B.11), likely due to the small sample size.

Pathway enrichment analysis

In order to better understand the functional role of the DEGs (FDR<0.05) between lesional and uninvolved epidermis from psoriasis patient skin biopsies, pathways enrichment analysis was conducted. A number of pathways were significantly enriched with FDR<0.01 and fold change>2 (Tables 4.10 and A.6).

These included pathways related to alterations in cell cycle and metabolic processes, such as hypoxia-inducible factor 1 (HIF-1) signalling, arginine and proline metabolism, glycolysis/gluconeogenesis and metabolism of carbohydrates. HIF-1 signalling has been found to be up-regulated in psoriatic skin, likely through hypoxia caused by increased cell proliferation rates and

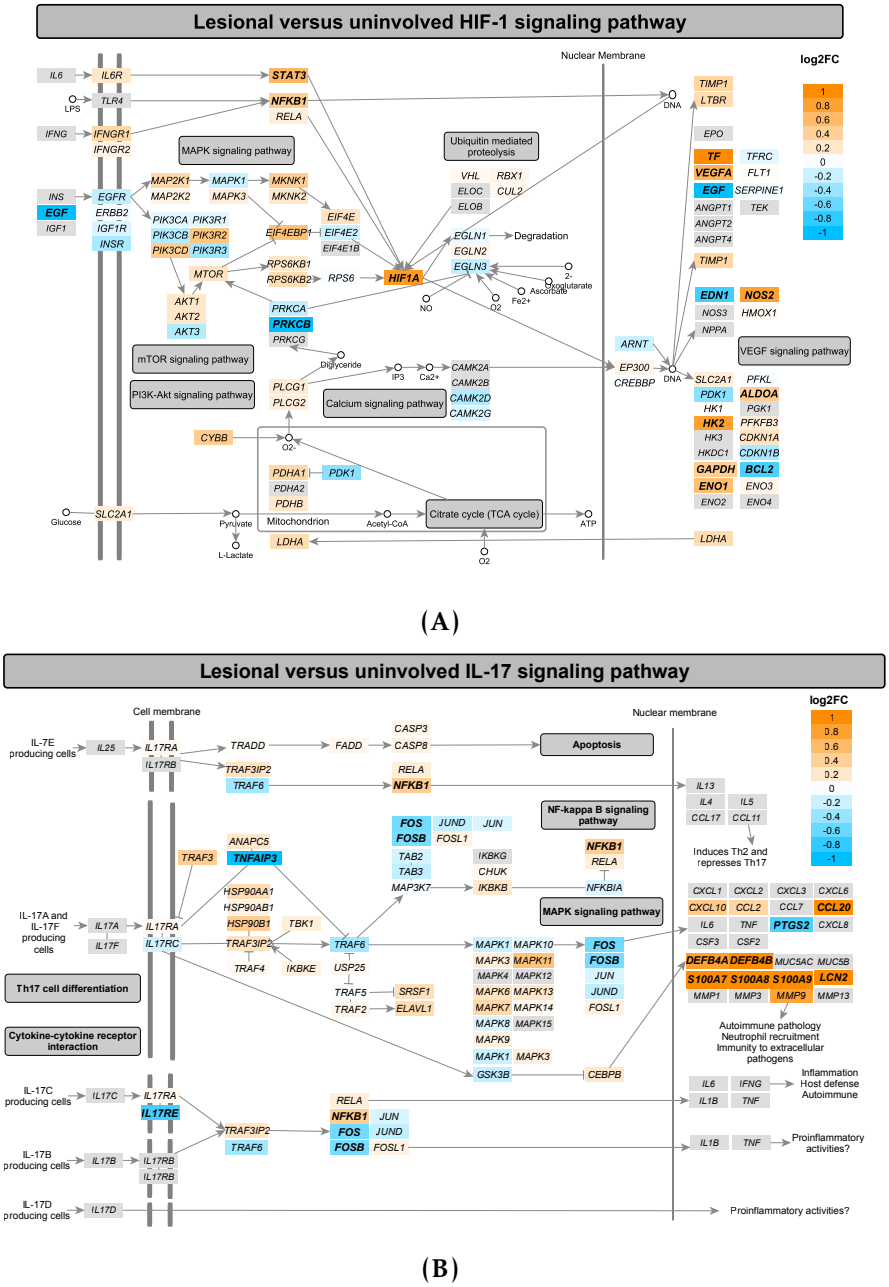
Pathway	FDR	Fold change
IFN-α/β/signalling	$5.4x10^{-5}$	3.65
Peroxisome proliferator-activated receptors (PPAR)	$1.5x10^{-4}$	3.42
NOD-like receptor signalling	$3.1x10^{-4}$	2.37
IL-17 signalling	$1.4x10^{-3}$	2.72
IL2-mediated signalling	$3.1x10^{-3}$	2.64
Hypoxia-inducible factor 1 (HIF-1) signalling	$5.0x10^{-3}$	2.31
Glycolysis/gluconeogenesis	$1.0x10^{-6}$	4.71
Cell cycle	$6.7x10^{-5}$	2.55
Apoptosis	$3.7x10^{-3}$	2.11
Arginine and proline metabolism	$7.5x10^{-6}$	5.28
AGE-RAGE signalling in diabetic complications	$8.3x10^{-3}$	2.17

Table 4.10: Most relevant pathways enriched for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies. Significant pathways for FDR<0.01 and fold change>2 are listed here and in Table A.6. The analysis was performed using as input significant DEGs (FDR<0.05). Enriched pathways had a minimum of ten members overlapping with DEGs.

epidermal thickening (Rosenberger *et al.* 2007). Here, up-regulation of *HIF1A*, *VEGFA*, *ENO1* and the GWAS gene *NOS2*, amongst others, contributed to the enrichment of this pathway (Figure 4.16A, in orange and bold).

Immune relevant pathways including IFN, IL-17 and NOD-like receptor signalling were also identified in this analysis. The NOD-like receptor pathway was enriched with 23 significantly DEGs and it is involved in pathogen recognition and generation of an innate immune responses through activation of the inflammasome or through NF-κB and MAPK with subsequent cytokine production and apoptosis. This included up-regulation of *NOD2*, *CARD6*, *IFI16*, *NLRX1*, *IRF7* and *NFKB1* (Figure 4.17 in orange and bold) and down-regulation of *TNFAIP3*, *TXNIP* and *BCL-2* (Figure 4.17 in blue and bold), amongst others.

Another relevant pathway enriched in DEGs between lesional and uninvolved epidermis was IL-17 signalling. Enrichment of this pathways was driven by up-regulation of the S100 protein family (*S100A7*, *S100A8* and *S100A9*) and chemokines such as *CCL20*, which binds the receptor CCR6 and is involved in Th-17 cell chemotaxis (Figure 4.16B in orange and bold).



(A)

(B)

Figure 4.16: Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the HIF-1 and IL-17 signalling pathways. The (A) HIF-1 and (B) IL-17 signalling pathways were sourced from KEGG, manually curated in a way that all member genes are maximised visually and then automatically color-coded by the $\log_2$fold change expression between the lesional and uninvolved epidermis. Significant DEGs (FDR<0.05) are highlighted in bold. These pathways were identified by enrichment analysis using DEGs FDR<0.05.

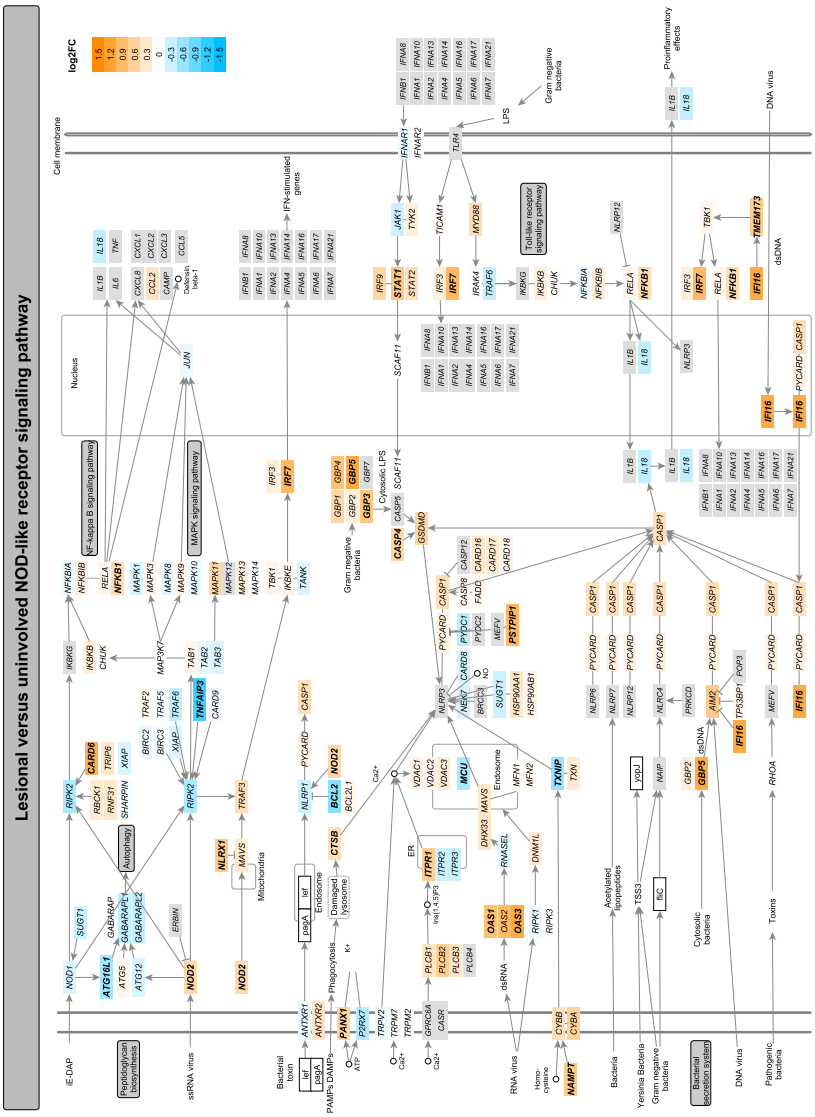


Figure 4.17: Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the NOD-like signalling pathway. This pathway was sourced from KEGG, manually curated in a way that all member genes are maximised visually and then automatically color-coded by the $\log_2$fold change expression between the lesional and uninvolved epidermis. Significant DEGs (FDR<0.05) are highlighted in bold. This pathway was identified by pathway enrichment analysis using only DEGs (FDR<0.05).

Notably, *IL17RE* which together with IL-17RA forms the receptor for IL-17C was down-regulated in lesional skin (Figure 4.16B in blue and bold). Enrichment of DEGs between lesional and uninvolved skin for the peroxisome proliferator-activated receptor (PPAR) signalling and AGE-RAGE signalling in diabetic complications highlighted the link between metabolic dysregulation (particularly oxidation of lipids, carbohydrates and aminocacids) and immunity. PPAR signalling involved up-regulation of the PPAR receptor δ (*PPARD*) (fold change=1.39), the stearoyl-CoA desaturase *SCD* (fold change=1.82), involved in fatty acid synthesis, and *CD36* (fold change=2.44), a fatty acid transport, all of them also DEGs in the Tsoi and Tervaniemi studies.

4.3.6 Comparison of systemic and tissue specific gene expression signatures in psoriasis

In order to investigate commonalities and differences in psoriasis gene expression in the affected tissue (skin) and at the systemic level (circulating immune cells), overlap between the lists of DEGs was performed. Only modest overlap was found between DEGs in lesional skin compared to uninvolved and the DEGs identified in circulating immune cells. $CD14^+$ monocytes and $CD8^+$ cells showed the greatest overlap; however many of these genes did not show a consistent direction of change in differential expression (Table 4.11). Examples included *TNFAIP3*, which was up-regulated in psoriasis $CD4^+$ and $CD8^+$ cells compared to controls but down-regulated in lesional epidermis in comparison to uninvolved. Similarly, early growth response genes were up-regulated in $CD14^+$ monocytes (*EGR3*) and $CD4^+$ T cells (*EGR1*, *EGR2* and *EGR3*) but showed down-regulation (*EGR2* and *EGR3*) in lesional epidermis compared to uninvolved.

Examples of genes changing in the same direction in immune cells and psoriasis skin included the up-regulation in $CD14^+$ monocytes of the nicotinamide phosphoribosyltransferase involved in stress response and

inflammation *NAMPT* (Présumey *et al.* 2013) and the down-regulation in $CD8^+$ cells of the protein kinase *PRKCB*.

DEGs overlapping with skin	Total overlap	Same direction	Opposite direction
$CD14^+$ monocytes	37	19	18
$CD4^+$	10	6	4
$CD8^+$	37	24	13
$CD19^+$	16	5	11

Table 4.11: Overlap between the DEGs in the four circulating immune cell types (psoriasis patients vs controls) and the DEGs in psoriasis patients skin biopsies (lesional vs uninvolved). DEGs based on FDR<0.05 for each of the comparisons

The limited overlap between circulating and skin DEGs was also reflected in the different enriched pathways identified for each analysis. $CD14^+$ and $CD8^+$ DEGs were mostly enriched for immune-related pathways, including chemokines, MAPK, IL-12 , TNF and NF-κB signalling, and no enrichment was found for metabolic, oxidative stress and cell cycle, which showed significant dysregulation in skin. Moreover, the DEGs contributing to the enrichment of immune-related pathways in psoriatic epidermis had a more pronounced pro-inflammatory signature, consistent with the skin being the main site of active inflammation.

4.3.7 Integration of chromatin accessibility and expression data for peripheral blood immune cells in psoriasis

The characterisation of the chromatin accessibility landscape and the transcriptome of circulating immune cells from psoriasis patients revealed a greater effect of disease status on gene expression than in chromatin accessibility. An integrated analysis combining evidence of differential accessibility and expression in the four investigated immune cell types was performed by overlapping proximal genes to DARs (≤5Kb) and DEGs. An overlap was only

found in $CD8^+$ cells, where 6 out of the 55 DARs colocalised with DEGs in this cell type (*ARL4A*, *ASCL2*, *ENTPD1*, *TIAM1*, *TRAT1* and *ZNF276*).

An example was T Cell lymphoma invasion and metastasis 1 (*TIAM1*), which activates IL-17 expression and T cell transendothelial migration during inflammation (Kurdi *et al.* 2016; Gérard *et al.* 2009) and showed up-regulated expression (log_2fold change 0.44) in $CD8^+$ cells from psoriasis patients (Figure 4.18A). Likewise, psoriasis $CD8^+$ cells presented greater chromatin accessibility compared to healthy controls (log_2fold change 0.41) in a region located in an intron of *TIAM1* annotated as an active enhancer according to the chromatin segmentation data in this cell type (Figure 4.18C). Common SNPs within this DAR were not identified as eQTLs in $CD8^+$ cells (Kasela *et al.* 2017); however promoter capture Hi-C data in unstimulated $CD8^+$ cells revealed physical interaction between this region and the promoter of *TIAM1* (Javierre *et al.* 2016).

Another two relevant genes in the immune response with overlapping ATAC and RNA-seq were the ectonucleoside triphosphate diphosphohydrolase 1 (*ENTPD1*), which hydrolyses the pro-inflammatory mediator ATP attenuating the inflammation, and the TCR-associated transmembrane adaptor 1 (*TRAT1*) gene, a positive regulator of TCR signalling (Antonioli *et al.* 2013; Valk *et al.* 2006). Both, *ENTPD1* and *TRAT1*, showed up-regulated expression (Figure 4.18B) and increased chromatin accessibility (Figure 4.18D) in psoriasis patients $CD8^+$ cells compared to healthy controls.

4.3.8 Fine-mapping of psoriasis GWAS loci and functional interpretation

Fine-mapping using summary statistics data

Due to the impossibility of accessing genotyping data, fine-mapping of psoriasis Immunochip GWAS loci was conducted using summary statistics of the Immunochip psoriasis GWAS study from Tsoi *et al.* 2012. This study was based

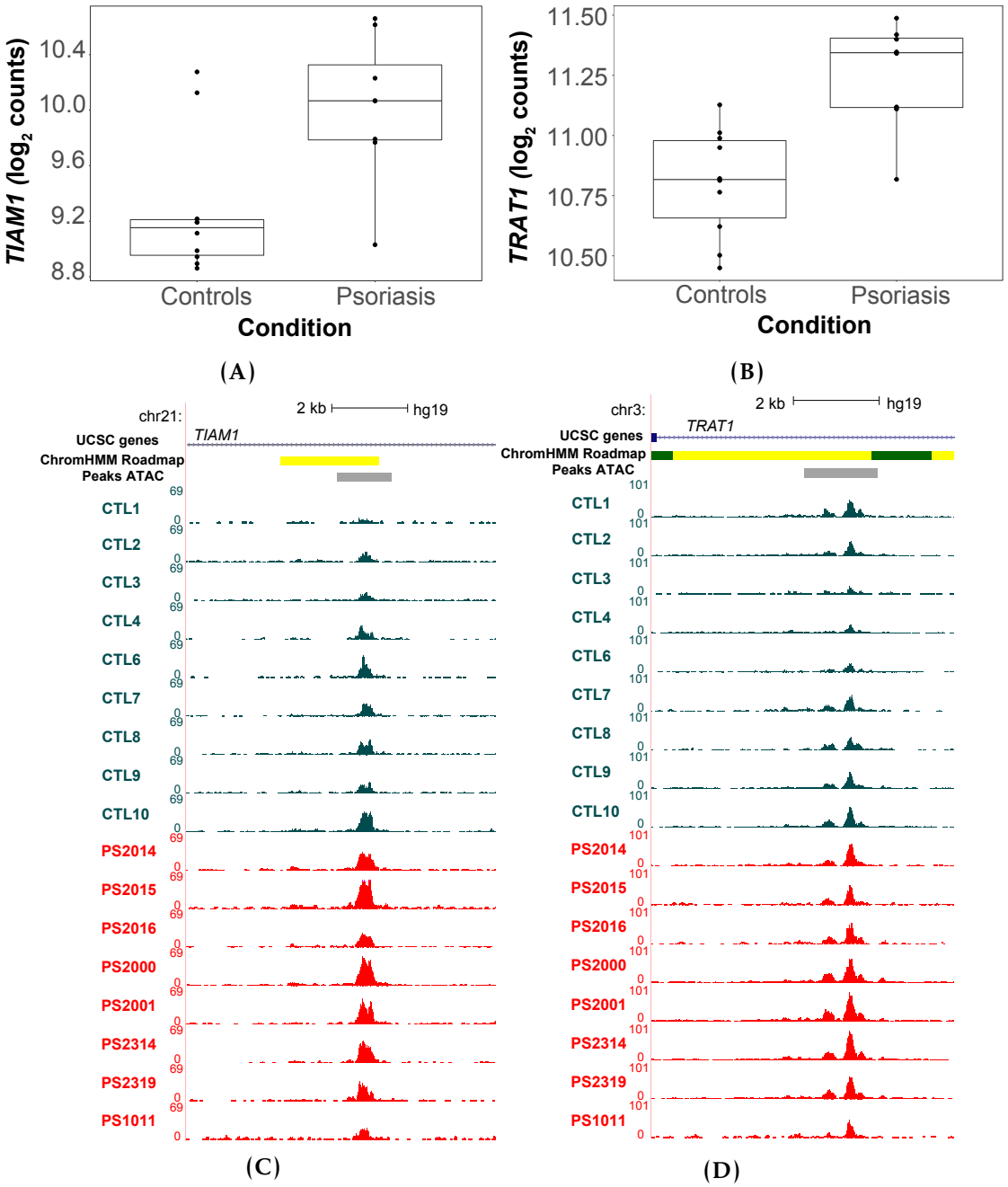


Figure 4.18: Differential gene expression and chromatin accessibility landscape for *TIAM1* and *TRAT1* genes in CD8$^+$ cells. Boxplots show RNA-seq $\log_2$ normalised counts for (A) *TIAM1* and (B) *TRAT1* in psoriasis and healthy controls CD8$^+$ cells. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) in introns of (C) *TIAM1* and (D) *TRAT1* genes (x-axis) in CD8$^+$ cells. Tracks are colour-coded by condition and assay: control (CTL)=dark turquoise and psoriasis (PS)=red. The Roadmap Epigenomics Project chromatin segmentation map for CD8$^+$ cells is shown above the ATAC tracks.

on two cohorts (GAPC and PAGE) and only summary statistics for the GAPC cohort (2,997 cases and 9,183 controls) were publicly available in ImmunoBase at the time of this analysis. As explained in Chapter 2, fine-mapping from summary statistics with DIST used the z-score statistics of each of the genotyped SNPs from the GAPC cohort to impute the z-scores for the missing SNPs based on the r^2 relationship from the 1000 Genome Project Version 3 (Lee *et al.* 2013). Following z-score imputation for the non-genotyped SNPs, ABF and PP were calculated using Wakefield approximation and 90% credible sets of SNPs were built for each GWAS signal (Tables 4.12).

Fine-mapping was performed for 25 of the Immunochip psoriasis GWAS loci reported by Tsoi *et al.* 2012, excluding the MHC and those loci which lead SNPs were in high LD with missense mutations having experimentally proved or highly confident predicted damaging effects. Out of the 25 regions, 8 loci did not reach $\log_{10}$ABF≥3 (cut-off used as in Bunt *et al.* 2015) with 90% credible sets ranging from 19 to 853 SNPs (Table 4.12 bottom). In 10 loci presenting $\log_{10}$ABF≥3, the fine-mapping lead SNP was in low LD with the Tsoi *et al.* 2012 GWAS lead SNP and/or was not included in the 90 or 50% credible set (Table 4.12 with *). This is likely due to the lack of power as only summary statistics from one cohort were available for this analysis.

Table 4.12: Summary results for the fine-mapping analysis of 26 psoriasis Immunochip GWAS loci. For the 17 Immunochip psoriasis GWAS loci showing $\log_1 0$ABF≥3, the table reports the closer gene(s), FM lead SNP, RAF, imputation status, z-score, OR for the GAPC cohort lead SNP, $\log_{10}$ABF for the FM lead SNP, PP, number of SNPs included in the 90% credible set and the Tsoi *et al.* 2012 GWAS lead SNP. Imputation status refers to whether the SNP was imputed (Yes) or genotyped by the Immunochip array (No). If a SNP is imputed, the z-score statistic was determined by imputation based on LD with other SNPs, as previously explained. The sign of the z-score indicates whether the RAF allele increases (+) or decreases (-) risk of psoriasis. In 10 of the loci with $\log_{10}$ABF≥3, the fine-mapping lead SNP was in low LD ($r^2 < 0.5$) with the psoriasis GWAS SNP and did not contain it that SNP in the credible set (labelled with *). Details of the 8 loci presenting $\log_1 0$ABF< 3 are also included. FM=fine-mapping; RAF= reference panel allele frequency; OR=odds ratio; ABF=approximate Bayes factor; PP=posterior probability.

chr	Closest gene	FM lead SNP	RAF	Imputed	z	OR GAPC	$\log_{10}$ABF FM lead SNP	PP	90% credible set	Tsoi lead SNP
1	*IL28RA*	rs61774731*	0.10	No	-7.68	1.21	11.5	0.99	1	rs7552167
2	*FLJ16341/REL*	rs6714339*	0.14	No	-6.82	1.17	9.0	0.99	1	rs62149416
2	*IFIH1*	rs2111485*	0.59	No	5.94	1.27	6.7	0.50	2	rs17716942
5	*TNIP1*	rs17728338	0.06	No	7.45	1.59	10.6	0.40	6	rs2233278
5	*IL12B/ADRA1B*	rs12188300	0.05	No	7.71	1.58	11.2	0.18	9	rs12188300
6	*TNFAIP3*	rs1416173*	0.85	No	-6.36	1.23	7.7	0.15	10	rs582757
14	*NFKBIA*	rs74243591	0.21	No	-5.23	1.16	5.0	0.30	12	rs8016947
17	*NOS2*	rs117094752*	0.02	No	-6.53	1.22	7.3	0.94	1	rs28998802
1	*SLC45A1/TNFRSF9*	rs425371	0.25	Yes	5.52	1.13	5.6	0.14	22	rs11121129
1	*RUNX3*	rs61774731 *	0.10	No	-7.68	1.13	11.5	0.99	1	rs7536201
2	*B3GNT2/TMEM17* (2p15)	rs9309343*	0.33	Yes	4.92	1.12	4.3	0.66	34	rs10865331
7	*ELMO1*	rs77840275*	0.10	No	-6.31	1.11	7.5	0.99	1	rs2700987
11	*ZC3H12C*	rs11213274	0.40	Yes	-4.78	1.14	4.0	0.05	69	rs4561177

11	*ETS1*	rs10893884	0.53	No	-4.51	1.15	3.5	0.26	19	rs3802826
16	*PRM3/SOCS1*	rs111251548*	0.02	No	-7.41	1.13	9.4	0.97	1	rs367569
17	*PTRF/STAT3*	rs963986	0.17	No	4.82	1.15	4.1	0.18	8	rs963986
19	*ILF3/CARM1*	rs34536443*	0.03	No	-7.43	1.17	9.7	0.93	1	rs892085

	No	fine-mapped								
chr	**Closer gene**	**FM lead SNP**	**RAF**	**Imputed**	**z**	**OR GAPC**	**$\log_{10}$ABF FM lead SNP**	**PP**	**90% credible set**	**Tsoi lead SNP**
10	*ZMIZ1*	rs1316431	-	-	-	1.09	2.4	-	401	rs1250546
11	*RPS6KA4/PRDX5*	rs58779949	-	-	-	1.06	0.3	-	334	rs645078
20	*RNF114*	rs13041638	-	-	-	1.11	2.9	-	116	rs1056198
6	*EXOC2/IRF4*	rs113866081	-	-	-	1.14	2.4	-	400	rs9504361
6	*TAGAP*	rs62431928	-	-	-	1.11	2.2	-	853	rs2451258
9	*DDX58*	rs7045087	-	-	-	1.05	0.4	-	167	rs11795343
9	*KLF4*	rs6477612	-	-	-	1.12	2.1	-	80	rs10979182
18	*POL1/STARD6*	rs11661229	-	-	-	1.11	1.6	-	121	rs545979

The remaining seven loci in high LD with the Tsoi *et al.* 2012 GWAS lead SNPs showed 90% credible set of SNPs ranging from 6 to 69 SNPs (Table 4.12). Of those *TNIP1*, *IL12B/ADRA1B* and *PTRF/STAT3* had 90% credible sets refined to fewer than 10 SNPs. Interestingly, only two of the seven loci showed the fine-mapping lead SNP to be the same as the Tsoi GWAS lead SNP, supporting the sensitivity of the association analysis to sample size. *TNIP1* and *IL12B/ADRA1B* loci had been previously fine-mapped by Das and colleagues using dense genotyping with a customised array followed by association analysis (Das *et al.* 2014). Notably, two (rs75851973 and rs2233278) and three SNPs (rs918519, rs918518 and rs733589) from the *TNIP1* and *IL12B/ADRA1B* 90% credible sets, respectively, were amongst the sets of significant variants and perfect near proxies ($r^2 > 0.9$) reported by Das *et al.* for those two same loci.

Integration with functional data

A total of 144 unique SNPs formed the union of 90% credible sets from the seven loci with fine-mapped lead SNPs presenting a $\log_{10}$ABF≥3 and including the Tsoi *et al.* 2012 GWAS lead SNP. None of the SNPs overlapped with DARs or differential H3K27ac regions identified in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ or $CD19^+$ cells. However, when overlapped with the consensus master list of ATAC peaks from each cell types, 17 unique SNPs from six loci were located at accessible chromatin in at least one cell type (*NFKBIA*[1 SNP], *PTRF/STAT3*[1 SNP], *SLC45A1/TNFRSF9*[4 SNPs], *TNIP1*[4 SNPs], *ZC3H12C*[6 SNPs] and *ETS1*[1 SNP]) and no overlap was found for the 9 SNPs in the *IL12B/ADRA1B* 90% credible set. $CD14^+$ monocytes showed the largest proportion of accessible chromatin regions containing SNPs from the credible sets (2.3%), followed by $CD19^+$, $CD4^+$ and $CD8^+$ cells (1.7, 1.6 and 1.05%, respectively) (Table 4.13). Altogether, integration of the SNPs from the credible sets with ATAC accessible regions in four cell types allowed further refinement of the number of genetic

variants with a putative functional role in psoriasis for six out of the seven analysed loci. Moreover, in the *PTRF/STAT3* and *SLC45A1/TNFRSF9* loci all the SNPs from the credible sets that were overlapping accessible chromatin were in cell-type specific peaks (Table 4.12 and 4.13). Out of the 8 SNPs in the *PTRF/STAT3* 90% credible set, the only overlapping accessible chromatin was $CD14^+$ monocyte-specific. Similarly, the *SLC45A1/TNFRSF9* locus presented only 4 out of the 22 SNPs from the 90% credible set within ATAC peaks, of which all were $CD14^+$ monocytes-specific.

For the two GWAS loci with proximal genes differentially expressed, *NFKBIA* and *ETS1*, the only SNP of each credible set overlapping ATAC peaks were found in $CD14^+$monocytes and $CD19^+$ cells (rs7155561) and in $CD4^+$ and $CD8^+$ T cells (rs12223943), respectively. Notably, rs12223943 was located upstream of the promoter of the shorter *ETS1* transcript variant (a TF involved in T cell differentiation) and within an intron of the longer one, and it overlapped accessible chromatin and H3K27ac modifications in $CD4^+$ and $CD8^+$ T cells from patients and controls (Figure B.12). Moreover, rs12223943 was not found to be an eQTL for *ETS1* in any of the immune cells eQTL publicly available datasets, including unstimulated $CD4^+$ and $CD8^+$ cells (Kasela *et al.* 2017).

ATAC cell type master list	90% credible set overlapping SNPs (number)	Cell-type specific overlap
$CD14^+$ monocytes	13	*PTRF/STAT3*(1), *SLC45A1/TNFRSF9*(4)
$CD4^+$	6	None
$CD8^+$	4	*TNIP1* (2)
$CD19^+$	5	None

Table 4.13: SNPs from the 90% credible set of the successfully fine-mapped psoriasis loci overlapping ATAC accessible chromatin in four cell types. The number of SNPs in the 90% credible set union (total 144 SNPs) from the seven successfully fine-mapped loci overlapping ATAC accessible chromatin in each cell type master list are reported. Additionally, the number of SNPs only found to overlap open chromatin in one cell type are indicated together with the locus in which the SNP was fine-mapped.

The functional landscape at the *SLC45A1/TNFRSF9* intergenic locus

SLC45A1/TNFRSF9 was one of the new intergenic GWAS association reported by Tsoi *et al.* 2012, a nearby independent GWAS signal has been identified for IBD and UC (Jostins *et al.* 2012; Anderson *et al.* 2011). This locus was successfully fine-mapped in this analysis (Figure 4.19) and integration of its 90% credible set with ATAC data further refined the number of candidate functional SNPs from 22 to 4.

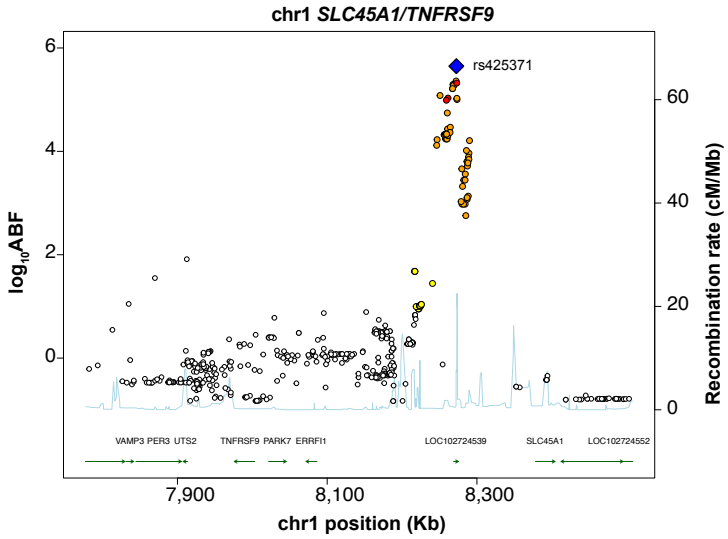


Figure 4.19: Association plot for the fine-mapping using summary statistics from the psoriasis GAPC cohort Immunochip GWAS results of the *SLC45A1/TNFRSF9* locus. For each of the SNPs (dot) the location (x-axis) and the $\log_{10}$ABF (left y-axis) from the fine-mapping association analysis are shown. The colour of each dot indicated the LD relationship (r^2) with the lead SNPs (blue diamond), being white=low LD (r^2 <0.2), yellow=weak LD (r^2 <0.5), orange=moderate LD (r^2 <0.8) and red=high LD ($r^2 \geq 0.8$). The recombination rates in this window of the genome is indicated in the right y-axis (light blue line).

Amongst the four SNPs overlapping $CD14^+$ monocyte-specific ATAC peaks was the fine-mapping lead SNP rs425371 (PP=0.14), located at an intergenic region, 269.3Kb downstream the *TNFRSF9* gene (Figure 4.20 top panel). The other three SNPs overlapping $CD14^+$ monocytes accessible chromatin were rs11121131 (PP=0.076), rs12745477 (PP=0.09) and rs417065 (PP=0.07) (Figure 4.20 top panel). Notably, the two ATAC peaks harbouring the four fine-

mapped SNPs showed variable chromatin accessibility across individuals (but it was not a DAR) (Figure 4.20 bottom panel), enrichment for H3K4me1 $CD14^+$ monocytes Roadmap Epigenomics Project data and a modest enrichment for in-house H3K27ac data in some of the samples (Figure 4.20 bottom panel). This was consistent with enhancer designation in these two regions in the chromatin segmentation map for $CD14^+$ monocytes. No accessible chromatin was found at the location of the four fine-mapped SNPs in $CD4^+$, $CD8^+$ and $CD19^+$ cells, in line with the classification of these two regions as heterochromatin or repressed chromatin by the chromatin segmentation maps in the corresponding cell types. Furthermore, rs11121131, rs12745477 and rs417065 also overlapped with TF binding sites identified by ENCODE ChIP-seq data in a number of cell types and rs11121131 was nearby a CpG island (Figure 4.20 bottom panel), altogether reinforcing a putative role of these two regions in regulating gene expression. Integration with eQTL datasets revealed that rs12745477 and rs42537 had a modest effect in regulating *PARK7* expression levels in $CD14^+$ monocytes treated with LPS for 24h (FDR=0.01) or in whole blood (FDR=0.04), respectively (Fairfax *et al.* 2012; Westra *et al.* 2013). Promoter capture Hi-C data showed physical interaction between a bait containing rs12745477 and the *ERRFI1* gene (Javierre *et al.* 2016). The gene expression data in this thesis did not find differences for *TNFRSF9*, *SLC45A*, *PARK7*, *ERRFI1* or any of the other proximal genes between psoriasis patients and healthy controls in $CD14^+$ monocytes.

4.3.9 Allele-specific differences in chromatin accessibility at the GWAS locus 2p15

The chr2p15 psoriasis risk locus (lead SNP rs10865331, OR=1.12) is another of the psoriasis GWAS associations identified by the Immunochip study from Tsoi *et al.* 2012 that is located in a large intergenic region. This locus is also shared with other chronic inflammatory diseases including AS and CD (Cortes *et*

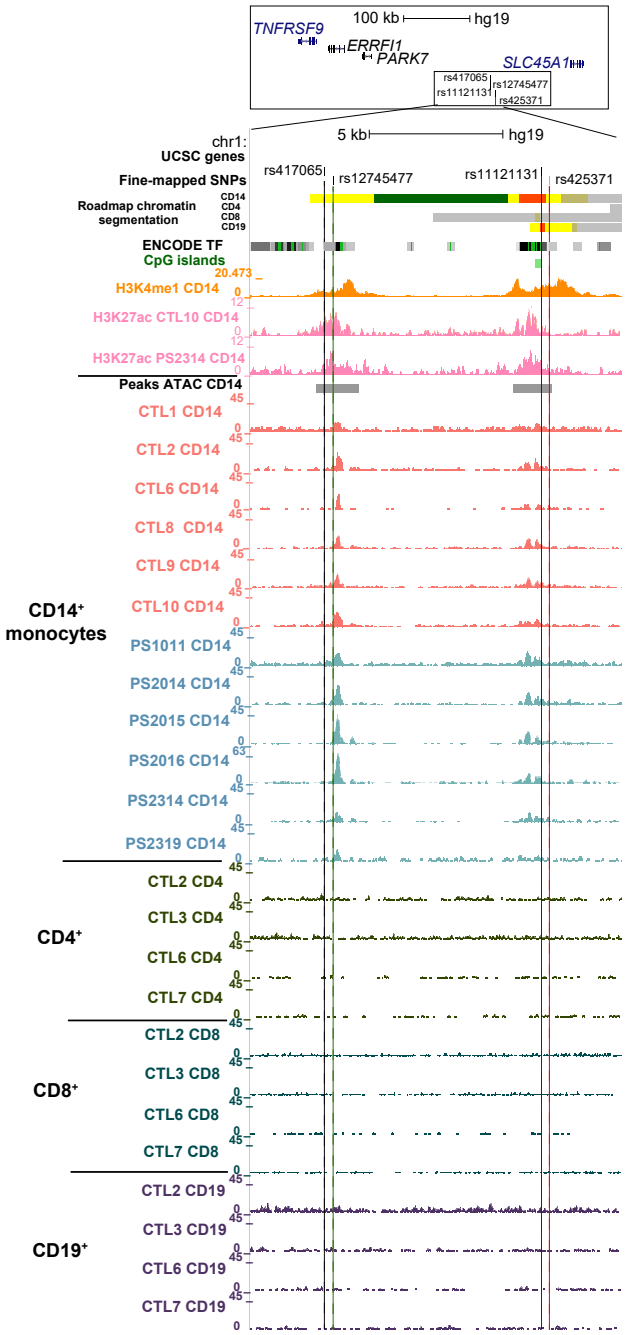


Figure 4.20: Epigenetic landscape at the location of the SNPs in the four SNPs in the 90% credible set for the *SLC45A1/TNFRSF9* psoriasis GWAS locus. The top panel illustrates the genomic location of the four SNPs in the *SLC45A1/TNFRSF9* fine-mapping 90% credible set overlapping ATAC accessible regions in CD14$^+$ monocytes and their distance to the nearest genes. The bottom panel represents the UCSC Genome Browser view illustrating normalised ATAC read density and H3K27ac fold-enrichment as well as publicly available epigenetic datasets (chromatin segmentation map, H3K4me1, ENCODE TF ChIP-seq and CpG islands) (y-axis) at the location of the four SNPs (rs425371, rs11121131, rs12745477 and rs417065) (x-axis) from the credible set. Only representative ATAC tracks from control samples are shown for CD4$^+$, CD8$^+$ and CD19$^+$ cells to illustrate the absence of accessible chromatin at the location of the four SNPs.

al. 2013; Jostins *et al.* 2012). The fine-mapping using summary statistics from the psoriasis GAPC Immunochip performed here identified a signal with the lead SNP (rs9309343) showing $\log_{10}$ABF=4.3; however it was not in LD with the psoriasis GWAS lead SNP from Tsoi *et al.* 2012. Fine-mapping analysis on the AS UK Immunochip performed by Dr Anna Sanniti refined the 95% credible set of this locus to three SNPs (rs4672505, rs6759298 and rs6759003), of which rs4672505 (risk allele rs4672505_A) demonstrated the greatest PP (0.4). Notably, rs4672505 was also identified as the lead SNP for an association signal in Ellinghaus and colleagues multi-disease meta-analysis shared by psoriasis, AS and CD (Ellinghaus *et al.* 2016).

rs4672505 overlaps a $CD8^+$ T cell-specific ATAC peak, not present in $CD14^+$ monocytes, $CD4^+$ and $CD19^+$ cells (Figure 4.21A), that was not differentially accessible between patients and controls in the differential analysis but did show marked variability across individuals (Figure 4.21B), with some individuals (PS2314 and CTL1) demonstrating no ATAC signal at this location.

Integration with publicly available ENCODE and Roadmap Epigenomics Project DHS data confirmed accessible chromatin at this site in Th-1, Th-2 and Th-17 cells and $CD8^+$ T cells, respectively (Figure 4.21A). Variability across individuals was also observed for H3K27ac enrichment as per ChIPm data generated in cohort 1B (Figure 4.21B). Although prediction of disrruption for TF binding was not found for rs4672505, ENCODE ChIP-seq data from GM12878 showed binding of RUNX3 at this location. RUNX3 is a TF involved in $CD8^+$ cell differentiation (Wong *et al.* 2011) and also a psoriasis and AS GWAS locus. In addition, *in silico* TF binding sites prediction using PROMO (Messeguer *et al.* 2002) and ENCODE genomic DNase-I footprint in GM128778 predicted STAT1 binding at a region spanning rs4672505. Altogether, integration of fine-mapping, ATAC and publicly available epigenetic data indicated that rs4672505 was a

likely causal functional variant accounting for the association of the 2p15 locus with psoriasis risk.

The genotype at rs4672505 of each individual was determined using Sanger sequencing. Amongst the eighteen samples (nine controls and eight psoriasis patients), one (PS2314) was homozygous for the risk allele (A, MAF=0.43), 11 were heterozygous and 6 were homozygous for the protective allele (G) (Figure 4.21B). Interestingly, PS2314, the only homozygous individual for the risk allele, showed complete absence of the ATAC peak at rs4672505. To further investigate the role of rs4672505 genotype in the variability of chromatin accessibility across individuals, normalised read counts at the ATAC peak chr2:62,559,749-62,561,442 were used as a dependent variable in linear model analysis based on rs4672505 genotype, using batch as a covariate.

Significant negative correlation (p-value=0.035 and β=-24.082) for the number of copies of the minor allele was found, showing allele-dependent chromatin accessibility (Figure 4.22A). Furthermore, allelic imbalance of ATAC reads at rs4672505 position was investigated in individuals identified as heterozygous by Sanger sequencing, for which 50% of the ATAC reads were expected to map to each of the alleles. This analysis demonstrated a larger percentage of ATAC reads (greater than the expected 50%) preferentially tagging the protective allele G (Figure 4.22B). This finding was not driven by mapping bias, since A was the reference allele in the hg19 build used in this analysis. Overall, these results showed evidence of greater chromatin accessibility in presence of the fine-mapped protective allele rs4672505(G) at the chr2p15 locus.

A major challenge with intergenic GWAS signals is the difficulty in determining the specific gene they may be modulating, for example through differential enhancer activity affecting gene expression. rs4672505 is located 140Kb downstream of *B3GNT2* and 150Kb upstream of *TMEM1*. Publicly available promoter capture Hi-C data from Javierre *et al.* 2016 in $CD8^+$

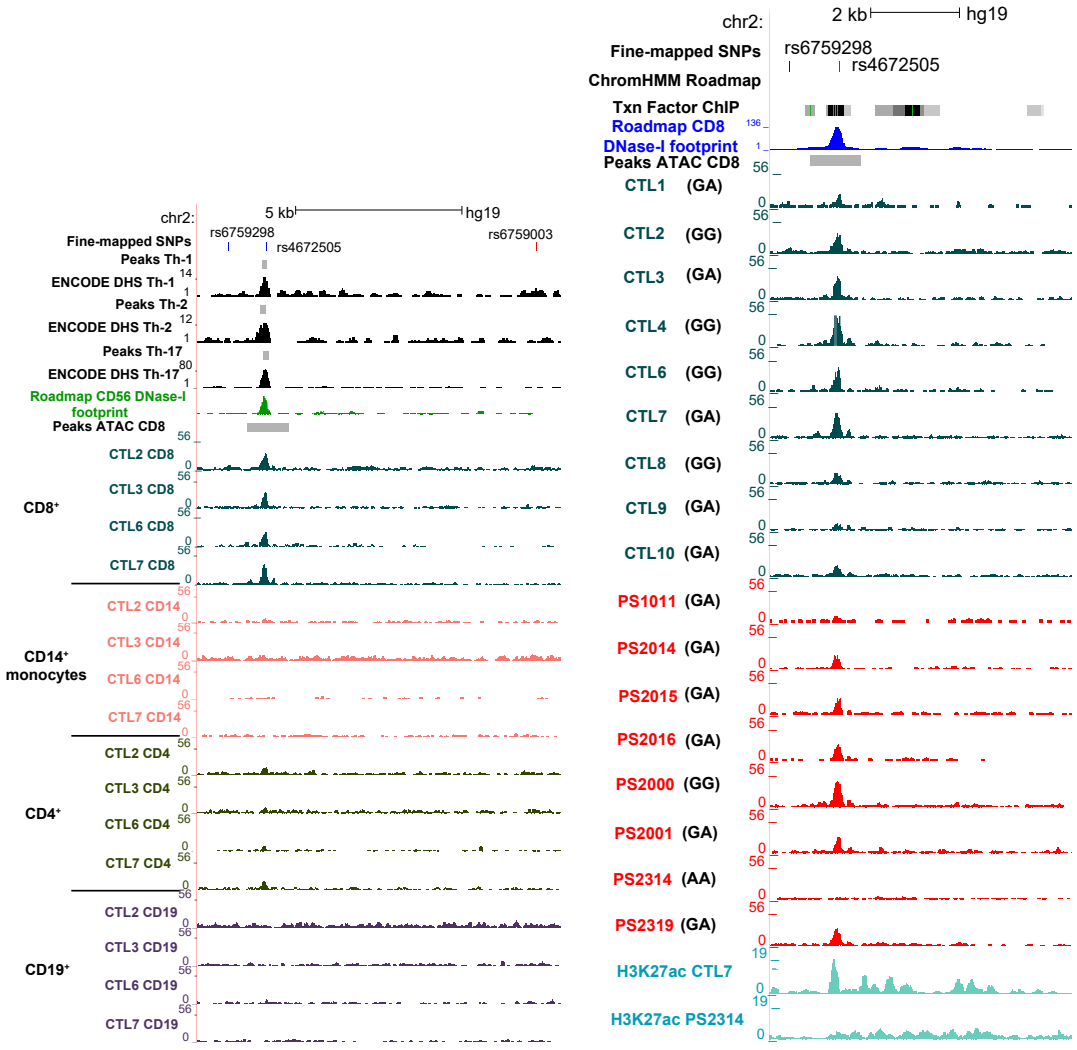


Figure 4.21: Epigenetic landscape at the location of the 95% credible set of AS fine-mapped SNPs for the chr2p15. (A) UCSC Genome Browser view illustrating normalised read density for in-house ATAC and a number of other publicly available epigenetic data (DHS, DNAse-I footprint, chromatin segmentation map) (y-axis) snapping the three SNPs (rs6759298, rs4672505 and rs6759003) (x-axis) from the 95% credible set obtained in the fine-mapping analysis of the chr2p15 GWAS association in AS. Representative ATAC data from the same four controls in the cohort and the four cell types included in this study are shown. (B) UCSC Genome Browser view illustrating the normalised read density for $CD8^+$ ATAC (x-axis) generated in psoriasis patients and healthy controls, in-house H3K27ac ChIPm, ENCODE TF ChIP-seq and DNase-I fooprint (y-axis) at the location of the SNP rs4672505 (y-axis). For each of the patients and controls of the cohort the Sanger sequencing genotype of rs4672505 is included (A=risk allele).

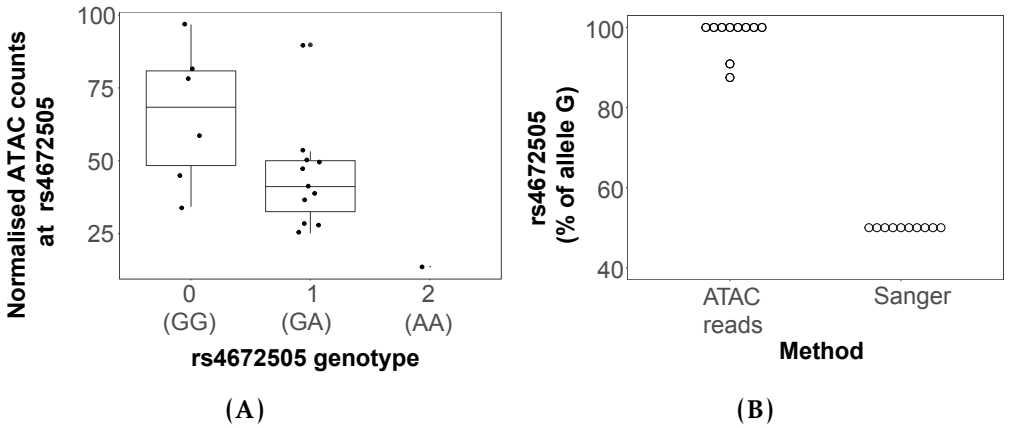


Figure 4.22: rs4672505 genotype and chromatin accessibility at chr2:62,559,749-62,561,442 in CD8$^+$ cells. (A) Boxplot illustrating the effect of the rs4672505 genotype on chromatin accessibility at the chr2:62,559,749-62,561,442 ATAC peak. Log_2 normalised ATAC read counts adjusted for batch effect (included as a covariate for the linear model) are plotted for each sample against the number of copies of the minor allele (GG=0, AG=1, AA=2). (B) Representation of the percentage of ATAC reads mapping to the major allele (G) at rs4672505 in comparison to the Sanger genotype results for the 11 heterozygous individuals.

revealed a significant physical interaction (CHiCAGO score=7.67) between the region containing rs4672505 and the promoter of *B3GNT2* (Figure 4.23A). This interaction was not found in any of the additional 16 human primary hematopoietic cell included in the study and no upstream interaction with *TMEM1* promoter was identified. Investigation of the publicly available T cell eQTL dataset from Kasela *et al.* and Raj *et al.* did not show a significant eQTL for rs4672505 or SNPs in high LD (r^2>0.8) either in CD8$^+$ or CD4$^+$ (Raj *et al.* 2014; Kasela *et al.* 2017). Similarly, no eQTL effect of rs4672505 was found in unstimulated or stimulated CD14$^+$ monocytes (Fairfax *et al.* 2014). However, a whole blood eQTL study from Jansen and colleagues revealed a significant *cis*-eQTL (FDR=1.34x10^{-5}) with moderate effect size (β=-0.16) for the minor allele (Jansen *et al.* 2017). Differential gene expression analysis in circulating immune cells (previously presented) showed significant down-regulation (FDR<0.05) of *B3GNT2* expression in psoriasis patients when compared to controls in CD8$^+$ cells (fold change=0.80) (Figure 4.23B).

Altogether, these data showed evidence of allele-specific chromatin accessibility at chr2p15 in $CD8^+$ T cells involving rs4672505, with chromatin conformation capture and gene expression data suggesting that this SNP may regulate the expression of *B3GNT2*, at a distance, and providing new insights into the possible mechanistic basis of the disease association seen in psoriasis, AS and CD for this locus.

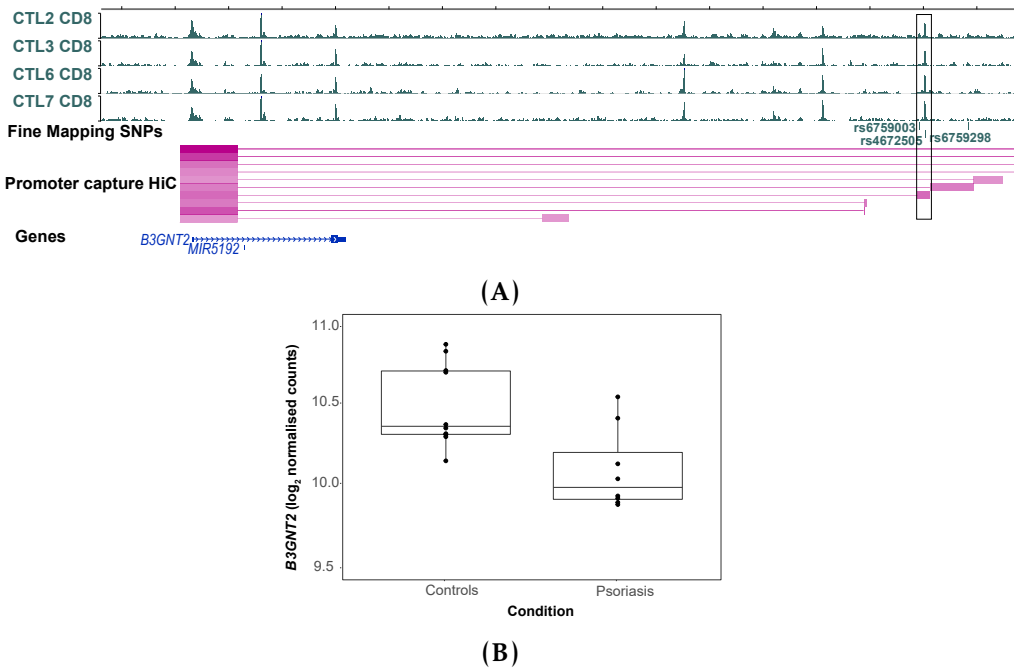


Figure 4.23: Potential role of rs4672505 in regulating *B3GNT2* gene expression. (A) WASHU Genome Browser track showing normalised ATAC read density in four of the healthy controls $CD8^+$ cells at the location of the three SNPs in the 95% credible set (AS fine-mapping analysis) and the promoter capture Hi-C data depicting the regions interacting with viewpoint at the *B3GNT2* promoter. (B) Boxplot illustrating the *B3GNT2* $\log_2$ normalised RNA-seq counts adjusted for batch effect in psoriasis and healthy controls.

4.4 Discussion

4.4.1 Chromatin accessibility and H3K27ac landscape in psoriasis immune cells

Epigenomic characterisation of circulating immune cells from psoriasis patients revealed several regions with altered chromatin accessibility and H3K27ac histone modifications, mainly in $CD14^+$ monocytes and $CD8^+$ T cells.

In psoriasis $CD8^+$ T cells greater chromatin accessibility was found in two regions proximal to *IL7R* and *TNFSF11* that overlapped FANTOM eRNAs in the same cell type. Both genes are well known for having a pro-inflammatory effect and to be involved in chronic inflammatory diseases (Gregory *et al.* 2007; Cortes and Brown 2011). For example, *TNFSF11* is proximal to one of the GWAS associations in CD (Franke *et al.* 2010). Its protein product RANKL was found to be overexpressed in epidermis from psoriasis patients and in activated T cells, where it leads to activation of the pro-inflammatory kinase MAPK8 (Toberer *et al.* 2011; Wong *et al.* 1997).

Overlap of the ATAC and H3K27ac ChIPm differential analysis found only one region in $CD8^+$ cells in an intron of *DTD1*, containing eQTL SNPs regulating *DTD1* expression in whole blood (GTEx). D-Tyr-tRNA deacylase is responsible for releasing D-aminoacids from the tRNAs, making them available to be loaded with L-aminoacids for the production of functional proteins (Bhatt *et al.* 2016). Aminoacyl-tRNA synthetases, responsible for the loading of L-aminoacids, have been described to have a role in modulation of inflammation and angiogenesis (Yao and Fox 2013); however, no evidence of *DTD1* involvement in chronic inflammation and psoriasis has yet been reported. Overall, general lack of overlap between DARs and differentially H3K27ac modified regions is not unexpected since chromatin accessibility is the result of a complex network

of interactions between a number of histone modifications, TFs, and structural proteins

The results in this chapter suggest that disease status does not involve global differences in chromatin accessibility and H3K27ac compared to controls in immune cells isolated from peripheral blood. Recent similar studies performing ATAC in B cells from SLE patients showed larger differences in the chromatin accessibility landscape between patients and controls (Scharer *et al.* 2016). Likewise, H3K27ac mapping in mCD4$^+$ cells isolated from juvenile idiopathic arthritis synovial fluid found approximately a thousand differential enhancers compared to healthy control circulating cells, whereas only small differences were found when comparing mCD4$^+$ from peripheral blood of patients and controls, (Peeters *et al.* 2015). This may highlight the specificity of the disease signature at the site of inflammation and the importance of studying the most disease relevant cell types and tissues.

4.4.2 Dysregulation of gene expression in psoriasis circulating immune cells

Comparison of gene expression between psoriasis and healthy controls in a cell-type specific manner identified larger numbers of DEGs compared to epigenetic changes. Similar to ATAC and ChIPm, CD14$^+$ monocytes and CD8$^+$ cells showed the largest number of transcriptomic changes in disease. This may suggest greater relevance of these two cell types in the systemic "footprint" of psoriasis. The more dysregulated gene expression in CD8$^+$ compared to CD4$^+$ may suggest that, as in skin, CD8$^+$ are the main effector cells upon induced-activation by CD4$^+$ cells (Nickoloff and Wrone-Smith 1999). Monocytes/macrophages are important sources of TNF-α, IL-12 and IL-23 at the site of inflammation and are present in the psoriasis skin lesions, where their TNF-α production, for example, contributes towards maintenance of

inflammation (Parameswaran and S 2010; Nickoloff *et al.* 2007; Wang *et al.* 2006). Amongst the key pro-inflammatory psoriasis genes, only *IL23A* showed dysregulation in $CD14^+$ monocytes showing down-regulation compared to controls, which could be the result of a feed-back loop.

The cell type-specific RNA-seq analysis conducted in my thesis identified significant enrichment of relevant biological processes, including MAPK and IL-12 signalling, in $CD14^+$ monocytes and $CD8^+$ T cells. Interestingly, some of the well-known pro-inflammatory genes contributing to the enrichment of these pathways were down-regulated in psoriasis compared to controls. For example, the down-regulation of *MAP3K4* in $CD14^+$ monocytes and $CD8^+$ T cells from psoriasis patients was consistent with its reduced expression in LPS stimulated PBMCs from CD patients, and has been identified as an immune-suppressive feature leading to reduced expression of the cytokine IL-1α (Kraan *et al.* 2012). IL-12 signalling pathway leads to Th-1 differentiation and IFN-γ production through activation of TFs from the STAT family, importantly STAT3 but also STAT1, STAT4 and STAT5 (Kusaba *et al.* 2005). Here, $CD14^+$ cells showed down-regulation of *STAT4* and *STAT5A* in psoriasis patients. *STAT2* has previously been found down-regulated in psoriasis PBMCs and in AS monocyte-derived macrophages but was not differentially expressed in my data (Coda *et al.* 2012; Smith *et al.* 2008). In other chronic inflammatory diseases such as T2D, persistent STAT5 phosphorylation has been found in circulating monocytes isolated from patients upon GM-CSF stimulation (Litherland *et al.* 2005). Further investigation to determine protein abundance, particularly of the phosphorylated forms of STAT4 and STAT5, will be required to determine whether the down-regulation at the transcript level observed in psoriasis $CD14^+$ monocytes is biologically relevant.

A very interesting observation in this data was the down-regulation of *IFNG* in psoriasis $CD8^+$ cells, a down-stream gene which undergoes up-regulation as

a results of IL-12 signalling. This has previously been reported in unstimulated and stimulated macrophages derived from AS patients, in synovial fluid from SpA patients compared to RA and in DCs from a SpA rat model (Smith *et al.* 2008; Fert *et al.* 2014). In AS monocytes and in rat SpA DCs, *IFNG* down-regulation was accompanied by an overall "inverse" transcriptional response of IFN-regulated genes characterised by up-regulation of IFN-γ supressed genes, which was not seen in my data. Reduced expression of *IFNG* in knock-out mice has been shown to increase activation of the IL-23/IL-17 axis, which is pivotal in psoriasis pathogenesis, resulting in a pro-inflammatory effect (Cañete *et al.* 2000; Chu *et al.* 2007). Nevertheless, neither IL-23 or IL-17 were up-regulated in the psoriasis cell types interrogated in this thesis compared to healthy controls.

In $CD8^+$ DEGs showed significant enrichment for three relevant pathophysiological pathways in psoriasis: NF-κB, TNF and chemokine signalling, with a number of dysregulated genes contributing to the three as a results of pathway cross-talk. Interestingly, the enrichment of these pathways involved up-regulation of pro-inflammatory genes (e.g *JUNB*, *ATF4*, *RELA* and *RELB*) but also increased expression of well-characterised immunoregulatory genes. These included *NFKBIA* and *TNFAIP3*, also up-regulated in $CD14^+$ monocytes and $CD4^+$ cells, respectively, and both associated with psoriasis risk and other chronic inflammatory diseases, including MS, RA, SLE and T1D (Vereecke *et al.* 2011). *NFKBIA* (coding for IκB-α) inhibits NF-κB by binding it and preventing its translocation to the nucleus. Similarly, *TNFAIP3* (A20) undergoes up-regulation upon inflammatory stimuli and NF-κB activation to inhibit the NF-κB TNF-mediated response and promote the return to homeostasis. *NFKBIA* and *TNFAIP3* were not found to be dysregulated in psoriasis PBMCs by Coda *et al.* 2012 , Lee *et al.* 2009 and Mesko *et al.* 2010 or in PBMCs from PsA patients (Dolcino *et al.* 2015). Interestingly, qPCR analysis in PBMCs from mild (PASI<4.84) and severe (PASI>4.84) psoriasis

vulgaris revealed a significant negative correlation between *TNFAIP3* expression and disease severity (Jiang *et al.* 2012). They also demonstrated *TNFAIP3* up-regulation in PBMCs from the mild group of patients, but not from the severe one, when compared to healthy controls. This would be consistent with my findings, with the caveat that all patients from my cohort would be classified as severe (PASI>4.84) according to Jiang and colleagues criteria. Moreover, in UC similar up-regulation of *TNFAIP3* in patients inflamed colon biopsies has been reported (Majumdar *et al.* 2017). Altogether, the up-regulated expression of *TNFAIP3* and *NFKBIA* in psoriasis patients immune cells may be the result of a feedback loop reflecting persistent inflammatory stimuli in peripheral blood and a still functional mechanism that limits the systemic inflammatory response to some extent (Idel *et al.* 2003).

Regarding chemokine signalling, none of the psoriasis relevant pro-inflammatory chemokines or their receptors showed differential expression in this analysis, and Coda *et al.* 2012 only demonstrated up-regulation for *CXCL5* in psoriasis PBMCs. In this thesis, differences in chemokine expression between patients and controls may have been masked by studying total $CD8^+$ cells instead of more discrete subpopulations such as memory T cells. Nevertheless, enrichment of this pathway was driven by up-regulation of the chemokine receptors *CCR10* and *CXCR4* in psoriasis $CD8^+$ cells, which were not differentially expressed in the studies using PBMCs (Coda *et al.* 2012; Lee *et al.* 2009; Mesko *et al.* 2010). In peripheral blood, *CCR10* expressing $mCD4^+$ and $mCD8^+$ T cells are preferentially recruited to inflamed skin, where *in vitro* keratinocytes express the *CCR10* ligand CCL27, induced upon TNF-α and IL-1β stimulation (Hudak *et al.* 2002; Homey *et al.* 2002). Interestingly, *in vivo* studies have reported the down-regulation of *CCL27* in lesional skin and lack of $CCR10^+$ expression by most infiltrated effector cells in skin, suggesting a putative role of this T cell subpopulation in maintenance of the skin homesotasis (Sahmatova *et*

al. 2017; Yang *et al.* 2016a). Therefore the increase on *CCR10* expression in $CD8^+$ cells may reflect the lack of recruitment of this subpopulation and therefore impaired homesotasis at the inflamed skin. Similarly, the CXCR4/CXCL12 is pivotal for recruitment of $CD8^+$ cells into the epidermis in a mice model of skin inflammation (Zgraggen *et al.* 2014). Altogether, the up-regulation of *CCR10* and *CXCR4* could suggest an increase of $CD8^+$ $CCR10^+$ CXCR4+ cells ready to migrate into the psoriatic lesional skin.

The simultanous up- and down-regulation of both, pro- and anti-inflammatory genes, observed in this analysis does not clearly point to an overall hyperinflammation or immunosuppression. Nevertheless, these data reveal differences in gene expression between psoriasis patients and controls in $CD14^+$ monocytes and $CD8^+$ T cells for genes involved in a number of pathophysiologically relevant pathways in psoriasis, demonstrating a systemic signature of disease.

Despite finding changes between healthy controls and patients, separating disease cause and consequence remains in general challenging in this type of study. To address this question, the integration of transcriptomic profiles with GWAS data can help to identify putative causal genes through their link to genetic variants predisposing to disease. eQTLs can be used for this purpose, so genes that are differentially expressed and mapping to a disease-related eQTL are more likely to be involved in disease aetiology (Yagil *et al.* 2005). Similarly, caQTLs could also be integrated together with eQTLs and the transcriptomic profiles to identify GWAS variants that are simultaneously affecting chromatin accessibility and expression. Nevertheless, this approach in order to separate causality from consequence is limited by lack of eQTL and caQTLs data in all the disease-relevant cell types and conditions (e.g patient-specific QTLs), the incomplete functional characterisation of many GWAS intergenic variants and the high number of susceptibility loci with small effect that have not yet been

identified. In addition to data integration, the design of longitudinal studies analysing the transcriptomic profiles in the same individuals before and after developing disease and comparing with controls could also provide valuable information to shed some light into disease aetiology (Petri *et al.* 2009).

4.4.3 Integration of chromatin accessibility and gene expression data

In this analysis, greater changes in gene expression than in chromatin accessibility were found when comparing psoriasis patients circulating immune cells to healthy controls. Strikingly, in $CD8^+$ cells, 687 transcripts were differentially expressed between psoriasis and healthy controls but only 55 regions showed differential chromatin accessibility and only six of the 687 DEGs were proximal to DARs. This lack of correspondence may be due to different sensitivities of the ATAC and RNA-seq assays when only small changes take place (e.g in absence of stimulation and not at the main site of inflammation). Additionally, the discordance could be explained by the complexity of gene regulation where chromatin accessibility is only one of the events, and may not be temporally correlated with gene expression. For example, in some cases differential chromatin accessibility in homeostasis only translate into changes in gene expression upon stimulation (Alasoo *et al.* 2018; Calderon *et al.* 2018). Similarly, lack of overlap between gene expression and protein levels have been extensively reported, illustrating the complex relationship between epigenetics, transcription and translation (Liu *et al.* 2016).

In addition to the genome-wide quantitative discrepancies between differential chromatin accessibility and gene expression, only seven DEGs were proximal to a DAR. Notably, no differential expression was observed for *DTD1*. Two relevant examples of DEGs nearby DARs were *TIAM1* and *TRAT1*, which showed increased chromatin accessibility and gene expression in psoriasis $CD8^+$

cells. *TRAT1* is a positive regulator for TCR signalling which has been shown to be down-regulated in Tregs (Birzele *et al.* 2011) and the T cell lymphoma invasion and metastasis 1 (*TIAM1*) is involved in IL-17 expression and cell migration into inflamed tissues (Kurdi *et al.* 2016; Gérard *et al.* 2009). Interestingly, the intronic *TIAM1* DAR, annotated as an enhancer by chromatin segmentation maps, was found to have frequent interactions with *TIAM* promoter in unstimulated $CD8^+$ cells (Javierre *et al.* 2016), suggesting a putative role for this region in regulating expression of *TIAM1*. Additional investigation will be required to formally establish a link between these two events.

4.4.4 Trancriptomic profiles in lesional and uninvolved psoriatic epidermis

Investigation of differences in the transcriptomic profiles between paired lesional and uninvolved skin was conducted for three psoriasis patients. Most previous transcriptional studies in psoriasis have used full-thickness skin biopsies, comprising a mix of cell types including fibroblasts, adipocytes, keratinocytes (from the epidermis and dermis) and infiltrated immune cells. Large differences in gene expression between full-thickness biopsies and FACS-isolated keratinocytes have been demonstrated and this may be masking keratinocyte-specific (and also immune-infiltrated) pathophysiological differences (Ahn *et al.* 2016). In this chapter, RNA-seq was conducted on epidermal sheets isolated from whole biopsies and a total of 1,227 DEGs were identified. Comparison with a contemporary study by Tervaniemi using lesional and uninvolved psoriatic epidermis split biopsies (mainly formed by epidermal keratinocytes) revealed an overlap of 359 out of the 1,227 DEGs (12.1% and 30% of Tervaniemi *et al.* and this thesis DEGs, respectively) (Tervaniemi *et al.* 2016). Despite similarity in the sample type, the Tervaniemi study used 5' RNA-seq library preparation and the quantification was based on synthetic spike-in

RNA-seq normalisation, which may also explain lack of a larger overlap with my data.

DEGs between lesional and uninvolved psoriatic skin in this data has also highlighted enrichment for HIF-1 and IL17-signalling, two closely related pathways in psoriasis pathogenesis. HIF-1α is a TF induced under hypoxic conditions that plays and important role in the expression of genes involved in angiogenesis (Forsythe *et al.* 1996). Although *HIF-1A* is expressed in normal skin, due to hypoxic conditions in homeostatic epidermis, it undergoes further up-regulation upon psoriatic lesion development and correlates with an increase in vascular endothelial growth factor (*VEGFA*) transcript levels (Rosenberger *et al.* 2007). This is consistent with the up-regulation of *HIF1A* and *VEGFA* observed in my data and supports enhanced angiogenesis in the psoriatic lesional compared to uninvolved epidermis. HIF-1 signalling is also implicated in regulating Th-17/Treg ratios and thus in IL-17 signalling (Dang *et al.* 2011; Shi *et al.* 2011). The observed enrichment for the IL-17 signalling pathway is consistent with the therapeutic role of IL-17A inhibitors in the treatment of psoriasis (Mahil *et al.* 2016; Coates *et al.* 2016b). IL-17 signalling in skin was recapitulated here by up-regulation of *IL6*, *IL36*, *STAT3*, the chemokine *CCL20* and genes from the *S100* family (involved in keratinocyte differentiation, calcium sensing and chemotaxis), all shown to be dysregulated by previous studies (Tsoi *et al.* 2015b; Tervaniemi *et al.* 2016). Dang *et al.* have shown that HIF-1α-STAT3 is required for activation of IL-17 transcription in naíve T cells. In keratinocytes, Th-17 cytokine secretion promotes keratinocyte proliferation, activation of NF-κB and STAT3, and downstream up-regulation of chemokines such as *CCL20* (Carrier *et al.* 2011; Harper *et al.* 2009). Additionally, up-regulation of *IL36A* (fold change=1.46) also observed in this data was in line with previous studies where this cytokine has demonstrated to augment Th-17 cytokines effect through a positive feedback loop in psoriasis (Keermann *et al.* 2015; Tervaniemi *et al.* 2016; Carrier *et al.* 2011).

Enrichment of the NOD-like signalling in DEGs was consistent with Tervaniemi *et al.* findings. NOD-like signalling involves signal transduction by NOD-like receptors, which can recruit and activate caspases into the inflammasomes for IL-1β maturation or trigger inflammation through activation of NF-κB and MAPKs (McCormack *et al.* 2009). NOD-like receptor signalling genes up-regulated in this data and in the Tervaniemi study included *CARD6*, *IFI16*, *NOD2* and *NLRX1*, amongst others. *NOD2* (also known as *CARD15*) is a particularly relevant member of this pathway for which genetic variability has been linked to inflammatory diseases including CD, atopic eczema, arthritis and potentially to psoriasis and PsA (Zhong *et al.* 2013; Zhu *et al.* 2012). Tervaniemi and colleagues demonstrated significant differences in DEGs between microarrays and RNA-seq experiments, and attributed the enrichment of NOD-like receptor signalling and other novel pathways to the greater sensitivity of RNA-seq. The DEGs from the Tsoi study, using RNA-seq in a larger cohort of full-thickness skin biopsies, did not show enrichment for NOD-like signalling and identified comparatively less DEGs belonging to this pathway than my data, highlighting the value of studying epidermis instead of full-thickness skin to uncover dysregulation of additional functionally relevant pathways in psoriatic keratinocytes. Notably, the importance of studying pure cell populations has been further supported by a recent scRNA-seq report identifying important phenotypic heterogeneity within and across epidermal keratinocyte from different anatomical locations, as well as within psoriatic lesional epidermis (Cheng *et al.* 2018). Interestingly, this scRNA-seq study identified the expansion in psoriasic epidermis of two keratinocyte subpopulations (named as channel and mitotic) characterised by the expression of pore and intercellular communication transcripts or DNA replication and cell division genes, respectively, consistent with the enrichment of cell cycle and mitotic processes identified in the bulk RNA-seq analysis presented here.

Lastly, pathways enriched in DEGs between lesional and uninvolved psoriatic epidermis highlighted dysregulation of metabolic processes, including amino acid metabolism, glycolysis, biological oxidation and PPAR signalling, some already identified in other studies (Coda *et al.* 2012; Gudjonsson *et al.* 2010; Aterido *et al.* 2016; Tervaniemi *et al.* 2016). Metabolic alterations can increase the disponibility of advanced glycation end products (AGEs), which are reported to be up-regulated in skin and serum of psoriasis patients (Papagrigoraki *et al.* 2017). In line with Tsoi and Tervaniemi studies, my data showed up-regulation of *S100A7* and *S100A12*, which shown to be ligands for AGE receptor (RAGE) (Eckert *et al.* 2004; Moser *et al.* 2007; Broome *et al.* 2003). Although up-regulation of *RAGE* has not been identified in in my data, the enrichment of the AGE-RAGE signalling pathway together with dysregulation of other metabolic processess revealed by pathway analysis suggests a contribution of this cascade to the transcriptional activation of NF-κB and STAT signalling observed here. Similarly, enrichment of the PPAR signalling represents another bridge between metabolic and innate immunity dysregulation in psoriasis. Following skin injury, *PPARG* and *PPARD* undergo up-regulation as a result of TNF-α and IFN-γ release (Tan *et al.* 2001). Up-regulation of *PPARD* has previously been reported in psoriatic lesional skin, in line with my data, and molecular studies have demonstrated a role of this PPAR in keratinocyte hyperproliferation through induction of heparin-binding EGF-like growth factor (HB-EGF) (Romanowska *et al.* 2008).

Comparison of the dysregulated genes in circulating immune cells vs psoriatic skin revealed very limited overlap, with $CD14^+$ monocytes and $CD8^+$ cells showing the greatest number of shared DEGs, also likely due to these cell types showing the largest number of DEGs in the analysis. Notably, almost half of the overlapping genes demonstrated opposite direction in differential expression, consistent with a report by Coda comparing DEGs in psoriasis PBMCs to DEGs in skin biopsies (Coda *et al.* 2012). Genes showing opposite

expression changes in circulating immune cells and in skin included the GWAS hit *TNFAIP3* and others such as *EGR2* and *EGR3*. *EGR2* up-regulation in skin may also increase keratinocyte proliferation as has been shown in certain types of cancer (Wu *et al.* 2010). Regarding enriched pathways differences between circulating immune cells and skin were found. Importantly DEGs in lesional psoriatic skin highlighted the role of IL-17, NOD-like recptor signalling and metabolic processes at the main site of inflammation.

4.4.5 LncRNAs in psoriasis

The largest number of differentially expressed lncRNAs between psoriasis patients and controls was found in $CD14^+$ monocytes and $CD8^+$ T cells (28 and 31, respectively for FDR<0.05). Similarly, 46 lncRNAs were also found differentially expressed between lesional and uninvolved skin (FDR<0.05) in this thesis. Several studies have contrasted lncRNAs in lesional, uninvolved and healthy skin (Li *et al.* 2014; Gupta *et al.* 2016; Ahn *et al.* 2016; Tsoi *et al.* 2015b). The role of lncRNAs in chronic inflammation has been studied in RA and SLE monocytes, AS hip joints and PsA PBMCs (Müller *et al.* 2014; Shi *et al.* 2014; Zhang *et al.* 2017; Dolcino *et al.* 2018). However, no study has been conducted to identify differentially expressed lncRNAs in a cell type-specific manner in peripheral blood from psoriasis patients. Although characterisation of lncRNA biological function is a developing field, some of them have been demonstrated to participate actively in the regulation of the immune response (Heward and Lindsay 2014).

My data showed up-regulation of *HOTAIRM1* in psoriasis $CD14^+$ monocytes, which was accompanied by down-regulation of its predicted target gene *UPF1*. *UPF1* is involved in nonsense-mediated decay and degradation of inflammation-related mRNAs (Mino *et al.* 2015) and its down-regulation

in psoriasis $CD14^+$ monocytes may suggest impairment of this homeostatic mechanism and resulting longer half-life of immune-related transcripts.

Another example of a relevant lncRNA up-regulated in $CD14^+$ monocytes was *NEAT1*, which has also been observed in SLE PBMCs and $CD14^+$ monocytes (Zhang *et al.* 2016) Knock-down of this lncRNA in THP1 cells demonstrated impairment of TLR4 signalling and down-regulation of inflammatory genes including IL-6 and CXCL10 (Zhang *et al.* 2016). This is in line with the positive correlation between *NEAT1* expression levels and PASI scores in my data, further supporting the role of this mIR in the exhacerbation of the inflammatory response in psoriasis.

MIR146A was found to be differentially expressed between lesional and uninvolved skin (up-regulated) and also when comparing psoriasis $CD8^+$ T cells to healthy controls (down-regulated) in my data. The up-regulation of *MIR146A* expression in lesional compared to uninvolved skin in epidermal and dermal compartments has also been observed in other studies (Tsoi *et al.* 2015b; Li *et al.* 2014). In RA, opposing findings have been reported regarding transcription and serum levels of miR-146 (Filková *et al.* 2014; Churov *et al.* 2015). In psoriasis, a study conducted by Xia and colleagues showed up-regulation of *MIR146* in lesional skin and PBMCs from psoriasis patients when compared to controls and demonstrated positive correlation with IL-17 expression levels (Xia *et al.* 2012). In Xia's study the miR-146 target gene *IRAK1* was down-regulated in lesional skin when compared to healthy controls; however up-regulation of *IRAK* was found in lesional skin compared to uninvolved from the same psoriasis patients despite down-regulation of *MIR146*, suggesting alteration in psoriatic skin of the miR-146 mechanism regulating *IRAK1*. In PsA, $CD14^+$ monocytes showed up-regulation of *MIR146* expression when compared to healthy controls, correlating with increased activation and bone reporption *in vitro*, a phenotype that was amelliorated upon its neutralisation (Lin *et al.* 2019). Despite *MIR146* being

involved in the down-regulation of *IRAK1*, a key molecules in NF-κB activation, data from the aforementioned studies demonstrate a pathogenic effect upon up-regulation of this miR that seems to be driven by IL-17, thus suggesting a more complex mechanisms of action of miR-146 in psoriasis that will require further investigation.

Lastly, *MIR31HG* was up-regulated in lesional epidermis. Notably, silencing miR-31hg in keratinocyte immortal cell line HaCaT induced cell cycle arrest and inhibited cell proliferation, consistent with two characteristic aspects dysregulated in psoriatic keratinocytes (Gao *et al.* 2018).

Overall, the lncRNA differential analysis conducted in this chapter gives an overview of dysregulation in blood and skin of psoriasis patients. A more comprehensive analysis could be performed by integrating the dysregulated lncRNAs targets reported by NPInter database and using those interactions to identify relevant biological processes through network and pathway analysis, similarly to the strategy used by Dolcino *et al.* 2018. However, such an analysis would likely require increased sample size to be appropriately powered.

4.4.6 Fine-mapping using summary statistics and integration with epigenetic data

Fine-mapping of GWAS loci is a strategy to narrow down putative functionally relevant variants identified by GWAS studies. Using summary statistics from the psoriasis Immunochip GWAS GAPC cohort (Tsoi *et al.* 2012), successful fine-mapping was achieved for 17 loci but for 10 of loci the fine-mapped lead SNPs were not in LD with the Tsoi and colleagues GWAS signal, likely due to the smaller sample size (GAPC cohort only), as previously demonstrated (Bunt *et al.* 2015). The analysis conducted here also showed a better ability to fine map to less than 10 SNPs those loci with larger effect size, notably *IL12B/ADRA1B* and *PTRF/STAT3*, consistent with other studies (Bunt *et*

al. 2015). It is important to note that statistical fine-mapping is sensitive to power and thus in some cases the size of the credible set may be larger than the one that would be obtained by using high LD (for example r^2 >0.8) as a proxy.

The seven successfully fine-mapped loci encompassed 126 SNPs out of which 16 co-localised with ATAC peaks in at least one cell type, and some of them with cell-type specific accessible chromatin. This integration with ATAC data further reduced the credible set of SNPs in the intergenic locus proximal to *SLC45A1/TNFRSF9* from 22 to 4, all of them overlapping $CD14^+$-specific ATAC peaks. eQTL and promoter capture Hi-C data were integrated to identify putative target genes which expression could be regulated by one or more of these four SNPs. Moderate regulation of *PARK7* expression by rs12745477 and rs425371 was observed in stimulated and unstimulated $CD14^+$ monocytes (Fairfax *et al.* 2014), whereas promoter capture Hi-C data in monocytes showed to an interaction of rs12745477 with the promoter of *ERRFI1* (Javierre *et al.* 2016). *ERRFI1* is a negative regulator of the epidermal growth factor receptor (*EGFR*) gene involved in skin morphogenesis (Ferby *et al.* 2006) and may have a more relevant role in skin; however it did not show differential expression between lesional and uninvolved psoriatic epidermis either in my data or previously (Tsoi *et al.* 2015b; Tervaniemi *et al.* 2016). *PARK7* is a sensor for oxidative stress causing early-onset of Parkison's disease (Bonifati *et al.* 2003). A role for *PARK7* in inflammation has been described in Tregs development and in bone marrow-derived macrophages, where *PARK7* impairs LPS response with implications in sepsis pathogen clearance (Singh *et al.* 2015; Amatullah *et al.* 2017). Further investigation will be required to determine the functional role of these two SNPs in psoriasis risk.

Another locus successfully fine-mapped in this analysis was *ETS1*, a TF involved in Th-1 inflammatory response (Grenningloh *et al.* 2005), which is also associated with other chronic inflammatory diseases including RA, SLE

and celiac disease (Okada *et al.* 2014; Trynka *et al.* 2011; Bentham *et al.* 2015). *ETS1* was up-regulated in psoriasis $CD8^+$ T cells and rs12223943 overlapped a $CD8^+$ specific ATAC peak, but no eQTL effect was found for this SNP in naïve $CD8^+$ cells. Further investigation incorporating genotype data may reveal genetic regulation of chromatin accessibility by this variant.

Efforts to integrate fine-mapping SNPs and tissue specific chromatin accessibility maps have led to successful prioritisation of putative causal variants in other diseases (Stefan *et al.* 2014). Moreover, tools such as Risk Variant Inference using Epigenomic Reference Annotation (RiVIERA) have been applied to perform fine-mapping using summary statistics and incorporate in the model fourty-three ENCODE and Roadmap annotation features most enriched for psoriasis GWAS loci (Li *et al.* 2016). For example, Li *et al.* study only reported two SNPs for the *SLC45A1/TNFRSF9* locus credible set in psoriasis but none was in LD with the GWAS signal from Tsoi *et al.* 2012 requiring careful interpretation of these results.

Allelic differences in chromatin accessibility at the chr2p15 locus

A particularly interesting psoriasis GWAS association is the 2p15 locus, where the lead SNP is located in an intergenic region 140Kb away from *B3GNT2* and 150Kb away from *TMEM1*, and is also associated with CD and AS (Jostins *et al.* 2012; Cortes *et al.* 2013). Fine-mapping at 2p15 in this thesis failed to identify a signal in LD with the psoriasis GWAS and the 90% credible set of SNPs did not contain the GWAS lead SNP. The results published by Li *et al.* using RiVIERA also failed to fine-map this locus in psoriasis. Conversely, fine-mapping analysis in AS implicated rs4672505, which is in high LD with the GWAS SNP (rs10865331), and was also reported to be the lead SNP for this locus in a cross-disease analysis (Ellinghaus *et al.* 2016). rs4672505 has also been shown as the 4^{th} variant prioritised by the fine-mapping algorithm published by Farh and colleagues

(Farh *et al.* 2015). Integration with my psoriasis-control cohort ATAC data showed that rs4672505 overlaps a $CD8^+$ specific peak. Chromatin accessibility at this location was observed to vary between individuals, and integration of genotyping data for rs4672505 with ATAC revealed allele-dependent chromatin accessibility, with loss of chromatin accessibility correlating with the number of copies for the risk allele (rs4672505_A).

Promoter capture Hi-C data linked rs4672505 to the *B3GNT2* promoter only in $CD8^+$ cells, suggesting the accessible chromatin at rs4672505 may harbour an enhancer interacting with *B3GNT2* promoter (Javierre *et al.* 2016). Moreover, RNA-seq data in this thesis also revealed down-regulation of *B3GNT2* in $CD8^+$ cells from psoriasis patients. rs4672505 was identified as an eQTL for *B3GNT2* in whole blood (Jansen *et al.* 2017) but not in $CD8^+$ cells (Kasela *et al.* 2017), potentially due to smaller sample size of the latter study or a missed stimuli-specific effect. To our knowledge, the genotypic effect of this SNP in chromatin accessibility in psoriasis patients and healthy controls has not previously been reported in the literature and related to the putative functional mechanisms of this GWAS risk loci in chronic inflammatory diseases.Similar examples have been found in other diseases, for example in T2D, where only one SNP from the credible set located at a *TCF7L2* intron overlapped accessible chromatin, with the risk allele showing greater abundance at open chromatin and increased enhancer activity (Gaulton *et al.* 2010; Stefan *et al.* 2014).

In terms of pathophysiology, *B3GNT2* is a major polylactosamine synthase involved in the post-translational N-glycosylation of carbohydrate chains, which are essential for cell-cell, receptor-ligand and carbohydrate-carbohydrate interactions. Interestingly, *B3GNT2* knock-out mice demonstrated more sensitive and strongly proliferating T cell and B cell responses to stimulation (Togayachi *et al.* 2010). In T cells, this effect was linked to a reduction of polylactosamine chains in co-stimulatory accessory molecules such as CD28, overall leading

to enhanced initiation of the immune response *in vitro*. The relevance of post-translational modifications and N-linked glycosylation in psoriasis has previously been highlighted through genome-wide pathway enrichment analysis (Aterido *et al.* 2016). In this study, Aterido and colleagues showed increased glycosylation levels in psoriasis patients $CD4^+$ and $CD8^+$ cells with the copy number for the protective allele of the lead gene in the post-translational modifications pathway. Under this scenario, the presence of the rs4672505 risk allele (A) at this putative stimulus-specific *B3GNT2* enhancer could increase risk of disease by reducing chromatin accessibility and *B3GNT2* expression in both, homozygous and heterozygous individuals. Reduced *B3GNT2* levels could lead to hyperactivation of $CD8^+$ cells through altered N-glycosylation levels in the context of psoriasis. Overall, establishing a more comprehensive model to further explain the functional role of rs4672505 in psoriasis susceptibility will require additional work.

4.4.7 Limitations in the approach

Although the work in this chapter has shed light on the chromatin landscape and gene expression in psoriasis in a cell type and tissue specific manner, a number of limitations are noted. Due to difficulties in optimising ATAC protocols to yield good quality data, mapping chromatin accessibility in lesional and uninvolved keratinocytes was not achieved. Other limitations in this study include its relatively small sample size, the batch effect introduced by the use of two different ATAC protocols, lack of genotyping data and skin biopsies only being available for three patients in the cohort. These limitations are intrinsic to time and resources constraints, and will be addressed as the study continues. Moreover, at the time of sample recruitment these patients did not present self-reported joint involvement that indicated transition into PsA;

however additional examination by a rheumatologist with follow-up visits would be required to ensure non-conversion into PsA phenotype by these patients.

Finally, further analysis could also be performed with the differentially expressed lncRNA integrating their experimentally validated targets to identify specific processed regulated by these RNA species in the context of disease.

4.5 Conclusions

In this chapter, use of cutting-edge methods for epigenetic profiling together with gene expression quantification has allowed the characterisation of the regulatory landscape in relevant cell types isolated from psoriasis patients and healthy individuals. Modest differences in chromatin accessibility and H3K27ac modifications between psoriasis and healthy controls have been identified in circulating immune cells. Conversely, a number of relevant biological processes dysregulated in the context of psoriasis have been shown at the transcriptional level both, in circulating cells and in psoriatic epidermis. Moreover, this chapter illustrates how GWAS signals may be interpreted through integration of multiple data types. Importantly, a $CD8^+$ specific genotypic effect in chromatin accessibility has been identified at the 2p15 psoriasis risk locus, with a putative effect of this region in regulating *B3GNT2* expression. Overall, the protocols established and data generated in this chapter provide a valuable resource that may be built upon in future work.

Chapter 5

Cross-tissue comparison of chromatin accessibility, gene expression signature and cytokine production in PsA

5.1 Introduction

5.1.1 The relevance of cell type and tissue specificity in the study of PsA

Consideration of cell-type specificity in the study of complex diseases is fundamental for the understanding of the disease pathophysiology. As previously reviewed (in Chapter 1), the dysregulated immune response in PsA is the result of the interaction between cellular components of the innate and adaptive immune response. Consequently, the molecular characterisation of the different immune cell types is pivotal not only for the understanding of the immune response but also to define disease state, comprehend the impact of genetic variants and identify drugs with optimal efficacy and specificity.

PsA is considered a systemic disease in which studies in PBMCs have demonstrated changes in cell type composition and cytokine production when compared to healthy individuals. For example, increased frequencies of circulating IL-17$^+$ and IL-22$^+$ CD4$^+$ T cells have been reported in peripheral blood from PsA patients compared to control individuals (Benham *et al.* 2013). Moreover, reduced percentages of pDCs and NK cells in peripheral blood have

also been observed in in PsA compared to controls (Jongbloed *et al.* 2006; Spadaro *et al.* 2004). In terms of cytokine production, stimulated PBMCs from PsA patients released greater levels of IL-17 and IL-22 than the healthy control counterparts (Benham *et al.* 2013).

PsA is characterised by involvement of the joints, where repeated local inflammatory response leads to joint destruction. Oligoarticular PsA (involving four or fewer joints) is commonly managed by joint aspiration followed by intra-articular steroid injection to relieve pain, facilitating the sample collection for research purposes (Kavanaugh and Ritchlin 2006). The importance of studying the synovium in PsA has been highlighted by differences in cell composition and cytokine production, amongst others, between peripheral blood and synovial fluid in PsA patients. For example, expansion of $mCD8^+$ and elevated proportion of T cells expressing the chemokine receptors $CCR6^+$ and $IL\text{-}23R^+$ was observed in synovial fluid when compared to peripheral blood in PsA paired samples (Ross *et al.* 2000; Benham *et al.* 2013). Moreover, the levels of TNF-α, IL-1, IL-6 and IL-18 in PsA synovial fluid were elevated when compared to those in peripheral blood, and they were within the same range of those reported for other rehumatic diseases such as RA (Kuijk *et al.* 2006).

5.1.2 Bulk transcriptomic studies in PsA and their limitations

Genome-wide transcriptomic studies in PsA have been mainly focused in characterising gene expression in bulk PBMCs, synovial fluid mononuclear cells (SFMCs) samples, and synovial membranes. Several studies have been conducted to better understand gene expression differences in blood between PsA and controls and also specific differences between peripheral blood and synovial fluid from the same PsA patients or differences with other arthritic diseases (Stoeckman *et al.* 2006; Batliwalla *et al.* 2005; Gu *et al.* 2002; Dolcino *et al.* 2015). A study conducted by Dolcino and colleagues revealed genes from

the Th-17 axis and type-I IFN signalling to be differentially expressed between PsA and healthy controls in synovial membranes and highlighted differences and commonalities with the changes in gene expression in peripheral blood when comparing patients vs healthy controls . Cytokine production assays have also been performed in sera and synovial fluid, revealing increased levels of TNF-α in both tissues when compared to controls and osteoarthritis patients, respectively (Ritchlin *et al.* 1998; Li *et al.* 2017).

Studies using mixed cell populations can be influenced by the relative proportion of the different cell populations within the sample (Whitney *et al.* 2003). Additionally, the importance of considering cell types to understand the impact of genetic variants in transcriptional regulation has been explored in a number of immune cells (Fairfax *et al.* 2012; Fairfax *et al.* 2014; Raj *et al.* 2014; Peters *et al.* 2016; Kasela *et al.* 2017). Expression analysis for a limited number of genes have been performed in specific cell type populations such as stimulated macrophages and Th-17 cells *in vitro* differentiated from peripheral blood and synovial fluid naïve $CD4^+$ cells isolated from PsA patients (Antoniv and Ivashkiv 2006; Leipe *et al.* 2010). *In vivo* investigation of the transcriptional profile of patients isolated discrete cell populations have yielded interesting findings in monocytes and Th-17 cells in AS, intestinal epithelial cells in CD, and fibroblast-like synoviocytes in RA amongst others (Al-Mossawi *et al.* 2017; Smith *et al.* 2008; Howell *et al.* 2018; Ai *et al.* 2016). Overall, achieving a detailed and precise understanding of complex diseases requires the study of sorted cell populations and when possible isolated from the affected tissue.

5.1.3 Transcriptomics and proteomics at the single cell resolution

In addition to the study of specific cell types, evidence for heterogeneity in the transcriptome within cells from the same population has accelerated the

development of strategies providing single cell resolution. The establishment of single cell RNA sequencing (scRNA-seq) and mass-cytometry techniques represent an unbiased way to characterise and identify cell subpopulations within the samples, avoiding the previous separation of particular cell types and thus providing a global overview of cell composition and interactions in the tissue of interest.

A wide range of approaches to study single-cell transcriptomics have been developed in the last few years, with Drop-seq, SmartSeq2 and 10X Chromium amongst the most widely used (Picelli *et al.* 2014; Ziegenhain *et al.* 2017). 10X Chromium technology is based on microfluidics where cells in suspension get directly encapsulated into nanoL droplets that incorporate the cell and transcript barcode identifiers (see Chapter 2). As a result, 10X Chromium technology does not require pre-sorting of single-cells into plates and enables higher throughput with less manipulation and variability than other scRNA-seq methods such as SmartSeq2 (Baran-Gale *et al.* 2017).

Likewise, mass cytometry represents the next generation of fluorescence based flow-cytometry analysis to interrogate expression of cell surface and intracellular molecules. Mass cytometry is a hybrid technique between mass spectrometry and flow cytometry, where the antibodies used to recognise the molecular intra- or extra-cellular markers have been labelled with stable isotopes instead of fluorophores (Bandura *et al.* 2009). The use of isotopes enables identifying up to 45 antibodies to profile cellular populations and assess molecular functions.

5.1.4 Multi-omics approach in the study of complex diseases

The interaction between genetics and the environment can shape the cellular epigenetic landscape and eventually result in the development of complex diseases. This dynamism of the epigenome involves cell and

context specific features which reinforce the importance of studying purified cell types instead of mixed populations. As previously mentioned, the epigenomic landscape has a pivotal role in understanding disease state and also contextualizing the role of putative genetic risk variants in the study of complex diseases. In this context, the implementation of multi-omics approaches in the study of complex diseases has enabled to better understand the relationship between the regulatory landscape, gene expression and protein translation in relevant cell populations and subpopulations of interest.

Incorporation of scRNA-seq and mass cytometry in addition to bulk RNA-seq and flow cytometry have led to a more detailed understanding of the immune system, accounting for the variability at the single-cell level in gene expression and protein translation (Jaitin *et al.* 2014; Villani *et al.* 2017; Bengsch *et al.* 2018). In complex diseases such as RA, scRNA-seq has revealed heterogeneity in the synovial fibroblast population and identified a potentially pathogenic cluster with a phenotype characterised by high proliferation and active secretion of pro-inflammatory cytokines (Mizoguchi *et al.* 2018). Similarly, mass cytometry analysis performed in RA identified an expanded $CD4^+$ T cell population promoting B cell response (Rao *et al.* 2017). Recently, Zhang and colleagues have published one of the most comprehensive available study integrating multi-omics (bulk RNA-seq, scRNA-seq and mass cytometry) in RA (Zhang *et al.* 2018). This study performed isolation of the main pathophysiological cell types infiltrated into RA synovial membranes, including T cells, B cells, monocytes, and fibroblasts, and identified 18 unique subpopulations by systematic correlation between transcriptional profiles and mass cytometry.

5.1.5 Integration of fine-mapping GWAS SNPs and functional data in PsA

Fine-mapping of GWAS signals is required in order to potentially reduce the putative number of causal SNPs whilts accounting for all the common genetic variants that may be involved in driving the association with disease. Fine-mapping using genotype level data incorporates a locus step-wise conditional analysis to identify independent secondary signals, prior to calculate PP and credible sets (Maller *et al.* 2012; Bunt *et al.* 2015). In some cases, the genomic region associated to disease is reduced and thus the number of putative causal SNPs that could be functionally relevant for the disease pathophysiology. Although fine-mapping decreases the number of putative causal SNPs from thousands to tens, additional integration of epigenetic and functional data as well as molecular assays are required to pinpoint the causal variant and the functional mechanisms driving disease association.

The PsA GWAS study conducted by Bowes and colleagues successfully performed fine-mapping for seventeen of the associated regions with p-value<0.01 (Bowes *et al.* 2015a). This study provided a summary table for the overlap of SNPs from the 90% credible set (list of SNPs explaining 90% of the PsA GWAS association) with genomic annotations and ENCODE features (cell lines and healthy donors primary cells), to further narrow down the set of putative causal SNPs for each associations as well as the most relevant cell type. Bowes and colleagues further investigated the PsA-specific association identified in this study at the 5q31 region and a pilot eQTL study confirmed the most significant correlation with *SLC22A5* expression for a SNP in high LD with the GWAS lead SNP (Bowes *et al.* 2015a).

Leveraging epigenetic data to further refine the candidate causal SNPs from fine-mapping studies would benefit from the generation of disease-specific chromatin regulatory maps in PsA affected tissue and, possibly, further

integration of scRNA-seq and mass cytometry from the same individuals. Altogether, disease-specific chromatin accessibility maps could represent an additional layer of information in the attempt to identify the causal variant driving GWAS associations with PsA and provide further insight into the disease pathophysiology.

5.2 Aims

This chapter aims to explore the differences in PsA between peripheral blood and synovial fluid by establishing and integrating cell type and tissue specific multi-omic datasets, including chromatin accessibility, transcriptomic profiling, and protein expression. The long term goal is to identify cell subsets contributing to pathophysiological relevant pathways in PsA and chronic inflammation, and facilitate the functional interpretation of PsA GWAS loci.

The specific aims for this chapter are:

1. To identify differences in the chromatin accessibility landscape between synovial fluid and peripheral blood in $CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells isolated from PsA patients.

2. To identify differentially expressed genes between synovial fluid and peripheral blood for relevant immune genes in $CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ from the same PsA patients.

3. To integrate differences in chromatin accessibility and changes in genes expression.

4. To further explore transcriptional differences at the single-cell level in cell types of interest and perform a basic integration with ATAC and mass cytometry data for cytokine production.

5. To conduct fine-mapping for a number of PsA GWAS loci using genotype data and integrate with cell and tissue-specific PsA chromatin accessibility maps, publicly available epigenetic and functional data, to further narrow down the putative causal SNPs driving such associations.

5.3 Results

5.3.1 PsA patients cohort description and datasets

In this study peripheral blood and synovial fluid were collected from a cohort of six PsA patients, with equal numbers of males and females (Table 5.1). All the patients presented oligoarticular joint involvement and skin lesions, having previously been diagnosed with psoriasis. The cohort presented a mean of 1.5 tender or swollen affected joints (TJC68 and SJC66), which is characteristic of the oligoarticular form of disease, involving four or fewer joints. Regarding global assessment, the mean scores for the patient and physician evaluation were 3.2 and 3, respectively, on a scale of 0 to 5. These four measurements including joint and global assessment define the PsARC assessment, used by clinicians as the main criteria to monitor response to treatment recommended by the National Institute for Health and Care Excellence (NICE) (detailed in Chapter 1). The mean age of the cohort at the time of diagnosis was 44.3 years old and the mean disease duration 8.8 years. The average C-reactive protein (CRP) level, a marker of inflammation, was 17.45 mg/L and was higher in PsA1719 and PsA1728 compared to the other patients. At the time of sample recruitment no patients were on active immunosupressive therapy.

Table 5.1: Description and metadata of the PsA patients cohort. PsARC disease activity score is composed of tender joint count 66 (TJC66) and swollen joint count 66 (SJC66), joint pain (4 point score) and self-patient and physician global assessment (5 point score). Joint pain and global assessment use a likert scale based on questionnaire answers that measure the level of agreement with each of statements included. C-reactive protein (CRP).

Sample ID	Sex	Age at diagnosis	Disease duration (months)	Type	TJC68/SJC66	Physician assessment	Patient assessment	CRP (mg/L)
PsA1718	Female	17	180	Oligo	2/2	3	3	6
PsA1719	Male	33	24	Oligo	1/1	3	4	36.6
PsA1607	Male	42	108	Oligo	1/1	4	3	8
PsA1728	Female	72	48	Oligo	2/2	3	4	43.2
PsA1801	Female	53	168	Oligo	2/2	3	3	9.9
PsA1505	Male	35	108	Oligo	1/1	2	2	1
Mean (±SD)	–	44.3 (±15.74)	104 (±59.6)	–	1.5(±0.5) /1.5(±0.5)	3 (±0.6)	3.2 (±0.8)	17.4 (±17.8)

For each of the patients, paired peripheral blood and synovial fluid data was generated from bulk or isolated cell types of interest (detailed in Table 5.2 and Chapter 2). Due to project constraints, Fast-ATAC, PCR gene expression array, scRNA-seq and mass cytometry were not generated for all six individuals in the cohort.

Sample ID	ATAC	RNA PCR array	scRNA-seq	Mass cytometry
PsA1718	Yes	No	No	Yes
PsA1719	Yes	Yes	No	Yes
PsA1607	Yes	Yes	Yes	Yes
PsA1728	No	Yes	No	Yes
PsA1801	No	No	Yes	Yes
PsA1605	No	No	Yes	Yes

Table 5.2: Datasets generated for each sample in the PsA cohort. Four types of data were generated in paired synovial fluid and peripheral blood from the same individual. Fast-ATAC data was generated for $CD14^+$, $mCD4^+$, $mCD8^+$ and NK cells. RNA expression by PCR array was performed only for $CD14^+$, $mCD4^+$ and $mCD8^+$ cells. scRNA-seq data was generated using 10X Chromium technology in bulk SFMCs and PBMCs.

5.3.2 Immune cellular composition of blood and synovial fluid in the PsA cohort

The immune cellular composition of three PsA samples (PsA1718, PsA1719 and PsA1607) was characterised in synovial fluid and peripheral blood using the mass cytometry antibody panel (detailed in Chapter 2). For both tissues, $mCD4^+$ (32.1-55.6%) constituted the most abundant cell type followed by $mCD8^+$ (16.9-24.9%) and $CD14^+$ "classical" monocytes (6.9-21.7%) (Figure 5.1). Consistent with previous studies, a trend of increased percentage of $mCD8^+$ pDCs and cDCs was observed in synovial fluid compared to peripheral blood (Ross *et al.* 2000; Jongbloed *et al.* 2006). This data also showed reduced percentage of synovial fluid NK cells compared to peripheral blood, in line with previous studies suggesting the role of impaired non-MHC-restricted cytotoxicity in PsA (Spadaro *et al.*

2004). Similarly, a tendency towards reduced proportions of B cells in synovial fluid compared to peripheral blood supports a lower contribution of the humoral immune response at the site of inflammation in PsA pathophysiology. The observed differences in cell composition between synovial fluid and peripheral blood were not statistically significant for any of the 12 analysed populations likely due to the small samples size (n=3) available for the analysis.

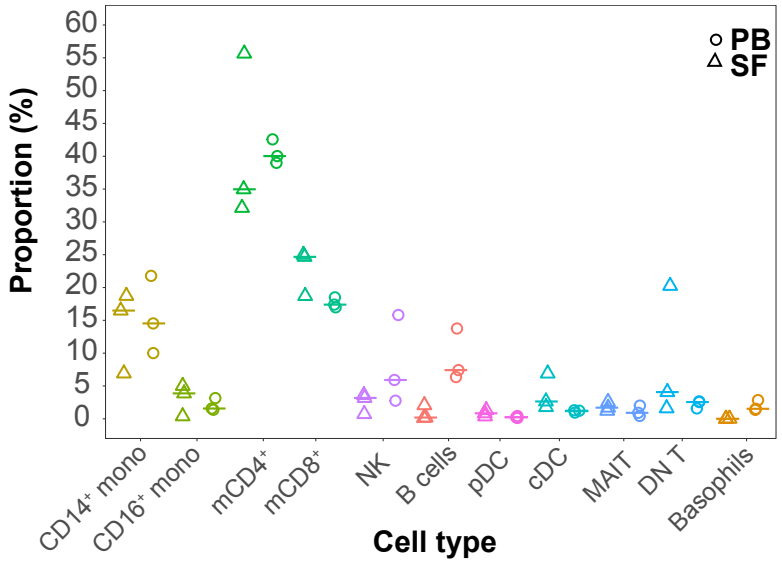


Figure 5.1: Comparative percentages of peripheral blood and synovial fluid immune cellular composition from the PsA cohort. Percentages of each of the twelve cell types identified by mass cytometry are shown by individual and tissue for PsA1718, PsA1719 and PsA1607. Horizontal line represents the median percentage for a particular cell type in the appropriate tissue (synovial fluid or peripheral blood). Each of the cell types is displayed in a different colour. cDC=central DC; MAIT=mucosal-associated invariant T; DN=double negative; SF=synovial fluid; PB=peripheral blood.)

5.3.3 Differential chromatin accessibility analysis in immune cells reveals differences between synovial fluid and peripheral blood

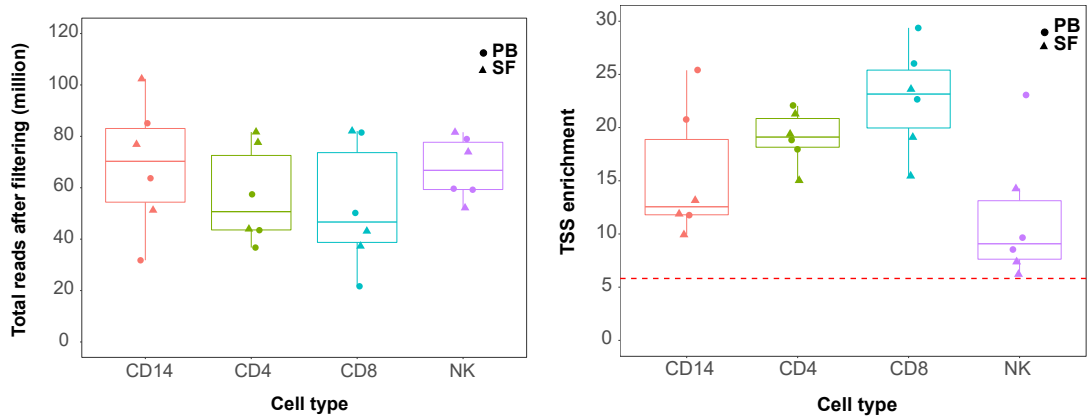
Data processing and quality control

Chromatin accessibility was studied in the four most abundant cell types in PsA synovium fluid, $CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells. The

24 ATAC samples (generated using Fast-ATAC protocol as detailed in Chapter 2) from four cell types were sequenced and processes using the in-house pipeline as previously detailed (Chapters 2 and 3). After filtering of duplicated and mitochondrial reads (Figure B.13A), the median total number of reads was between 46.6 and 70.2 millions (Figure 5.2A), depending on cell type (Figure 5.2A).

TSS enrichment analysis showed differences in the levels of background noise across cell types and highlighted the variability of ATAC performance (Figure 5.2B), where $mCD4^+$ and $mCD8^+$ showed the best signal-to-noise ratios. In contrast, NK was the cell type with the lowest TSS enrichment values with fold enrichments for PsA1719 and PsA1607 close to the 6 (ENCODE threhold cut-off). Given the limited cohort size, these samples were not excluded, but it is worth noting that they could be contributing to the noise and thus reduce the power of the differential analysis. The number of accessible regions (peaks after filtering with p-values based on IDR analysis) ranged between $24x10^3$ and $97x10^3$ (Figure B.13B), showing dependence with sample quality and cell type but no obvious outliers.

A combined consensus list of called ATAC peaks across all the samples and cell types (named here as CP_all) was built (as previously explained in Chapters 2 and 3) and PCA was conducted on the normalised counts mapping to each of those peaks.



(A) (B)

Figure 5.2: Quality control assessment of ATAC data generated in four immune cell types isolated from peripheral blood and synovial fluid of PsA patients samples. For each of the cell types and samples, boxplots representing (A) million of reads after filtering and (B) values for fold-enrichment of ATAC fragments across the Ensembl annotated TSS. For each point, colour codes for cell type and shape for tissue (SF=synovial fluid; PB=peripheral blood).

In this analysis, 65% of the variability (PC1) in the chromatin landscape correlated with cell type, leading to sample separation into 4 clusters (Figure 5.3), demonstrating the ability to capture cell-type specific global differences in the regulatory landscape. The myeloid ($CD14^+$ monocytes) and lymphoid ($mCD4^+$ and $mCD8^+$) clusters were the most distinct ones based on the ATAC profile, consistent with biological expectations. Modest separation between synovial fluid and peripheral blood samples was also observed in $mCD8^+$ and NK cells clusters in the combined analysis (Figure 5.3).

Characterisation of the differential accessible regions between synovial fluid and peripheral blood

In order to perform differential chromatin accessibility analysis between synovial fluid and peripheral blood, a consensus list of ATAC peaks (accessible chromatin regions) was built for each cell types (named here as CP_CD14, CP_mCD4, CP_mCD8, and CP_NK) (detailed in Chapter 2). Differential chromatin accessibility analysis between synovial fluid and peripheral blood

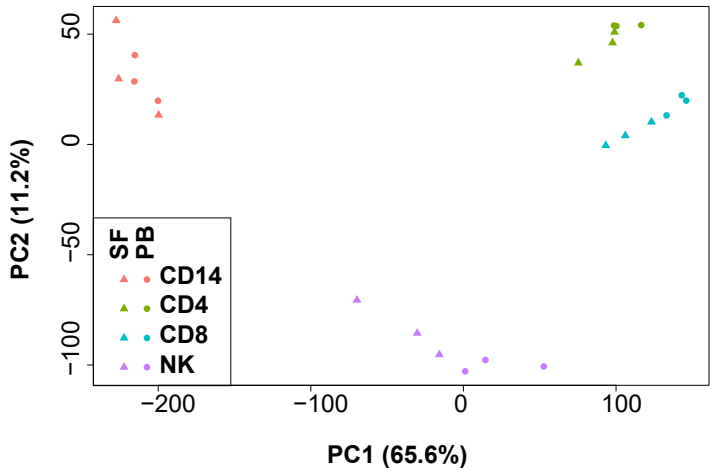


Figure 5.3: PCA based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and SF. The analysis was performed using the normalised counts mapping to each of the peaks included in the consensus list of called peaks (CP_all) across all four cell types (CD14$^+$ monocytes, mCD4$^+$, mCD8$^+$ and NK cells) and two tissues of interest (total of 24 samples). The first two PCs (x-axis and y-axis, respectively) for the ATAC peaks included in the ML_all are plotted. Each point represents a sample, where colour indicates cell type and shape tissue (SF=synovial fluid; PB=peripheral blood). The proportion of variation explained by each principal component is indicated.

was conducted on normalised counts retrieved for each of the cell type master lists using DESeq2 with a paired design (Table 5.3). Applying an 80% cut-off for background noise (Chapter 3), CD14$^+$ monocytes and NK cells showed the greatest total number (5,285 and 2,314, respectively) and proportion of DARs (9.1 and 8.9%, respectively) with FDR<0.01 and fold change>1.5.

Considering the direction of the differential chromatin accessibility, DARs were divided in DARs more open in synovial fluid compared to peripheral blood (named here synovial fluid open DARs) and DARs less open in synovial fluid compared to peripheral blood (named here peripheral blood open DARs). In CD14$^+$ monocytes the number of synovial fluid open DARs were notably larger than the number of peripheral blood open DARs (3,779 and 1,506 DARs, respectively) (Table 5.3). In contrast, a similar proportion of synovial fluid and peripheral blood open DARs were found for the other three cell types.

Cell type	Total DARs	Proportion DARs (%)	Synovial fluid open DARs	Peripheral blood open DARs
$CD14^+$	5,285	9.1	3,779	1,506
$CD4^+$	1,329	4.3	621	708
$CD8^+$	1,570	4.5	807	763
NK	2,314	8.9	1,223	1,091

Table 5.3: Summary results of the differential chromatin accessibility analysis between synovial fluid and peripheral blood in PsA samples. For each of the cell types the total number of DARs and the proportion represented by DARs over all the regions included in the differential analysis are reported. The total number of DARs are further divided in those more accessible in synovial fluid (synovial fluid open DARs) when compared to peripheral blood and those less accessible in synovial fluid when compared to peripheral blood (peripheral blood open DARs) using FDR<0.01 and fold change¿1.5.

Permutation analysis using the unique ten possible permutations (given the small number of samples included in the analysis) did not show by chance a greater number of DARs than the ones identified for the true groups, reinforcing the robustness of the differential analysis results in all four cell types (Figure B.14).

Genomic annotation of the DARs identified in each the cell types revealed that 80% or more of all regions with differential accessibility were located at intronic and intergenic regions (Figure 5.4A).

Universal promoter regions were the third most represented genomic feature, accounting for the annotation of approximately 5 to 15% of the DARs in each cell type (Figure 5.4A). For all four cell types, between 44.96 and 72.11% of the DARs were annotated as enhancers based on cell-type specific chromatin segmentations maps, the category accounting for the largest proportion of DARs and the most significantly enriched based on FDR (Figures B.15 and 5.4). The over-representation of enhancers was consistent with large percentage of introns and intergenic regions found for the genomic features annotation, as those are the preferred location for enhancer elements. Modest percentages of DARs were also annotated as heterochromatin and repetitive regions, which partly

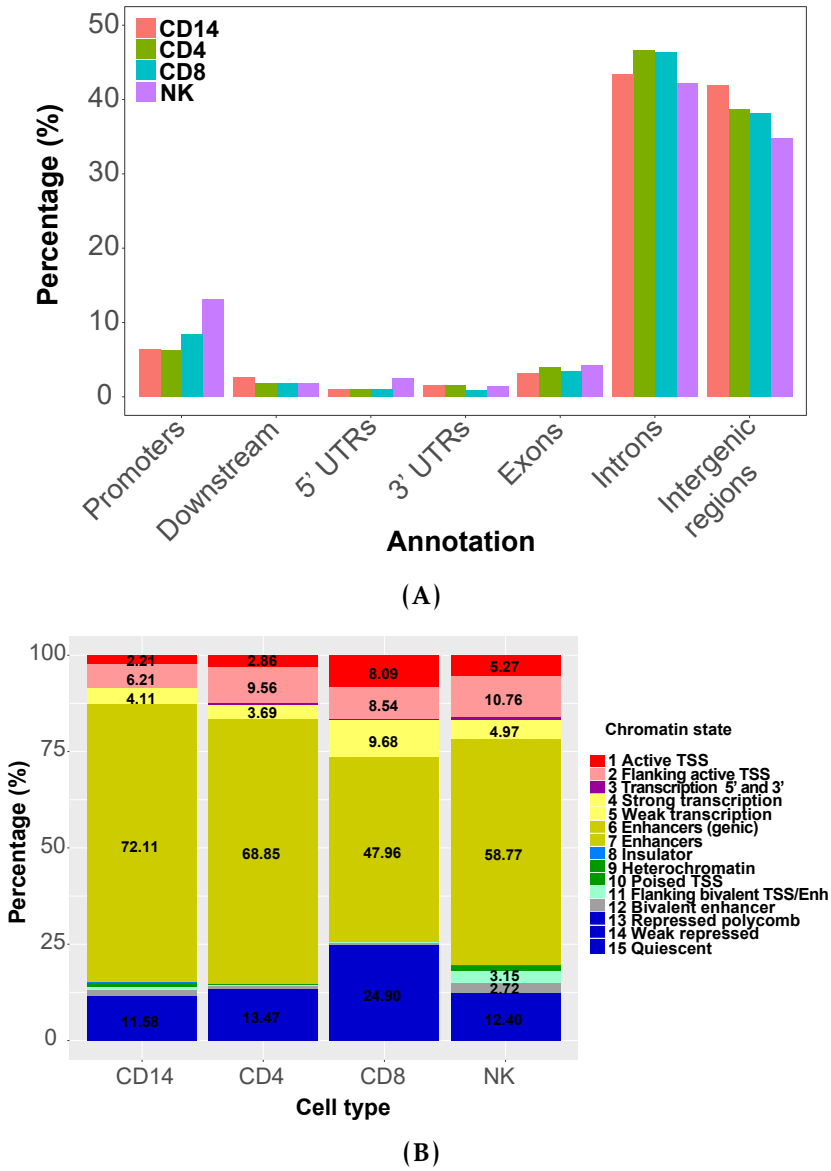


Figure 5.4: Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states. (A) Barplot illustrating the percentage of nucleotides within DARs for each cell type that are annotated as promoters, downstream (regions at ≤1,000bp to a promoter), exons, introns, 5' or 3'UTR and intergenic regions. (B) Stacked barplot representing the percentage of DARs annotated for each of the 15 chromatin states defined in each of the four relevant cell types by Roadmap Epigenomics Project chromatin segmentation maps ($CD14^+$ peripheral blood isolated monocytes, $mCD4^+$, $mCD8^+$ and NK cells).

represent intrinsic technical noise but could also reflect specific synovial fluid features that have not been captured by the chromatin segmentation maps generated using peripheral blood immune cells. The functional relevance of the differential chromatin accessibility in terms of regulation of gene expression was further investigated by integration of the eRNA data from the FANTOM5 project. Statistically significant enrichment for robust and permissive enhancers was found for the DARs in all four cell types (Figure 5.5). Moreover, DARs from all four cell types also showed amongt the top most significant enriched enhancer sets the corresponding to the relevant cell type. The proportion of DARs overlapping the appropriate cell type set of expressed eRNAs ranged between 198 (122 synovial fluid open and 76 peripheral blood open) in $mCD4^+$ T cells and 924 DARs (690 synovial fluid open and 234 open in peripheral blood) in $CD14^+$ monocytes.

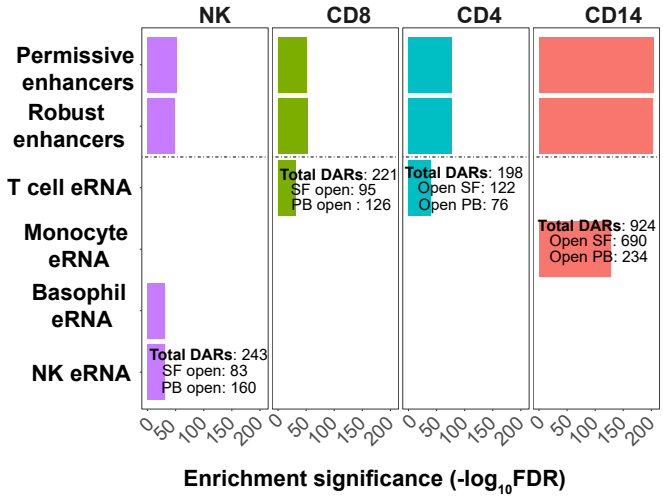


Figure 5.5: Enrichment of PsA DARs for the FANTOM5 eRNA dataset. Robust enhancers have been defined as those detected at the genome-wide significant level in at least one primary cell type or tissue. Permissive enhancers are all detected eRNAs but not passing genome-wide filtering criteria (Andersson *et al.* 2014). Robust enhancers represent a subset of the permissive enhancers. Total number of DARs overlapping a robust eRNA detected in the corresponding cell type are indicated. Significant enrichment is considered for FDR<0.01.

The differential analysis demonstrated that a number of DARs were overlapping a gene body (Table 5.4). Interestingly, the majority were located

within introns instead of untranslated regions (UTRs) and were annotated as weak or strong enhancers according to the cell-type specific chromatin segmentation map previously illustrated.

Cell type	DARs in gene body	Gene with more than one DAR	Enhancers	Introns
$CD14^+$	2,357	744	1,775	1,920
$CD4^+$	700	99	504	577
$CD8^+$	831	118	503	666
NK	1,246	235	782	937

Table 5.4: Characterisation of the DARs located within genes in each of the four cell types from PsA samples. The number of DARs overlapping a gene body for each of the cell types are indicated together with those genes harbouring more than one DARs. Further details about those regions includes specification of the number located at introns and those annotated as enhancers according to the Roadmap Epigenomics Project chromatin segmentation maps of each appropriate cell type.

For example, differential chromatin accessibility analysis in NK cells identified a DAR located in an intron of the *VAV3* gene that was also more accessible in peripheral blood compared to synovial fluid, and also significantly expressed as an eRNA (Figure 5.6A). By contrast, an example of a gene harbouring two $CD14^+$ monocytes DARs was *IL7R* 5.6B). The DARs were located at the 5' and 3' UTRs of this gene, respectively, and both were more accessible in synovial fluid compared to peripheral blood (Figure 5.6B).

Additionally, DARs in all four cell types were shown to overlap or be proximal (≤5Kb) with genes nearby psoriasis and PsA GWAS loci. $CD14^+$ monocytes presented the largest number of overlaps (13), followed by $mCD8^+$ (9), NK^+ (8) and $mCD4^+$ (4). For example, DARs proximal to *ELMO1* and *RUNX3* were found for all the cell types. Moreover, LD-based enrichment analysis was conducted between psoriasis and PsA GWAS catalogue LD blocks and the DARs identified in each of the cell types. This analysis compared the observed to the expected overlap between a null distribution of LD blocks from all common SNPs, respecting allele frequencies and proximity to genes of the GWAS lead

SNP of each block (see Chapter 2). Significant enrichment of GWAS LD blocks for DARs between synovial fluid and peripheral blood in $CD14^+$ monocytes (empirical p-value=0.043) was observed using a total of 20,000 permutations.

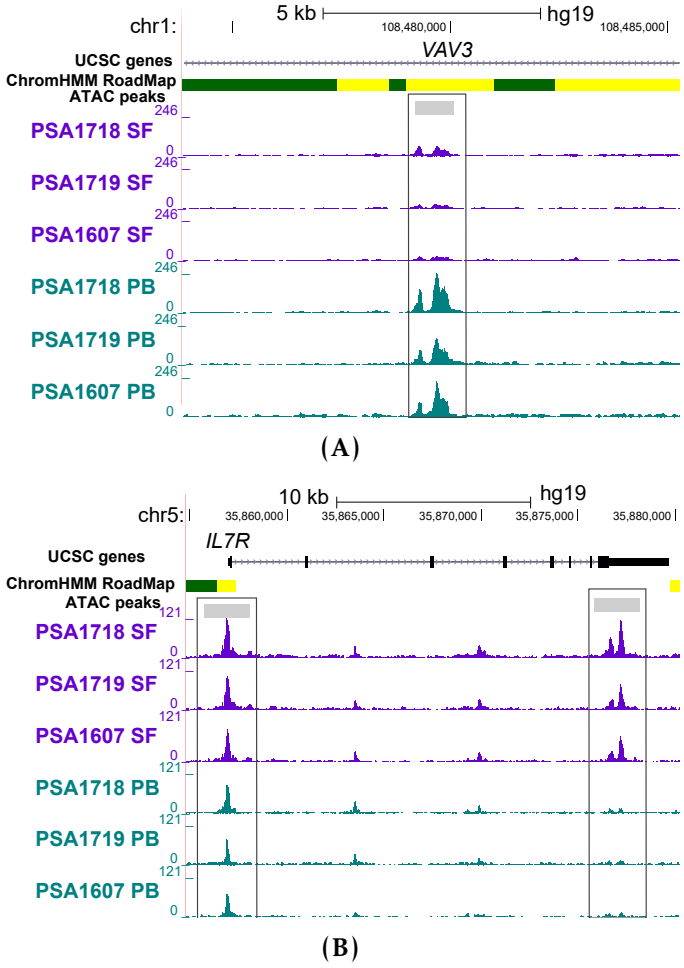


Figure 5.6: Differentially accessible regions located within gene bodies in $CD14^+$ monocytes and NK cells from PsA patients. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) in (A) DAR located at an intron of *VAV3* gene (x-axis) in NK (less accessible in synovial fluid compared to peripheral blood) and (B) two DARs mapping to the 5' and 3'UTR of the *IL7R*, respectively, in $CD14^+$ monocytes (both more accessible in synovial fluid compared to peripheral blood). Tracks are colour-coded by tissue (SF=purple; PB=turquoise). The Roadmap Epigenomics Project chromatin segmentation track for the appropriate cell type are also shown. All DARs were significant based on FDR<0.01 and fold change>1.5.SF=synovial fluid; PB=peripheral blood.

5.3.4 Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility

To explore the potential function of specific accessibility changes of synovial fluid (synovial fluid open DARs) and peripheral blood (peripheral blood open DARs) specific accessibility changes in each cell type, gene annotation of each set of DARs was performed based on location (by physical proximity) as detailed in Chapter 2, followed by pathway enrichment analysis. Despite common pathways enriched in both, a number of functionally relevant pathways appeared significantly enriched (FDR<0.01 or 0.05) only for synovial fluid open DARs and not for peripheral blood open DARs and vice versa (Figure 5.7). In $CD14^+$ monocytes, synovial fluid open DARs showed enrichment for pathways involved in hemostasis, integrin interactions, calcium signalling, and importantly the NF-κB pathway (Figure 5.7A in purple). Synovial fluid open DARs in $CD14^+$ monocytes were also enriched for cytokine related pathways, including IL-2 and IL-3, IL-5 and granulocyte-macrophage colony stimulating factor (GM-CSF) signalling. By contrast, peripheral blood open DARs in $CD14^+$ monocytes were not enriched for any of these pathways, showing enrichment for DAP12 interactions and negative regulation of the PI3K/AKT network (Figure 5.7A in turquoise).

$mCD4^+$ synovial fluid open DARs, in contrast to peripheral blood, showed enrichment for TCR signalling as well as chemokine signalling,including DARs in proximity to IFN-γ or *CXCL13* and *CXCR6*, respectively (Figure 5.7B in purple). Peripheral blood open DARs in this cell type were enriched for signalling by receptor tyrosine kinases and focal adhesion members, also involved in the T cell activation (Dustin 2001). Enriched pathways for synovial fluid or peripheral blood open DARs in $mCD8^+$ were only significant when using an FDR<0.05 threshold. The G protein coupled receptor (GPCR) signalling, with varied roles in regulation of inflammation and mediation of the chemotactic recruitment

of T cells to the inflamed tissue, was enriched for mCD8$^+$ peripheral blood open DARs, as well as the Wnt signaling pathway (Figure 5.7C in turquoise). mCD8$^+$ synovial fluid open DARs showed enrichment for chemokine signalling and regulation of stem cell pluripotency (Figure 5.7C in purple).

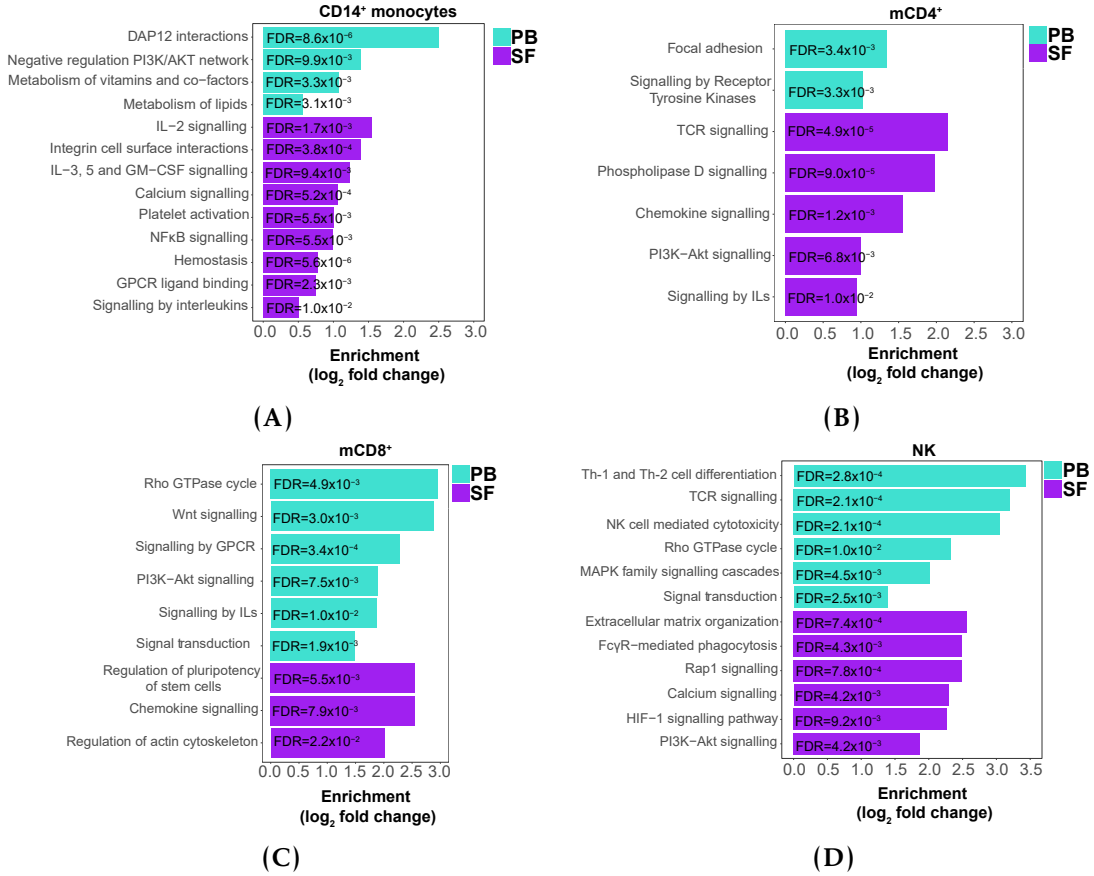


Figure 5.7: Functional pathways significantly enriched and unique for either synovial fluid open or peripheral blood open DARs in CD14$^+$ monocytes, CD4m$^+$,CD8m$^+$ and NK cells. Enrichment analysis was performed separately for genes annotating synovial fluid open DARs and peripheral blood open DARs. Log_2 fold change of unique functionally relevant significant pathways (FDR <0.01 or 0.05) enriched for synovial fluid open DARs or peripheral blood open DARs in (A) CD14$^+$ monocytes, (B) mCD4$^+$, (C) mCD8$^+$ and (D) NK cells are shown. SF=synovial fluid; PB=peripheral blood.

NK synovial fluid open DARs showed enrichment for extracellular matrix organisation and RAP-1 signalling as well as for Fcγ receptor (FCγR)-mediated phagocytosis (Figure 5.7D in purple). Members of the HIF-1 pathway involved in oxygen homeostasis were also enriched in NK synovial fluid open DARs, in line with the hypoxic environment found in joint inflammation. Interestingly,

enrichment of open peripheral blood DARs in the proximity of genes involved in NK-mediated toxicity was found (Figure 5.7D in turquoise).

5.3.5 Differential gene expression analysis in paired circulating and synovial immune cells

Immune-relevant gene expression by qPCR

In order to contextualise the ATAC data in the frame of gene expression, qPCR gene expression analysis for 370 key genes in the inflammatory and autoimmune response was conducted in $CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ cells isolated from synovial fluid and peripheral blood of three PsA patients (Table 5.2).

The PCR array represented a cost-effective approach to study gene expression between peripheral blood and synovial fluid focusing on a relevant subset of immune genes, given the inflammatory component in PsA. For each cell types, fold change in expression was calculated pair-wise for synovial fluid respect to peripheral blood within each sample for each individual genes (detailed in Chapter 2). Fold change in gene expression showed differences in magnitude and reproducibility across the three cell types (Figure 5.8). Genes showing marked up-regulation (fold change>1.5) in the three cell types when comparing synovial fluid vs peripheral blood included, for example *FN1*, *SPP1* or *CCL2*, amongst others (Figure 5.8 orange box). On the other hand, genes presenting reduced expression in synovial fluid (fold change<1.5) in at least one of the three cell types were also observed, including for example *FOS*, *IL16*, *PPBP* and *TPST1* (Figure 5.8 purple box). Genes demonstrating consistent changes in the same direction in the three $CD14^+$ monocyte samples but not in T cells were also identified (Figure 5.8 dark yellow box). These included up-regulation of *CXCL10* and *IL7R* in the synovial fluid of $CD14^+$ monocytes compared to peripheral blood that were inconsistent in $mCD4^+$ and $mCD8^+$ T cells. Moreover,

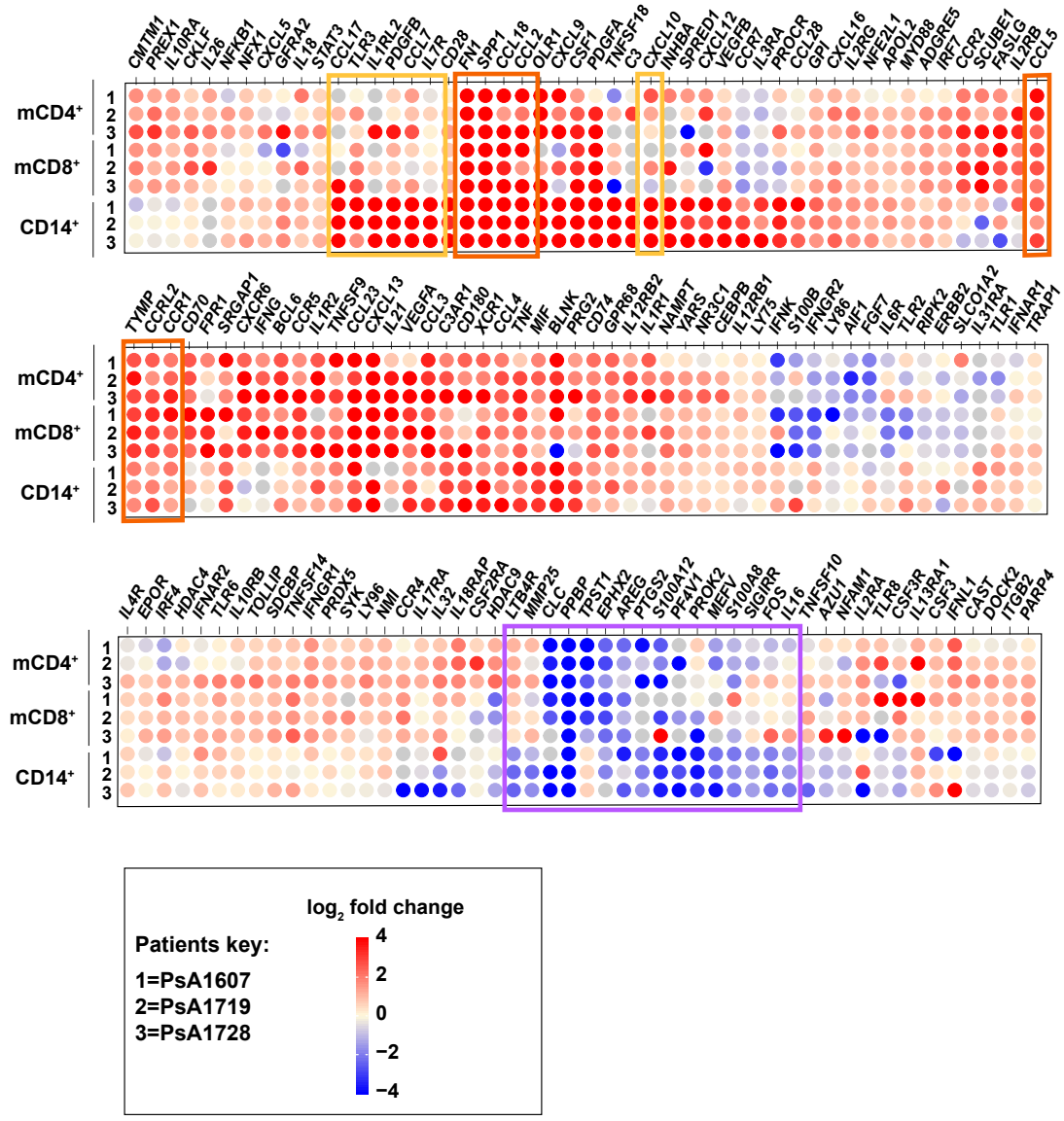


Figure 5.8: Heatmap illustrating fold changes in gene expression between synovial fluid and peripheral blood across CD14$^+$ monocytes and mCD4$^+$ and mCD8$^+$ T cells. Amongst the 370 genes measured by qPCR, the fold change in gene expression between the synovial fluid and peripheral blood for each pair of samples is represented only for those genes which were consistently modulated across the 3 PsA samples (p-value<0.05) in at least one of the cell types. Each column represents a genes and each row a pair of synovial fluid-peripheral blood PsA samples. The $\log_2$ fold change in gene expression between synovial fluid and peripheral blood is colour-coded. For visualisation purposes the heatmap is shown in 3 distinct blocks. PsA identity is included.

differences in fold change were also observed for genes modulated in the same direction across the three cell types, for instance *VEGFB* and *CCR1* (Figure 5.8).

Changes in expression did not reach significance after multiple testing correction (FDR<0.05 or 0.1) for the majority of the genes assayed by the qPCR array, likely due to the small sample size. Following this, filtering using a p-value<0.05 (one sample t-test) as the statistical threshold for significance and a mean fold change>1.5, $CD14^+$ monocytes and $mCD8^+$ showed greater number of significantly modulated genes between synovial fluid and peripheral blood (72 and 77, respectively) compared to $mCD4^+$ cells (46 genes) (Figure 5.9A, B and C). In all three cell types, most of the dysregulated immune genes were up-regulated in the synovial fluid compared to peripheral blood (Figure 5.9A, B and C). For example, 56 out of the 70 significantly modulated genes in $CD14^+$ monocytes showed mean up-regulation of fold change>1.5 vs the 14 genes with mean down-regulation fold change<1.5 (Figure 5.9A).

Gene expression in peripheral blood of three PsA patients was also compared to peripheral blood of the three healthy controls for $CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ cells. P-values and fold changes were calculated between the peripheral blood expression in each of the three PsA patients and the mean expression of the three healthy controls for each of the genes in the array (as detailed in Chapter 2). Integration of the differentially expressed (and therefore modulated genes) between synovial fluid and peripheral blood in PsA and those changing expression between peripheral blood of PsA and healthy controls allowed the identification of three group of genes (Figure 5.10).

Firstly, the genes solely modulated in peripheral blood between controls and PsA were designated as systemic genes (Figure 5.10 green dots). These genes were not significantly expressed between synovial fluid and peripheral blood in PsA patients, and could then be considered as the circulating disease "footprint". For example *CCL24* and *CCL27* in $CD14^+$ monocytes or *CCR10*

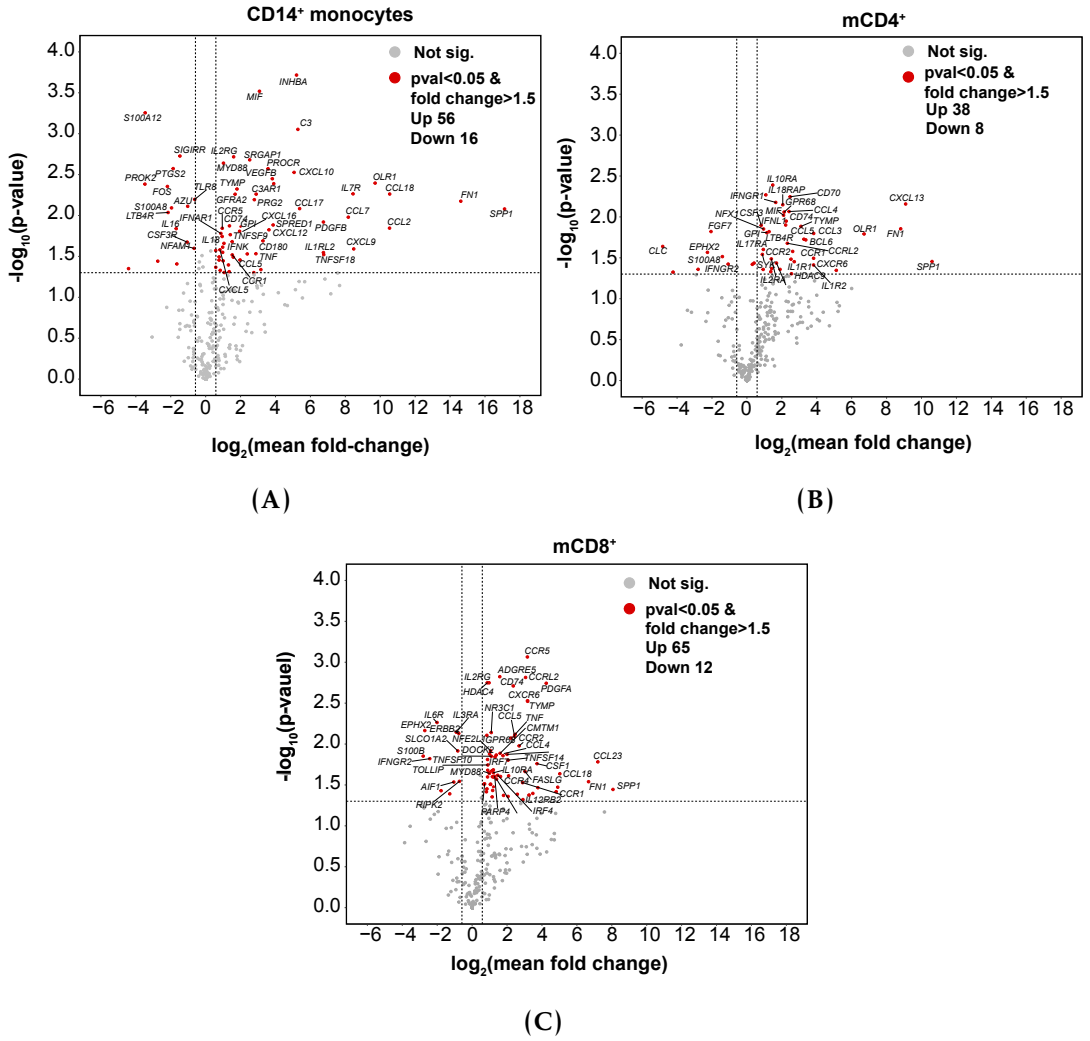


Figure 5.9: Gene expression changes in immune-relevant genes between synovial fluid and peripheral blood in CD14$^+$ monocytes, mCD4$^+$ and mCD4$^+$ cells. Volcano plots showing differences in gene expression measured by qPCR array between synovial fluid and peripheral blood for (A) CD14$^+$ monocytes, (B) mCD4$^+$ and (C) mCD8$^+$ cells. The significance ($\log_{10}$p-value) of the modulation in gene expression between the two tissues (y-axis) is plotted against the $\log_2$ of the mean fold change across the three PsA patients. Positive fold change indicates higher expression in synovial fluid. Genes showing p-value<0.05 and mean fold change>1.5 in red, with the most significant genes labelled. orange=up-regulated in the three cell types; purple=down-regulated in at least one cell type; yellow=consistent changes in CD14$^+$ monocytes only.

in mCD8$^+$ cells. A second group of genes were designated as tissue-specific, since they were significantly modulated between synovial fluid and peripheral blood in PsA patients but did not show significant changes between controls and PsA at the circulating level (Figure 5.10 red dots). This group included *SPP1* and *FN1* in the three cell types. Both genes demonstrated the largest fold changes in expression between synovial fluid and peripheral blood and presented no statistically significant differential expression in peripheral blood when comparing PsA and healthy controls. The third category comprised genes significantly modulated for each cell type between controls and PsA patients in peripheral blood, as well as between synovial fluid and peripheral blood in PsA patients. These genes were defined as putative disease-specific genes (Figure 5.10 blue dots), for example *GPR68* in mCD4$^+$ cells (Figure 5.10 B). Nevertheless, this analysis is clearly limited by the lack of comparison to synovial fluid from a non-related disease as well as by the small sample size and the quantification of a limited number of genes.

Correlation between gene expression and chromatin accessibility

Overlap between differentially modulated genes (in synovial fluid vs peripheral blood) and proximal DARs was observed in the three cell types (Table 5.5); however this overlap was only found to be significant in CD14$^+$ monocytes (Fisher exact test p-value=0.028) and not in mCD4$^+$ and mCD8$^+$ T cells (Fisher exact test p-value=0.466 and 0.173, respectively).

In CD14$^+$ monocytes, 13 out of the 56 significantly up-regulated genes in synovial fluid were also found to be proximal to a synovial fluid open DAR. One example being *IL7R*, where its increased expression in synovial fluid correlated with increased chromatin accessibility at the 5' and 3' UTR of this gene (Figure 5.6B). Another relevant example was the *FN1* gene, for which up-regulated expression in PsA synovial biopsies have been previously reported accompained

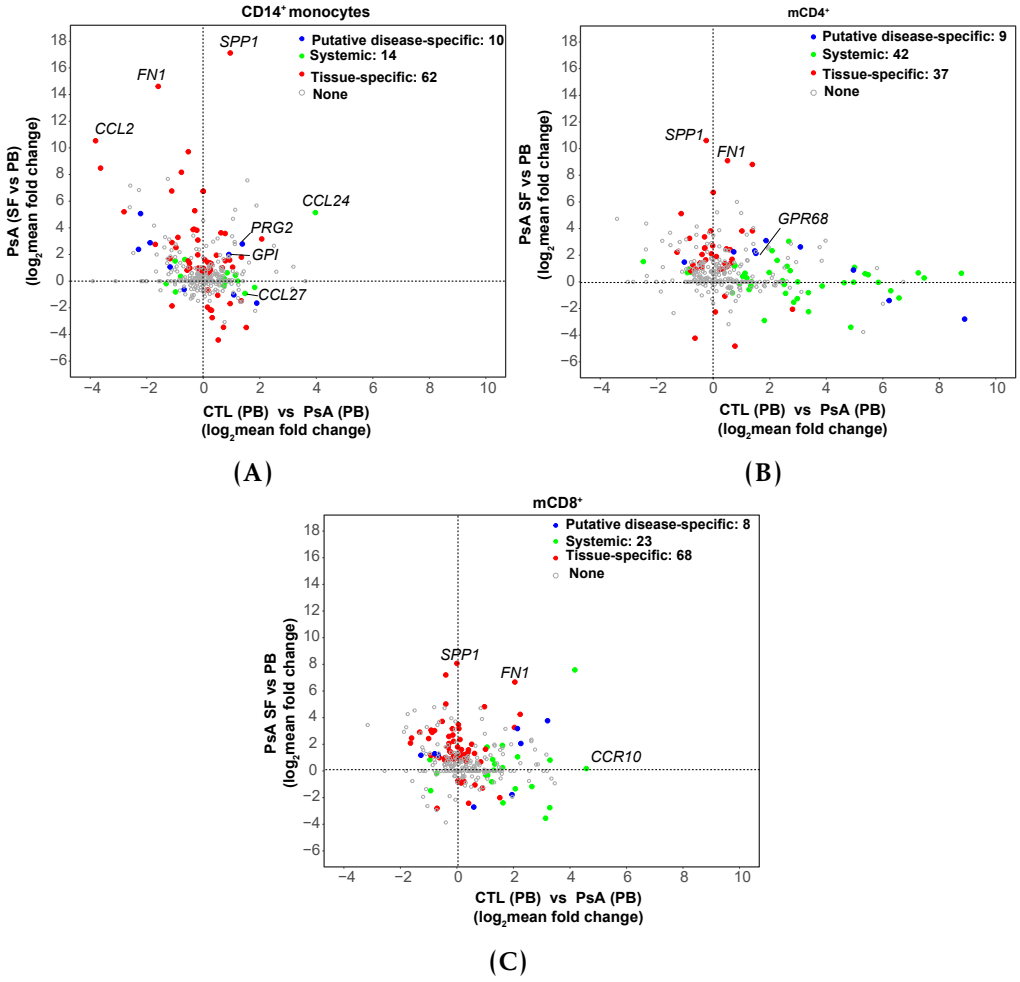


Figure 5.10: Comparison of immune-relevant gene expression modulation across PsA tissues (synovial fluid vs peripheral blood) and in PsA patients vs healthy controls. The qPCR $\log_2$ mean fold change for each of the genes in the PsA synovial fluid vs peripheral blood contrast are plotted against the $\log_2$ mean fold change for the same genes in the comparison peripheral blood PsA vs healthy control for (A) $CD14^+$ monocytes, (B) $mCD4^+$ and (C) $mCD8^+$ cells. The genes are colour-coded according to the three categories of genes made by comparison of changes in gene modulation between the two contrasts: systemic genes, tissue-specific and putative disease-specific. SF=synovial fluid; PB=peripheral blood. In grey genes which differential expression was not significant (p-value<0.05) for either of the two contrasts.

Cell type	Genes up-regulated and overlapping open chromatin in SF	Genes down-regulated and overlapping closed chromatin in SF
$CD14^+$ monocytes	13 (*BLNK*, *CCL2**, *CCR1**, *CD180*, *CXCL10*, *FN1*, *IL18*, *IL31RA**, *IL7R**, *NFKB1**, *PRG2*, *SRGAP1*, *STAT3*)	2 (*FOS*, *PROK2**)
$mCD4^+$	3 (*CXCL13*, *CXCR6**, *IL2RA*)	0
$mCD8^+$	6 (*CCL3*, *CCR2*, *CCR5* ,*IRF4* *TNFSF10*, *YARS*)	1 (*EPHX2*)

Table 5.5: Immune genes with significant modulated expression in synovial fluid and proximal to a DAR in ATAC. An overlap is defined by significant change in expression (p-value<0.05) of a particular gene where there is also a proximal DAR showing changes in chromatin accessibility in the same direction. * indicates that the proximal DAR overlapping an eRNA identified by FANTOM5 project in that particular cell type (see subsection Characterisation of the differential accessible chromatin regions). SF=synovial fluid.

by no changes in the counterpart comparison in blood (Dolcino *et al.* 2015). In this cohort, *FN1* expression was up-regulated in synovial fluid for all three cell types, with the highest fold change found in $CD14^+$ monocytes (Figure 5.9A), concomitantly with more accessible chromatin at the promoter and downstream the 3' UTR of the gene (Figure 5.11).

Pathway enrichment analysis highlights the role of synovial $CD14^+$ monocytes in cytokine and chemokine signalling

To identify relevant pathways amongst the modulated genes between synovial fluid and peripheral blood, enrichment analysis was performed for each individual cell type. Up-regulated and down-regulated genes showing mean fold change>1.5 and p-value<0.05 were used as input for the enrichment analysis. Interestingly, the modulated genes between synovial fluid and peripheral blood in $CD14^+$ monocytes were significantly enriched (FDR<0.05) for chemokine, NOD-like signalling and TLR signalling pathways (Table 5.6). Significant enrichment was also found for "Innate immune system" and "Neutrophil

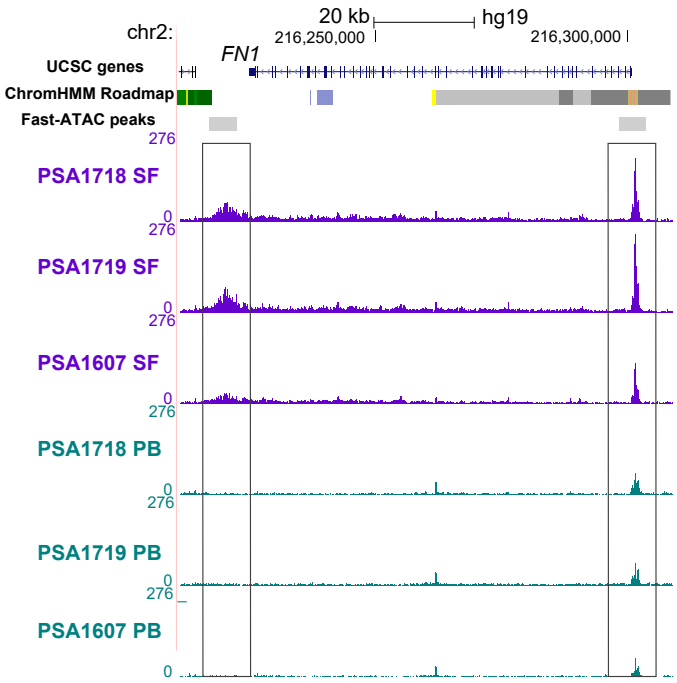


Figure 5.11: Chromatin accessibility landscape at the qPCR differentially expressed *FN1* gene in CD14$^+$ monocytes. UCSC Genome Browser view illustrating the ATAC normalised read density (y-axis) in two DARs located at the promoter and downstream the 3' UTR of the *FN1* gene (x-axis) in CD14$^+$ monocytes from synovial fluid and peripheral blood in 3 PsA patients. Both DARs were more accessible in synovial fluid when compared to peripheral blood. Tracks are colour-coded by tissue (SF=purple and PB=turquoise). The Roadmap Epigenomics Project chromatin segmentation track for peripheral blood isolated CD14$^+$ monocytes is also shown. All DARs were significant based on FDR<0.01 and fold change>1.5. SF=synovial fluid; PB=peripheral blood.

degranulation" pathways. Of particular relevance in monocytes were the chemokine, NOD-like and TLR signalling pathways, involved in the activation of cytokines and chemokines gene expression, leading to T cell recruitment and perpetuation of the adaptive inflammatory response. Some of the genes, such as *CCL5* and *NFKB1*, contributed to the enrichment of these all three relevant pathways.

The TLR signalling pathways enrichment involved *FN1* and *SPP1*, two of the top differentially expressed genes found in this pilot study, and up-regulation of the TLR receptors *TLR1* and *TLR2* (Table 5.6). Similarly to *FN1*, *SPP1* was also highly up-regulated (mean fold change>16) in the three cell types but showed the greatest fold change in $CD14^+$ monocytes (Figure 5.9A).Genes such as *TNF*, *IRF7* and *MYD88* highlighted the cross-link between the NOD-like and the TLR signalling pathways in the activation of the NFκB pathway.

The enrichment for the chemokine pathway in $CD14^+$ monocytes (Table 5.6) included genes highly up-regulated (mean fold change>16) in synovial fluid compared to peripheral blood (e.g *CCL18* and *CCL2*) for all three cell types as well as genes only consistently modulated between synovial fluid and peripheral blood in $CD14^+$ monocytes (e.g *CCL28* and *CXCL10*, Figure 5.8).

The modulated genes between synovial fluid and peripheral blood in $mCD4^+$ T cells only showed significant enrichment for genes in the IL-10 signalling pathway (Table 5.6), whereas $mCD8^+$ modulated genes between the two tissues were not enriched for any pathway.

Table 5.6: Pathway enrichment analysis for the modulated genes between synovial fluid and peripheral blood in CD14$^+$ and mCD4$^+$. The analysis was performed using only those genes showing p-value<0.05 and mean fold change>1.5. Reported enriched pathways were significant at an FDR <0.05 and fold change is specified in brackets.

Cell type	Pathway	Genes
CD14$^+$	Chemokine signalling	*CCL17, CCL18, CCL2, CCL28, CCL5, CCL7, CCR1, CCR5,CXCL10*
	(fold change=1.53)	*CXCL12, CXCL16, CXCL5, CXCL9, NFKB1, PPBP, PF4V1, STAT3, XCR1*
	NOD-like receptor signalling	*CCL2, CCL5, IFNAR1, IL18, IRF7, MEFV, MYD88, NFKB1,*
	(fold change=1.85)	*NAMPT, TNF*
	TLR signalling	*CCL5, CXCL10, CXCL9, IFNAR1, IRF7, MYD88, NFKB1, SPP1,FOS,*
	(fold change=1.64)	*TLR1, TLR2, TLR8, TNF*
mCD4$^+$	IL-10 signaling	*CCL3, CCL4, CCL5, CCR1, CCR2, CSF1, CSF3, IL10RA,*
	(fold change=2.13)	*IL1R1, IL1R2*

Pathway analysis using the significantly modulated genes between peripheral blood of healthy controls and PsA patients only showed significant enrichment for the Reactome immune system pathway in $CD14^+$ monocytes and $mCD4^+$ cells. This suggests some tissue-specificity of the pathways enriched for DEGs between synovial fluid and peripheral blood in this analysis, highlighting a more pronounced inflammatory phenotype of the pathological $CD14^+$ monocytes at the inflamed joints.

Overall, the integration of the chromatin accessibility and immune transcriptional data reinforced a relevant role of synovial $CD14^+$ monocytes in the production of cytokines and chemokines, contributing to the activation of the innate immune response and the recruitment of T cells to this site of inflammation.

5.3.6 Characterisation of $CD14^+$ monocyte heterogeneity in PsA using scRNA-seq

According to the analysis of chromatin accessibility and immune-related gene expression in this pilot cohort, $CD14^+$ monocytes showed the largest number of significant DARs and the most reliable modulation of expression, highlighting changes in expression between peripheral blood and synovial fluid for pro-inflammatory chemokines and cytokines. Monocytes are very plastic cells which initiate differentiation into macrophages at the site of inflammation. Therefore, exploring differences at the single-cell level may identify subpopulations with particular phenotypes of interest and may also help to understand the differences in the immune response driven by this cell type in circulation and at the inflamed synovium.

Exploration of monocyte subsets in synovial fluid and peripheral blood

First, scRNA-seq analysis was performed in paired PBMCs and SFMCs isolated from three PsA patients (Table 5.2). scRNA-seq data from each of the PBMCs and SFMCs samples were downsampled to 3,000 cells and filtered to remove genes that were expressed only in few cells and those cells expressing low number of genes (as described in Chapter 2). The monocyte population sorted by FACS was characterised for expressing the surface marker $CD14^+$. For consistency, the scRNA-seq monocyte population for downstream analysis was identified based on cluster analysis followed by expression of *CD14* and *LYZ*, two canonical markers of $CD14^+$ monocytes (Zhao *et al.* 2009)(Figures 5.12A and B and B.16A and B). After additional filtering to remove cells with abundant mitochondrial reads and excessive number of unique molecular identifiers (UMI) a total of 2,798 cells across the six samples were identified as $CD14^+$ monocytes (1,537 and 1,261 from synovial fluid and peripheral blood, respectively). This represented approximately 15% of the bulk SFMCs and PBMCs cells included in the analysis and in line with the proportion of $CD14^+$ monocytes previously reported using cell surface markers by mass cytometry (Figure 5.1). Interestingly, expression of the low-affinity immunoglobulin G Fcγ receptor (FcγRIIa) $CD16^+$ by the synovial fluid $CD14^+$ monocytes suggested a more intermediate phenotype ($CD14^+$ $CD16^+$) in this tissue compared to peripheral blood monocytes showing a more classical phenotype ($CD14^+$ $CD16^-$) (Figure 5.12C). Subsequent visualisation using uniform manifold approximation and projection (UMAP) clearly separated the synovial fluid and peripheral blood $CD14^+$ monocyte populations (Figure 5.12D).

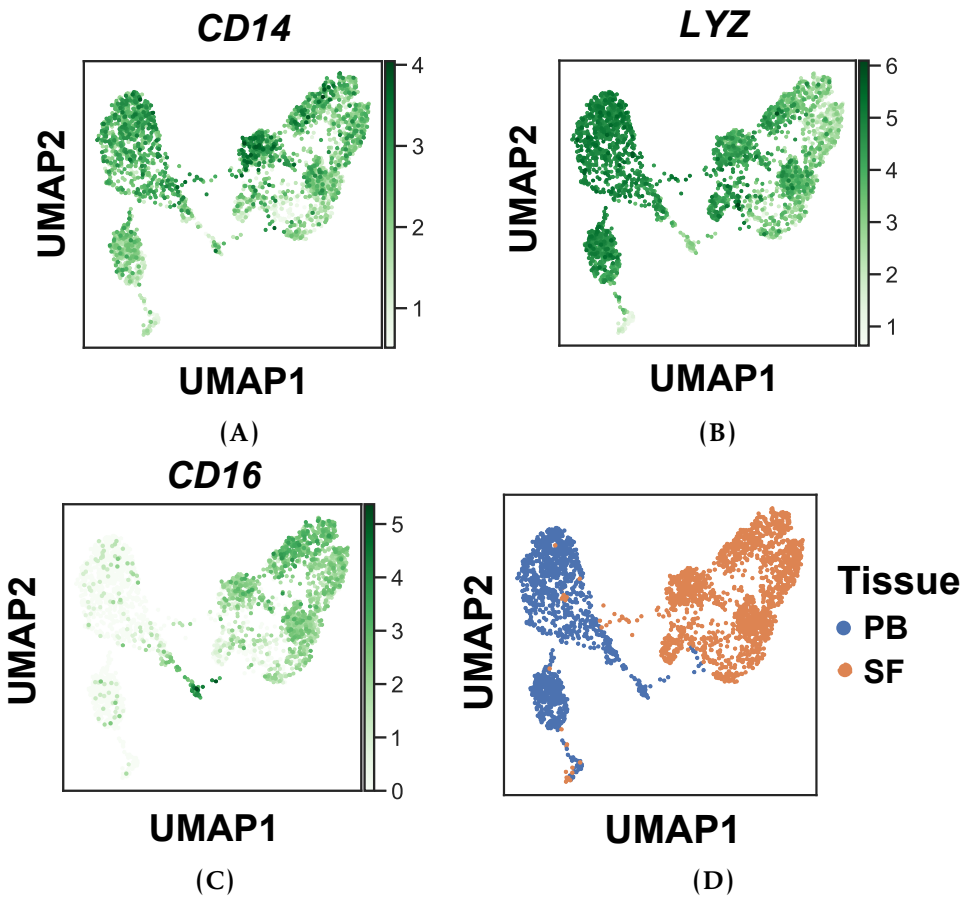


Figure 5.12: Visualisation using UMAP dimensional reduction of the synovial fluid and peripheral blood CD14$^+$ monocytes. Visualisation using UMAP dimensional reduction of (A) *CD14*, (B) *LYZ* and (C) *CD16* expression intensities in the combined synovial fluid and peripheral blood CD14$^+$ monocytes cells. (D) Overlaid colour-coded tissue identity for each of the CD14$^+$ monocytes cells from the combined population (orange or blue for synovial fluid and peripheral blood, respectively).

Cluster analysis was conducted in the combined population of synovial fluid and peripheral blood CD14$^+$ monocytes using standard default resolution of the Leiden clustering algorithm (detailed in Chapter 2). Using these parameters, 8 CD14$^+$ monocyte clusters (MC) were identified, with MC-1 to MC-4 predominantly formed by synovial fluid monocytes and clusters MC-5,6 and 8 being mostly of peripheral blood origin (Figure 5.13A). MC-1 and MC-6 were the largest synovial fluid and peripheral blood clusters, respectively, representing 40.7 and 65.8% of the total number of monocytes of each tissue (Figure 5.13A).

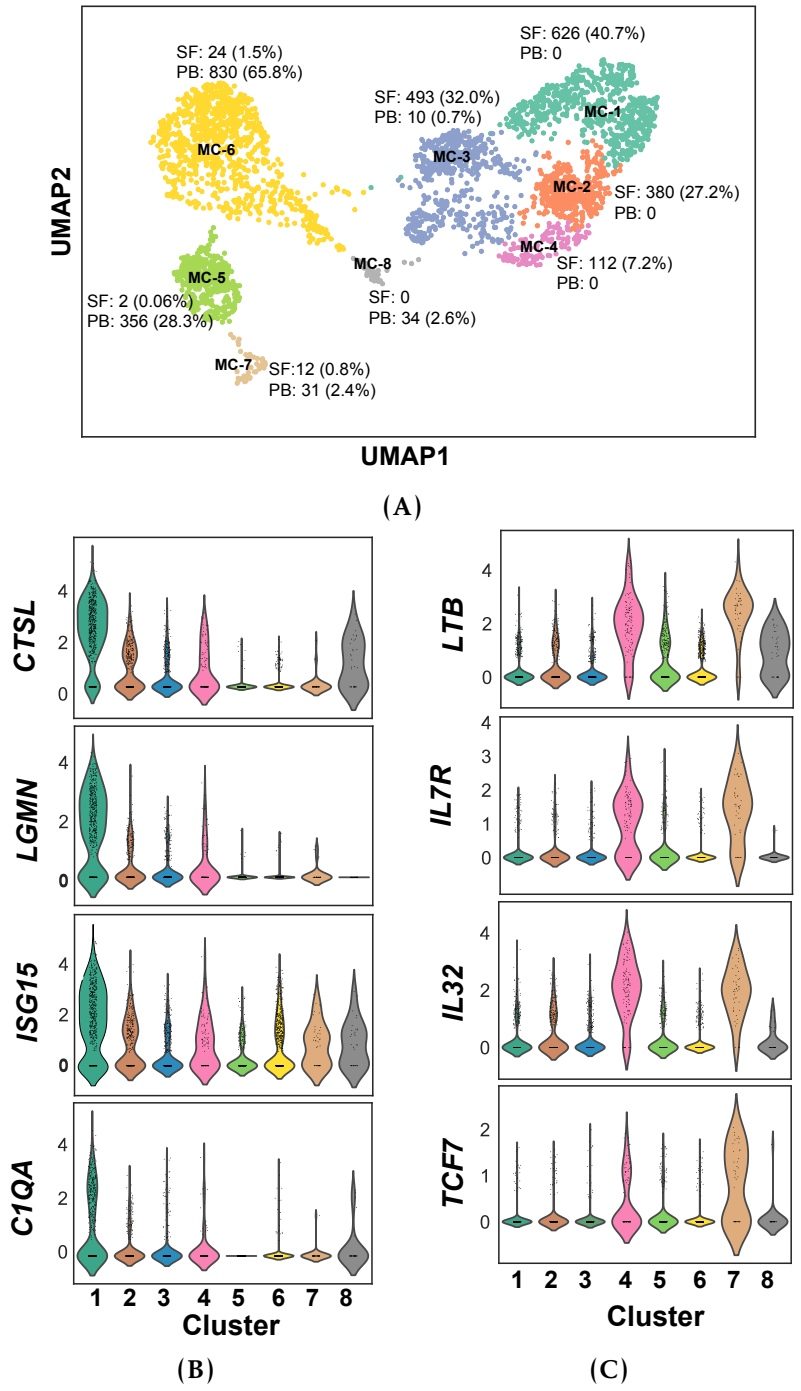


Figure 5.13: Identification and characterisation of CD14$^+$ monocytes subpopulations in synovial fluid and peripheral blood. (A) Visualisation using UMAP dimensional reduction of the CD14$^+$ monocytes subpopulations identified in the combined synovial flui and peripheral blood CD14$^+$ monocytes. For each cluster the number of cells and the percentage over the total number of cells from the corresponding tissue are indicated. Violin plots representing for MC-1 to MC-8 (x-axis) the probability density of expression (y-axis) for marker genes charcateristic of the (B) MC-1 and (C) MC-4 subpopulations. SF=synovial fluid; PB=peripheral blood.

MC-7 was the smaller cluster and was characterised by the most even contribution of $CD14^+$ monocytes from the two tissues (0.8 and 2.4% from synovial fluid and peripheral blood, respectively). MC-7 may be a real cluster or the result from remaining unfiltered dropouts or cell missclassification by the clustering algorithm. Additional sample size will be required to confirm this cluster identity.

Each of the 8 $CD14^+$ monocytes clusters was characterised by a number of marker genes, a set of top significant up-regulated genes by the cells of that particular subpopulation when compared to the rest of the cells (Figure 5.14). The heatmap illustrating expression levels of the top marker genes in each of the clusters only showed moderate differences across the subpopulations formed by monocytes of the same tissue but displayed a very distinct pattern of expression across tissues (Figure 5.14).

Functional exploration of synovial fluid and peripheral blood $CD14^+$ monocytes subpopulations

Amongst the synovial fluid clusters, MC-1 and MC-4 appeared to be of particular interest based on their distinctive marker genes. MC-1 presented significant up-regulated expression of IFN-related genes such as *CTSL*, *ISG15*, *IFI6* or *IFI27*, genes involved in antigen presentation such as *LGMN* and members of the complement cascade, especially the *C1QA* when compared to all the other clusters. Gene ontology analysis using all the significant up-regulated genes (FDR<0.01 and fold change>1.5) unique for this cluster demonstrated enrichment of type-I IFN (and IFN-γ) mediated signalling, response to virus, chemokine-mediated signalling and complement cascade, GO terms that were not found to be enriched for any of the remaining 7 clusters when performing the same analysis (Figure 5.15A).

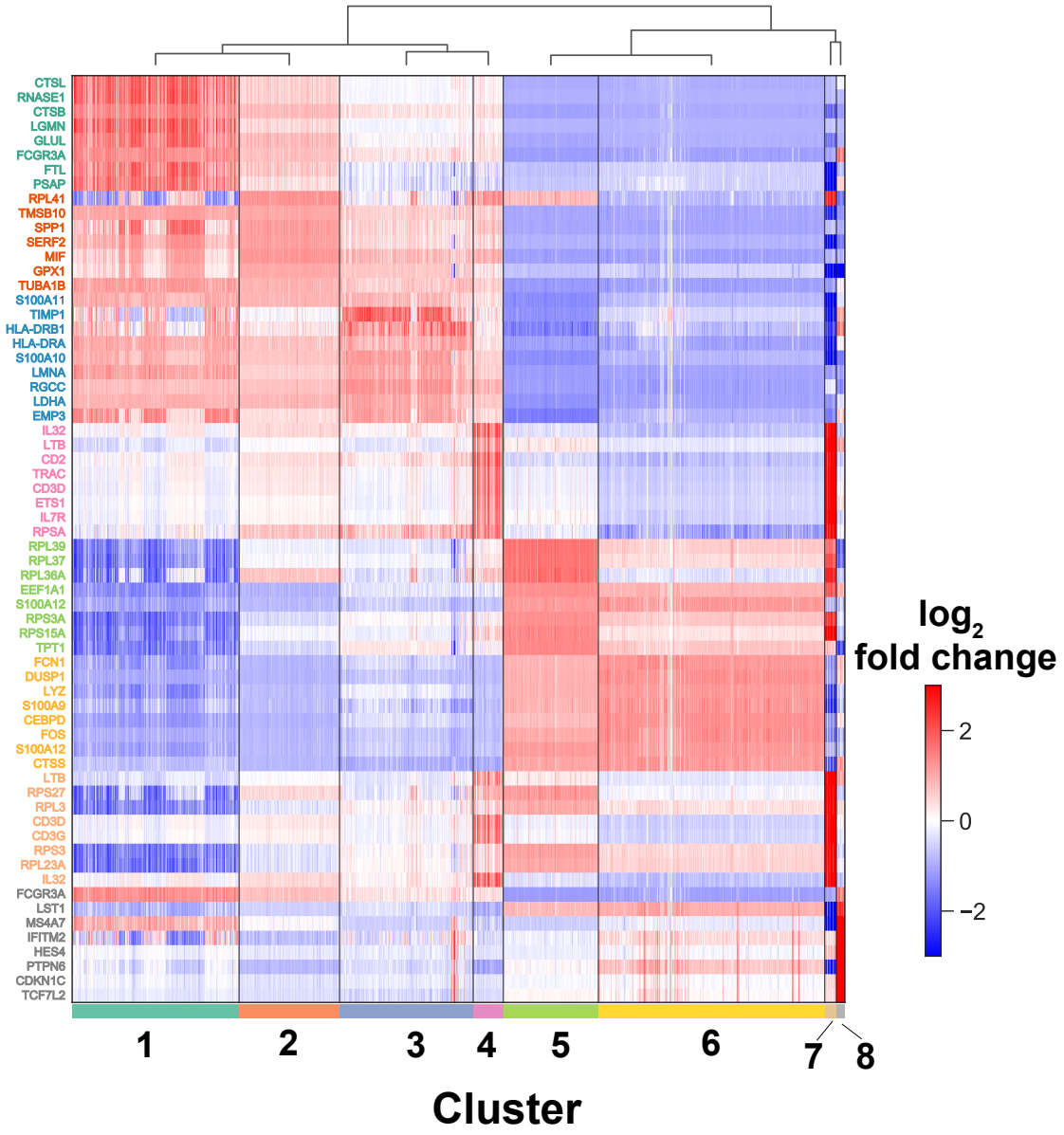


Figure 5.14: Heatmap and hierarchical clustering for the top eight marker genes of the CD14$^+$ monocyte subpopulations identified in synovial fluid and peripheral blood from three PsA patients. The rows represent the genes and the columns correspond to each of the cells analysed in the experiment. The heatmap illustrates the relative expression ($\log_2$fold change in a particular cell of the cluster compared to the average expression of all the cells from the other clusters) for the top eight marker genes defining each of the eight clusters. Dendogram indicating hierarchical relationship between clusters is indicated at the top and cluster identity for each of the cells is specified at the bottom.

MC-4 represented 7.2% of synovial fluid $CD14^+$ monocytes and was characterised by high expression of *IL7R*, the IL-7 regulated cytokines *LTB*, the TF *TCF7*, the TNF-α induced pro-inflammatory cytokine *IL-32* (Kim *et al.* 2005) and the chemokine *CCL5* (Figures 5.13C and 5.14). These genes were also up-regulated by the $CD14^+$ monocyte subpopulation expressing $IL7R^+$ upon *in vitro* LPS stimulation conducted by Al-Mossawi and colleagues (in revision, Al-mossawi *et al.*,2018). In addition to the top cluster marker genes, MC-4 cluster showed differential expression of 340 genes (228 and 112 up- and down-regulated, respectively) when compared to the rest of the monocytes from the other clusters. These included the immunoregulatory non-coding RNA *MALAT1* and the stabiliser of the TCR *TRAT1* (also up-regulated in psoriasis $mCD8^+$ cells), amongst others (Figure 5.15B). The DEGs (FDR<0.01) identified for the MC-4 subpopulation showed enrichment for the IL-7 responsive genes identified in $IL7R^+$ $CD14^+$ LPS-stimulated monocytes in the aforementioned study (Al-Mossawi *et al.* 2018) (overlap of 116 genes, p-value=$1.9x10^{-13}$, Fisher exact test). Moreover, this set of DEGs showed enrichment for similar GO terms, including ribosomal biogenesis and protein translation, as the ones revealed by the up-regulated IL-7 $CD14^+$ monocytes signature from Al-Mossawi and colleagues (Figure 5.15C).

Interestingly, the mixed synovial fluid and peripheral blood MC-7 subpopulation (12 and 32 synovial fluid and peripheral blood cells, respectively) also showed up-regulation for some of MC-4 marker genes (Figures 5.14 and 5.13B). MC-7 only represented 2.4% of the peripheral blood $CD14^+$ monocytes, suggesting that the subpopulation of $CD14^+$ monocytes expressing *IL7R* (amongst the other markers) may be predominant and distinctive of the PsA synovial $CD14^+$ monocytes. This is consistent with the synovial fluid open DARs found at the 5' and 3' UTR of *IL7R* gene as well as its up-regulated expression observed in synovial fluid monocytes by qPCR (Figures 5.6A and 5.9A).

Synovial fluid CD14$^+$ monocytes, particularly the MC-1 and MC-3 subpopulations expressed significant higher levels of a number of MHC-II genes, including *HLA-DRB1*, *HLA-DPB1* or *HLA-DRB5*, involved in the monocytes antigen presentation and CD4$^+$ T cell activation, in line with the intermediate monocytes phenotype (Figures B.17A and 5.14). Moreover, peripheral blood CD14$^+$ monocytes showed greater expression of some of the phagocytic marker such as *CD93*, *CD36* and *CLEC4E*, mostly driven by the MC-5 and MC-6 subpopulations and in line with their more classic phenotype compared to the synovial fluid counterparts (Figure B.17B), as well as genes from the S100 anti-microbial family, including *S100A8* or *S100A12* (Figure B.18A). Nevertheless, the unique marker genes specific for other subpopulations did not reveal unique relevant functional roles for these clusters probably due to only moderate differences in the transcriptional profiles of the clusters within each tissue, and additional sample size may help to characterise these populations or find additional ones.

5.3.7 Identification of additional pathways dysregulated between synovial fluid and peripheral blood CD14$^+$ monocytes using scRNA-seq data

Differential gene expression between the CD14$^+$ monocytes from the two tissues was conducted to explore genome-wide differences that could complement the restricted qPCR analysis and serve as validation, since two out of the three PsA patients included in the single-cell analysis were different from the ones used for the qPCR analysis. A total of 1,275 genes were differentially expressed at an FDR<0.01 and fold change>1.5 between synovial fluid and peripheral blood (778 and 497 up- and down-regulated, respectively (Figure B.18A). Out of the 72 DEGs identified by qPCR, 38 were also differentially expressed (FDR<0.01) by scRNA-seq analysis (30 when further restricting on fold

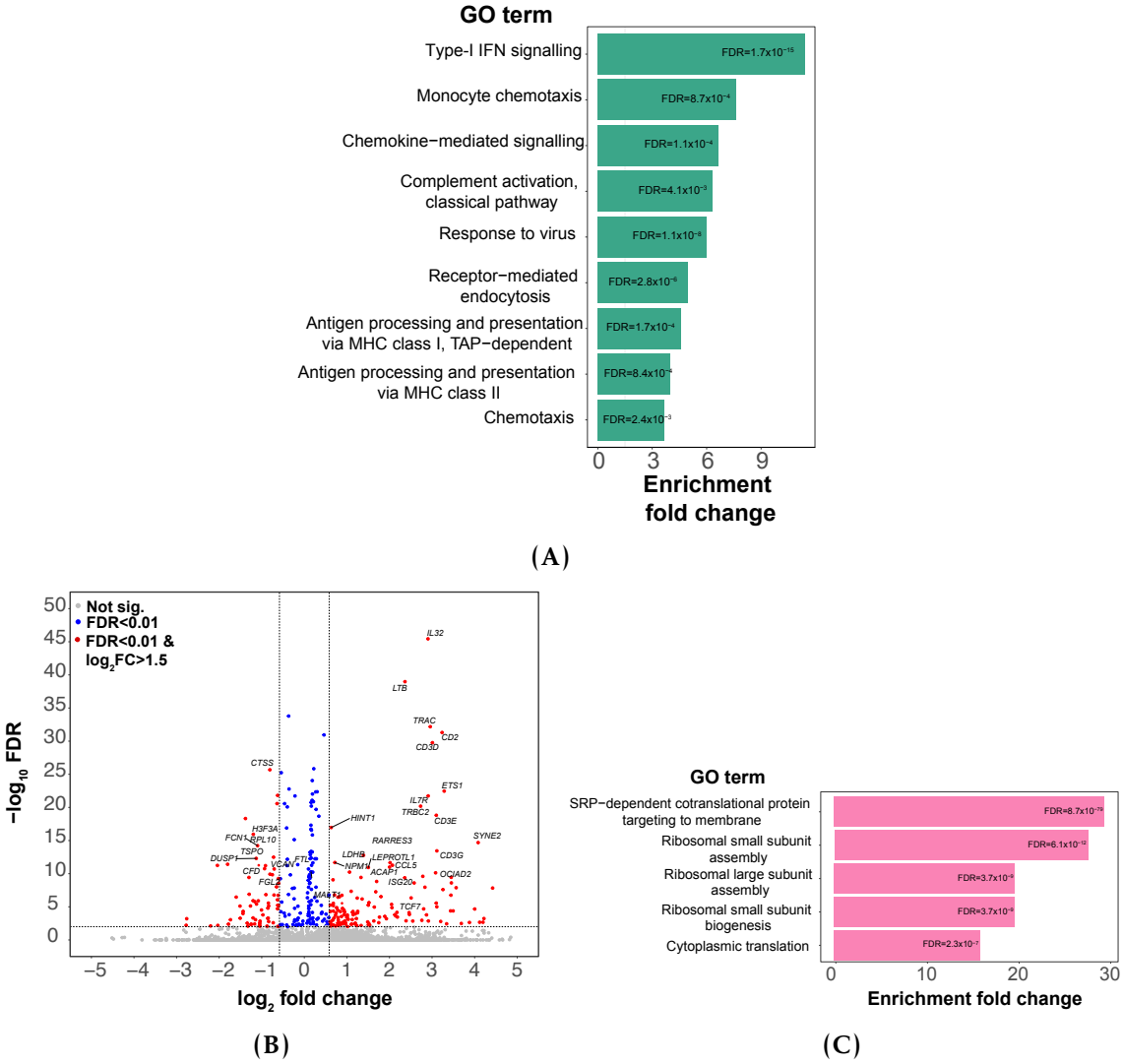


Figure 5.15: Functional characterisation of the MC-1 and MC-4 CD14$^+$ PsA monocyte subpopulations. (A) Barplot illustrating the top Gene Ontology (GO) terms enriched for significantly up-regulated (FDR<0.01 and fold change>1.5) unique genes in the MC-1 cluster. (B) Volcano plots showing the differentially expressed genes in the MC-4 cluster when compared to the rest of CD14$^+$ monocytes interrogated by scRNA-seq. The significance ($\log_{10}$FDR) of the differential expression (y-axis) is plotted against the $\log_2$fold change. Genes are coloured based on FDR and $\log_2$fold changes and some of the top significant genes are labelled. (C) Barplot illustrating the top GO terms enriched for the significantly up-regulated genes (FDR<0.01) identified in the MC-4 cluster.

change>1.5), including those showing the largest changes between tissues, such as *SPP1*, *FN1*, *S100A12* or *FOS* (Figure B.18B).

Pathway enrichment analysis using the scRNA-seq significant DEGs (FDR<0.01 and fold change>1.5) between synovial fluid and peripheral blood PsA monocytes confirmed enrichment for the chemokine signalling and NOD-like receptor pathways, previously reported by the qPCR array data (Table A.7). Up-regulation in synovial fluid $CD14^+$ monocytes of chemokines (including *CCL18*, *CCL2*, *CCL3*, *CCL4*, *CCL8* and *CXCL10*) and chemokine receptors (*CCR1* and *CCR5*) was consistent with the qPCR analysis and also revealed new members dysregulated between the two tissues. Additional genes of NOD-like receptor signalling pathways not interrogated by qPCR array that contributed to enrichment in this analysis included down-regulation of the negative regulators of the inflammasome *CARD8* and *NLRP12* (also a negative regulators of NF-κB) (Mao *et al.* 2018; Williams *et al.* 2005). Interestingly, the NOD-like receptor *NLRP1* and caspase-1 (*CASP1*), required for assembly of the inflammasome were down-regulated in PsA synovial fluid $CD14^+$ monocytes compared to peripheral blood.

Other relevant enriched pathway identified in this analysis included IFN and IFN-γ signalling, MHC-II antigen presentation, complement cascade, apoptosis and osteoclast differentiation (Table A.7). The relevance of IFN signalling in the synovium was consistent with the up-regulation of IFN regulated genes in the MC-1 cluster, and also with the up-regulation of *STAT1* in synovial fluid $CD14^+$ monocytes. Enrichment of the apoptosis pathway was driven by down-regulation of positive regulators, such as the TFs *FOS* and *APAF1*, *PRKCB* and *TNFSF10*, and up-regulation of inhibitors, including *BCL2*, *BCL2A1* and *BCL2L1*, in synovial fluid compared to peripheral blood $CD14^+$ monocytes. This may suggest an increased resistance to apoptosis by PsA synovial fluid monocytes. Moreover, $CD14^+$ monocytes from the synovium

showed large up-regulation of the *TNFRSF11A* (also known as *RANK*) (fold change=26.5, Figure B.18A) involved in monocytes differentiation to osteoclasts and NF-κB activation (Mensah *et al.* 2008), and down-regulation of *NCF1*, which has been shown to impair the extracellular degradation of inflammatory mediators by neutrophil proteases (Schauer *et al.* 2014).

Lastly, genome-wide comparison between the $\log_2$fold changes for all the genes analysed by scRNA-seq expression and all the regions interrogated for differential chromatin accessibility between synovial fluid and peripheral blood only showed low correlation (R=0.162, p-value=2x10^{-16}) (Figure B.18B), with 252 and 75 genes changing expression and chromatin accessibility in the same or opposite direction, respectively (significant enrichment, p-value=6.264x10^{-11} Fisher exact test). For example, *CCL2*, *FN1* and *FABP5* showed transcriptional up-regulation in scRNA-seqand were in the proximity of a synovial fluid open DAR (Figure B.18C).

5.3.8 Integration of mass cytometry differences in protein expression with chromatin accessibility and gene expression data in PsA CD14$^+$ monocytes.

To determine the effect of chromatin accessibility and genes expression differences in CD14$^+$ monocytes between synovial fluid and peripheral blood at the protein level, mass cytometry was conducted by Dr Hussein Al-Mossawi and Dr Nicole Yager in collaboration with UCB. This analysis included the six samples with either paired ATAC/qPCR data or scRNA-seq analysis presented in this chapter (Table 5.2) and an additional four PsA patients recruited only for mass cytometry processing. Intra-cellular staining for relevant cytokines and chemokines was performed before and after treatment with BFA (blocker of protein transport) and enables identification of cells actively producing the

protein of interest in the closest state to the physiological, without additional inflammatory stimuli (see Chapter 2).

Amongst the measured proteins, TNF-α, osteopontin (protein product of *SPP1*), MCP-1 (protein product of *CCL2*) and IP-10 (protein product of *CXCL10*) were of particular interest as they showed transcriptional up-regulation in synovial fluid CD14$^+$ monocytes compared to peripheral blood and have a key role in the pathological implications of monocytes/macrophages in PsA. In the three PsA patients processed for ATAC and qPCR mass cytometry, expression in CD14$^+$ of TNF-α, osteopontin, MCP-1 and IP-10 demonstrated a greater increase in cell staining for each cytokine after 6h of BFA treatment in CD14$^+$ monocytes from synovial fluid when compared to peripheral blood (Figure 5.16).

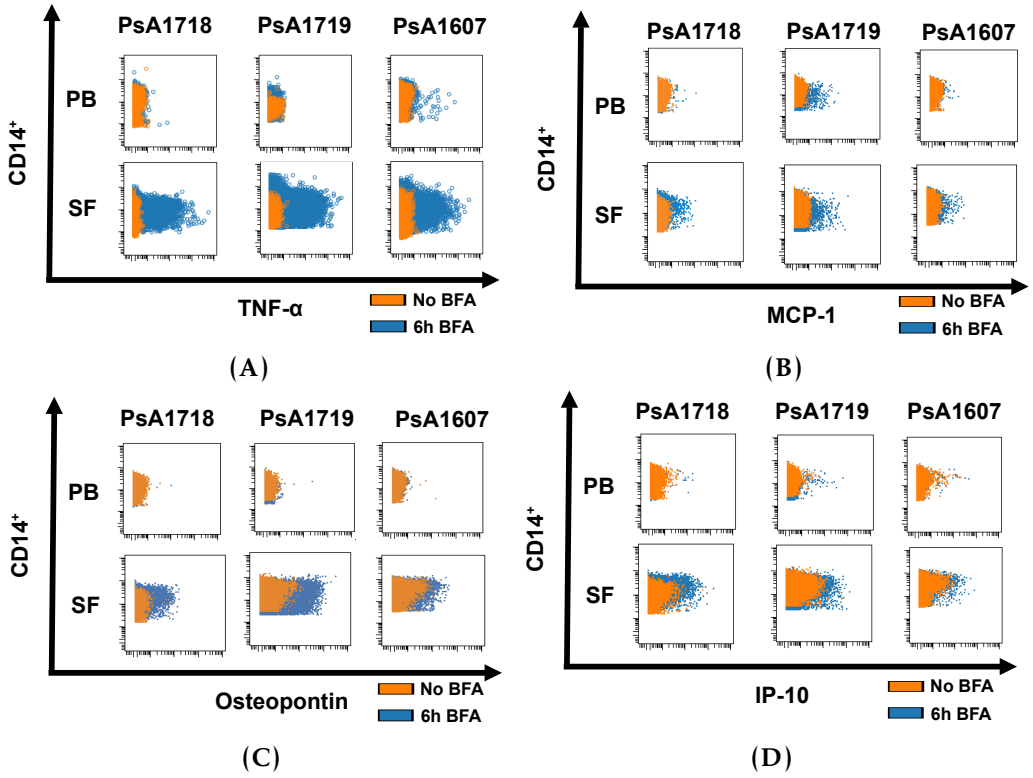


Figure 5.16: Biaxial representation of CD14$^+$ monocytes expressing TNF-α, MCP-1, osteopontin and IP-10 in synovial fluid and peripheral blood before and after protein transport blockade with BFA using mass cytometry. For each of the three PsA patients with ATAC and expression data, expression of the monocytes marker CD14$^+$ vs (y-axis) vs (A) TNF-α, (B) MCP-1, (C) osteopontin or (D) IP-10 (x-axis) in matched synovial fluid and peripheral blood without protein transport blockade (blue dots) or after 6h of treatment with BFA (orange dots).

The increment in percentage of $CD14^+$ monocytes producing each of these four inflammatory mediators after blocking protein transport with BFA treatment was determined for each of the PsA patients in each compartment (peripheral blood and synovial fluid). For each immune mediator and compartment, this increment was calculated as the difference between the percentage of cells with positive staining after 6h BFA treatment and the percentage of positive staining cells at 0h (in absence of BFA incubation).The increment in the percentage of $CD14^+$ monocytes producing TNF-α was significantly greater in synovial fluid compared to peripheral blood (Wilcoxon signed-rank test, p-value=0.048) (Figure 5.17A), in line with enrichment of accessible chromatin nearby genes of the NF-κB signalling pathway, the enrichment of DEGs for the TLR and NOD-like signalling pathways and the qPCR up-regulated expression of *TNF* in synovial $CD14^+$ monocytes. Likewise, greater percentage of $CD14^+$ monocytes producing osteopontin, MCP-1 and IP-10 (Wilcoxon signed-rank test p-value=0.001, p-value=0.003 and p-value=0.001, respectively) after BFA treatment were observed in synovial fluid compared to peripheral blood (Figure 5.17B, C and D). The greater increase in the percentage of cells producing MCP-1 and osteopontin in synovial fluid monocytes was in line with the up-regulated expression of these two genes in synovial fluid compared to peripheral blood by qPCR and scRNA-seq analysis. Importantly, PsA1807 showed particularly higher percentage of synovial fluid $CD14^+$ monocytes producing all four molecules after BFA treatment when compared to the other patients in the cohort (Figure 5.17A, B and C in pink). Nevertheless, the differences in the increment of positive cells after BFA treatment between synovial fluid and peripheral blood still remained significant (p-value¡0.05) when removing PsA1807 from the analysis, also in the case of TNF-α and MCP-1 (showing more moderate differences than IP-10 and osteopontin).

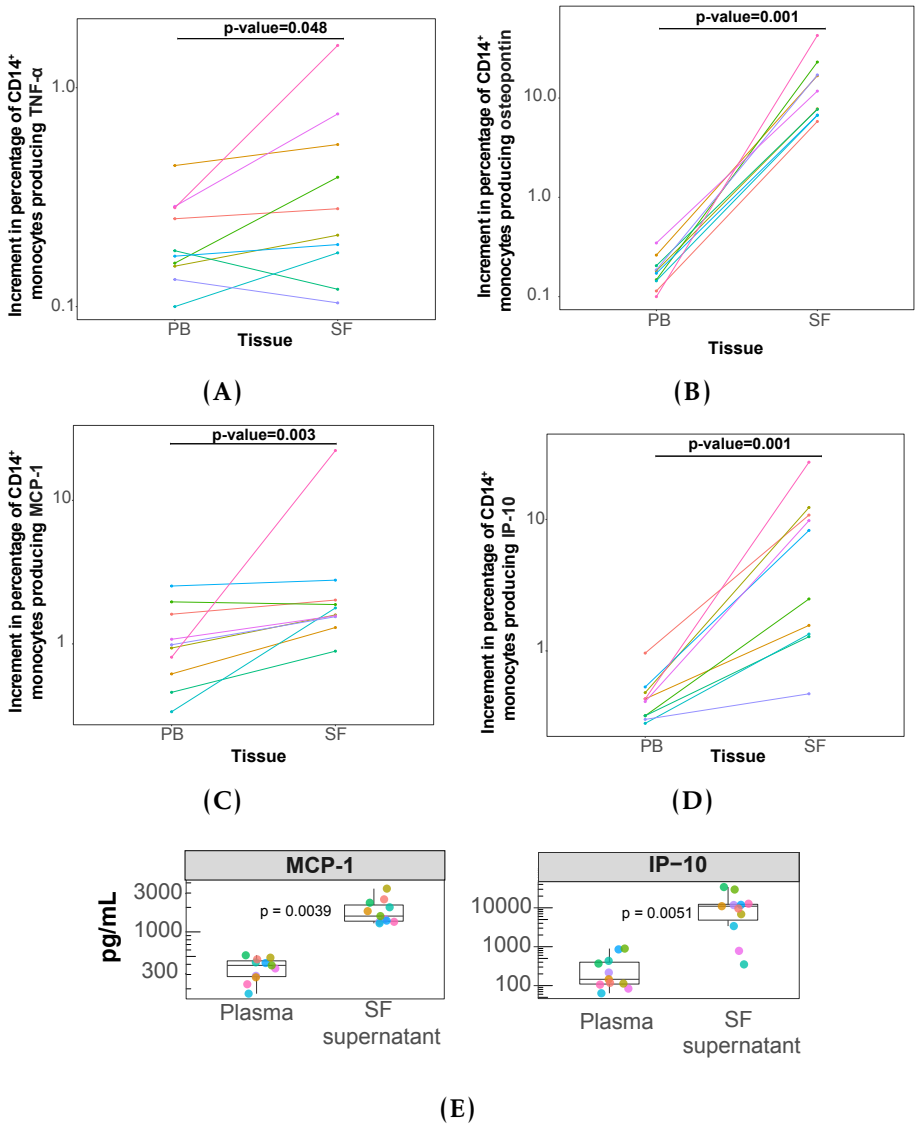


Figure 5.17: Increment in the percentage of CD14$^+$ monocytes producing TNF-α, osteopontin, MCP-1 and IP-10 after BFA treatment in synovial fluid and peripheral blood. The increment in the percentage of CD14$^+$ monocytes producing (A) TNF-α, (B) osteopontin , (C) MCP-1 and (D) IP-10 in synovial fluid and peripheral blood are shown for each of the ten PsA samples assayed by mass cytometry. In each sample and tissue, the increment in the percentage of cells producing each of the cytokines is calculated as the difference in positive cells for the relevant cytokine before and after BFA treatment (6h minus 0h). Axis are transformed into a $\log_{10}$ scale. (E) Boxplots illustrating the concentration in pg/mL (x-axis $\log_{10}$ scale) of MCP-1 and IP-10 measured in supernatant from synovial fluid and plasma from peripheral blood of ten PsA patients using enzyme-linked immunosorbent assay (ELISA). The statistical significance of the differences between synovial fluid and peripheral blood was determined using Wilcoxon signed-rank test. SF=synovial fluid; PB=peripheral blood.

The differences in the percentage of CD14$^+$ monocytes producing MCP-1 and IP-10 between synovial fluid and peripheral blood were of particular

interest due to correlating changes in chromatin accessibility, gene expression and protein production. A statistically significant synovial fluid open DAR was found in CD14$^+$ monocytes upstream *CCL2* gene (Figure 5.18A). Similarly, two synovial fluid open DARs were also identified upstream and at the promoter of the *CXCL10* gene (Figure 5.18B)

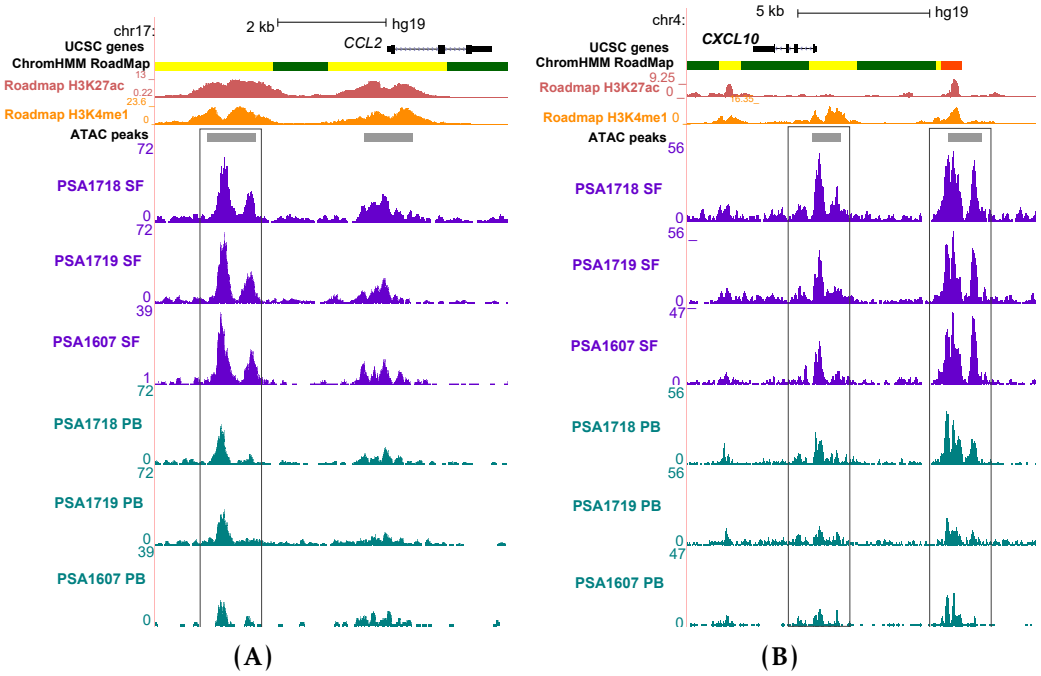


Figure 5.18: Chromatin landscape in CD14$^+$ monocytes in the proximity to *CCL2* and *CXCL10*. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) for (A) a DAR located upstream *CCL2* gene and (B) two DARs at the promoter and upstream the *CXCL10* gene (x-axis) in synovial fluid and peripheral blood CD14$^+$ monocytes from three PsA patients. In (A) and (B) DARs (black rectangles) were significant for FDR<0.01 and fold change>1.5 and showed presenting greater accessibility in synovial fluid when compared to peripheral blood. Tracks are colour-coded by tissue (SF=purple and PB=turquoise). Additional epigenetic tracks including chromatin segmentation maps and H3Kme1 and H3K27ac histone modifications from CD14$^+$ monocytes of the Roadmap Epigenomics Project are included.

The DAR proximal to *CCL2* was annotated as an enhancer according to Roadmap Epigenomics Project chromatin segmentation maps and overlaps an eRNA reported by FANTOM5 in CD14$^+$ monocytes. The DAR upstream *CXCL10* co-localised with H3K4me1 and H3K27ac, reinforcing the enhancer properties of this region. In addition to the up-regulated transcriptional and protein expression in PsA CD14$^+$ monocytes, enzyme-linked immunosorbent assay

(ELISA) quantification of MCP-1 and IP-10 levels conducted by collaborators revealed increased concentration of both chemokines in synovial fluid supernatant compared to plasma in the same ten PsA patients (Figure 5.17E). This further supported the observed differences between synovial fluid and peripheral blood in chromatin accessibility, transcription and protein expression for *CCL2* and *CXCL10*.

5.3.9 Prioritisation and interpretation of PsA GWAS SNPs

Bayesian fine-mapping using genotyping data

Baysian fine-mapping for the PsA Immunochip GWAS signals identified by Bowes and colleagues (Bowes *et al.* 2015a) was conducted using genotyping data from 1,103 patients and 8,900 controls. These samples only included the UK cohort from Bowes *et al.*, 2015 PsA Immunochip. In the manuscript, Bowes and colleagues also performed fine-mapping for a number of associated loci with genome-wide or nominal evidence of association, reporting the number of independent signals as well as the number of SNPs in the 99% credible set (SNPs accounting for 99% of the association in a particular locus) but not the identity of each of those SNPs. Thus *de novo* fine-mapping was conducted here to obtain the identity and the location of the SNPs in each credible set, and to further integrate them with other functional data.

Fine-mapping was performed in 36 loci reported in Bowes *et al.*, all showing at least nominal significance (p-value<0.05) in their GWAS study. Out of the 36 loci, 22 showed $\log_{10}$ABF<3 (cut-off used in Bunt *et al.*, 2015) for the lead SNP in the fine-mapping association analysis (Table A.8). The majority corresponded to those signals with lowest significance in the GWAS association analysis (p-value< 10^{-4}, e.g *RSPH3/TAGAP* and *ELMO1*) (Table A.8).

Table 5.7: Summary table of the PsA GWAS loci presenting log_{10}BF≥3 for the fine-mapping lead SNP. For 12 PsA loci log_{10}BF of the fine-mapping lead SNP was 3 or greater. In 5 of those loci (*) the fine-mapping lead SNP was in low LD (r^2 <0.5) with the PsA GWAS SNP and/or did not contain it in the credible set. The effect size (OR) is reported relative to minor allele. FM=fine-mapping; MAF= minor allele frequency; OR=odds ratio; BF=Bayes factor; PP=posterior probability.

chr	Closest gene	FM lead SNP	MAF cases/controls	OR FM	log_{10}BF FM lead SNP	PP	90% credible set	Bowes FM lead SNP	Bowes 90% credible set
2	*IFIH1*	rs13406089*	0.28/0.33	0.78	4.58	0.48	2	rs35667974	4
5	*IL12B/ADRA1B*	rs2546890	0.42/0.48	0.76	6.53	0.6	23	rs4921482	3
5	*IL13*	rs2069616*	0.5/0.44	1.25	5.16	0.05	55	NA	NA
1	*IL23R*	rs12044149	0.32/0.25	1.41	9.83	0.14	29	rs12044149	34
1	*IL28RA/GRHL3*	rs2135755*	0.45/0.50	1.20	3.06	0.03	49	NA	NA
19	*ILF3*	rs11085727*	0.25/0.30	0.79	3.83	0.22	35	NA	NA
17	*PTRF/STAT3*	rs730086*	0.29/0.34	0.81	3.39	0.39	400	NA	NA
1	*RUNX3*	rs6600250	0.45/0.50	1.20	3.07	0.03	48	rs7523412	52
12	*STAT2/IL23A*	rs12368739	0.04/0.06	1.70	4.04	0.02	110	rs2020854	121
6	*TRAF3IP2*	rs33980500	0.11/0.07	1.71	8.26	0.87	2	rs33980500	7
19	*TYK2*	rs11085727	0.25/0.30	0.79	3.83	0.21	32	rs34725611	5
5	*CSF2/P4HA2*	rs11242104	0.48/0.43	1.24	5.31	0.07	58	rs715285	35

Fine-mapping for GWAS signals with greater statistical significance (p-value< 10^{-4}), for example *DDX58*, also failed and were likely discarded by Bowes and colleagues fine-mapping based on the marker density in the region being lower than 100 SNPs.

Amongst the loci that also failed to be fine-mapped in this thesis ($\log_{10}$BF<3), four of them (*B3GNT2*, *NOS2A*, *REL* and *TNIP1*) had been successfully fine-mapped by Bowes and colleagues (Table A.8). This is likely due to the well established influence of sample size on the association analysis (Bunt *et al.* 2015).

Of the 12 loci passing the $\log_{10}$BF≥3 cut-off, five presented a fine-mapped lead SNPs in low LD ($r^2 < 0.5$) with the PsA GWAS lead SNP and/or did not contain the GWAS lead SNP in the credible set (Table 5.7 labelled with *). This was the result of the association analysis identifying a different signal, in some cases the signal from a proximal locus, reinforcing again the sensitivity of the association analysis for sample size. For example, the fine-mapping signal of the *IL13* locus was confounded by the *TYK2* signal, both in proximity at chr19. The additional seven signals successfully fine-mapped in this study contained the GWAS lead SNP and also the fine-mapping lead SNPs from Bowes and colleagues analysis in their credible sets. For the seven successfully fine-mapped loci, the size of credible sets ranged between two and 110 SNPs, with *TRAF3IP2* being the only locus refined to fewer than ten SNPs and showing $\log_{10}$BF>5, generally considered the genome-wide significant cut-off for this type of analysis (Table 5.7). Furthermore, for *IL23R* and *TRAF3IP2*, the fine-mapping lead SNPs in this analysis was the same as the reported by Bowes and colleagues. In the step-wise conditional analysis a secondary independent signal was also identified for the *IL12B/ADRAB1* locus; however the lead SNP in the analysis of this thesis (rs12653520) was different from the one reported by Bowes and colleagues (rs12188300).

Enrichment of fine-mapped SNPs with ATAC and eQTL data

The 90% credible sets of the seven loci successfully fine-mapped in this analysis included a total of 292 unique SNPs. Overlapping epigenetic and expression data generated in clinical samples with the credible set of SNPs from PsA GWAS loci may be informative for refining putative functional causal variants in non-coding or intergenic GWAS associations compared to integrating epigenetic data from cell lines or healthy controls. Moreover, this may also reveal the location of some of those variants at DARs across tissues. Co-localisation of the genomic position of the credible set of SNPs with the significant differentially accessible regions (DARs with FDR<0.01 and fold change>1.5) identified between synovial fluid and peripheral blood in the four cell types did not reveal any overlap. Enrichment of fine-mapped SNPs in the proximity of DARs (10Kb window) was explored using as background sets of SNPs preserving the same LD structure, allele frequencies and distance to genes as the fine-mapped SNPs (detailed in Chapter 2). Significant enrichment was only identified proximal to DARs in NK cells (Figure 5.19A) but not in the other three cell types (p-value>0.05).

Additional overlap was performed between the 292 SNPs of the credible sets from the seven successfully fine-mapped loci and the consensus list of ATAC peaks for each of the four cell types studied in this chapter (CP_CD14, CP_mCD4, CP_mCD8 and CP_NK). The largest number of overlaps between ATAC peaks and SNPs from the 90% credible sets (17 SNPs) was found in $CD14^+$ monocytes, followed by $mCD8^+$, $mCD4^+$ and NK cells (Table 5.8) with significant enrichment found in the four cell types when compared to matched backgrounds of SNPs (Figure 5.19B, C, D and E).

The 43 unique SNPs from the 90% credible sets located within ATAC peaks were distributed across the *CSF2* (8), *IL12B* (3), *IL23R* (4), *RUNX3* (6), *STAT2* (14), *TRAF3IP2* (1) and *TYK2* (7) loci. A number of these SNPs were found to

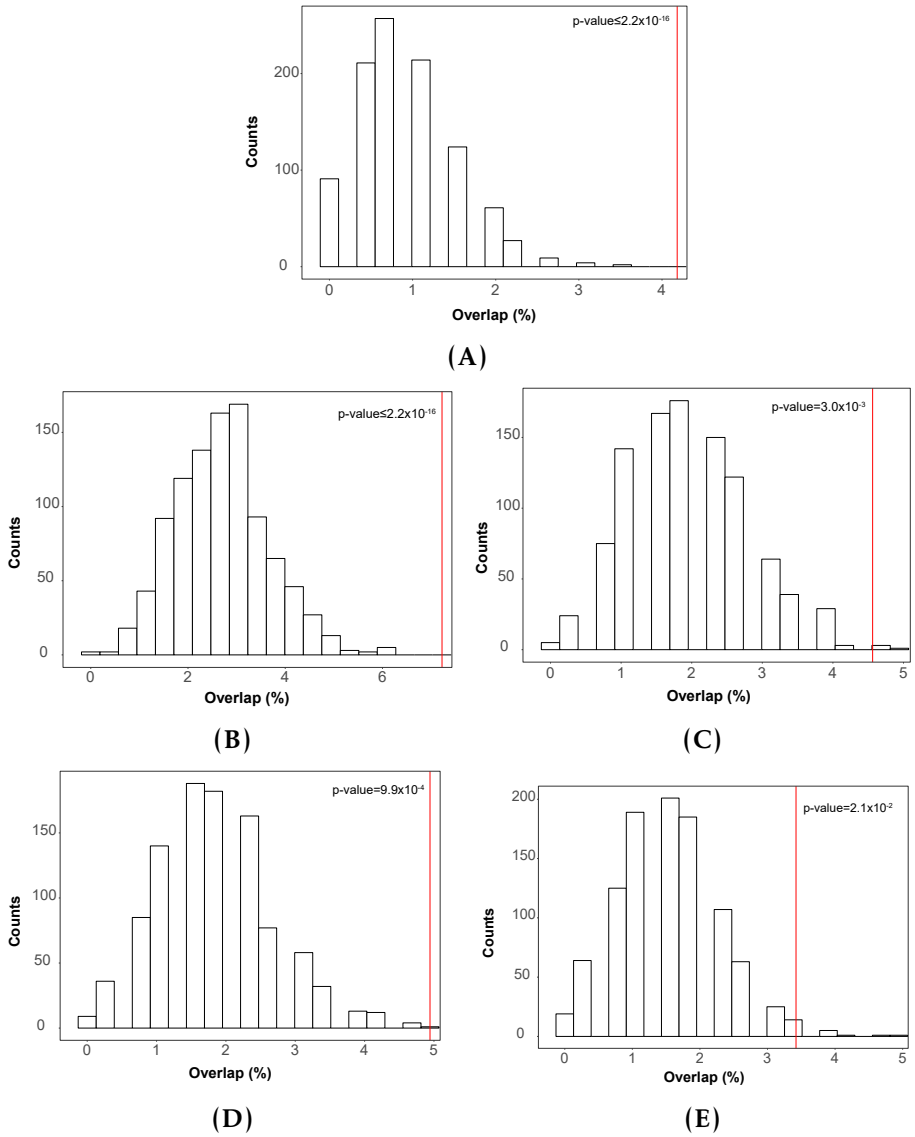


Figure 5.19: Enrichment of PsA GWAS fine-mapped SNPs for DARs and ATAC peaks in CD14$^+$ monocytes, mCD4$^+$, mCD8$^+$ and NK cells. (A) Histogram illustrating the percentage of SNPs from the 90% credible set of the 7 successfully fine-mapped PsA loci overlapping a 10Kb window around DARs in NK cells (red line) compared to the same overlap found for each of the 1,000 sets of background SNPs. (B) to (E) Histograms showing the percentage of the same fine-mapped SNPs overlapping ATAC peaks from the consensus list (red line) of (B) CD14$^+$ monocytes, (C) mCD4$^+$, (D) mCD8$^+$ and (E) NK cells vs the distribution of frequencies of overlaps identified in each of the 1,000 sets of background SNPs. Significance of the enrichment is indicated as the p-value calculated using binomial test.

ATAC cell type master list	90% credible set overlapping SNPs (number)	Cell-type specific overlap of fine-mapped SNPs
$CD14^+$ monocytes	32	*STAT2* (5), *TYK2*(2) *RUNX3*(1), *TRAF3IP2*(1)
$mCD4^+$	29	*CSF2*(1), *IL23R*(1)
$mCD8^+$	28	*RUNX3* (1)
NK	19	*TYK2* (1)

Table 5.8: PsA fine-mapped SNPs from the 90% credible sets overlapping accessible chromatin identified by ATAC in four cell types. The number of SNPs in the 90% credible set union from each of the seven fine-mapped loci overlapping ATAC peaks from each cell type consensus list are reported. The number of SNPs only found to overlap open chromatin in one cell type are indicated together with the locus in which the SNP was fine-mapped.

only overlap accessible chromatin in one particular cell type. Interestingly, for the *TRAF3IP2* locus, which was fine-mapped with the greatest resolution, none of the two SNPs of the 90% credible set overlapped accessible chromatin in any of the four studied cell types.

Integration of epigenetic and functional data at the *RUNX3* locus

Another of the successfully fine-mapped loci was *RUNX3*, with the 90% credible set containing 56 SNPs (Table 5.7 and Figure 5.20). The integration with ATAC data from the four investigated cell types further narrowed down the number of putative causal variants to 6 SNPs. rs11249213 was one of the SNPs overlapping in $CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ cells a consensus ATAC peak annotated as an enhancer according to the corresponding chromatin segmentation maps of the three cell types. rs11249213 is located at an intergenic region, approximately 20Kb downstream the *RUNX3* gene and 1Kb away from Bowes GWAS lead SNP rs7523412.

In $mCD8^+$ the peak harbouring rs11249213 was identified as a DAR when using a threshold for significance of FDR<0.05, showing greater accessibility in

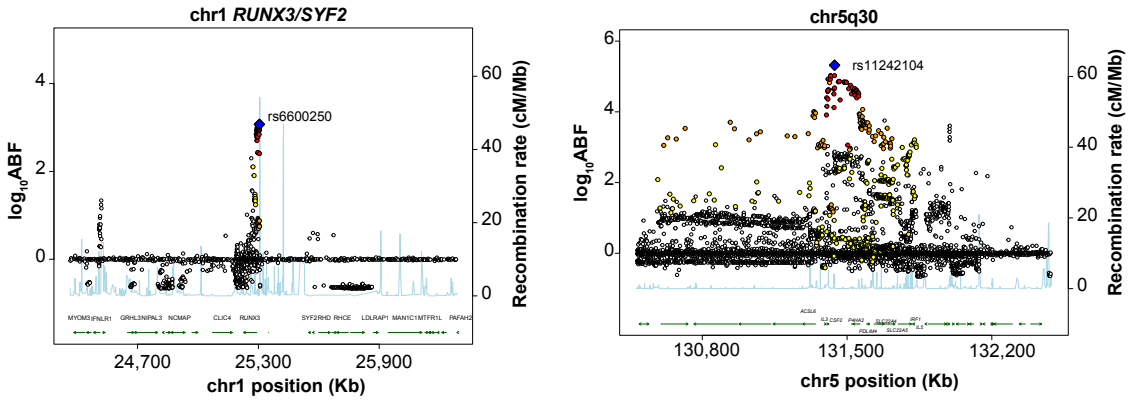


Figure 5.20: Association plot for the Bayesian fine-mapping results of the 5q31 (*CSF2/P4HA2*) and *RUNX3/SYF3* PsA GWAS signal. Association plots for the loci (A) 5q31 (*CSF2/P4HA2*) and (B) *RUNX3/SYF3*. For each of the SNPs (dot) the location (x-axis) and the $\log_{10}$BF (left y-axis) from the fine-mapping association analysis are shown. The colour of each dot indicated the LD relationship (r^2) with the lead SNPs (blue diamond), being white=low LD (r^2 <0.2), yellow=weak LD (r^2 <0.5), orange=moderate LD (r^2 <0.8) and red=high LD ($r^2 \geq 0.8$). The recombination rates in this window of the genome is indicated in the right y-axis (light blue line).

synovial fluid when compared to peripheral blood (fold change=1.7) (Figure 5.21 black box). In mCD4$^+$ the same ATAC peak was only found in cells from synovial fluid only in two of the patients (PsA1618 and PsA1619). In contrast, in CD14$^+$ monocytes all 3 PsA patients showed accessible chromatin at the location of the SNP in both tissues. No significant eQTL signal was found for rs11249213 in immune-relevant cell types, including unstimulated CD4$^+$ and CD8$^+$ cells or unstimulated and treated monocytes (LPS or IFN-γ). For the other five SNPs of the 90% credible set within ATAC peaks no overlap with significant eQTLs was found for any gene in the immune-related or GTEx datasets. Altogether, the observations based on in-house ATAC data may suggest a role of rs11249213 in regulating chromatin accessibility in the context of synovial inflammation only in mCD8$^+$ and mCD4$^+$ T cells. Further investigation will be required to determine whether rs11249213 has an specific effect on chromatin accessibility in the inflamed synovium.

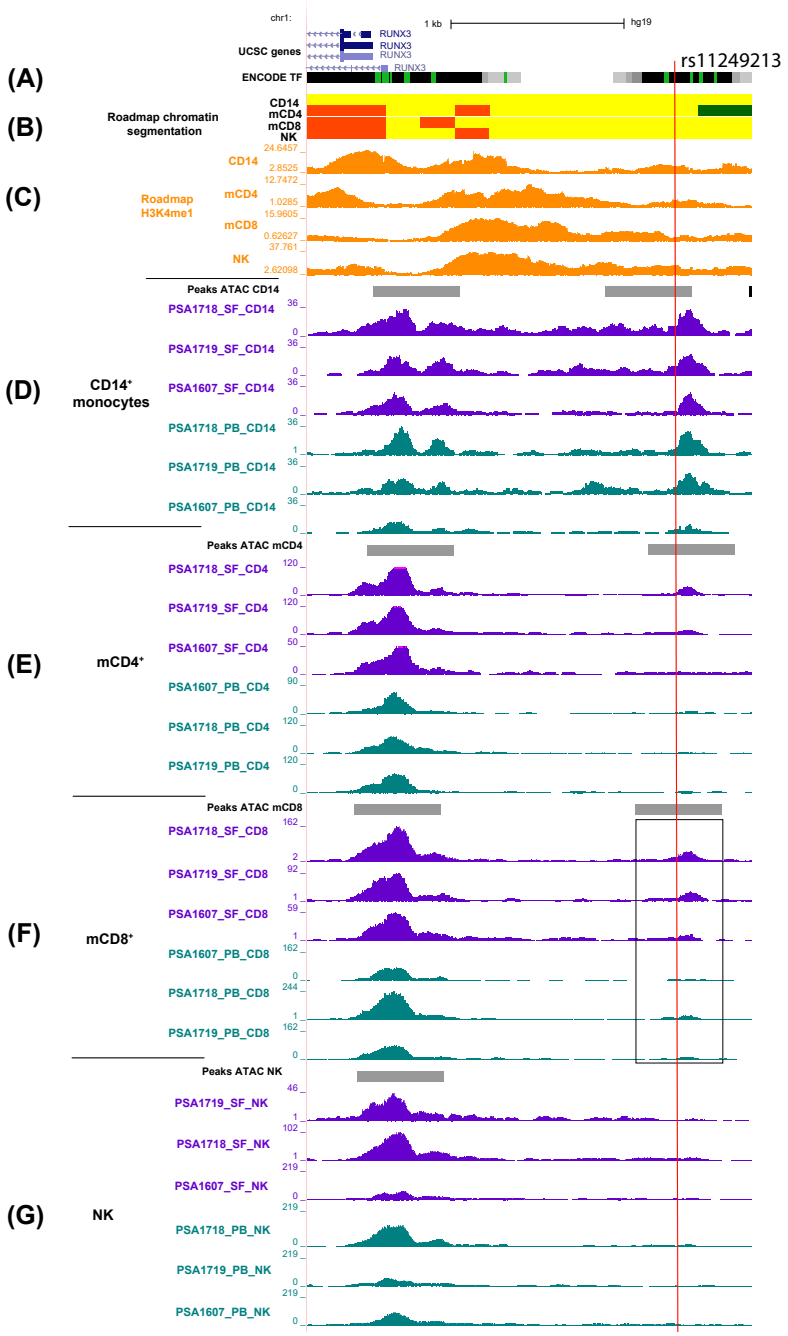


Figure 5.21: Epigenetic landscape at the genomic location of the fine-mapped SNP rs11249213 at the *RUNX3* PsA GWAS signal. UCSC visualisation of the epigenetic landscape at rs11249213 (red), within a mCD8$^+$ DAR (black box) when using a significance threshold of FDR<0.05. The track includes publicly available data for (A) ENCODE cluster for TF binding, (B) chromatin segmentation maps, (C) H3K4me1 (relative fold-enrichment) and in-house ATAC tracks (normalised counts) from 3 PsA patients for (D) CD14$^+$ monocytes, (E) mCD4$^+$, (F) mCD8$^+$ and (G) NK cells isolated from synovial fluid and peripheral blood.

Integration of epigenetic and functional data at 5q31 PsA-specific locus

Fine-mapping was successfully performed for the 5q31 locus nearby *CSF2/P4HA2* genes (Figure 5.20), the only PsA-specific GWAS association identified in the Immunochip GWAS study from Bowes and colleagues (Bowes *et al.* 2015a). The fine-mapping conducted here revealed rs11242104 as the lead SNP for this locus (Table 5.7, $\log_{10}$BF=5.31 and PP=0.07), differing from the fine-mapping lead SNP reported by Bowes and colleagues, likely due to the differences in the sample size.

Integration of the in-house chromatin accessibility data with the 58 SNPs in the 90% credible set showed only 8 variants overlapping ATAC peaks in at least one of the four cell types analysed in this chapter (Table 5.9 and Figure 5.22 top panel). Neither the lead SNP of this analysis or the lead fine-mapped SNP from Bowes and colleagues overlapped ATAC peaks. Nevertheless, three of the variants highlighted by Bowes and colleagues as functionally relevant based on ENCODE data (rs10065787, rs27437 and rs7721882) were also within an ATAC peak of at least one of the cell types isolated from PsA samples (Table 5.9).

Two relevant SNPs from the 90% credible set in this analysis were rs2069803 and rs4705908, both overlapping ATAC peaks in all four cell types (Figure 5.22 panel 1 black and brown lines, respectively). Both SNPs also overlapped enriched regions for H3Kme1 from the Roadmap Epigenomics Project for some of the same cell types (Figures 5.22 panel 1, B and D) and rs2069803 was within a region annotated as enhancer in mCD4$^+$, mCD8$^+$ and NK cells, according to chromatin segmentation maps (Figure 5.22 panel 1, C). Although both SNPs overlapped accessible chromatin in all four cell types, the most significant *cis*-eQTL signals was found for regulation of *SLC22A5* expression in CD4$^+$ or CD8$^+$ T cells (Table 5.9), in line with Bowes and colleagues finding in their pilot eQTL study). Promoter capture Hi-C data in naïve and total

SNP	Cell type ATAC overlap	Top eGene	Cell type (adj p-value)
rs10065787	$CD14^+$, $mCD4^+$	*P4HA2*	$CD14^+$ LPS2 ($2.4x10^{-15}$), $CD14^+$ LPS24 ($5.9x10^{-23}$) $CD14^+$ IFN-γ ($4.1x10^{-20}$)
		SLC22A5	$CD14^+$ UT ($9.3x10^{-33}$) $CD14^+$ IFN-γ ($6.6x10^{-26}$) $CD4^+$ ($1.6x10^{-3}$) $CD8^+$ ($8.0x10^{-3}$)
		SLC22A4	$CD14^+$ UT ($2.3x10^{-11}$)
rs11242104	All	NA	
rs11242105	All	NA	
rs2069803	All	*SLC22A5*	$CD4^+$ ($<2.2x10^{-308}$) $CD8^+$ ($4.5x10^{-4}$)
rs27437	$CD14^+$, $mCD4^+$	*SLC22A5*	$CD4^+$ ($1.9x10^{-3}$) $CD8^+$ ($2.6x10^{-2}$)
rs4705908	All	*SLC22A5*	$CD4^+$ ($<2.2x10^{-308}$) $CD8^+$ ($8.7x10^{-4}$)
rs2089855	All	*P4HA2*	$CD14^+$ LPS2 ($3.5x10^{-25}$) $CD14^+$ LPS24 ($1.1x10^{-34}$) $CD14^+$ IFN-γ ($3.1x10^{-34}$)
		SLC22A5	$CD14^+$ UT ($8.8x10^{-51}$) $CD14^+$ IFN-γ ($4.1x10^{-36}$) $CD4^+$ ($<2.2x10^{-308}$) $CD8^+$ ($<2.2x10^{-308}$)
rs7721882	$mCD4^+$	*SLC22A5*	$CD4^+$ ($<2.2x10^{-308}$) $CD8^+$ ($1.5x10^{-4}$)

Table 5.9: Publicly available and unpublished *cis*-eQTL datasets reporting an effect for the PsA 5q31 GWAS locus fine-mapped SNPs (90% credible set) overlapping ATAC accessible regions. For each of the SNPs, the cell type for the ATAC overlap, the gene which expression is reported to be regulated by the SNP (eGene) and the cell type where the eQTL study was conducted are specified. The eQTLs datasets included in the analysis were $CD14^+$ monocytes (UT, LPS 2h, LPS 24h, IFN-γ) (Fairfax *et al.* 2014), B cells (Fairfax *et al.* 2012), total $CD4^+$ and total $CD8^+$ (Kasela *et al.* 2017). adj p-value refers to the corrected p-value for multiple testing reported by each of the studies.

$CD4^+$ $CD8^+$ cells also revealed interaction of the *IL3* promoter bait containing rs2069803 with the promoter of *SLC22A5*.

rs2089855 was another of the fine-mapped SNPs overlapping ATAC in all cell types showing extremely significant eQTL effect in regulating *SLC22A5* expression in $CD4^+$ and $CD8^+$ cells (Table 5.9). This SNP is proximal to rs11955347, reported by Bowes *et al.* as the most significant SNP correlating with expression of *SLC22A5* in $CD4^+$ and $CD8^+$ in a pilot study (Figure 5.22A in green). rs2089855 and rs11955347 were also *cis*-eQTL SNPs for *SLC22A5* and *P4HA2* expression in unstimulated and stimulated monocytes (LPS or IFN-γ), respectively (Table 5.9). Similarly, rs10065787 mapped to an ATAC peak in $CD14^+$ monocytes and $mCD4^+$ cells with moderate enrichment for the enhancer histone mark H3K4me1 in $mCD4^+$ (Figure 5.22 panel 2 red line, D, F, and G) and showed an eQTL effect for *SLC22A5* and *P4HA2* expression in unstimulated and stimulated monocytes (LPS2, LPS24, IFN-γ), respectively (Table 5.9). Chromatin conformation data using promoter capture HiC (Javierre *et al.* 2016) in unstimulated monocytes does not clearly reveal interaction for rs10065787 with the promoter of *P4HA2*, which may be only found under stimulation.

Altogether, the integration of in-house and publicly available epigenetic and functional data with the SNPs from the 90% credible set obtained by fine-mapping suggests a role for the 5q31 PsA-specific GWAS association in regulating expression of *P4HA2* and *SLC22A5*, not only in T cells but also in monocytes under different pro-inflammatory conditions.

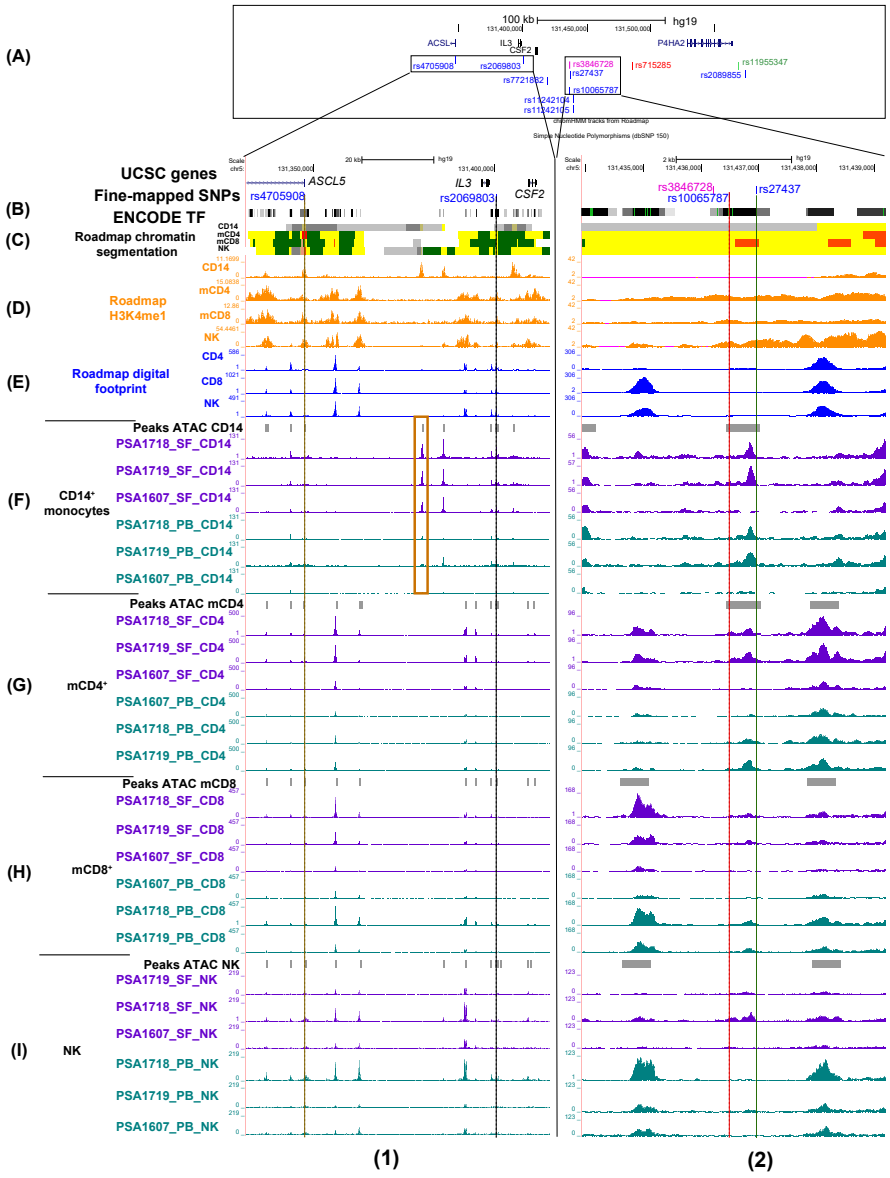


Figure 5.22: Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal. (A) Location of the 6 SNPs from the 5q31 locus 90% credible set overlapping ATAC peaks from the consensus list of at least one of the four cell types (blue). Relevant SNPs Bowes *et al.* study, including the GWAS lead SNP (red), eQTL SNP showing the best correlation with the GWAS lead SNP (green) and one credible set SNP overlapping several ENCODE annotation features but not PsA ATAC data (purple). UCSC visualisation of the epigenetic landscape for rs4705908 (brown) and rs2069803 (black) in panel (1) and rs1006587 (red) and rs27437 (green) panel (2) representing 4 relevant fine-mapped SNPs overlapping PsA ATAC peaks and eQTLs signals. Also shown are publicly available data for (B) ENCODE cluster for TF binding, (C) chromatin segmentation maps, (D) H3K4me1 (relative fold-enrichment), (E) digital footprint and in-house ATAC tracks (normalised counts) from 3 PsA patients for (F) $CD14^+$ monocytes, (G) $mCD4^+$, (H) $mCD8^+$ and (I) NK cells isolated from synovial fluid and peripheral blood. The yellow box highlights a DAR found in $CD14^+$ monocytes near rs2069803 and rs4705908.

5.4 Discussion

5.4.1 Characterising chromatin accessibility in PsA samples

The epigenetic landscape of immune cells isolated from PsA patients peripheral blood and synovial fluid has been characterised and compared in this chapter, representing the first study to date of chromatin accessibility from patients samples in this disease. The interrogated cell types ($CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK) are key players in the innate and adaptive immune responses dysregulated in PsA pathophysiology. In particular, expansion of $mCD4^+$ and $mCD8^+$ cells in the synovium of PsA patients has been previously reported, and GWAS have highlighted significant associations within or nearby genes involved in T cell biology (Taams 2018).

Differential chromatin accessibility analysis in this chapter showed that in all four cell types the majority of DARs were located at intergenic and intronic regions annotated as enhancers and showed enrichment for relevant eRNAs identified in the FANTOM project (Forrest *et al.* 2014). Notably, $CD14^+$ monocytes showed the largest number of DARs between synovial fluid and peripheral blood, with 9.1% of the interrogated ATAC regions being differentially accessible. Pathway enrichment analysis using genes proximal to DARs revealed differences between synovial fluid and peripheral blood in all four cell types. In synovial fluid, enrichment for the NF-κB pathway was found in $CD14^+$ monocytes, consistent with the TNF-α signature and the efficacy of anti-TNF therapies in PsA (Mahil *et al.* 2016). Interestingly, open chromatin in peripheral blood $mCD8^+$ T cells was enriched for Wnt signaling pathway, which is involved in many biological processes and its dysregulation is associated with a number of autoimmune diseases including RA (Miao *et al.* 2013). In this data, the significantly increased chromatin accessibility near Wnt signalling genes in peripheral blood $mCD8^+$ cells may be related to an increased recall proliferation

capacity of the circulating memory cells compared to the tissue-resident ones (Boudousquié *et al.* 2014). In NK cells from peripheral blood increased chromatin accessibility was found at genes involved in NK-mediated cytotoxicity. NK $CD56^{bright}$ cells resident at sites of inflammation are more specialised in cytokine production than in cytotoxic response (Michel *et al.* 2016). Notably, flow-cytometry data in these patients have shown a reduced proportion of $CD56^{bright}$ cells in peripheral blood compared to synovial fluid, which would be consistent with this enrichment of NK peripheral blood open DARs for genes involved in the cytotoxic response.

Overall, this approach has identified robust differences in chromatin accessibility between synovial fluid and peripheral blood in PsA patients, in line with similar studies comparing patients and controls conducted in other diseases (Scharer *et al.* 2016; Wang *et al.* 2018; Corces *et al.* 2016). Although these findings suggest meaningful functional differences between circulating and affected tissue immune cells, a limitation of pathway analysis in ATAC is linking open chromatin to genes. Annotation of ATAC peaks to nearby genes has been extensively used in the literature (Scharer *et al.* 2016; Ackermann *et al.* 2016; Corces *et al.* 2016; Wang *et al.* 2018); however, this approach fails to evaluate long range regulatory effects and assumes that accessible chromatin regulates neighbouring regions.

5.4.2 Bulk qPCR expression profiling and integration with paired chromatin accessibility data

Recent reports have investigated differences in the transcriptional profiles between PBMCs, bulk T cells, SFMCs and synovial tissue from PsA patients (Dolcino *et al.* 2015; Fiocco *et al.* 2015), but a comparative analysis in matched discrete cell subpopulations from synovial fluid and peripheral blood of the same patient was lacking. The pilot study presented in this chapter has characterised

the expression of relevant immune genes in $CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ cells isolated from synovial fluid and peripheral blood.

In this study, $CD14^+$ monocytes and $mCD8^+$ showed the largest number of statistically significant DEGs between synovial fluid and peripheral blood. Integration of transcriptome profiling with paired ATAC data revealed that a small number of the DEGs between synovial fluid and peripheral blood were nearby DARs showing changes in the same direction.

Pro-inflammatory genes, including *TNFA*, *CXCL13* or *CCL18*, were found to be up-regulated in synovial fluid compared to peripheral blood in at least one cell type. These genes were previously reported by Dolcino and colleagues as DEGs in synovial membranes of PsA patients compared to controls (Dolcino *et al.* 2015), suggesting the up-regulation of those genes to be a feature of PsA joints. Genes such as *CXCL10*, *CCL7*, and *CCL17* were only up-regulated in $CD14^+$ monocytes, consistent with their role in the migration of monocytes from bone marrow to tissues (Tsou *et al.* 2007). In contrast, genes involved in Th-17 cell response such as *CXCL13* and *IL26* (Takagi *et al.* 2008) only showed up-regulation in synovial fluid $mCD4^+$ and/or $mCD8^+$ when compared to peripheral blood.

A preliminary analysis comparing DEGs between PsA patients and controls in peripheral blood with the genes differentially expressed between synovial fluid and peripheral blood in patients identified a number of systemic, tissue-specific and putative disease specific DEGs. In $CD14^+$ monocytes, a much larger number of immune-relevant genes were differentially expressed between the two tissues from patients than between PsA patients and healthy controls in peripheral blood. This is likely due to the plasticity of monocytes and their ability to differentiate into macrophages at the inflamed synovium (Yoon *et al.* 2014; Park *et al.* 2016). A particularly interesting systemic DEG in $mCD8^+$ cells was *CCR10*, a chemokine receptor co-expressed by a subset of memory cells that preferentially migrate to skin, which was also up-regulated in $CD8^+$

cells in the psoriasis cohort (Chapter 4) and in patients with atopic dermatitis (Hijnen and Nijhuis 2005). In mCD4$^+$, *GPR68*, a G protein-coupled receptor (GPCR) was up-regulated in peripheral blood of PsA patients when compared to controls with further increased expression in synovial fluid. *GPR68* undergoes activation through pH acidification, which is a characteristic feature of synovial tissue (Biniecka *et al.* 2016; Saxena *et al.* 2012). Another member of the acid-sensing GPCR family, *GPR65*, has been associated with other immune mediated diseases, including AS, CD and MS, and is a marker of pathogenic Th-17 cells (Cortes *et al.* 2013; Lassen *et al.* 2016; Wirasinha *et al.* 2018; Gaublomme *et al.* 2015; Al-Mossawi *et al.* 2017).

SPP1 and *FN1* were the most significantly up-regulated genes in synovial fluid compared to peripheral blood of PsA patients, in line with previous findings (Dolcino *et al.* 2015) and interestingly no significant changes in peripheral blood between PsA patients and controls was observed. However, Dolcino and colleagues did detect a nore moderate up-regulation of *SPP1* in peripheral blood in PsA patients when compared to controls. *FN1* encodes fibronectin-1, a main component of the cartilage matrix, involved in cell adhesion, migration, growth and differentiation and found to be highly expressed in RA inflamed synovium (Chang *et al.* 2005). *FN1* has also shown to induce bone resorption mediated by pro-inflammatory mediators, such as nitric oxide and IL-1β (Gramoun *et al.* 2010). Consistent with the up-regulation in synovial fluid CD14$^+$ monocytes, synovial fluid open DARs were found at the promoter and downstream of *FN1* in CD14$^+$ monocytes only.

Pathway analysis using significant qPCR DEGs in PsA between synovial fluid and peripheral blood revealed enrichment for chemokine, TLR and NOD-like signalling pathways in CD14$^+$ monocytes. This was consistent with the well-known role of TLR and NOD-like receptors in rheumatic diseases (McCormack *et al.* 2009). NOD-like signalling has also been highlighted as one of the

dysregulated pathways in the genome-wide trancriptomic analysis between lesional and uninvolved psoriatic skin shown in Chapter 4. The significanr down-regulation of the TLR inhibitor *SIGIRR* (Ueno-Shuto *et al.* 2014) and up-regulation of *TLR1*, *TLR2*, *MYD88* detected by qPCR array may reflect the increased activation of NFκB TF in synovial fluid CD14$^+$ monocytes. This is consistent with the significant reduction of NF-κB activation upon etanercept treatment in PsA synovial biopsies (Lories *et al.* 2008). The enrichment of synovial fluid open DARs in the proximity of genes within the NF-κB pathway is in line with the transcriptional up-regulation of downstream genes such as *TNFA*, *CCL2* and *CCL5* and subsequent enrichment of chemokine signalling in this cell type. In mCD4$^+$, enrichment of the IL-10 signalling pathway was interesting in the context of the expansion of Tregs in synovial fluid observed by flow cytometry analysis in the same samples (data not shown). Tregs are characterised by the expression of anti-inflammatory cytokines, importantly IL-10 (O'Garra *et al.* 2004) and qPCR data showed a significant increase of the IL-10 receptor subunit α *IL10RA* and a similar trend for *IL10* in synovial fluid mCD4$^+$ and mCD8$^+$ compared to peripheral blood. This may suggest that inflammation in PsA is refractory to the immunomodulatory effects of IL-10 in synovial fluid or counterbalanced by its immunostimulatory properties, in line with failure of IL-10 agonist therapy in the treatment of CD and psoriasis (Marlow *et al.* 2013; Kimball *et al.* 2002). Interestingly, a study in macrophages from RA and SpA patients showed that IL-10 only retains the ability to activate a subset of IL-10 induced genes shared with IFN-γ through activation of *STAT1*.

5.4.3 The relevance of monocytes in the PsA synovial fluid

In this exploratory study, CD14$^+$ monocytes demonstrated the largest number of differences in chromatin accessibility and gene expression between synovial fluid and peripheral blood, with enrichment for disease relevant

functional pathways. The work presented in this chapter is part of a collaborative multi-omics pilot study and in the light of chromatin accessibility and qPCR results for the cohort size available at the time of analysis, $CD14^+$ monocytes were chosen as the cell type to further explore by scRNA-seq and integrate with mass cytometry data. $mCD4^+$ and $mCD8^+$ also showed relevant changes in chromatin accessibility and gene expression, and have been identified as the two cell types undergoing greater expansion in the synovium of PsA patients. Colleagues have since performed differential TCR clonality analysis between synovial fluid and peripheral blood using the same three PsA samples processed for scRNA-seq in this chapter. Interestingly, a significant number of $mCD8^+$ TCR clones with potentially shared antigen recognition across patients have been found to be enriched in synovial fluid compared to peripheral blood (manuscript in preparation) and future work will be conducted to further integrate it with changes in chromatin accessibility.

5.4.4 Characterisation of synovial and peripheral blood $CD14^+$ monocytes by scRNA-seq

Monocytes are a very plastic cell population known for being precursors of DCs, macrophages and osteoclasts and importantly undergo phenotypic changes at the site of inflammation (Qu *et al.* 2014; Rivollier *et al.* 2012). Therefore, interrogation at the single-cell level has the potential to reveal distinct functionally relevant subpopulations. In addition, scRNA-seq was also used to perform differential gene expression of synovial fluid vs peripheral blood complementary to the qPCR array and validate and/or unveil additional relevant pathways dysregulated between the two tissues.

Combined scRNA-seq analysis of PsA $CD14^+$ monocytes from synovial fluid and peripheral blood unsurprisingly showed that the largest differences were driven by the tissue of origin. Synovial fluid monocytes appeared as more

intermediate monocytes expressing *CD14* and *CD16* in contrast to the peripheral blood cells showing a more classical phenotype lacking expression of *CD16*. This was consistent with mass cytometry data showing expansion of $CD14^+$ $CD16^+$ monocytes in the synovium of the same ten patients presented in Section 5.3.8 and is consistent with the more pro-inflammatory phenotype characteristic of intermediate monocytes and the more accute pro-inflammatory environment at the joints (Gren *et al.* 2015). Expansion of intermediate monocytes has also been seen in the synovium of RA patients as well as in blood from PsA patients and other inflammatory diseases such as sepsis and SLE when compared to controls (Chiu *et al.* 2010; Yoon *et al.* 2014; Mukherjee *et al.* 2015).

The eight subpopulations (MC-1 to 8) of monocytes identified by cluster analysis were mostly formed by cells from one tissue, being MC-7 the only one with a mixed contribution, reinforcing the large variability across monocytes from the two tissues.

Of particular interest was the MC-4 subpopulation of synovial fluid monocytes, comprising 7.2% of the monocytes in this tissue and showing distinctive expression of *IL7R*, *IL32* and *CCL5*. The MC-7 cluster also expressed a number of markers characteristic of the MC-4 subpopulation, but only included 2.7% of the peripheral blood monocytes. This is consistent with increased gene expression and chromatin accessibility at the *IL7R* locus in synovial fluid monocytes of PsA patients also observed in this thesis. Similarly, flow cytometry analysis conducted by Al-Mossawi and colleagues showed up to 35% of the total synovial fluid $CD14^+$ monocytes being IL-$7R^+$ in PsA and AS patients vs approximately the 1% in peripheral blood (Al-Mossawi*et al.*, 2018, in revision).

The IL-$7R^+$ monocyte subpopulation, predominantly found in synovial fluid, may be of functional relevance since *IL7R* polymorphisms are associated with a number of chronic inflammatory diseases, including AS and MS (Gregory *et al.* 2007; Cortes and Brown 2011). Despite the main role for IL-7 and IL7R

having been described in maintenance of homeostasis and survival of naïve and memory T cells, the expression of *IL7R* and production of soluble IL7R (sIL7R) is genetically regulated by the MS risk allele (rs6897932) under LPS and TNF-α stimulation in monocytes but not T cells (Kondrack *et al.* 2003; Fairfax *et al.* 2014; Al-Mossawi *et al.* 2018). These findings presented by Al-Mossawi and colleagues suggests that the increased percentage of $CD14^+$ $IL7R^+$ monocyte in synovial fluid is the result of expansion of this population upon inflammatory stimuli in the joints. Interestingly, the synovial fluid open DAR found at the 3' UTR of the *IL7R* gene is nearby this risk variant and is a tissue and monocyte specific peak. *In vitro* work conducted by Al-Mossawi and colleagues has also demonstrated that $IL7R^+$ monocytes are transcriptionally responsive to IL-7 and the main source of sIL7R, which may suggest a role of monocytes in prolonging IL-7 response in T cells (Lundström *et al.* 2013). Furthermore, the functional relevance of $CD14^+$ $IL7R^+$ monocyte could be related to a role for IL-7 in osteoclastogenesis previously described in the literature through the production of soluble osteoclastogenic cytokines by T cells and potentially increased ability of this subpopulation to either undergo differentiation or enhance osteoclastogenesis (Weitzmann *et al.* 2000). Nevertheless, additional work will be needed to fully characterise this monocyte subpopulation.

Amongst the other subpopulations, MC-1 was most functionally distinct and was characterised by enrichment of uniquely marker genes involved in IFN signalling, viral response, antigen presentation and chemokine signalling. A recent study in RA has also identified a subset of monocytes expressing *IFI6*, amongst other markers, and showing enrichment for IFN signalling and the IFN-γ mediated pathways (Zhang *et al.* 2018). Moreover, synovial fluid monocytes, and particularly those from the MC-1 and MC-2 subpopulations, also showed increased expression for MHC-II *HLA-DR*, *HLA-DP* and *HLA-DQ* contributing to enrichment of the MHC-II antigen presentation pathway between tissues.

Increased expression of MHC-II genes is characteristic of the pro-inflammatory intermediate and non-classical ($CD14^{-}$ $CD16^{++}$) monocytes with a role in T cell recruitment and activation through antigen presentation (**Grage2001**).

The enrichment of synovial fluid vs peripheral blood DEGs for the chemokine signalling pathway by scRNA-seq not only recapitulated the qPCR results but also reinforced the increased chemotaxis and T cell activation driven by IFN-induced up-regulation of *CCL3*, *CCL2* and *CXCL10* in synovial fluid monocytes (Baba and Mukaida 2014). Altogether, this suggests that the inflammatory cytokine milieu in the joints leads to an enhanced IFN activated phenotype of synovial monocytes, importantly of the MC-1 subpopulation, with functional relevance in chemotaxis, antigen presentation and T cell activation. This has been demonstrated in RA, where intermediate synovial fluid monocytes showed ability to promote Th-1 and Th-17 pathogenic response in autologous $mCD4^{+}$ T cells (Yoon *et al.* 2014).

The pathway enrichment analysis conducted using the scRNA-seq data also recapitulated dysregulation of the NOD-like receptor signalling pathway. Surprisingly, the receptor *NLRP1* and *CASP1*, key genes for inflammasome activation were down-regulated in synovial fluid monocytes but no significant down-regulation of *IL-1B* has been identified by scRNA-seq when compared to peripheral blood $CD14^{+}$ monocytes. Interestingly, a recent study has shown that type I Interferon can inhibit NLRP1-mediated inflammasome activation via STAT1 (which is up-regulated in synovial fluid monocytes) and may be a feedback loop resulting from exacerbated inflammation in the joints (Guarda *et al.* 2011). Additional pathways to the ones identified by qPCR were also unveiled by this analysis. For example, up-regulation of inhibitors and down-regulation of positive regulators of the apoptosis pathway may suggest resistance to apoptosis by synovial fluid monocytes. This is consistent with a functional study in RA and PsA synovial fluid monocytes showing resistance to the extrinsic and

intrinsic apoptosis pathway *in vitro* which suggested that monocytes in inflamed synovium may be more prone to macrophage differentiation (Rajasekhar *et al.* 2017; Srivastava *et al.* 2010).

DEGs between synovial fluid and peripheral blood monocytes also highlighted enrichment for the osteoclast differentiation pathway, with a significant up-regulation in synovial fluid for the RANKL receptor *TNFRSF11A*, a key gene in osteoclast differentiation from monocyte precursors (Mensah *et al.* 2008). In PsA, osteoclast differentiation has been found to occur preferentially from intermediate and non-classical monocytes, with *CD16* expression being a marker of increased bone erosion activity in patients (Chiu *et al.* 2010). In line with these previous studies, my data supports the expansion in the synovium of monocytes with an "intermediate-like" transcriptomic phenotype and up-regulation of key genes involved in osteoclast differentiation, thus likely to be contributing to pathological altered bone remodelling.

Although two of the dysregulated pathways highlighted by qPCR gene expression analysis were also recapitulated by the scRNA-seq differential analysis, the overlap of DEGs when comparing synovial fluid vs peripheral blood monocytes by qPCR and scRNA-seq was only moderate. This could be caused by differences in the way of selecting monocytes (FACS based on $CD14^+$ for qPCR vs *CD14* expression in scRNA-seq), the sensitivity of the two techniques and patient heterogeneity in the two assays (Table 5.1), requiring validation in a larger cohort.

5.4.5 Integrating mass cytometry data with chromatin accessibility and gene expression changes in PsA $CD14^+$ monocytes

Single-cell mass cytometry data allowed investigation of expression of inflammatory mediators in synovial fluid and peripheral blood $CD14^+$ monocytes in a total of ten PsA patients revealing an increased percentage of monocytes producing TNF-α, MCP-1, osteopontin and IP-10 in synovial fluid.

This was consistent with the transcriptional up-regulation of these four genes detected by qPCR and/or scRNA-seq in PsA $CD14^+$ monocytes synovial fluid compared to peripheral blood. Increased production of TNF-α in PsA synovial fluid compared to osteoarthritis has been previously reported, consistent with the therapeutic effect of TNFi (Partsch *et al.* 1997).

Interestingly, osteopontin, MCP-1 and IP-10 can be induced in monocytes upon IFN stimulation, in line with the enrichment of this pathways between synovial and peripheral blood cells as well as the functional biological processes characterising the MC-1 subpopulation (Li *et al.* 2003; Kopydlowski *et al.* 1999). Osteopontin is involved in promoting cell migration, adhesion and cell-mediated immune response through Th-17 differentiation and has been demonstrated to have a role in a number of chronic inflammatory and autoimmune diseases (Morimoto *et al.* 2010; Rittling and Singh 2015). In this pilot cohort, the osteopontin coding gene *SPP1* has already been highlighted as one of the top two DEGs between synovial fluid and peripheral blood, particularly in $CD14^+$ monocytes. *CCL2*/MCP-1 and *CXCL10*/IP-10 are two examples of pro-inflammatory molecules whose differences between synovial fluid and peripheral blood were consistent in $CD14^+$ monocytes for chromatin accessibility, transcriptomic and protein expression levels. Up-regulation of these two genes in synovial fluid monocytes could be driven by the simultanous up-regulation of *STAT1* found by scRNA-seq differential analysis, since the promoters of both genes contain binding sites for this TF (Rauch *et al.* 2013). Dolcino's study did not identify up-regulation of *CCL2* or *CXCL10* expression in PsA synovial membranes when compared to controls, likely due to the mixed composition of cells in this tissue. The open synovial fluid DARs upstream *CCL2* and *CXCL10* in $CD14^+$ monocytes highlighted changes in chromatin accessibility that are likely driven by the pro-inflammatory environment in the joint. These two differential accessible sites were not found to interact with distal promoters

in unstimulated monocytes and, given the proximity to each of the genes, both regions may exert a local regulatory effect in the expression of *CCL2* and *CXCL10*, respectively (Javierre *et al.* 2016). Interestingly, increased secretion of these two chemokines has been observed in anti-apoptotic monocytes upon miR-155 overexpression and their up-regulation in PsA synovial fluid $CD14^+$ monocytes may be related to the aforementioned anti-apoptotic transcriptional profile when compared to the peripheral blood cells.

In terms of pathophysiology, MCP-1 is one of the key cytokines mainly involved in monocytes migration from peripheral blood into inflamed tissues and also in chemotaxis of T cells to the site of inflammation (Tsou *et al.* 2007; Shadidi *et al.* 2003). Up-regulation of MCP-1 expression in plasma compared to controls has previously been demonstrated and shown to correlate with levels of infiltrated T cells in the synovium of PsA patients (Ross *et al.* 2000). Nevertheless, correlation of chromatin accessibility, gene expression and protein production of *CCL2*/MCP-1 in PsA $CD14^+$ monocytes had not been investigated. The transcriptional up-regulation of the receptor of MCP-1, *CCR2*, in PsA synovial fluid $mCD8^+$ vs peripheral blood in the same samples highlights a biological role of *CCL2-CCR2* in mediating T cell infiltration in the synovium and the expansion of $mCD8^+$ T cells in this tissue. *CXCL10* is also a pro-inflammatory chemokine secreted by several immune and non-immune related cells and participates into the recruitment of T cells and monocytes to sites of inflammatio (Antonelli *et al.* 2014). Significant up-regulation of *CXCL10* by qPCR and scRNA-seq was found when comparing $CD14^+$ monocytes between synovial fluid and peripheral blood whereas it was down-regulated in PsA peripheral blood vs healthy controls. A recent conference abstract studying immune cell populations isolated from PsA has also demonstrated greater expression levels of *CXCL10* in monocytes compared to T cells (Muntyanu *et al.* 2017). Interestingly, mass cytometry data from $mCD4^+$ and $mCD8^+$ cells of the same ten PsA patients did not

show significant differences in the percentage of cells producing this chemokine between synovial fluid and peripheral blood, reinforcing the $CD14^+$ monocyte specificity of *CXCL10* dysregulation between the two tissues in PsA.

5.4.6 Use of PsA functional data to inform GWAS fine-mapping

Integration of epigenetic data with fine-mapped SNPs from GWAS studies has been widely demonstrated as a powerful tool to further narrow down candidate causal variants, particularly for intergenic or intronic signals not driven by LD with missense coding SNPs (Bunt *et al.* 2015; Farh *et al.* 2015). Similarly to the Immunochip psoriasis GWAS fine-mapping presented in Chapter 4, fine-mapping ($\log_{10}$BF>3) was only achieved for a subset of the loci successfully fine-mapped by Bowes and colleagues due to reduced sample size. Moreover, only the *TNFAIP3* locus, which presented the largest OR (1.71), was refined to fewer than 10 SNPs in the 90% credible set in line with the previously commented dependency between ability to narrow down the number of SNPs in the credible set and the effect size of the association (Bunt *et al.* 2015).

DARs identified when comparing synovial fluid vs peripheral blood did not show any overlap with SNPs from the credible sets, which could be partly due to DARs not reaching significance at the chosen FDR threshold in this cohort size. For example, rs11249213 at the *RUNX3/SYF2* locus overlapped a DAR only significant in $mCD8^+$ when using FDR<0.05 and showing the same trend in $mCD4^+$. Similar results regarding differential chromatin accessibility have been found by colleagues in the group comparing chromatin accessibility at this location in unstimulated and activated total $CD4^+$. This, together with the absence of eQTL signal for this SNP in unstimulated T cells may indicate a putative role for this SNP in regulating gene expression, potentially *RUNX3*, only in T cells in the context of inflammation. *RUNX3* is one of the key TFs in $CD8^+$ T cells development and activation, and its role in chromatin remodelling and

$CD8^+$ differentiation into memory cells upon TCR stimulation has been recently demonstrated (Woolf *et al.* 2003; Lotem *et al.* 2013; Wang *et al.* 2018). The *RUNX3* locus is also associated with AS and celiac disease risk, where GWAS lead SNPs are in high LD (r^2=0.99) with the PsA GWAS lead SNP (rs7523412) and also with rs11249213 (Reveille *et al.* 2010; Dubois *et al.* 2010). RNA-seq data in AS patients has not revealed differential expression for *RUNX3* when compared to controls (unpublished). Interestingly, Apel and colleagues identified a PsA genome-wide significant association for rs4649038, in an intron of *RUNX3*, independent from AS and CD GWAS lead SNP ($r^2 \leq 0.5$) and not significantly associated in psoriasis vulgaris (Apel *et al.* 2013), suggesting different signals with independent contribution to PsA risk near *RUNX3*.

Integration of the ATAC data generated in this thesis at the 5q31 PsA-specific signal reduced the 58 SNPs of the 90% credible set to eight variants, all at open chromatin in T cells from PsA patients and four showing highly significant eQTL effects on *SLC22A5* T cell expression (Kasela *et al.* 2017). This was consistent with the role of 5q31 association in regulating *SLC22A5* expression in total $CD4^+$ and $CD8^+$ cells suggested by Bowes and colleagues in a pilot eQTL study (Bowes *et al.* 2015a). *SLC22A5* is a cell membrane transporter of carnitine involved in fatty acids metabolism prioritised by an in-house pipeline as the third most promising druggable candidate for the treatment of psoriasis (Fang *et al.* 2019, in revision). *SLC22A5* has already been suggested as a potential druggable therapeutic target in asthma (Mo *et al.* 2011). Data integration also suggested a role of fine-mapped SNPs overlapping PsA $CD14^+$ monocytes accessible chromatin in regulating *P4HA2* only under pro-inflammatory stimuli (Fairfax *et al.* 2014). Silencing of the prolyl-4 hydroxylase-2 *P4HA2* in a myocardial infarction mouse model has shown reduced inflammation and production of MCP-1, whose role in PsA synovial $CD14^+$ monocytes has also been highlighted in this thesis (Natarajan *et al.* 2007). Here, scRNA-seq

differential analysis has not revealed differential expression for *SLC22A5* or *P4HA2* between synovial fluid and peripheral blood monocytes but found up-regulation of the nearby gene *IRF1*, an IFN-dependent TF that plays an important role in the amplification of myeloid cell response to IFN-γ (Hu and Ivashkiv 2009). Although eQTL data in monocytes showed only modest significance for rs10065787 and rs2089855 in regulating *IRF1* expression in unstimulated and LPS stimulated cells, this gene could be a plausible target of the 5q31 PsA locus in this cell type.

5.4.7 Limitations of the study

The work presented in this chapter is an exploratory study and a proof of principle for the implementation of a multi-omics approach in PsA clinical samples from affected joints, which is a very powerful strategy to dissect disease pathophysiology in a cell type and context specific manner. Nevertheless, a number of limitations and challenges need to be taken into account to contextualise these results. The main limitations included small sample size, lack of paired data and peripheral blood from healthy individuals across all techniques and also an absence of synovial fluid from another autoimmune or non-inflammatory joint disease as used by others as an additional control (Mizoguchi *et al.* 2018; Dolcino *et al.* 2015; Zhang *et al.* 2018). This would help to pinpoint changes in chromatin accessibility, transcription and protein expression that are specific to the PsA inflamed synovium. The use of RNA-seq instead of the biased qPCR array would have allowed a genome-wide overview of bulk transcriptomic differences between tissues in the same samples interrogated by ATAC and a more comprehensive pathway enrichment analysis. Additionally, RNA-seq could have also been integrated with the scRNA-seq analysis conducted in monocytes and used to validate this analysis, as it has been shown by others (Zhang *et al.* 2018).

Another challenge in this study relates to the analysis and integration of scRNA-seq and mass cytometry data. Both techniques are still emerging fields where there is no "gold" standard to combine samples across patients to account for batch effect or in downstream data analysis. In this study imputation of expression to reduce the number of drop outs has been conducted to compensate for the limited sample size. However, these imputation methods may introduce bias and may be avoided if a larger cohorts is available (Andrews 2018). Cluster identification is another aspect that lacks widely accepted analysis strategy and the results can be affected by the resolution parameters and the algorithm of choice as well as the sample size (Kiselev *et al.* 2019). This work represents an attempt to identify potential relevant subpopulations of monocytes in synovial fluid from PsA patients. Nevertheless, the identification of robust and stable subpopulations through cluster analysis would benefit from the implementation of algorithms designed for cluster validation such as Silhouettes, which has recently been successfully applied to the field of single-cell analysis (Rousseeuw 1987; Zhang *et al.* 2018). Similarly, different methodologies for cluster identification, annotation and cytokines quantification by mass cytometry are available and additional data analysis is in progress to identify distinct subpopulations of monocytes based on cytokine and chemokine expression.

Finally, fine-mapping analysis presented in this chapter was limited by the sample size, explaining the inability to fine-map a number of loci reported by Bowes and colleagues when using the entire PsA Immunochip cohort.

5.5 Conclusions

The strategy and the analysis presented in this chapter illustrate how a multi-omics approach can be applied to clinical samples in a cell type and tissue-specific manner. The study of chromatin accessibility and gene expression in immune cells from synovial fluid and peripheral blood of PsA

patients has demonstrated differences across the two tissues and shown specific enrichment for relevant pathophysiological pathways. In this study, $CD14^+$ monocytes were identified as the cell type presenting the largest number of significant changes in chromatin accessibility coupled with modulation of gene expression between synovial fluid and peripheral blood, with enrichment of DEGs in pathways leading to NF-κB activation and subsequent up-regulation in chemokine production. scRNA-seq analysis of PsA $CD14^+$ monocytes from additional patients has reinforced the differences between cells of the two tissues and identified additional pathways that were not unveiled by qPCR analysis. Moreover subpopulations of monocytes have been identified, particularly the $IL7R^+$ and the IFN-activated subsets, highlighting the heterogeneity of the monocyte population in PsA, importantly in the synovium. Pilot integration of mass cytometry data has confirmed increased ability of synovial fluid $CD14^+$ monocytes to produce a number of cytokines and chemokines involved in T cell recruitment and supported differences between synovial fluid and peripheral blood monocytes in PsA. Lastly, integration of the fine-mapping results with context specific epigenetic data has not revealed substantial overlap with DARs (maybe)due to the chosen FDR threshold for significance) but it has allowed contextualisation of genetic variation in a disease and tissue specific way. Overall, this chapter has shown that in PsA the pro-inflammatory environment at the site of inflammation drives changes in chromatin accessibility, gene expression and protein production that distinguishes circulating cells from those infiltrating the involved tissue.

Chapter 6

General discussion

In this thesis, I have characterised, for the first time, cell type and context specific differences in chromatin accessibility and gene expression between patients and controls, and across tissues, from psoriasis and PsA clinical samples. This required the establishment of methodological and analytical pipelines for ATAC and assessment of the effect that methods used to preserve clinical samples could have on the chromatin landscape. The use of low-input epigenetic techniques together with gene expression, single-cell analysis and the incorporation of GWAS fine-mapping data provides an integrated approach for the understanding of previously reported finding in psoriasis and PsA, adding cell and context specificity in physiological conditions. Although the results presented here have not unveiled new pathophysiological mechanisms, they represent a unique resource establishing the epigenetic and transcriptomic profiles of relevant cells in psoriasis and PsA, that could be further expanded and investigated at the functional level.

This chapter will give an overview of the key findings of this thesis in relationship to the general aims and discuss the future work required to maximise the value of this data and its future potential clinical application.

6.1 The importance of cell type and context specificity in disease

The importance of cell type and context specificity in understanding regulation of gene expression has been widely demonstrated (Fairfax *et al.* 2012; Fairfax *et al.* 2014). In psoriasis and PsA, the enrichment of prioritised GWAS SNPs in putative enhancers and regulatory regions has highlighted various immune and non-immune cell types and tissues, including Th-1, Th17, memory $CD4^+$ and $CD8^+$ cells, monocytes, keratinocytes, B cells and adipose tissue (Farh *et al.* 2015; Tsoi *et al.* 2017; Bowes *et al.* 2015a; Ellinghaus *et al.* 2016; Lin *et al.* 2018).

In this thesis, due to logistic and budget constraints, only a number pathophysiologically relevant cell types were investigated in psoriasis ($CD14^+$ monocytes, total $CD4^+$ and $CD8^+$ T cells and $CD19^+$ B cells) and PsA ($CD14^+$ monocytes, $mCD4^+$ and $mCD8^+$ T cells and NK cells). The variation in the number of differential targets and enriched pathways in chromatin accessibility and gene expression across cell types supports the need to expand this study to other cell types and discrete subpopulations to fully understand disease pathophysiology and aetiology. Of particular interest would be investigating DCs, key players in the skin inflammatory process that have been shown to be predominant at the psoriasis lesional skin, and for which psoriasis GWAS SNPs have shown context specific effects in their regulation of gene expression (Lee *et al.* 2014; Cheng *et al.* 2018).

Moreover, the epigenetic and transcriptional changes observed here between synovial fluid and peripheral blood of PsA patients demonstrates the importance of the environment at the site of inflammation to better understand disease pathophysiology, and highlights the need to isolate skin immune-

infiltrated cells using FACS or single-cell technologies for further investigation (Ahn *et al.* 2017)

6.2 The epigenetic landscape in psoriasis and PsA clinical samples: systemic vs affected tissue

The epigenetic landscape in psoriasis revealed only marginal differences in chromatin accessibility and the active enhancer H3K27ac mark between patients and controls for peripheral blood isolated immune cell types, with total $CD8^+$ cells demonstrating the largest number of differences. In contrast, the comparison of synovial fluid and peripheral blood in PsA revealed a larger number of differentially accessible regions in all the cell types, with $CD14^+$ monocytes showing greatest extent of changes in the chromatin accessibility landscape. Pathway enrichment analysis annotating the differential accessible regions with proximal genes showed that those changes were enriched for disease functional relevant processes. The contrasting results suggest that the physiological environment at the PsA inflamed joints drives chromatin accessibility changes of greater effect size than those observed in circulating cells between patients and control, and reinforce the need to study the chromatin accessibility landscape of skin in psoriasis. This is also in line with other studies where larger changes in chromatin accessibility or H3K27ac modifications have been found when directly interrogating the main pathological cell type or affected tissue (Scharer *et al.* 2016; Peeters *et al.* 2015; Wang *et al.* 2018; Liu *et al.* 2018).

Additional differences between circulating cells from psoriasis patients vs controls may also be unveiled by increasing sample size and studying more specific subsets within well defined cell populations, for example investigating memory and regulatory subsets within $CD4^+$ and $CD8^+$ populations. The

success in assessing chromatin accessibility by Omni-ATAC in NHEKs (Chapter 3) and the development of scATAC-seq opens the door to interrogate lesional skin keratinocytes and mapping the chromatin landscape of small numbers of infiltrated immune cells isolated from skin biopsies (Buenrostro *et al.* 2015; Corces *et al.* 2017). Lastly, in this thesis only one genomic region in $CD8^+$ cells showed paired changes in chromatin accessibility and H3K27ac between psoriasis patients and healthy controls, which could be due to H3K27ac being only one of the many mechanisms involved in shaping chromatin accessibility. This may suggest that differences in this histone modification between patients and controls may not be relevant in driving the observed changes in chromatin accessibility. However, the overall small sample size in this study and the differences in sizes of the cohorts interrogated for ATAC and H3K27ac provided additional challenges to draw robust conclusion.

6.3 Cell type and tissue specific transcriptional profiles in clinical cohorts

The majority of the transcriptional studies published in psoriasis and PsA have been conducted using bulk PBMCs or SFMCs, in a particular cell type undergoing exogenous inflammatory stimuli, and/or for a limited number genes. This has provided a very broad overview of dysregulated gene expression within mixed cell populations. Similarly, the study of lesional skin from psoriasis patients has been mostly carried out in whole skin biopsies containing not only keratinocytes from the different layers of the skin and infiltrated immune cells, but also fibroblasts and adipose tissue.

This thesis represents to my knowledge the first comprehensive study conducting a genome-wide transcriptomic analysis in four relevant immune cell types isolated from peripheral blood of patients and controls without

additional stimulation, and paired with epigenetic data. $CD14^+$ and $CD8^+$ were demonstrated as the main cell types undergoing dysregulation of gene expression, highlighting immune-relevant pathways with both up- and down-regulation of pro-inflammatory genes, and cell-type specific differentially expressed lncRNAs. Particularly unexpected was the up-regulation of two anti-inflammatory genes, *TNFAIP3* and *NFKBIA*, which are also psoriasis GWAS risk loci. *TNFAIP3* dysregulation has also been reported in other chronic inflammatory diseases including UC, CD, cystic fibrosis and pediatric IBD, demonstrating disease specificity in the direction of the change (Jiang *et al.* 2012; Majumdar *et al.* 2017; Sahlol *et al.* 2019). Interestingly, my data may also support a cell type and tissue specific effect in dysregulation of *TNFAIP3* in psoriasis, since this gene is down-regulated in PBMCs and lesional skin from patients but shows up-regulated expression in peripheral blood $CD4^+$ and $CD8^+$ T cells (Jiang *et al.* 2012; Sahlol *et al.* 2019).

The transcriptional study in epidermis sheets from lesional and uninvolved skin in patients highlighted larger changes in gene expression compared to the study in peripheral blood and the direction of the changes suggest a more pro-inflammatory signature, reinforcing again the role of the physiological inflammatory environment at the main disease affected tissue in driving functional changes. Moreover, the transcriptomic results also demonstrate the importance of aiming for pure cell populations, particularly of looking at purer keratinocyte samples from skin biopsies to unveil enrichment of not previously reported pathways. The use of epidermal sheets demonstrated enrichment of DEGs for the NOD-like pathways, which has not been reported by other studies using whole thickness skin biopsies, a finding that was further supported by a similar study published during this project (Tervaniemi *et al.* 2016).

The qPCR transcriptional study in three relevant cell types isolated from synovial fluid and peripheral blood allowed, for the first time, comparison of

cell-type specific dysregulation of inflammatory genes with changes in chromatin accessibility in PsA. My data showed the largest number of DEGs in $CD14^+$ monocytes, and similarly to the transcriptional results in skin, DEGs between synovial fluid and peripheral blood monocytes highlighted dysregulation of the NOD-like receptor signalling pathway by qPCR and scRNA-seq data, reinforcing the previously reported role of this pathway at the site of inflammation and in other chronic inflammatory diseases (McCormack *et al.* 2009).

The use of scRNA-seq to further dissect cell heterogeneity within a population represents a groundbreaking advance in the study of disease. This is particularly important for the phenotypic chracterisation of very plastic cells such as monocytes, which can differentiate into macrophages, DCs and osteoclast upon different environmental stimuli (Qu *et al.* 2014; Rivollier *et al.* 2012). In this thesis, a pilot exploratory study of subpopulations of monocytes in the synovium and peripheral blood of PsA was conducted for the first time and suggested the existence of functionally relevant subpopulations within the joints. Colleagues involved in this collaborative project are also exploring clonal expansion in T cells, with results suggesting relevant $mCD8^+$ clonal expansion in the joints compared to peripheral blood (manuscript in preparation). scRNA-seq has also been used to explore heterogeneity of pathphysiologically relevant cell populations in RA, MS and CD (Mizoguchi *et al.* 2018; Zhang *et al.* 2018; Schafflick *et al.* 2018). A recent study analysing healthy skin from different anatomical locations and lesional psoriatic epidermis has been published, revealing important differences across sites of sampling and showing expansion in the lesional epidermis of particular subpopulations also found in normal skin, further dissecting the complex aetiology of skin inflammation (Cheng *et al.* 2018). Although this study described the composition of immune-infiltrated cells, further investigation and characterisation of these cells will be required. Recently, spatial transcriptomics has also yielded cell-type specific

gene expression information from intact tissue sections of brain biopsies from MS patients or mice skin biopsies, representing an exciting avenue in the study of PsA joints and psoriasis skin samples (Itoh *et al.* 2018; Joost *et al.* 2016).

Although a systematic comparison between psoriasis and PsA is beyond the scope of this thesis, differences in gene expression for some of the targets were of interest. For example *CCR10* was up-regulated in total $CD8^+$ cells from psoriasis when compared to controls and also in $mCD8^+$ cells from synovial fluid in PsA. *CXCL10* presented lower expression in PsA peripheral blood $CD14^+$ monocytes when compared to controls, showing a monocyte-specific up-regulation of gene and protein expression only in the joint. No changes were found in psoriasis vs healthy controls for *CXCL10* gene expression in peripheral blood $CD14^+$ monocytes, in line with a recent conference abstract (Muntyanu *et al.* 2017). A role of *CXCL10* as potential PsA biomarker involved in the initiation and progression of joint inflammation previously hypothesised by another study conducted in blood sera (Abji *et al.* 2016).

6.4 Data integration in multi-omic studies

The ability to generate different layers of data including chromatin accessibility, histone modifications, transcript quantification and protein expression represents a challenge when it comes to integration and interpretation.

One of the major difficulties in the interpretation of ATAC data relates to the challenge of linking accessible chromatin to the regulated target gene. Enhancers and other regulatory regions do not always exert their effect on the most proximal gene, and cell and context specific long-distance chromatin interactions play a crucial role in shaping the regulatory landscape (Smemo *et al.* 2014). In this thesis annotation of the DARs is based on genes in physical proximity, in line with other current publications (Qu *et al.* 2015; Scharer *et*

al. 2016; Wang *et al.* 2018) and could be refined by incorporating chromatin interaction, eRNA and additional functional data. Recently, a publication by Liu and colleagues characterising the regulatory landscape of osteoarthritis used genome-wide integration of eRNA and chromatin segmentation maps to predict potential target genes of differentially accessible regions and correlate them with gene expression (Cao *et al.* 2017; Liu *et al.* 2018). Moreover, differential chromatin accessibility due to genetic variation has been shown to precede changes in transcript levels, and this should also be taken into account when trying to integrate epigenetic and transcriptional changes (Alasoo *et al.* 2018; Calderon *et al.* 2018). Furthermore, the analysis of differential chromatin accessibility and histone modifications in psoriasis and PsA would benefit from incorporating genotyping data, which could eventually be expanded to perform chromatin accessibility QTL in disease (Mo *et al.* 2018; Alasoo *et al.* 2018). This may allow the identification of cell type specific accessible regions similar to the ATAC peak harbouring rs4672505 at the 2p15 GWAS locus, where chromatin accessibility appears to depend on the genotype in a cell-type specific manner.

The integration of scRNA-seq and mass cytometry data in the PsA pilot study rremains challenging and it was limited by the sample size, technical aspects and time scale, still providing a platform for future validation studies. To better understand the correspondence between transcriptional profiles and protein in discrete subpopulations of cells, a larger sample size plus a more systematic approach for data integration will be required. Currently Zhang and colleagues have presented the most exhaustive methodology to integrate bulk-RNA-seq, scRNA-seq, mass cytometry and flow cytometry into multi-modal transcriptomic and proteomics profiles (Zhang *et al.* 2018). This strategy has revealed disease-specific functional expanded subpopulations amongst the most relevant cell types in RA pathophysiology. Moreover, generation of scATAC-seq

data will allow further charactersisation and correlation of changes in chromatin accessibility and gene expression at the single-cell level (Duren *et al.* 2018).

6.5 GWAS fine-mapping and integration with patients-derived epigenetic data

Statistical fine-mapping is a widely accepted strategy for refining putative causal SNPs for downstream functional follow-up. Fine-mapping using Bayesian methods have been conducted in this thesis using Immunochip summary statistics and genotyping data for psoriasis and PsA, respectively. In both cases the power of the analysis was limited by the incomplete cohort size of the original work, which explained the variation of results when compared to the fine-mapping conducted by Bowes and colleagues in PsA (Bowes *et al.* 2015a). Further refinement of the credible set of SNPs was achieved by integration of the in-house ATAC data generated in psoriasis and PsA samples, and highlighted two good candidates as putative causal variants at the *B3GNT2* and *RUNX3* loci in psoriasis and PsA, respectively. The overlap of rs11249213 at the *RUNX2* locus with a mCD8$^+$ synovial fluid open DAR which was significant for an FDR<0.05 but not <0.01, suggests the potential for finding additional overlaps with DARs of smaller effect size that may become significant in larger studies.

The step-wise Bayesian fine-mapping strategy implemented here performs under the assumption of only one SNP being causal for a particular association (Maller *et al.* 2012; Bunt *et al.* 2015). In contrast, other methods such as Bayesian evolutionary stochastic search allow to test for haplotype models where several variants may be driving the association (Wallace *et al.* 2015) and should be take into account for further analysis .

Notably, in this thesis overlap of psoriasis and PsA GWAS fine-mapped SNPs with significant eQTL SNPs (FDR<0.05) from relevant cell types or tissues

was reported. These overlap were not necessarily between the lead SNPs of each of the signals, and colocalisation analysis using tools such as Coloc (Giambartolomei *et al.* 2014) could be used to formally determine whether the fine-mapping and eQTL signals are the same and thus increase evidence of non-coding GWAS SNPs regulating expression of a particular gene.

6.6 Clinical relevance and translation

Technological advances in sequencing and the latest low-input and time-efficient epigenetic techniques, such as ATAC, enable implementing molecular tools based on gene expression and chromatin accessibility in a clinical setting. Genotyping, epigenetic and expression data could be used in the frame of personalised medicine to build disease, cell type, genotype and context-specific chromatin segmentation maps to help in diagnosis, patient stratification, prognosis, treatment decisions and drug design. Nevertheless, clinical translation of findings are challenging due to variability in the results across studies, influenced by sample size, age, sex and patient heterogeneity (S and Furman 2013). For example, interindividual heterogeneity has been assessed at the chromatin accessibility level in T cells, where gender was the main driver of variation and variable regions presented significant enrichment at autoimmune disease loci (Qu *et al.* 2015).

In cancer, gene expression profiling has revealed utility for diagnosis and chromatin accessibility has also been investigated in chronic lymphocytic leukemia, amongst others (Whitney *et al.* 2003; Rendeiro *et al.* 2016). In inflammatory diseases such as eosinophilic esophagitis, the expression of 94 genes in esophageal biopsies has been successfully used to diagnose disease and assess treatment response (Dellon *et al.* 2018). Currently, no molecular diagnosis based on gene expression or epigenetic profiles are formally used for psoriasis or PsA. A microarray study using PBMCs identified levels of

MAP3K3 and *CACNA1S* expression as the best combination of genes to correctly classify PsA and RA patients (Batliwalla *et al.* 2005). Expression levels in skin biopsies of a set of seven genes, including *CXCL8*, *CXCL10* and *CCL17*, appropriately discriminated psoriasis from atopic dermatitis, allergic contact dermatistis and irritant contact dermatitis (Zeeuwen *et al.* 2008). Additionally, a pilot metabolomic study showed that PsA patients had decreased level of alpha ketoglutaric acid and an increased level of lignoceric acid compared to psoriasis, which could potentially be used to distinguish between the two pathologies in blood sera (Armstrong *et al.* 2014). Further analysis and validation of the results generated in this thesis may contribute to progress towards the clinical translation of epigenetic profiles, transcriptomics and protein expression in psoriasis and PsA.

6.7 Future work

Extended recruitment and sample types

As previously mentioned, there are a number of factors limiting the identification of differences in chromatin accessibility and gene expression that are disease-specific and not due to general inflammation. The small sample size and unavailability of samples from the most relevant affected tissue are two major limitations. Additionally, patient heterogeneity and uncertainty due to conversion of psoriasis patients into PsA also needs to be taken into account to interpret these results.

In order to overcome limitations due to sample size, recruitment of additional psoriasis and PsA patients samples would need to be conducted. Expansion of the psoriasis recruitment should include peripheral blood, lesional and uninvolved skin from patients not undergoing biological treatment and also matched samples from those individuals undergoing biological therapy such as

TNFi. Regarding skin samples, paired scATAC-seq and scRNA-seq could be performed on the infiltrated immune cell types at the site of inflammation and compared to their counterparts in peripheral blood. PsA naíve-treated patient recruitment would also aim to include skin biopsies and samples following different types of biologic therapies. Moreover, peripheral blood and synovial fluid from healthy controls and individuals suffering from arthritic pathologies, such as RA and osteoarthritis, could also be included to enable comprison between chronic inflammation and PsA specific features. In the extended recruitment, unification in the methodology of processing samples will also be implemented to enable comparison across psoriasis and PsA patients. Lastly, additional clinical data and follow-up evaluation of the psoriasis patients by a rheumatologist would reinforce appropriate classification of patients as psoriasis non-converters into PsA.

Further exploration of chromatin accessibility and histones modification analysis

The successful implementation of the optimised Omni-ATAC protocol in NHEKs (Chapter 5) will allow future generation of chromatin accessibility data in keratinocytes isolated from lesional and uninvolved skin biopsies from psoriasis and PsA samples. Moreover, ChIPm would enable interrogation of additional histone marks, including H3K4me1 (enhancer) and H3K4me3 (promoter), RNA-pol II and TFs such as CTCF. Additionally, performing ATAC and ChIPm in the same relevant cell types isolated from patients and controls, followed by stimulation could focus the analysis on a set of promoters and regulatory regions undergoing modifications in inflammation (Peeters *et al.* 2015).

In terms of analysis, *de novo* methods not based on peak calling could be used to further explore differences in chromatin accessibility and histone marks (Shen *et al.* 2013). Moreover, genotyping of all the subjects would

allow the exploration of changes in chromatin and histone modifications in the context of genotype, similarly to the chr2p15 example illustrated for psoriasis. Furthermore, this type of analysis could be expanded to generate chromatin accessibility and histone modification QTLs in a tissue, cell type and disease status specific manner.

Interpretation of genomic regions data, importantly ATAC will be improved using a tool currently under development known as Atlas and Analysis of Pathways (A2), created by Dr Hai Fang in collaboration with Dr Anna Sanniti and myself. This tool will incorporate pathway-centric interpretation of genomic regions through identification of crosslinked genes based on experimentally-derived evidence, including eQTLs, chromatin interaction or enhancer-promoter co-expression. This analysis will improve and refine the interpretation of the differential chromatin accessibility analysis in PsA and any genomic-region based data.

Additional gene expression analysis and allele specific expression

In order to have a better understanding of the genome-wide changes in gene expression between PsA immune cells from synovial fluid and peripheral blood, RNA-seq could be performed in the same cell types currently interrogated by qPCR, and also in additional cell populations of interest. Allele-specific expression (ASE) analysis could be performed to assess the *cis* effects of genetic variants in gene expression and removing confounders such as *trans*-acting environmental effects. Moreover, miRNAs could also be interrogated in future to further investigate the role of non-coding RNAs in disease.

Data integration

Integration of the histone marks and chromatin accessibility changes in a larger cohort may reveal co-regulated changes in a disease or genotype specific

manner, revealing additional information about the mechanism of action of risk variants. Additionally, ASE data could also be integrated with the cell-type specific epigenetic maps from the same individuals, fine-mapping analysis and co-localisation of signals to help informing the mechanisms of action of putative causal variants. Better understanding of disease pathophysiology and genetic associations would be aided by incorporating psoriatic skin methylation data and other publicly available chromatin accessibility QTLs and methyl-QTLs generated in different tissues (Chandra *et al.* 2018; Pelikan *et al.* 2018; Kumasaka *et al.* 2015; Kumasaka *et al.* 2018; Chen *et al.* 2016). Integration of single-cell RNA-seq and mass cytometry data will also benefit from improvements in the methods for validating cluster identities, one of the most challenging aspects in single-cell analysis (Kiselev *et al.* 2019). Additionally, systematic integration of single-cell mass-cytometry and RNA-seq data should be considered to find correlation between relevant subpopulations in each assay. This could be achieved by algorithms based on Canonical Correlation Analysis that have been implemented in the study of RA (Zhang *et al.* 2018).

Functional follow-up

The epigenetic and gene expression data generated in this thesis, in addition to publicly available functional data, has pritoritised variants at GWAS loci that merit further investigation. One such example is rs4672505, which is hypothesised to be regulating *B3GNT3* expression in $CD8^+$ T cells. Genomic-editing using CRISPR-cas9 (Ding *et al.* 2013) could be used to delete the ATAC peak colocalising with rs4672505 in $CD8^+$ cells and determine the effect in the regulation of *B3GNT2* expression in homeostasis and also under anti-CD3 and anti-CD28 stimulation. Chromatin interaction experiments using capture-C could also be conducted in untreated and stimulated cells to assess the effect of rs4672505 genotype in the physical interaction with the *B3GTN2* promoter

seen by Javierre and colleagues (Javierre *et al.* 2016). Additionally, molecular assays to interrogate allelic differences in affinity for TF binding, including electrophoretic mobility shift assay (EMSA) and also ChIP using antibodies for TF of interest in $CD8^+$ cells harbouring different rs4672505 genotypes could be performed (Vernes *et al.* 2007).In terms of genome-wide screening, self-transcribing active regulatory region sequencing (STARR-seq) (Arnold *et al.* 2013) could be implemented to create a genome-wide quantitative enhancer map that ccould be colocalised with epigenetic data and non-coding fine-mapped SNPs.

6.8 Concluding remarks

The overall aim of this thesis was to investigate the epigenetic regulatory landscape and the transcriptional profile of psoriasis and PsA patients in a cell type and tissue specific manner to further help understanding of disease mechanism, requiring establishment of methodological techniques and analytical pipelines to interrogate chromatin accessibility. The data generated in this thesis has enabled a genome-wide comparison of the chromatin landscape and gene expression in disease vs healthy controls and also at the site of inflammation. Integration with fine-mapping GWAS and other publicly available datasets has further refined the number of putative causal variants for functional follow-up and reinforced the importance of integrating multiple cell type and context specific datasets to interpret GWAS results. As the datasets generated here continue to be expanded and experimental and analytical techniques are further refined, the understanding of disease mechanisms governing psorasis and PsA in important cell types and states will no doubt continue.

Based on the assays and cell types here presented, the results of this thesis have prioritised the role of $CD14^+$ monocytes and $CD8^+$ cells in psoriasis and PsA. At the systemic level $CD14^+$ monocytes and $CD8^+$ cells clearly showed the largest number of differentially expressed genes and highlighted pathologically

relevant pathways, suggesting that these two cell types may be driving the systemic inflammatory signature in psoriasis. The prominent role of $CD8^+$ cells observed in this thesis is consistent with the latest studies supporting the role of $CD8^+$ in the aetiology of skin lesions. Therefore, additional sample recruitment to perform a more extensive RNA-seq analysis of this cell type (including also lncRNAs and miRs) and a disease-specific eQTL analysis should be prioritised. Nevertheless, the larger changes observed in skin highlights the need to also perform differential gene expression and chromatin accessibility analysis in keratinocytes and immune infiltrated cells from lesional, uninvolved and healthy skin biopsies.

Regarding PsA, the results in this thesis have demonstrated $CD14^+$ monocytes to show the larger number of changes based on the study of chromatin accessibility, gene expression and protein production of key inflammatory molecules between blood and synovial fluid. Moreover, they have shown functional relevant synovial fluid specific subpopulations by scRNA-seq. In PsA, $CD14^+$ monocytes have not been described as key cells in the disease aetiology but they have been described as relevant cells in the perpetuation of the inflammatory response, importantly as precursors of osteoclast, which have a critical role in the dysregulation of bone remodelling characteristic in PsA. The current pathophysiological knowledge highlights $mCD8^+$ cells as drivers of of joint inflammation, where they represent the most expanded cell population. This has been further confirmed by Penkava and colleagues scRNA-seq analysis of synovial fluid $mCD8^+$ and the identification of tissue specific clones with relevant pathophysiological characteristics in PsA. Thus, $CD14^+$ monocytes and $mCD8^+$ T cells should be prioritised to conduct ATAC-seq (together with scATAC-seq) and scRNA-seq in a larger cohort for validation of this thesis' results and for further integration with genotyping data to explore tissue specific expression and chromatin accessibility QTLs.

Appendices

A Tables **307**

A.1 Chapter 3 Tables 307

A.2 Chapter 4 Tables 308

A.3 Chapter 5 Tables 312

B Additional figures **315**

B.1 Chapter 3 Figures 315

B.2 Chapter 4 Figures 318

B.3 Chapter 5 Figures 323

Bibliography **369**

List of Tables

A.1 Enrichment of ATAC-seq reads across the TSS for the $CD14^+$ monocytes and $CD4^+$ samples fresh, frozen and fixed. 307

A.2 Size of the master lists generated by DiffBind for the H3K27ac differential analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. . . . 308

A.3 Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChIPm assay. 309

A.4 Functional interactions and overlap with another study for the differentially expressed lncRNAs in each cell type. 309

A.5 Additional enriched pathways for DEGs between psoriasis and healthy controls in $CD14^+$ monocytes and $CD8^+$ cells. 310

A.6 Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies. 311

A.7 Enriched pathways for the scRNA-seq DEGs between synovial fluid PsA $CD14^+$ monocytes. 312

A.8 PsA GWAS Immunochip loci presenting $\log_{10}$BF< 3 for the fine-mapping lead SNP in the association analysis. 313

List of Figures

B.1 Distribution of the background read counts from all the master list peaks absent in each sample. 315
B.2 Pre-sequencing profiles of relative abundance of DNA library fragment sizes for a psoriatic lesional keratinocytes ATAC 1 library generated using 40 min of transposition. 315
B.3 Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples. . . . 316
B.4 Fast-ATAC and Omni-ATAC NHEK pre-sequencing profiles of relative abundance of DNA library fragment sizes. 316
B.5 Differences in TSS fold-enrichment and percentage of mitochondrial between the ATAC-seq and Fast-ATAC libraries from all the samples generated in the thesis. 317
B.6 Mitochondrial reads and called peaks after IDR p-value filtering in the ATAC libraries generated in psoriasis patients and healthy control samples. 318
B.7 PCA analysis illustrating batch effect in ATAC and RNA-seq samples.319
B.8 Mapping rate and total reads after filtering (million) mapping to Ensembl genes in all the RNA-seq samples from psoriasis patients and controls in four cell types. 319
B.9 Mapping quality control for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. 320
B.10 Distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles for all 72 samples. 320
B.11 Correlation in gene expression between the miRs *MIR31HG* and the target gene *HIF-1A* in lesional and uninvolved skin. 321
B.12 Epigenetic landscape at the location of the fine-mapped SNP rs12223943 at the *ETS1* psoriasis GWAS locus. 322
B.13 Assessment of the percentage of mitochrondrial and duplicated reads and the number of called peaks after filtering in the ATAC data from peripheral blood and synovial fluid of PsA patients samples. 323
B.14 Permutation analysis for the DARs found in synovial vs peripheral blood in $CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells. 324
B.15 Enrichment of DARs in the 15 chromatin states of the cell type specific Roadmap Epigenomic Project segmentation maps. 325
B.16 Identification of the $CD14^+$ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. 325
B.17 Expression levels of antigen presentation and phagocytosis characteristic genes in synovial fluid and peripheral blood $CD14^+$ monocytes. 326

B.18 Differential gene expression between synovial fluid and peripheral blood PsA $CD14^+$ monocytes and correlation with qPCR and chromatin accessibility 327

Appendix A

Tables

A.1 Chapter 3 Tables

Cell type	Condition	TSS enrichment CTL1	CTL2	CTL3
CD14	Fresh	17.4	19.6	14.11
	Frozen	26.3	25.2	27.1
	Fixed	2.5	16.5	22.4
CD4	Fresh	5.3	5.6	7.7
	Frozen	17.9	14.1	16.1
	Fixed	7.9	23.0	14.3

Table A.1: Enrichment of ATAC-seq reads across the TSS for the $CD14^+$ monocytes and $CD4^+$ samples fresh, frozen and fixed.

A.2 Chapter 4 Tables

Cell type	Master list size genome-wide	Master list size enhancers
$CD14^+$	99,862	60,962
$CD4^+$	110,353	56,282
$CD8^+$	137,194	51,607
$CD19^+$	199,014	88,722

Table A.2: Size of the master lists generated by DiffBind for the H3K27ac differential analysis between psoriasis patients and healthy controls in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells. In the genome-wide analysis, the master list size refers to the number of H3K27ac enriched sites included in the consensus list built using DiffBind to perform the differential analysis. In the analysis restricted to enhancers, the size of the master list was reduced to only those sites from the genome-wide master list annotated as enhancers (weak and strong) according to the chromatin segmentation map for each particular cell type.

Sample ID	NRF	PBC1/ PBC2	Sample ID	NRF	PBC1/ PBC2
PS2000 CD14	0.77	0.60/2.5	PS2000 CD8	0.77	0.76/4.5
PS2001 CD14	0.84	0.70/3.0	PS2001 CD8	0.74	0.74/4.0
PS2314 CD14	0.81	0.60/1.8	PS2314 CD8	0.74	0.75/4.1
PS2319 CD14	0.79	0.60/2.2	PS2319 CD8	0.72	0.75/4.0
CTL7 CD14	0.81	0.65/2.2	CTL7 CD8	0.32	0.32/1.5
CTL8 CD14	0.83	0.66/2.3	CTL8 CD8	0.70	0.70/3.3
CTL9 CD14	0.80	0.60/2.3	CTL9 CD8	0.73	0.73/3.7
CTL10 CD14	0.83	0.65/2.1	CTL10 CD8	0.68	0.65/2.9
PS2000 CD4	0.84	0.75/3.4	PS2000 CD19	0.38	0.42/1.9
PS2001 CD4	0.82	0.72/2.9	PS2001 CD19	0.71	0.71/3.7
PS2314 CD4	0.82	0.71/2.8	PS2314 CD19	0.29	0.34/1.8
PS2319 CD4	0.82	0.73/3.2	PS2319 CD19	0.76	0.78/4.8
CTL7 CD4	0.78	0.68/2.5	CTL7 CD19	0.74	0.69/3.1
CTL8 CD4	0.81	0.71/2.9	CTL8 CD19	0.68	0.67/3.2
CTL9 CD4	0.81	0.74/3.3	CTL9 CD19	0.75	0.76/4.6
CTL10 CD4	0.77	0.61/1.9	CTL10 CD19	0.61	0.59/2.6

Table A.3: Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChIPm assay. NRF, PBC1 and PBC2 are the three measures used according to the ENCODE standards as referred in Chapter 2. 0.5≤NRF<0.8 acceptable; 0.8≤NRF≤0.9 compliant; NRF>0.9 ideal; 0.5≤PBC1<0.8 and 1≤PBC2<3 moderate bottlenecking; 0.8≤PBC1<0.9 and 3≤PBC2<10 mild bottlenecking. NRF= non-redundant fraction; PBC= PCR bottleneck coefficient.

Cell type	LncRNAs with functional interactions	LncRNAs overlapping Dolcino *et al.*,2018
$CD14^+$	24	4 (*HOTAIRM1**, *ILF3-AS1**, *MMP24-AS1*, *RP11-325F22.2*)
$CD4^+$	10	1 (*MMP24-AS1*)
$CD8^+$	21	1(*CTB-25B13.12*)
$CD19^+$	5	0

Table A.4: Functional interactions and overlap with another study for the differentially expressed lncRNAs in each cell type. For each cell type the number of differentially expressed lncRNAs (FDR<0.05) for which a functional interaction has been experimentally validated based on NPInter database is shown. NPInter documents functional interactions between noncoding RNAs (except tRNAs and rRNAs) and biomolecules (proteins, RNAs and DNAs) which have published experimental validation. This table also records the number of differentially expressed lncRNAs overlapping with the Dolcino *et al.*, 2018 study, where PBMCs from PsA patients and healthy controls are contrasted.(*) indicates dysregulation in the opposite direction between this data and Dolcino *et al.*.

CD14$^+$ monocytes Pathway	FDR	Fold change
Generic transcription	$1.1x10^{-6}$	2.66
RNA transport	$4.5x10^{-5}$	2.98
GnRH signalling	$2.4x10^{-4}$	3.61
Ribosome biogenesis in eukaryotes	$3.0x10^{-4}$	3.42
Neurotrophin signaling	$4.0x10^{-4}$	2.96
Spliceosome	$3.2x10^{-3}$	2.51
Autophagy	$8.9x10^{-3}$	2.18
Protein processing endoplasmic reticulum	$8.9x10^{-3}$	2.05

CD8$^+$ Pathway	FDR	fold change
Epstein-Barr virus infection	$1.5x10^{-2}$	2.33
RNA-pol I and III and	$3.1x10^{-2}$	2.79
Apoptosis	$4.5x10^{-3}$	2.41

Table A.5: Additional enriched pathways DEGs between psoriasis and healthy controls in CD14$^+$ monocytes and CD8$^+$ cells. Significant pathways for FDR<0.01. All the enriched pathways contained a minimum of ten DEGs FDR<0.05 from the analysis.

Lesional vs uninvolved Pathway	FDR	Fold change
Genes encoding extracellular matrix proteins	$5.1x10^{-9}$	3.3
Serine/threonine-protein kinase	$8.1x10^{-7}$	4.6
Genes encoding secreted soluble factors	$8.1x10^{-7}$	3.42
FOXM1 TF network	$2.3x10^{-6}$	4.76
Phase 1 functionalisation compounds	$8.1x10^{-6}$	4.29
Biological oxidations	$2.9x10^{-5}$	3.02
G2/M Checkpoints	$5.4x10^{-5}$	3.86
Aurora B signaling	$8.1x10^{-5}$	3.92
Chemical carcinogenesis	$2.0x10^{-4}$	3.86
Serotonergic synapse	$2.0x10^{-4}$	3.86
Drug metabolism-cytochrome P450	$2.6x10^{-4}$	3.45
Mitotic M-M/G1 phases	$3.0x10^{-4}$	2.16
MicroRNAs in cancer	$6.4x10^{-3}$	2.18
Glycosaminoglycan metabolism	$1.9x10^{-2}$	2.36
E2F TF network	$2.0x10^{-3}$	2.49
p73 TF network	$2.3x10^{-3}$	2.45
Fc-ϵR I signalling in mast cells	$4.2x10^{-3}$	2.54
Tight junction	$4.2x10^{-3}$	2.13
Orc1 removal from chromatin	$4.7x10^{-3}$	2.38
Extra-cellular matrix receptor interaction	$9.3x10^{-3}$	2.24
Transport of inorganic ions and aminoacis	$5.6x10^{-3}$	2.30

Table A.6: Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies. Significant pathways for FDR<0.01. All the enriched pathways contained a minimum of ten DEGs FDR>0.05 from the analysis.

A.3 Chapter 5 Tables

Pathway	FDR	Fold change
Antigen processing and presentation	$2.5x10^{-6}$	3.4
Chemokine signalling	$1.8x10^{-5}$	2.3
GPCR ligand binding	$1.9x10^{-5}$	2.2
Platelet activation and aggregation	$1.7x10^{-7}$	2.1
NOD-like receptor signalling	$4.0x10^{-3}$	1.8
IFN/IFNγ signalling	$1.2x10^{-3}$	2.4
Cytokine and interleukin sig.	$5.5x10^{-4}$	1.7
Complement cascade	$3.8x10^{-10}$	6.7
MHC-II antigen presentation	$2.8x10^{-5}$	2.6
Osteoclast differentiation	$1.2x10^{-3}$	2.0
Apoptosis	$9.9x10^{-3}$	1.8
Phagosome	$1.2x10^{-12}$	3.5
Cell adhesion molecules (CAMs)	$2.4x10^{-5}$	2.7
PPAR signalling	$3.3x10^{-3}$	2.7
VEGF signalling	$2.0x10^{-4}$	2.4
Lymphoid and non-lymphoid cells interaction	$5.5x10^{-4}$	2.3
IL-4 and 13 signalling	$1.7x10^{-3}$	2.2
PI3K-Akt signaling pathway	$9.9x10^{-6}$	2.1
Leukocyte transendothelial migration	$4.6x10^{-3}$	2.1
FoxO signalling	1.710^{-3}	2.1
Cytokine-cytokine receptor interaction	5.910^{-3}	1.8
Signalling by receptor tyrosin kinases	5.510^{-4}	1.6

Table A.7: Enriched pathways for the scRNA-seq DEGs between synovial fluid PsA CD14$^+$ monocytes. All the enriched pathways contained a minimum of ten DEGs (FDR<0.01 and fold change>1.5) from the analysis and were significant at an FDR<0.01.

Table A.8: PsA GWAS Immunochip loci presenting $\log_{10}$BF< 3 for the fine-mapping lead SNP in the association analysis. For each of the signals chromosome (chr), genes nearby, $\log_{10}$BF< 3 for the fine-mapping (FM) lead SNP, the PsA GWAS lead SNP including p-value in the study and the number of SNPs in the 99% credible set reported by Bowes *et al.* for that signal. NA refers to the locus reported as fine-mapped by Bowes *et al.* 2015. OR=odds ratio

chr	Closest gene	FM lead SNP	$\log_{10}$ABF FM lead SNP	GWAS lead SNP (p-value)	GWAS lead OR	Bowes FM lead SNP	Bowes 99% credible set
2	*B3GNT2/TMEM17*	2:62501912(INS)	1.8	rs6713082 ($4.59\text{x}10^{-5}$)	1.2	rs6713082	22
17	*CARD14*	rs11150848	0.8	rs11652075 (0.014)	1.1	NA	NA
9	*DDX58*	rs138398872	0.5	rs1133071 ($3.36\text{x}10^{-5}$)	1.2	NA	NA
7	*ELMO1*	rs10279209	1.1	rs73112675 (0.0041)	1.1	NA	NA
6	*ERAP1/ERAP2*	rs58711860	2.7	rs62376445 (0.00017)	1.4	NA	NA
1	*SLC45A1/TNFRSF9*	rs113677773	1.7	rs11121129 (0.00093)	1.1	NA	NA
11	*ETS1/FLI1*	rs7935286	0.6	rs4936059 (0.0014)	1.1	NA	NA
1	*LCE3B/LCE3A*	rs11205042	2.8	rs6693105 (0.0028)	1.1	NA	NA
22	*LOC150223*	rs371643642	1.2	rs2298428 ($4.38\text{x}10^{-5}$)	1.2	NA	NA
11	*ZC3H12C*	rs1648153	0.2	rs4561177 (0.0037)	1.1	NA	NA
9	*LOC392382*	rs36015268	0.8	rs12236285 (0.038)	1.2	NA	NA
17	*NOS2A*	rs4795067	1.9	rs4795067 ($1.94\text{x}10^{-7}$)	1.2	rs4795067	2

2	*PAPOLG/REL*	rs60685986	2.0	rs1306395 ($2.99x10^{-5}$)	1.2	rs1306395	32
18	*POLI*	18:51926806	0.3	rs602422 (0.0047)	1.1	NA	NA
14	*NFKBIA*	rs35309046	0.9	rs8016947 ($9.65x10^{-5}$)	1.2	NA	NA
11	*RPS6KA4*	rs146881600	1.3	rs645078 (0.00086)	1.1	NA	NA
6	*RSPH3/TAGAP*	rs11754601	1.3	rs1973919 (0.018)	1.1	NA	NA
6	*TNFAIP3*	rs1890370	2.0	rs610604 (0.00032)	1.1	NA	NA
5	*TNIP1/ANXA6*	rs75851973	2.8	rs76956521 ($4.98x10^{-9}$)	1.5	rs76956521	24
10	*ZMIZ1*	rs2395526	0.9	rs1972346 (0.0082)	1.1	NA	NA
20	*ZNF313*	rs73129222	1.6	rs6063454 ($2.90x10^{-5}$)	1.2	NA	NA
16	*ZNF668*	rs9939243	0.9	rs7197717 (0.0035)	1.1	NA	NA

Appendix B

Additional figures

B.1 Chapter 3 Figures

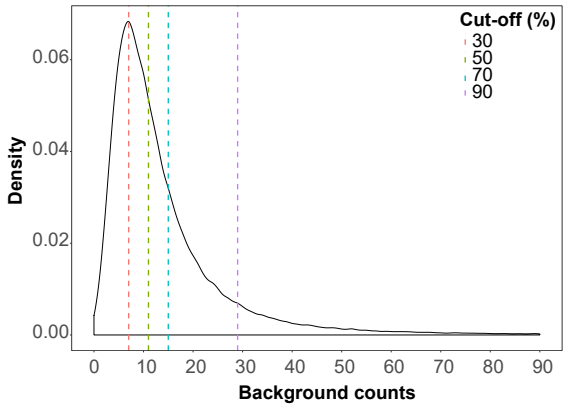


Figure B.1: Distribution of the background read counts from all the master list peaks absent peaks in each sample. Each cut-off corresponds to the number of background counts showed by a particular percentage of the total number of absent peaks.

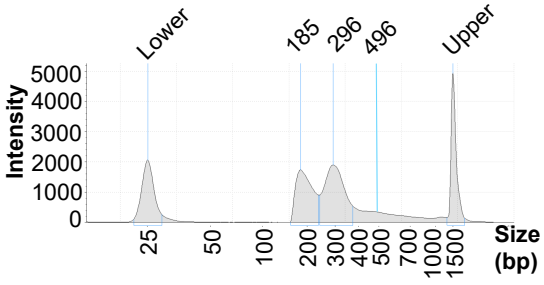


Figure B.2: Pre-sequencing profiles of relative abundance of DNA library fragment sizes for a psoriatic lesional keratinocytes ATAC 1 library generated using 40 min of transposition. Pre-sequencing quantification of DNA fragment sizes from the ATAC library generated using 50,000 keratinocytes isolated from a psoriatic lesional skin biopsy and the ATAC 1 protocol (detailed in Table 3.3) and transposition for 40 min.

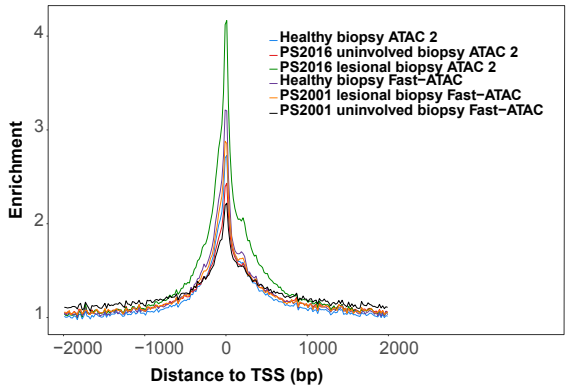


Figure B.3: Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples. Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from the different librariesgenerated using skin isolated keratinocytes from psoriasis patients (lesional and uninvolved) or healthy controls biopsies using ATAC 2 and Fast-ATAC protocols detailed in Table 3.3.

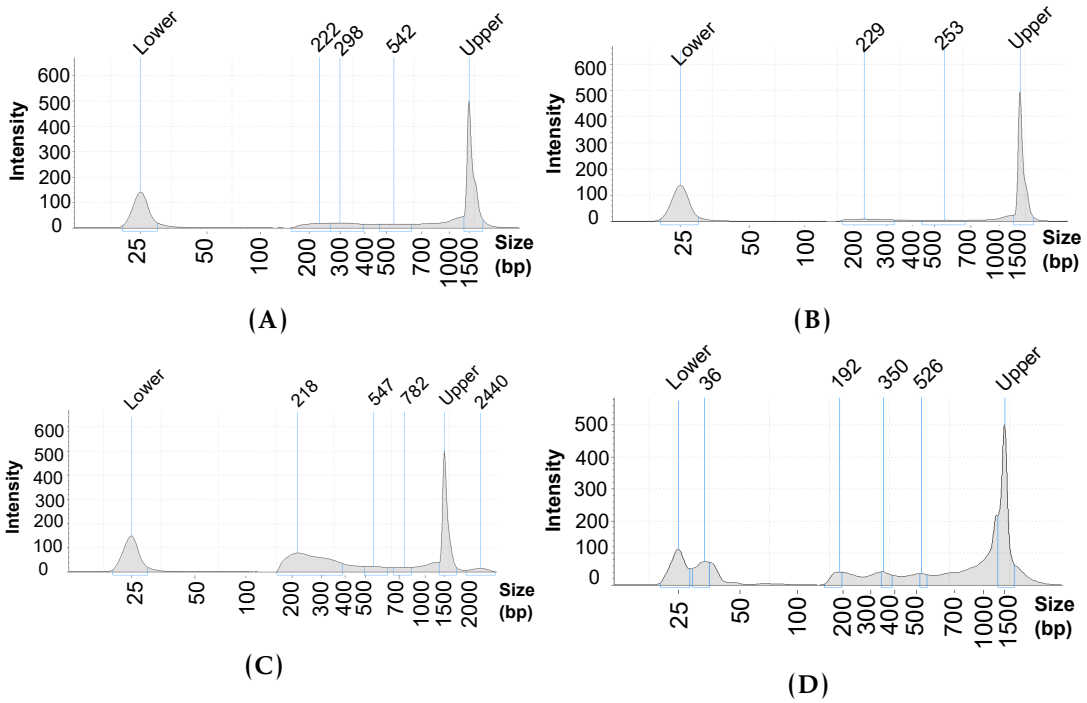
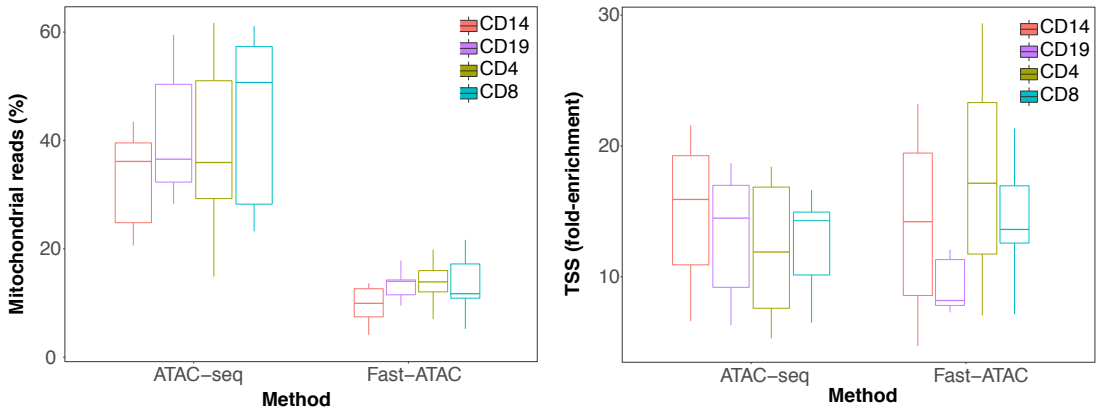


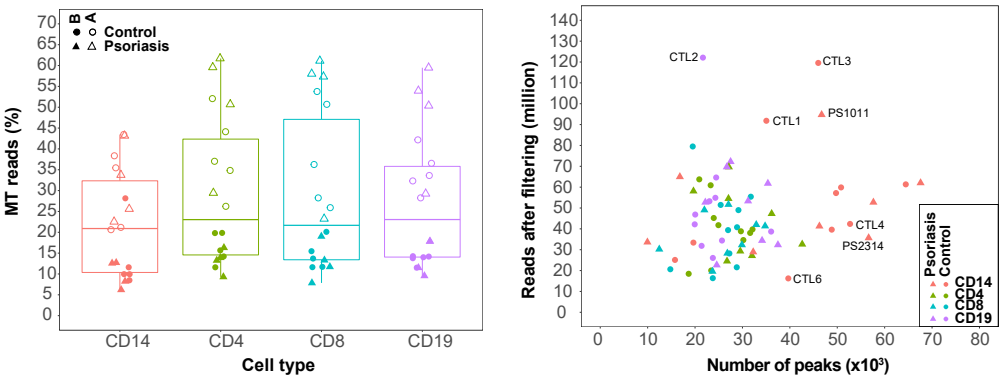
Figure B.4: Fast-ATAC and Omni-ATAC NHEK pre-sequencing profiles of relative abundance of DNA library fragment sizes. Pre-sequencing quantification of DNA fragment sizes from the libraries generated using the (A) C2, (B) C3, and (C) C4 versions of the Fast-ATAC protocol based on modifications in the detergent and Tn5 concentration and (D) Omni-ATAC. C2, C3 and C4 detergent and Tn5 concentrations are detailed in Table 3.3.



(A) (B)

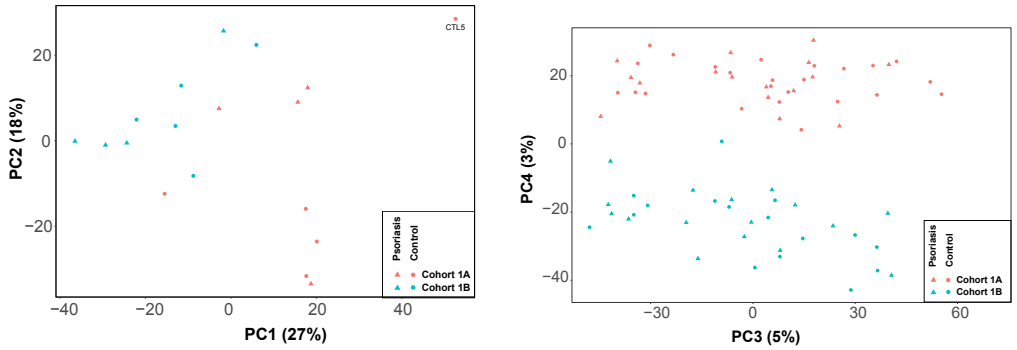
Figure B.5: Differences in TSS fold-enrichment and percentage of mitochondrial between the ATAC-seq and Fast-ATAC libraries from all the samples generated in the thesis. Boxplox illustrating differences between ATAC-seq and Fast-ATAC by cell type for (A) TSS fold enrichment and (B) mitochondrial reads (percentage).

B.2 Chapter 4 Figures



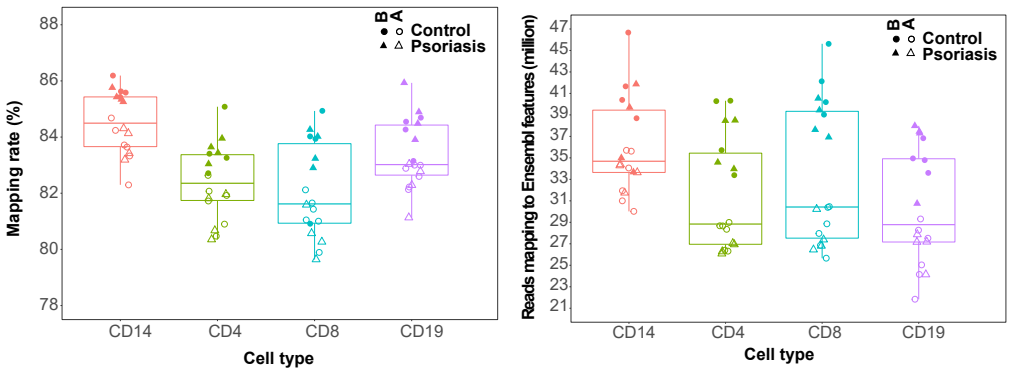
(A) **(B)**

Figure B.6: Miitochondrial reads and called peaks after IDR p-value filtering in the ATAC libraries generated in psoriasis patients and healthy control samples.. (A) Percentage of MT reads in the ATAC-seq samples generated in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ isolated from psoriasis patients and healthy controls. Samples from cohort 1A (open circles and triangles) were generated with the standard ATAC-seq protocol from Buenrostro *et al.*, 2013 whereas samples from cohort 1B (filled circles and triangles) were processed using FAST-ATAC (Corces *et al.* 2016). (B) For each of the ATAC libraries, representation of the number of significant peaks based on IDR optimal p-value filtering versus the total million of reads after filtering.



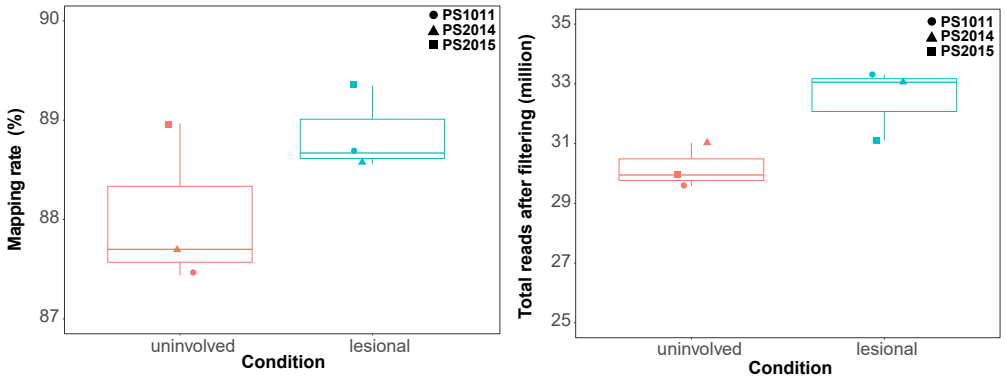
(A) (B)

Figure B.7: PCA analysis illustrating batch effect in ATAC and RNA-seq samples. (A) First and second component of the PCA analysis performed using the normalised ATAC counts in a list of consensus peaks across all the combined $CD8^+$ samples from psoriasis patients and healthy controls. (B) Third and fourth component of the PCA analysis performed on the normalised number of reads mapping to the Ensmbl list of mRNAs and lncRNAs detected in $CD8^+$ cells from psoriasis patients and healthy controls.



(A) (B)

Figure B.8: Mapping rate and total reads after filtering (million) mapping to Ensembl genes in all the RNA-seq samples from psoriasis patients and controls in four cell types. (A) The mapping rate refers to the percentage of total sequenced reads from each sample that uniquely mapped to a particular site of the genome. (B) The total number of reads after filtering for non-uniquely mapped and duplicated reads that mapped to Ensembl features, including coding protein genes and lncRNAs.



(A) (B)

Figure B.9: Mapping quality control for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. (A) Mapping rate calculated as the proportion of sequencing reads mapping uniquely to a particular region of the genome. (B) The total number of reads mapping to an Ensembl feature (including protein coding genes and lncRNAs) after removing the non-uniquely mapped and duplicated reads.

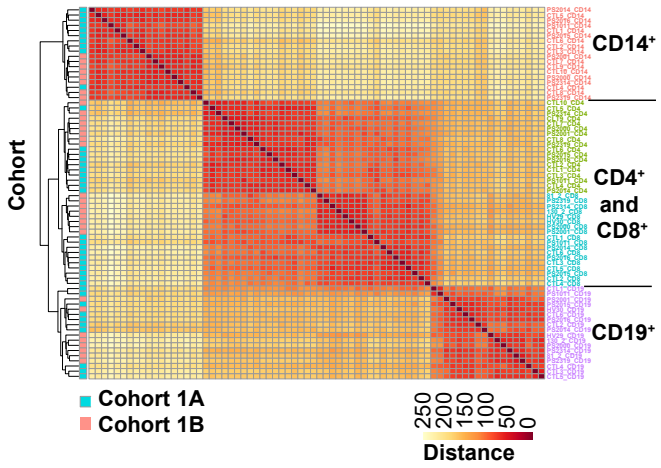


Figure B.10: Distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles for all 72 samples. Distance matrix built based on normalised read counts mapping to 20,493 Ensembl featured remaining after appropriate filtering followed by hierarchical clustering of the samples. Additional annotation of the clustering based on cohort identity is included on the left hand side.

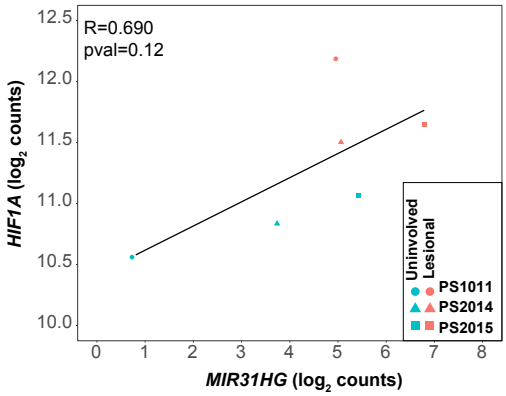


Figure B.11: Correlation in gene expression between the miRs *MIR31HG* and the target gene *HIF-1A* in lesional and uninvolved skin. Plot showing the correlation in $\log_2$ normalised read counts for *MIR31HG* and *HIF-1A*. Pearson correlation values (R) and significance (p-value) are included. Each of the dots represents one samples, where colour represents condition (lesional or uninvolved) and shape corresponds to the patient ID.

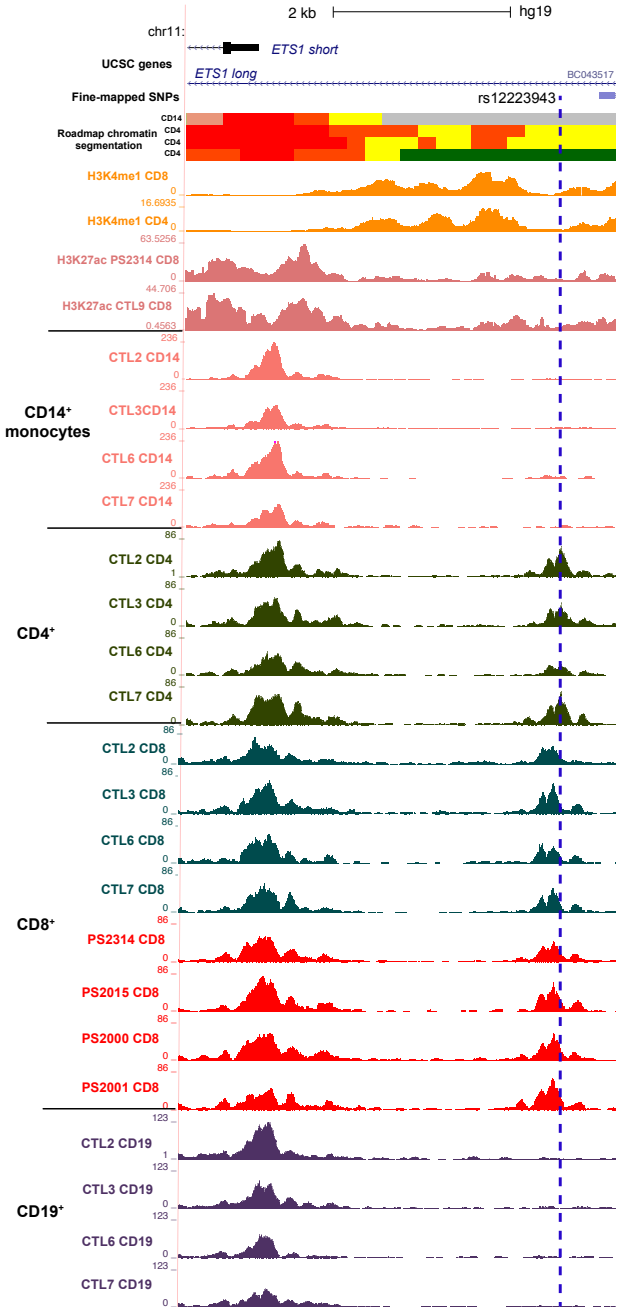
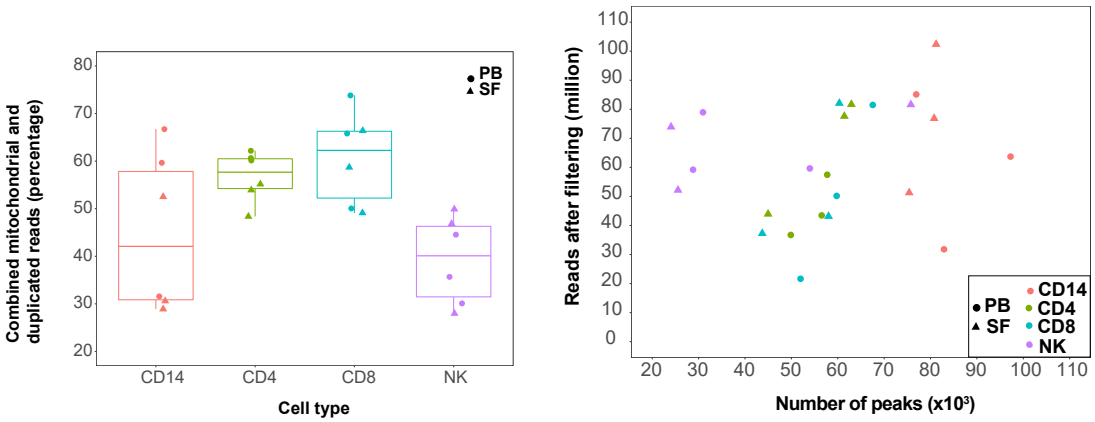


Figure B.12: Epigenetic landscape at the location of the fine-mapped SNP rs12223943 at the *ETS1* psoriasis GWAS locus. UCSC Genome Browser view illustrating normalised ATAC read density and H3K27ac fold-enrichment as well as publicly available epigenetic datasets (chromatin segmentation map and H3K4me1) (y-axis) at the location (x-axis) of the fine-mapped SNP rs12223943 (blue dashed line). Only representative ATAC tracks from control samples are shown for $CD14^+$ monocytes, $CD4^+$ and $CD19^+$ cells.

B.3 Chapter 5 Figures



(A) (B)

Figure B.13: Assessment of the percentage of mitochrondrial and duplicated reads and the number of called peaks after filtering in the ATAC data from peripheral blood and synovial fluid of PsA patients samples. For each of the cell types and samples (A) boxplot representing percentage of duplicated and mitochondrial reads combined and (B) representation of the number of significant peaks after filtering based on IDR optimal p-values versus the total million reads after filtering for each of the samples. For each point, colour codes for cell type and shape for tissue (SF=synovial fluid; PB=peripheral blood).

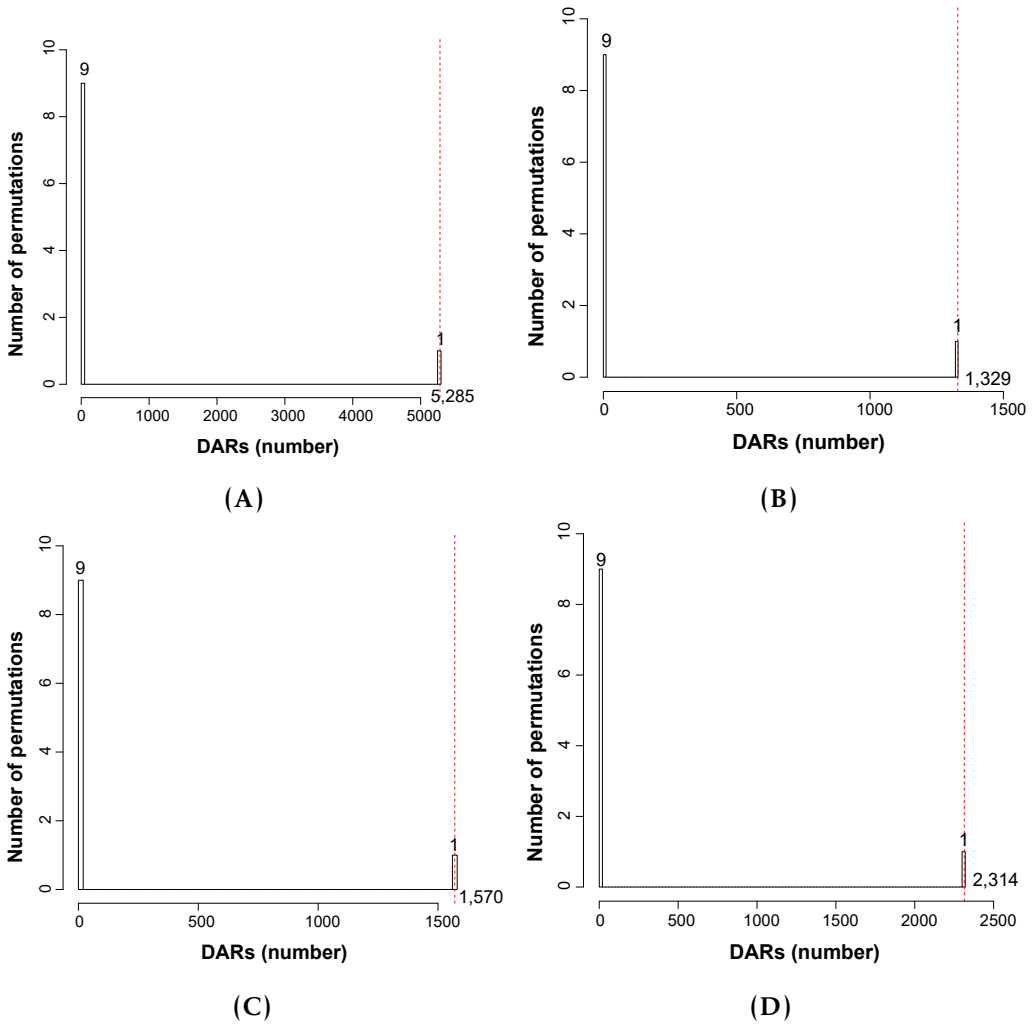


Figure B.14: Permutation analysis for the DARs found in synovial vs peripheral blood in CD14$^+$ monocytes, mCD4$^+$, mCD8$^+$ and NK cells. Sample labels were permuted within each cell type to achieve the ten unique possible combinations and differential analysis was performed. The number of significant DARs (FDR<0.01) across all permutations in plotted for (A) CD14$^+$ monocytes, (B) mCD4$^+$, (C) mCD8$^+$ and (D) NK cells, demonstrating that the true observation (dashed red line) is significantly more than expected by chance (p-value<0.1, the lowest p-value for the maximum number of permutations that can be conducted with this sample size) in all four cell types.

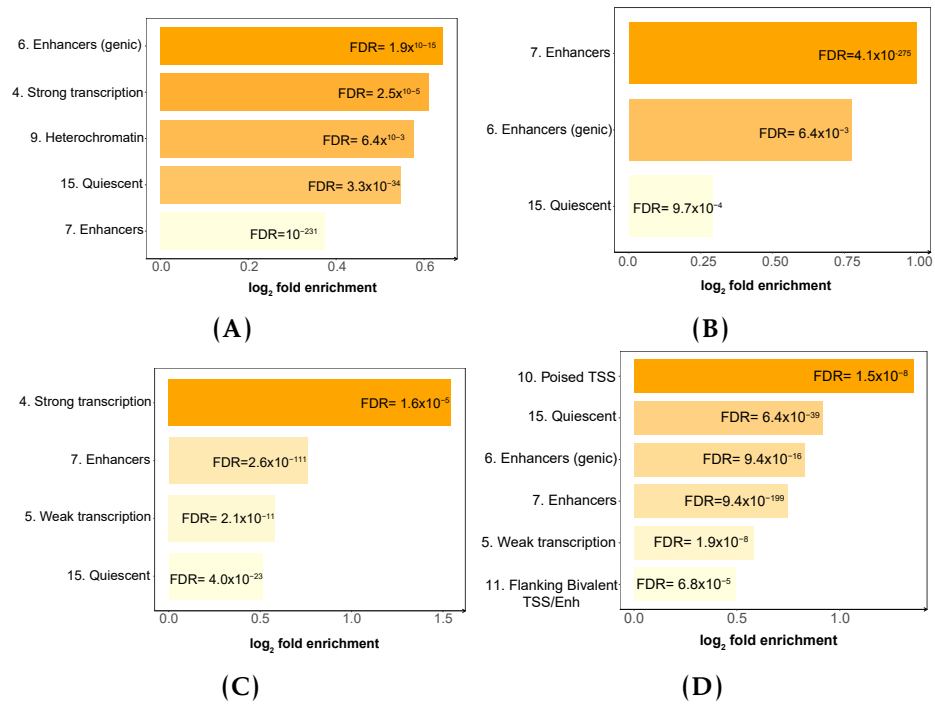


Figure B.15: Enrichment of DARs in the 15 chromatin states of the cell type specific Roadmap Epigenomic Project segmentation maps. Barplots illustrating the chromatin states significantly enriched (FDR<0.01) in DARs from the differential chromatin accessibility analysis between synovial fluid and peripheral blood in (A) $CD14^+$ monocytes, (B) $mCD4^+$, (C) $mCD8^+$ and (D) NK cells. For each of the enriched chromatin states the statistical significance (FDR) and the magnitude (log_2 fold change) of the enrichment is included.

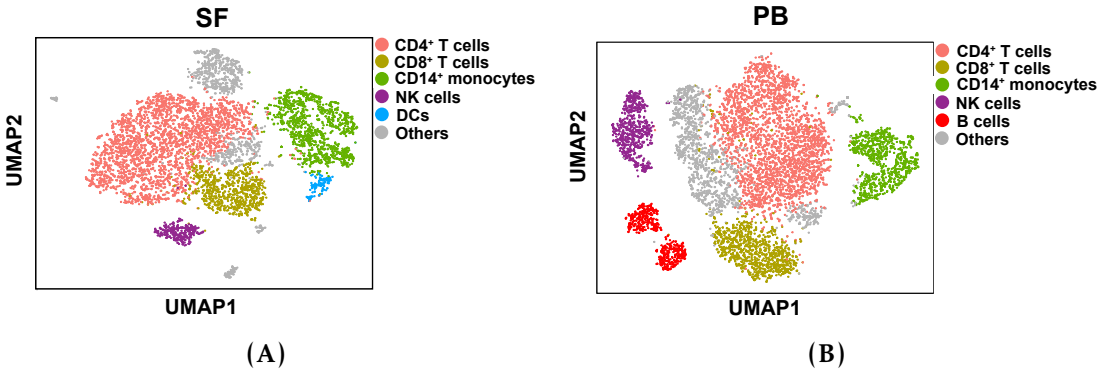


Figure B.16: Identification of the $CD14^+$ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. Visualisation using UMAP dimensional reduction of the cell type populations identified in (A) SFMCs and (B) PBMCs for a representative PsA sample. Clustering performed using default resolution allowed to identify $CD4^+$ (pink), $CD8^+$ (khaki), $CD14^+$ monocytes (green), NK (purple), DCs (blue), B cells (red) and others (grey).

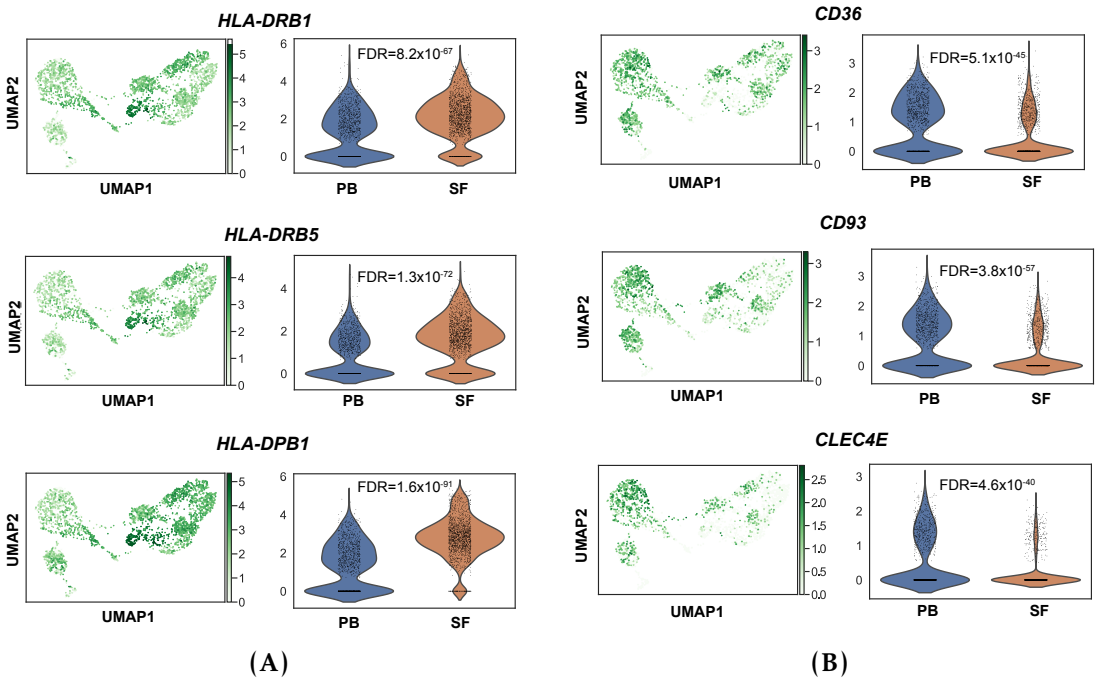


Figure B.17: Expression levels of antigen presentation and phagocytosis characteristic genes in synovial fluid and peripheral blood CD14$^+$ monocytes. Expression intensities of (A) MHC-II genes *HLA-DRB1*, *HLA-DRB5* and *HLA-DPB1* and (B) the phagocytic markers *CD36*, *CD93* and *CLEC4E*. On the left expression of each genes is overlaid on the UMAP dimensional reduction for the combined synovial fluid and peripheral blood CD14$^+$ monocytes analysed by scRNA-seq. On the right, violin plots illustrate for synovial fluid and peripheral blood CD14$^+$ monocytes (x-axis) the probability density of expression (y-axis) for each of the genes. SF=synovial fluid; PB=peripheral blood.

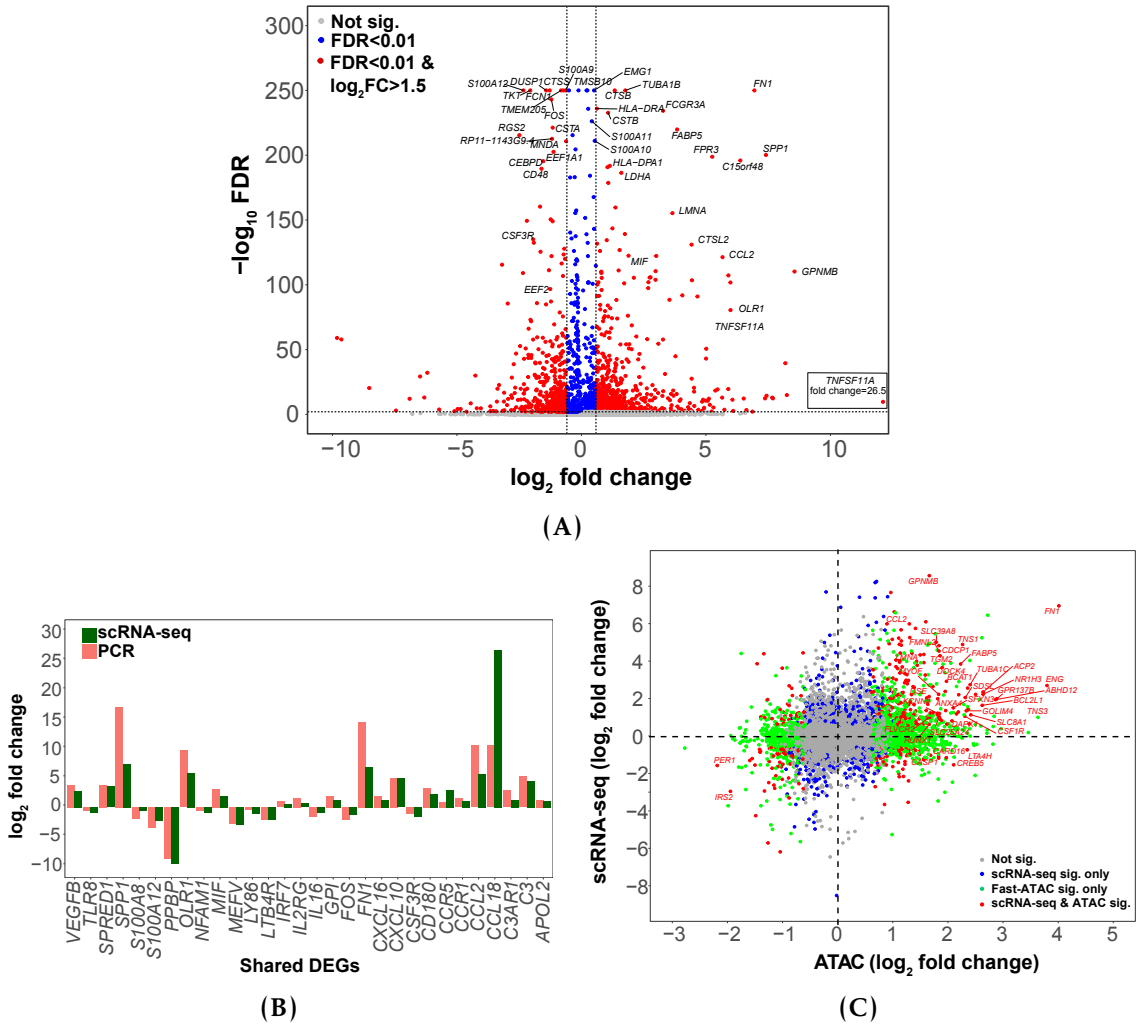


Figure B.18: Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14$^+$ monocytes. (A) Volcano plots showing differences in gene expression between synovial fluid and peripheral blood CD14$^+$ monocyte using scRNA-seq data. The significance ($\log_{10}$FDR) of the differential expression (y-axis) is plotted against the $\log_2$ fold change, where a positive fold change indicates higher expression in CD14$^+$ monocytes from synovial fluid compared to peripheral blood and viceversa. Genes are coloured based on FDR and $\log_2$fold changes and some of the top significant genes are labelled. (B) Comparison of the fold-changes for the shared DEGs between scRNA-seq and qPCR analysis of synovial fluid vs peripheral blood CD14$^+$ monocytes. The barplot illustrates the $\log_2$fold change for the 30 DEGs identified by scRNA-seq (based on FDR<0.01 and fold change>1.5) and qPCR array (based on p-value<0.05) analysis. (C) Correlation plot comparing synovial fluid and peripheral blood differences in scRNA-seq expression and ATAC chromatin accessibility in CD14$^+$ monocytes. The $\log_2$fold changes for scRNA-seq differential expression of all transcripts in CD14$^+$ monocytes are plotted against the $\log_2$fold change for total CD14$^+$ monocytes ATAC differential chromatin accessibility analysis in regions proximal (≤5Kb) to the same genes. Blue colouring indicates significant differential expression in scRNA-seq only; green represents ATAC significant DAR only; red indicates significant differential expression and chromatin accessibility; grey indicates no significant differential expression or chromatin accessibility in CD14$^+$ monocytes. Significance is based on FDR and fold change thresholds (FDR<0.01 and fold change>1.5) in both assays.

Bibliography

Abecasis, G. R., Auton, A., Brooks, L. D., *et al.* (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489: pp. 57–74.

Abji, F, Pollock, RA, Liang, K, *et al.* (2016). "Brief report: CXCL10 is a possible biomarker for the development of psoriatic arthritis among patients with psoriasis". *Arthritis & Rheumatology* 68: pp. 2911–2916.

Abraham, Gad and Inouye, Michael (2014). "Fast principal component analysis of large-scale genome-wide data". *PloS one* 9: e93766.

Abramson, SR. and Khattri, S (2016). "A Review of Current Evidence for TNF Inhibitor Switching in Psoriatic Arthritis". *Journal of Psoriasis and Psoriatic arthtritis* 50: pp. 613–622.

Ackermann, A. M., Wang, Z., Schug, J., *et al.* (2016). "Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes". *Molecular Metabolism* 5: pp. 233–244.

Adams, D., Altucci, L., and Antonarakis S. E. and, Ballesteros -J. (2012). "BLUEPRINT to decode the epigenetic signature written in blood". *Nature Biotechnology* 30: pp. 224–226.

Adey, Andrew, Morrison, Hilary G., Asan, *et al.* (2010). "Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition". *Genome Biology* 11: R119.

Ahn, R, Gupta, R, Lai, K, *et al.* (2016). "Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs". *BMC ...* 17: p. 841.

Ahn, RS, Taravati, Keyon KL, Lee, KM, *et al.* (2017). "Transcriptional landscape of epithelial and immune cell populations revealed through FACS-seq of healthy human skin". *Scientific Reports* 7: p. 1343.

Ai, R., Hammaker, D., Boyle, D. L., *et al.* (2016). "Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes". *Nature Communications* 10: p. 11849.

Al-Mossawi, H., Yager, N., Taylor, C., *et al.* (2018). "Context-specific regulation of monocyte surface IL7R expression and soluble receptor secretion by a common autoimmune risk allele". *bioRxiv*: p. 262410.

Al-Mossawi, M. H., Chen, L., Fang, H., *et al.* (2017). "Unique transcriptome signatures and GM-CSF expression in lymphocytes from patients with spondyloarthritis". *Nature Communications* 8: p. 1510.

Alasoo, K., Rodrigues, J., and Mukhopadhyay S. and, Knights A. (2018). "Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response". *Nature Genetics* 50: pp. 424–431.

Allen, M. and Barker, J. (2003). "The International Psoriasis Genetics Study: assessing linkage to 14 candidate susceptibility loci in a cohort of 942 affected sib pairs". *The American Journal of Human Genetics* 73: pp. 430–437.

Alrefai, H, Muhammad, K, and Rudolf, R (2016). "NFATc1 supports imiquimod-induced skin inflammation by suppressing IL-10 synthesis in B cells". *Nature Communication*.

Amatullah, H, Shan, Y, and Beauchamp, B. (2017). "DJ-1/PARK7 impairs bacterial clearance in sepsis". *American journal of respiratory and critical care medicine* 195: pp. 889–905.

Ameres, S. L., Horwich, M. D., Hung, J. H., *et al.* (2010). "Target RNA–directed trimming and tailing of small silencing RNAs". *Science* 328: pp. 1534–1539.

Anders, S., Pyl, PT., and Huber, W. (2015). "HTSeq—a Python framework to work with high-throughput sequencing data". *Bioinformatics* 31: pp. 166–169.

Andersen, William M. Gordon, Michael, D. Zeller, Rachel, H. Klein, *et al.* (2014). "A GRHL3-regulated repair pathway suppresses immune-mediated epidermal hyperplasia". *PLOS Genetics*.

Anderson, Carl A, Boucher, Gabrielle, Lees, Charlie W, *et al.* (2011). "Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47". 43: p. 246.

Andersson, Robin, Gebhard, Claudia, Miguel-Escalada, Irene, *et al.* (2014). "An atlas of active enhancers across human cell types and tissues". *Nature* 507: p. 455.

Andrews, Simon (2010). "FastQC: a quality control tool for high throughput sequence data." *http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Andrews, TS. (2018). "False signals induced by single-cell imputation". *F1000Research*.

Antonelli, Alessandro, Ferrari, Silvia, Giuggioli, Dilia, *et al.* (2014). "Chemokine (CXCL motif) ligand (CXCL) 10 in autoimmune diseases". *Autoimmunity Reviews* 13: pp. 272–280.

Antonioli, L., Pacher, P., Vizi, E. S., *et al.* (2013). "CD39 and CD73 in immunity and inflammation". *Trends in molecular medicine* 16: pp. 355–367.

Antoniv, Taras T and Ivashkiv, Lionel B (2006). "Dysregulation of interleukin 10dependent gene expression in rheumatoid arthritis synovial macrophages". 54: pp. 2711–2721.

Apel, M., Uebe, S., and Bowes J. Giardina, E. (2013). "Variants in RUNX3 contribute to susceptibility to psoriatic arthritis, exhibiting further common ground with ankylosing spondylitis". *Arthritis & Rheumatology* 65: pp. 1224–1231.

Appel, H., Maier, R., and Wu, P. (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 research and therapy* 13: R95.

Arend, W. P., Palmer, G., and Gabay, C. (2008). "IL-1, IL-18, and IL-33 families of cytokines". *Immunological reviews* 223: pp. 20–38.

Armstrong, AW, Harskamp CT and, Dhillon JS., and EJ., Armstrong (2014). "Psoriasis and smoking: a systematic review and meta-analysis". *British journal of Dermatology* 170: pp. 229–484.

Arnold, Cosmas D, Gerlach, Daniel, Stelzer, Christoph, *et al.* (2013). "Genome-wide quantitative enhancer activity maps identified by STARR-seq." *Science* 339: pp. 1074–7.

Assassi, S., Reveille, J. D., Arnett, F. C., *et al.* (2010). "Whole-blood gene expression profiling in ankylosing spondylitis shows upregulation of toll-like receptor 4 and 5". *The Journal of Rheumatology* 38: pp. 87–98.

Aterido, Adrià, Julià, Antonio, Ferrándiz, Carlos, *et al.* (2016). "Genome-Wide Pathway Analysis Identifies Genetic Pathways Associated with Psoriasis." *J. Invest. Dermatol.* 136: pp. 593–602.

Attar, M., Sharma, E., Li, S., *et al.* (2018). "A practical solution for preserving single cells for RNA sequencing". *Scientific reports*.

Axel, A. (1975). "Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease". *Biochemistry* 14: pp. 2921–2925.

Baba, T and Mukaida, N. (2014). "Role of macrophage inflammatory protein (MIP)-1/CCL3 in leukemogenesis". *Molecular & cellular oncology* 15: e29899.

Baechler, EC., Batliwalla, FM., Karypis., *et al.* (2003). "Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus". *Proc Natl Acad Sci U S A* 100: pp. 2610–2615.

Baeten, D., Breban, M., Lories, R., *et al.* (2013). "Are spondylarthritides related but distinct conditions or a single disease with a heterogeneous phenotype?" *Arthritis & Rheumatism* 65: pp. 12–20.

Bai, F, Zheng, W, Dong, Y, *et al.* (2018). "Serum levels of adipokines and cytokines in psoriasis patients: a systematic review and meta-analysis". *Oncotarget* 9: pp. 1266–1278.

Baker, K. F. and D., Isaacs -J. (2017). "Novel therapies for immune-mediated inflammatory diseases: What can we learn from their use in rheumatoid arthritis, spondyloarthritis, systemic lupus erythematosus, psoriasis, Crohn's disease and ulcerative colitis?" *Annals of the rheumatic diseases* 77: pp. 175–187.

Bandura, D. R., Baranov, V. I., and Analytical …, Ornatsky -O. I. (2009). "Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry". *Analytical …*

Bannister, AJ. and Kouzarides, T. (2011). "Regulation of chromatin by histone modifications". *Cell research* 21: pp. 381–395.

Bannister, AJ, Zegerman, P, Partridge, JF, *et al.* (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain". *Nature* 410: pp. 120–124.

Bao, Xiaomin, Rubin, Adam J, Qu, Kun, *et al.* (2015). "A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63". 16: p. 284.

Baran-Gale, J., Chandra, T., and Kirschner, K. (2017). "Experimental design for single-cell RNA sequencing". *Briefings in functional genomics* 17: pp. 233–239.

Barski, A., Cuddapah, S., Cui, K., *et al.* (2007). "High-resolution profiling of histone methylations in the human genome". *Cell* 129: pp. 823–837.

Batliwalla, F. M., Li, W., Ritchlin, C. T., *et al.* (2005). "Microarray analyses of peripheral blood cells identifies unique gene expression signature in psoriatic arthritis". *Molecular Medicine* 11: pp. 21–29.

Belasco, J, Louie, JS, Gulati, N, *et al.* (2015). "Comparative genomic profiling of synovium versus skin lesions in psoriatic arthritis". *Arthritis & ...*

Bell, A. C. and Felsenfeld -G. (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene". *Nature* 405: pp. 482–485.

Bengsch, B., Ohtani, T., Herati, R. S., *et al.* (2018). "Deep immune profiling by mass cytometry links human T and NK cell differentiation and cytotoxic molecule expression patterns". *Journal of Immunological Methods* 453: pp. 3–10.

Benham, H., Norris, P., Goodall, J., *et al.* (2013). "Th17 and Th22 cells in psoriatic arthritis and psoriasis". *Arthritis Research & Therapy* 15: R136.

Benjamin, Michael and McGonagle, Dennis (2009). *The enthesis organ concept and its relevance to the spondyloarthropathies*. The enthesis organ concept and its relevance to the spondyloarthropathies. Springer: pp. 57–70.

Bentham, J, Morris, DL, Graham, DSC, *et al.* (2015). "Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus". *Nature Genetics* 47: pp. 1457–1464.

Bergboer, J. G. M., Tjabringa, G. S., Kamsteeg, M., *et al.* (2011). "Psoriasis risk genes of the late cornified envelope-3 group are distinctly expressed compared with genes of other LCE groups". *The American journal of Pathology* 178: pp. 1470–1477.

Bernstein, B. E., Stamatoyannopoulos J. A.and Costello, JF., Ren, B., *et al.* (2010). "The NIH roadmap epigenomics mapping consortium". *Nature Biotechnology* 20: pp. 1045–1048.

Bhatt, Tarun K, Soni, Rani, and Sharma, Drista (2016). "Recent Updates on DTD (D-Tyr-tRNATyr Deacylase): An Enzyme Essential for Fidelity and Quality of Protein Synthesis". 4: p. 32.

Biniecka, M., Canavan, M., McGarry, T., *et al.* (2016). "Dysregulated bioenergetics: a key regulator of joint inflammation". *Annals of the Rheumatic Diseases* 75: pp. 2192–2200.

Birzele, F, Fauti, T, and Stahl, H andLenter MC (2011). "Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in Human". *Nucleic acids research* 39: pp. 7946–7960.

Black, A. P. B. and Ardern-Jones, M. R. (2007). "Human keratinocyte induction of rapid effector function in antigen-specific memory CD4+ and CD8+ T cells". *European journal of immunology* 37: pp. 1485–1493.

Blais, M. E., Dong, T., and Rowland-Jones, S. (2011). "HLA-C as a mediator of natural killer and T-cell activation: spectator or key player?" *Immunology* 133: pp. 1–7.

Blauvelt, A., Papp, K. A., Griffiths, C. E. M., *et al.* (2017). "Efficacy and safety of guselkumab, an anti-interleukin-23 monoclonal antibody, compared with adalimumab for the continuous treatment of patients with moderate to severe psoriasis: Results from the phase III, double-blinded, placebo- and active comparator-controlled VOYAGE 1 trial." *Journal of the American Academy of Dermatology* 76: pp. 405–417.

Bonifati, V, Rizzu, P, Baren, Van MJ, *et al.* (2003). "Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism". *Science* 299: pp. 256–9.

Borgato, L., Puccetti, A., Beri, R., *et al.* (2002). "The T cell receptor repertoire in psoriatic synovitis is restricted and T lymphocytes expressing the same TCR are present in joint and skin lesions". *J Rheumatol* 29: 1914–19.

Bouaziz, JD, Yanaba, K, Venturi, GM., *et al.* (2007). "Therapeutic B cell depletion impairs adaptive and autoreactive CD4+ T cell activation in mice". *Proc Natl Acad Sci U S A* 104: pp. 20878–20883.

Boudousquié, Caroline, Danilo, Maxime, Pousse, Laurène, *et al.* (2014). "Differences in the transduction of canonical Wnt signals demarcate effector and memory CD8 T cells with distinct recall proliferation capacity". *Journal of immunology (Baltimore, Md. : 1950)* 193: pp. 2784–2791.

Bowes, J., Ho, P., Flynn, E., *et al.* (2012). "Comprehensive assessment of rheumatoid arthritis susceptibility loci in a large psoriatic arthritis cohort". *Annals of the Rheumatic Diseases* 71: pp. 1350–1354.

Bowes, J., Budu-Aggrey, A., Huffmeier, U., *et al.* (2015a). "Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis". *Nature communications* 6: p. 6046.

Bowes, J., Loehr, S., and Budu-Aggrey, A. (2015b). "PTPN22 is associated with susceptibility to psoriatic arthritis but not psoriasis: evidence for a further PsA-specific risk locus". *Annals of the Rheumatic Diseases* 74: pp. 1882–1885.

Bowes, J., Ashcroft, J., Dand, N., *et al.* (2017). "Cross-phenotype association mapping of the MHC identifies genetic variants that differentiate psoriatic arthritis from psoriasis". *Annals of the Rheumatic Diseases* 76: pp. 1774–1779.

Boyle, A. P., Song, L., Lee, B. K., *et al.* (2010). "High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells". *Genome Research* 21: pp. 456–464.

Boyle, A. P., Hong, E. L., Hariharan, M., *et al.* (2012). "Annotation of functional variation in personal genomes using RegulomeDB". *Genome Res* 22: pp. 1790–7.

Boyman, O., Hefti, H. P., Conrad, C., *et al.* (2004). "Spontaneous development of psoriasis in a new animal model shows an essential role for resident T cells and tumor necrosis factor-alpha". *J Exp Med* 199: pp. 731–6.

Braun, J. E., Huntzinger, E., Fauser, M., *et al.* (2011). "GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets". *Molecular cell* 44: pp. 120–133.

Brest, P., Lapaquette, P., Souidi, M., *et al.* (2011). "A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease". *Nature Genetcis* 43: pp. 242–245.

Broome, AM, Ryan, D, and, Eckert RL (2003). "S100 protein subcellular localization during epidermal differentiation and psoriasis". *Journal of Histochemistry of cytochemistry* 51: pp. 675–85.

Buck, M. J., Raaijmakers, L. M., Ramakrishnan, S., *et al.* (2014). "Alterations in chromatin accessibility and DNA methylation in clear cell renal cell carcinoma". *Oncogene* 33: pp. 4961–65.

Buenrostro, Jason D., Paul, G. Giresi, Lisa, C. Zaba, *et al.* (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". *Nature Methods* 10: pp. 1213–1218.

Buenrostro, Jason D., Wu, Beijing, Chang, Howard Y., *et al.* (2015). "ATAC seq: A Method for Assaying Chromatin Accessibility Genome Wide". *Current Protocols in Molecular Biology* 109: pp. 1–9.

Bunt, M. van de, Cortes, A., Brown, M. A., *et al.* (2015). "Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci". *PLoS Genet* 11: pp. 1–14.

Burgos-Pol, R. and Martínez-Sesmero, J. M. (2016). "The cost of psoriasis and psoriatic arthritis in 5 European countries: a systematic review". *Actas Dermosifiliograficas* 107: pp. 577–590.

Burton P., WTCCC2 (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls". *Nature* 447: pp. 661–678.

Calderon, D, Nguyen, MLT, Mezger, A, *et al.* (2018). "Landscape of stimulation-responsive chromatin across diverse human immune cells". *bioRxiv*.

Cameron, A. L. and Kirby, B. (2003). "Circulating natural killer cells in psoriasis". *British Journal of Dermtology* 149: pp. 160–164.

Cañete, JD, Martínez, SE, Farrés, J, *et al.* (2000). "Differential Th1/Th2 cytokine patterns in chronic arthritis: interferon is highly expressed in synovium of rheumatoid arthritis compared with seronegative ..." *Annals of the Rheumatic Diseases* 59: pp. 263–268.

Cañete, JD, Celis, R, Hernández, V, *et al.* (2010). "Synovial immunopathological changes associated with successful abatacept therapy in a case of severe refractory psoriatic arthritis". *Annals of the Rheumatic Diseases* 69: pp. 935–936.

Cao, Qin, Anyansi, Christine, Hu, Xihao, *et al.* (2017). "Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines." *Nature Genetics* 49: pp. 1428–1436.

Capon, Francesca (2017). "The Genetic Basis of Psoriasis". *International Journal of Molecular Sciences* 18: p. 2526.

Capon, Francesca, Bijlmakers, Marie-José J., Wolf, Natalie, *et al.* (2008). "Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene". *Human molecular genetics* 17: pp. 1938–1945.

Cargill, Michele, Schrodi, Steven J., Chang, Monica, *et al.* (2007). "A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes". *The American Journal of Human Genetics* 80: pp. 273–290.

Carrier, Y, Ma, HL, Ramon, HE, *et al.* (2011). "Inter-regulation of Th17 cytokines and the IL-36 cytokines in vitro and in vivo: implications in psoriasis pathogenesis". *Journal of Investigative Dermatology* 131: pp. 2428–2437.

Carrieri, Claudia, Cimatti, Laura, Biagioli, Marta, *et al.* (2012). "Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat". *Nature* 491: pp. 454–457.

Carter, K. W., Pluzhnikov, A., Timms, A. E., *et al.* (2007). "Combined analysis of three whole genome linkage scans for ankylosing spondylitis". *Rheumatology* 46: pp. 763–771.

Chandra, Aditi, Senapati, Swapan, Roy, Sudipta, *et al.* (2018). "Epigenome-wide DNA methylation regulates cardinal pathological features of psoriasis." *Clin Epigenetics* 10: p. 108.

Chandran, Vinod, Schentag, Catherine T., Brockbank, John E., *et al.* (2009). "Familial aggregation of psoriatic arthritis". *Annals of the rheumatic diseases* 68: pp. 664–667.

Chang, C. C., Chow, C. C., Tellier, LC., *et al.* (2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets". *Gigascience*: pp. 4–7.

Chang, JC, Smith, LR, and . . ., Froning KJ of the (1994). "CD8+ T cells in psoriatic lesions preferentially use T-cell receptor V beta 3 and/or V beta 13.1 genes". *Proc Natl Acad Sci U S A*.

Chang, X, Yamada, R, Suzuki, A, *et al.* (2005). "Citrullination of fibronectin in rheumatoid arthritis synovial tissue". *Rheumatology Oxford* 44: pp. 1374–1382.

Chen, Xingqi, Shen, Ying, Draper, Will, *et al.* (2016). "ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing". *Nature methods* 13: p. 1013.

Cheng, Jeffrey B, Sedgewick, Andrew J, Finnegan, Alex I, *et al.* (2018). "Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution." *Cell Rep* 25: pp. 871–883.

Chiu, YG, Shao, T, Feng, C, *et al.* (2010). "CD16 (FcRIII) as a potential marker of osteoclast precursors in psoriatic arthritis". *Arthritis Research and Threapy* 12: R14.

Chou, W. C., Zheng, H. F., Cheng, C. H., *et al.* (2016). "A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples". *Scientific reports*.

Chu, CQ, Swart, D, Alcorn, D, *et al.* (2007). "Interferon- regulates susceptibility to collagen-induced arthritis through suppression of interleukin-17". *Arthritis & Rheumatisms* 56: pp. 1145–1151.

Churov, A. V., Oleinik, E. K., and Knip, M. (2015). "MicroRNAs in rheumatoid arthritis: altered expression and diagnostic potential". *Autoimmunity reviews* 14: pp. 1029–1037.

Cid, De R, E, Riveira-Munoz, Zeeuwen, PL., *et al.* (2009). "Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis". *Nature Genetics* 41: pp. 211–215.

Ciocon, D.H. and Kimball, A.B (2007). "Psoriasis and psoriatic arthritis: separate or one and the same?" *British Journal of Dermatology* 157: pp. 850–860.

Clegg, Daniel O., Reda, Domenic J., Mejias, Edwin, *et al.* (1996). "Comparison of sulfasalazine and placebo in the treatment of psoriatic arthritis. A Department of Veterans Affairs Cooperative Study". *Arthritis & Rheumatology* 39: pp. 2013–2020.

Coates, L. C., FitzGerald, O., Helliwell, P. S., *et al.* (2016a). "Psoriasis, psoriatic arthritis, and rheumatoid arthritis: Is all inflammation the same?" *Seminars in arthritis and rheumatism* 46: pp. 291–304.

Coates, Laura C., Kavanaugh, Arthur, Mease, Philip J., *et al.* (2016b). "Group for research and assessment of psoriasis and psoriatic arthritis 2015 treatment recommendations for psoriatic arthritis". *Arthritis & rheumatology* 68: pp. 1060–1071.

Coda, A. B., Icen, M., Smith, J. R., *et al.* (2012). "Global transcriptional analysis of psoriatic skin and blood confirms known disease-associated pathways and highlights novel genomic "hot spots" for . . . " *Genomics* 100: pp. 18–26.

Cohen, A. D., Sherf, M., Vidavsky, L., *et al.* (2008). "Association between psoriasis and the metabolic syndrome". *Dermatology*.

Constant, S, Schweitzer, N, West, J., *et al.* (1995). "B lymphocytes can be competent antigen-presenting cells for priming CD4+ T cells to protein antigens in vivo." *The Journal of Immunology* 155: pp. 3734–3741.

Corces, MR, Trevino, AE, Hamilton, EG, *et al.* (2017). "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues". *Nature Methods* 14: pp. 959–962.

Corces, Ryan M., Buenrostro, Jason D., Wu, Beijing, *et al.* (2016). "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution". *Nature Genetics* 48: pp. 1193–1203.

Cornell, T. T., Rodenhouse, P., Cai, Q., *et al.* (2010). "Mitogen-activated protein kinase phosphatase 2 regulates the inflammatory response in sepsis". *Infection and Immunity* 78: pp. 2868–2876.

Cortes, A., Pulit, S. L., Leo, P. J., *et al.* (2015). "Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1". *Nature Communications*.

Cortes, Adrian and Brown, Matthew A. (2011). "Promise and pitfalls of the Immunochip". *Arthritis research & therapy* 13: p. 101.

Cortes, Adrian, Hadler, Johanna, Pointon, Jenny P., *et al.* (2013). "Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci". *Nature genetics* 45: p. 730.

Cosma, M. P., Tanaka, T., and Nasmyth -K. (1999). "Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle–and developmentally regulated promoter". *Cell* 97: pp. 299–311.

Costantino, F., Talpin, A., Evnouchidou, I., *et al.* (2015). "ERAP1 gene expression is influenced by nonsynonymous polymorphisms associated with predisposition to spondyloarthritis". *Arthritis and Rehumatology* 67: pp. 1525–1534.

Costello, P., Bresnihan, B., and of . . ., O'Farrelly -C. (1999). "Predominance of CD8+ T lymphocytes in psoriatic arthritis". *The Journal of . . .*

Creyghton, M. P., W., Cheng -A., Welstead, GG., *et al.* (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state". *Proc Natl Acad Sci U S A* 107: pp. 21931–21936.

Cui, Y., Li, G., Li, S., *et al.* (2010). "Designs for linkage analysis and association studies of complex diseases". *Statistical Methods in Molecular Biology* 620: pp. 219–242.

Cusanovich, Darren A, Daza, Riza, Adey, Andrew, *et al.* (2015). "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing". 348: pp. 910–914.

Dalbeth, N and Callan, MF. (2002). "A subset of natural killer cells is greatly expanded within inflamed joints". *Arthritis & Rheumatism* 46: pp. 1763–1772.

Dand, N., Mucha, S., Tsoi, L. C., *et al.* (2017). "Exome-wide association study reveals novel psoriasis susceptibility locus at TNFSF15 and rare protective alleles in genes contributing to type I IFN signalling". *Human molecular genetics* 26: pp. 4301–4313.

Dang, EV, Barbi, J, Yang, HY, *et al.* (2011). "Control of TH17/Treg Balance by Hypoxia-Inducible Factor 1". *Cell* 146: pp. 772–784.

Das, Sayantan, Stuart, Philip E., Ding, Jun, *et al.* (2014). "Fine mapping of eight psoriasis susceptibility loci". *European Journal of Human Genetics* 23: pp. 844–853.

Davies, James O., Oudelaar, A. M., Higgs, Douglas R., *et al.* (2017). "How best to identify chromosomal interactions: a comparison of approaches". *Nature methods* 14: pp. 125–134.

De Santa, Francesca, Barozzi, Iros, Mietton, Flore, *et al.* (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers". *PLoS biology* 8.

Delaneau, O., Marchini, J., and F., Zagury -J. (2012). "A linear complexity phasing method for thousands of genomes". *Nature methods* 9: pp. 179–181.

Dellon, Evan S, Selitsky, Sara R, Genta, Robert M, *et al.* (2018). "Gene expression-phenotype associations in adults with eosinophilic esophagitis." *Dig Liver Dis* 50: pp. 804–811.

Derrien, Thomas, Johnson, Rory, Bussotti, Giovanni, *et al.* (2012). "The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression". *Genome Research* 22: pp. 1775–1789.

Dijk, Van D, Sharma, R, Nainys, J, *et al.* (2018). "Recovering Gene Interactions from Single-Cell Data Using Data Diffusion". *Cell* 174: pp. 716–729.

Diluvio, L., Vollmer, S., Besgen, P., *et al.* (2006). "Identical TCR -chain rearrangements in streptococcal angina and skin lesions of patients with psoriasis vulgaris". *The Journal of Immunology* 176: pp. 7104–7111.

Dimas, Antigone S., Deutsch, Samuel, Stranger, Barbara E., *et al.* (2009). "Common regulatory variation impacts gene expression in a cell typedependent manner". *Science* 325: pp. 1246–1250.

Ding, Qiurong, Regan, Stephanie N., Xia, Yulei, *et al.* (2013). "Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs". *Cell stem cell* 12: pp. 393–394.

Dobin, A., Davis, C.A., Schlesinger, F., *et al.* (2013). "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics* 29: pp. 15–21.

Dolcino, M, Pelosi, A, Fiore, PF, *et al.* (2018). "Long non coding RNAs play a role in the pathogenesis of Psoriatic Arthritis by regulating microRNAs and genes involved in inflammation and metabolic syndrome." *Frontiers in Immunology* 9: p. 1533.

Dolcino, Marzia, Ottria, Andrea, Barbieri, Alessandro, *et al.* (2015). "Gene Expression Profiling in Peripheral Blood Cells and Synovial Membranes of Patients with Psoriatic Arthritis." *PLoS ONE* 10: e0128262.

Doyle, MS., Collins, ES., FitzGerald, OM., *et al.* (2012). "New insight into the functions of the interleukin-17 receptor adaptor protein Act1 in psoriatic arthritis". *Arthritis Research and Therapy* 14: p. 226.

Dozmorov, M. G., Dominguez, N., Bean, K., *et al.* (2015). "B-cell and monocyte contribution to systemic lupus erythematosus identified by cell-type-specific differential expression analysis in RNA-seq data". *Bioinformatics and biology insights* 9: pp. 9–11.

Dubois, PCA, Trynka, G, Franke, L, *et al.* (2010). "Multiple common variants for celiac disease influencing immune gene expression". *Nature genetics* 42: pp. 295–302.

Duffy, D. L., Spelman, L. S., and G., Martin -N. (1993). "Psoriasis in Australian twins". *Journal of the American Academy of Dermatology* 29: pp. 428–434.

Duren, Zhana, Chen, Xi, Zamanighomi, Mahdi, *et al.* (2018). "Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations." *Proc Natl Acad Sci U S A* 115: pp. 7723–7728.

Dustin, Michael L. (2001). "Role of adhesion molecules in activation signaling in T lymphocytes". *Journal of clinical immunology* 21: pp. 258–263.

Eckert, RL, Broome, AM, Ruse, M, *et al.* (2004). "S100 proteins in the epidermis". *Journal of Investigative Dermatology* 123: pp. 23–33.

Eder, L, Law, T, Chandran, V., *et al.* (2011). "Association between environmental factors and onset of psoriatic arthritis in patients with psoriasis". *Arthritis care & Research* 63: pp. 1091–1097.

Eder, L., Chandran, V., Pellet, F., *et al.* (2012). "Human leucocyte antigen risk alleles for psoriatic arthritis among patients with psoriasis". *Annals of the Rheumatic Diseases* 71: pp. 50–55.

Edwards, Alexis C, Bigdeli, Tim B, Docherty, Anna R, *et al.* (2015). "Meta-analysis of positive and negative symptoms reveals schizophrenia modifier genes". 42: pp. 279–287.

Edwards, S. L., Beesley, J., French, J. D., *et al.* (2013). "Beyond GWASs: illuminating the dark road from association to function". *Am J Hum Genet* 93: pp. 779–97.

Eedy, D., Burrows, D., Bridges, J., *et al.* (1990). "Clearance of severe psoriasis after allogenic bone marrow transplantation". *BMJ* 300.

Elkayam, O., Ophir, J., Yaron, M., *et al.* (2000). "Psoriatic arthritis: interrelationships between skin and joint manifestations related to onset, course, and differentiation." *Clinica Rheum* 19. 301–5.

Ellinghaus, David, Jostins, Luke, Spain, Sarah L., *et al.* (2016). "Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci". *Nature Genetics* 48: pp. 510–518.

Ellinghaus, E., Ellinghaus, D., Stuart, P. E., *et al.* (2010). "Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2". *Nature genetics* 42: pp. 991–996.

Ellinghaus, Eva, Stuart, Philip E., Ellinghaus, David, *et al.* (2012). "Genome-wide meta-analysis of psoriatic arthritis identifies susceptibility locus at REL". *Journal of Investigative Dermatology* 132: pp. 1133–1140.

ENCODE (2007). "Identification and analysis of functional elements in 1the human genome by the ENCODE pilot project". *Nature* 447: p. 799.

Eppinga, H. and Konstantinov, SR. (2014). "The microbiome and psoriatic arthritis". *Current rheumatology reports* 16: p. 407.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., *et al.* (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types". *Nature* 473: pp. 43–49.

Ernst, Jason and Kellis, Manolis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome". *Nat Biotechnol* 28: pp. 817–825.

Evans, D. M., Spencer, C. C. A., Pointon, J. J., *et al.* (2011). "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility". *Nature Genetics* 43: pp. 761–767.

Eyerich, S, Eyerich, K, and ..., Pennino D of (2009). "Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling". *The Journal of ...*

Fabregat, Antonio, Jupe, Steven, Matthews, Lisa, *et al.* (2018). "The reactome pathway knowledgebase". *Nucleic acids research* 46.

Faghihi, Mohammad A., Modarresi, Farzaneh, Khalil, Ahmad M., *et al.* (2008). "Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase". *Nature medicine* 14: pp. 723–730.

Fagny, M., Paulson, J. N., L., Kuijjer -M., *et al.* (2017). "Exploring regulation in tissues with eQTL networks". *Proc Natl Acad Sci U S A* 114: e7841–7850.

Fairfax, Benjamin P., Makino, Seiko, Radhakrishnan, Jayachandran, *et al.* (2012). "Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles". *Nature genetics* 44: pp. 502–510.

Fairfax, Benjamin P., Humburg, Peter, Makino, Seiko, *et al.* (2014). "Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression". *Science* 343: p. 1246949.

Fang, Hai, Knezevic, Bogdan, Burnham, Katie L., *et al.* (2016). "XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits". *Genome Medicine* 8: p. 129.

Farber, E. M., Nall, M. L., and Watson -W. (1974). "Natural history of psoriasis in 61 twin pairs". *Archives of dermatology* 109: pp. 207–211.

Farh, Kyle, Carrasco-Alfonso, J., Dita, Mayer, *et al.* (2015). "Genetic and epigenetic fine mapping of causal autoimmune disease variants". *Nature* 518: 337–343.

Feil, Robert and Fraga, Mario F. (2012). "Epigenetics and the environment: emerging patterns and implications". *Nature Reviews Genetics* 13: p. 97.

Feldmeyer, L, Keller, M, Niklaus, G, *et al.* (2007). "The inflammasome mediates UVB-induced activation and secretion of interleukin-1 by keratinocytes". *Current Biology* 17: pp. 1140–1145.

Ferby, Ingvar, Reschke, Markus, Kudlacek, Oliver, *et al.* (2006). "Mig6 is a negative regulator of EGF receptormediated skin morphogenesis and tumor formation". 12: p. 568.

Ferlazzo, G, Pack, M, and Thomas, D (2004). "Distinct roles of IL-12 and IL-15 in human natural killer cell activation by dendritic cells from secondary lymphoid organs". *Proc Natl Acad Sci U S A* 101: pp. 16606–16611.

Fert, I., Cagnard, N., and Glatigny S. and, Letourneur -F. (2014). "Reverse interferon signature is characteristic of antigen-presenting cells in human and rat spondyloarthritis". *Arthritis & Rheumatology* 66: pp. 841–851.

Filková, M., Aradi, B., Šenolt, L., *et al.* (2014). "Association of circulating miR-223 and miR-16 with disease activity in patients with early rheumatoid arthritis". *Annals of the rehumatic diseases* 73: pp. 1898–1904.

Finlay, Andrew (2005). "Current severe psoriasis and the rule of tens". *British Journal of Dermatology* 152: pp. 861–867.

Finzel, S., Sahinbegovic, E., and Kocijan -R. (2014). "Inflammatory bone spur formation in psoriatic arthritis is different from bone spur formation in hand osteoarthritis". *Arthritis & Rheumatology* 66: pp. 2968–2975.

Fiocco, Ugo, Martini, Veronica, Accordi, Benedetta, *et al.* (2015). "Transcriptional network profile on synovial fluid T cells in psoriatic arthritis". *Clin Rheumatol* 34: pp. 1571–1580.

Forrest, Alistair R. R., Kawaji, Hideya, Rehli, Michael, *et al.* (2014). "A promoter-level mammalian expression atlas". *Nature* 507.

Forsythe, Jo A, Jiang, Bing-Hua, Iyer, Narayan V, *et al.* (1996). "Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1." 16: pp. 4604–4613.

Franke, A, DPB, McGovern, Barrett, JC, *et al.* (2010). "Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci". *Nature Genetics* 42: pp. 1118–1125.

Fredriksson, T. and Pettersson, U. (1978). "Severe psoriasis–oral therapy with a new retinoid". *Dermatology* 157: pp. 238–244.

Friedman, R. C., Farh, K. K. H., and B., Burge -C. (2008). "Most mammalian mRNAs are conserved targets of microRNAs". *Genome research* 19: pp. 92–105.

Fujiwara, S., Baek, S, Varticovski, L, *et al.* (2019). "High Quality ATAC-Seq Data Recovered from Cryopreserved Breast Cell Lines and Tissue". *Scientific reports*.

Furue, K., Ito, T., and Furue, M. (2018). "Differential efficacy of biologic treatments targeting the TNF-/IL-23/IL-17 axis in psoriasis and psoriatic arthritis". *Cytokine* 111: pp. 182–188.

Gandhapudi, S. K. and Murapa, P. (2013). "HSF1 is activated as a consequence of lymphocyte activation and regulates a major proteostasis network in T cells critical for cell division during stress". *The Journal of Immunology* 191: pp. 4068–4079.

Gao, J., Chen, F., Hua, M., *et al.* (2018). "Knockdown of lncRNA MIR31HG inhibits cell proliferation in human HaCaT keratinocytes". *Biological Research*.

Gardembas-Pain, M., Ifrah, R., Foussard, C., *et al.* (1990). "Psoriasis after allogeneic bone marrow transplantation". *Archives of dermatology* 126: p. 1523.

Gaublomme, JT, Yosef, N, Lee, Y, *et al.* (2015). "Single-cell genomics unveils critical regulators of Th17 cell pathogenicity". *Cell* 163: pp. 1400–1412.

Gaulton, K. J., Nammo, T., Pasquali, L., *et al.* (2010). "A map of open chromatin in human pancreatic islets". *Nature* . . .

Gelfand, J. M., Gladman, D. D., Mease, P. J., *et al.* (2005). "Epidemiology of psoriatic arthritis in the population of the United States". *Journal of the American Academy of Dermatology* 53: p. 573.

Gelfand, J. M., Neimann, A. L., Shin, D. B., *et al.* (2006). "Risk of myocardial infarction in patients with psoriasis". *Jama network* 296: pp. 1735–1741.

Giambartolomei, Claudia, Vukcevic, Damjan, Schadt, Eric E, *et al.* (2014). "Bayesian test for colocalisation between pairs of genetic association studies using summary statistics." *PLoS Genet.* 10: e1004383.

Giresi, P. G., Kim, J., McDaniell, R. M., *et al.* (2006). "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin". *Genome Research* 17: pp. 877–885.

Gladman, D. D., Anhorn, K. A., Schachter, R. K., *et al.* (1986). "HLA antigens in psoriatic arthritis". *The Journal of Rheumatology* 13: pp. 586–592.

Gladman, D. D., Antoni, C., Mease, P., *et al.* (2005). "Psoriatic arthritis: epidemiology, clinical features, course, and outcome". *Annals of the rheumatic diseases* 64: pp. 14–17.

Glessner, J. T., Wang, K., Cai, G., *et al.* (2009). "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes". *Nature* 459: pp. 569–573.

Gobbi, De M., Viprakasit, V., Hughes, J. R., *et al.* (2006). "A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter". *Science* 312: pp. 1215–1217.

Golden, JB, Groft, SG, and Squeri, MV. (2015). "Chronic psoriatic skin inflammation leads to increased monocyte adhesion and aggregation". *The Journal of Immunology* 195: pp. 2006–2018.

Goldman, L and Schafer, AI (2011). "Goldman's Cecil Medicine E-Book".

Gong, Chenguang and Maquat, Lynne E. (2011). "lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements". *Nature* 470: pp. 284–288.

Goodnow, Christopher C., Sprent, Jonathon, Groth, Barbara, *et al.* (2005). "Cellular and genetic mechanisms of self tolerance and autoimmunity". *Nature* 435: p. 590.

Gossec, L., Smolen, J. S., Ramiro, S., *et al.* (2016). "European League Against Rheumatism (EULAR) recommendations for the management of psoriatic arthritis with pharmacological therapies: 2015 update". *Annals of the rheumatic diseases* 75: pp. 499–510.

Gottlieb, AB, Matheson, RT., Lowe, N., *et al.* (2003). "A randomized trial of etanercept as monotherapy for psoriasis". *Archives of Dermatology* 139: pp. 1627–1632.

Gradman, R. J., Ptacin, J. L., Bhasin -A., *et al.* (2008). "A bifunctional DNA binding region in Tn5 transposase". *Molecular Microbiology* 67: pp. 528–540.

Gramoun, Azza, Azizi, Natoosha, Sodek, Jaro, *et al.* (2010). "Fibronectin inhibits osteoclastogenesis while enhancing osteoclast activity via nitric oxide and interleukin-1-mediated signaling pathways." *J. Cell. Biochem.* 111: pp. 1020–34.

Greb, JE., Goldminz, AM., Elder, JT., *et al.* (2016). "Psoriasis". *Nature Reviews Disease Primers* 2.

Gregory, Adam P., Dendrou, Calliope A., Attfield, Kathrine E., *et al.* (2012). "TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis". *Nature* 488: p. 508.

Gregory, Simon G, Schmidt, Silke, Seth, Puneet, *et al.* (2007). "Interleukin 7 receptor chain (IL7R) shows allelic and functional association with multiple sclerosis". 39: p. 1083.

Gren, ST, Rasmussen, TB, and one, Janciauskiene S (2015). "A single-cell gene-expression profile reveals inter-cellular heterogeneity within human monocyte subsets". *PloS one*.

Grenningloh, Roland, Kang, Bok, and Ho, I-Cheng (2005). "Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses". 201: pp. 615–626.

Griffiths, C. E. M. and Barker, JN. (2007). "Pathogenesis and clinical features of psoriasis". *The Lancet* 370: pp. 263–271.

Griffiths, C. E. M., Reich, K., Lebwohl, M., *et al.* (2015). "Comparison of ixekizumab with etanercept or placebo in moderate-to-severe psoriasis (UNCOVER-2 and UNCOVER-3): results from two phase 3 randomised ..." *The Lancet* 386: pp. 541–551.

Grjibovski, AM., Olsen, AO., Magnus, P., *et al.* (2007). "Psoriasis in Norwegian twins: contribution of genetic and environmental effects". *Journal of the European Academy of Dermatology and Venereology* 21: pp. 1337–1343.

Gu, J., Märker-Hermann, E., Baeten, D., *et al.* (2002). "A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients". *Rheumatology Oxford* 41: pp. 759–766.

Guan, Tianxia, Dominguez, Claudia X., Amezquita, Robert A., *et al.* (2018). "ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8+ T cell fates". *Journal of Experimental Medicine* 215: pp. 1153–1168.

Guarda, G, Braun, M, Staehli, F, *et al.* (2011). "Type I interferon inhibits interleukin-1 production and inflammasome activation". *Immunity* 34: pp. 213–223.

Gudjonsson, J. E. and Karason, A. (2003). "Psoriasis patients who are homozygous for the HLA-Cw* 0602 allele have a 2· 5-fold increased risk of developing psoriasis compared with Cw6 heterozygotes". *British journal of Dermatology* 148: pp. 233–235.

Gudjonsson, J. E., Ding, J., Johnston, A., *et al.* (2010). "Assessment of the psoriatic transcriptome in a large sample: additional regulated genes and comparisons with in vitro models". *J Invest Dermatol* 130: pp. 1829–40.

Gupta, R., Lai, K., and Chopra, N. (2016). "Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs". *BMC Genomics* 17: p. 841.

Gusev, Alexander, Lee, Hong S., Trynka, Gosia, *et al.* (2014). "Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases". *The American Journal of Human Genetics* 95: pp. 535–552.

Gutowska-Owsiak, D., Schaupp, AL., Salimi, M, *et al.* (2012). "IL-17 downregulates filaggrin and affects keratinocyte expression of genes associated with cellular adhesion". *Experimental dermatology* 21: pp. 104–110.

Gérard, A., Kammen, R. A. van der, and Janssen -H. (2009). "The Rac activator Tiam1 controls efficient T-cell trafficking and route of transendothelial migration". *Blood* 113: pp. 6138–6147.

Hackshaw, Allan (2008). "Small studies: strengths and limitations".

Haddad, Amir and Chandran, Vinod (2013). "Arthritis Mutilans". *Current Rheumatology Reports* 15: p. 321.

Haftek, M, Viac, J, Cordier, G, *et al.* (1986). "Flow cytometry for separation of keratinocyte subpopulations from the viable epidermis". *Journal of investigative dermatology* 87: pp. 480–484.

Hammer, R. E., Maika, S. D., Richardson, J. A., *et al.* (1990). "Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human 2m: an animal model of HLA-B27-associated human disorders". *Cell* 63: pp. 1099–1112.

Hansen, K. H., Bracken, A. P., Pasini, D., *et al.* (2008). "A model for transmission of the H3K27me3 epigenetic mark". *Nature cell . . .*

Hanson, A. L., Cuddihy, T., Haynes, K., *et al.* (2018). "Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile". *Arthritis& Rheumatology* 70: pp. 255–265.

Hao, Yajing, Wu, Wei, Li, Hui, *et al.* (2016). "NPInter v3. 0: an upgraded database of noncoding RNA-associated interactions". *Database*.

Harper, EG, Guo, C, Rizzo, H, *et al.* (2009). "Th17 cytokines stimulate CCL20 expression in keratinocytes in vitro and in vivo: implications for psoriasis pathogenesis". *Journal of Investigative Dermatology* 129: pp. 2175–2183.

Harris, DP, Haynes, L, Sayles, PC, *et al.* (2000). "Reciprocal regulation of polarized cytokine production by effector B and T cells". *Nature Immunology* 1: pp. 475–482.

Harris, T. J., Grosso, J. F., Yen, H. R., *et al.* (2007). "Cutting edge: An in vivo requirement for STAT3 signaling in TH17 development and TH17-dependent autoimmunity". *The Journal of Immunology* 179: pp. 4313–4317.

Hecht, I, Toporik, A, Podojil, JR, *et al.* (2018). "ILDR2 is a novel B7-like protein that negatively regulates T cell responses". *The Journal of Immunology* 202.

Heintzman, N. D., Stuart, R. K., Hon, G., *et al.* (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". *Nature Genetics* 39: pp. 311–318.

Henseler, T. and Christophers, E. (1985). "Psoriasis of early and late onset: characterization of two types of psoriasis vulgaris". *J Am Acad Dermatol* 13: pp. 450–6.

Herman, J. G. and Baylin, S. B. (2003). "Gene silencing in cancer in association with promoter hypermethylation". *New England Journal of Medicine* 349: pp. 2042–2054.

Hesselberth, J. R., Chen, X., Zhang, Z., *et al.* (2009). "Global mapping of protein-DNA interactions in vivo by digital genomic footprinting". *Nature Methods* 6: pp. 283–289.

Heward, James A and Lindsay, Mark A (2014). "Long non-coding RNAs in the regulation of the immune response". 35: pp. 408–419.

Hijnen, DJ and Nijhuis E and. . ., Bruin-Weller - M (2005). "Differential expression of genes involved in skin homing, proliferation, and apoptosis in CD4+ T cells of patients with atopic dermatitis". *Journal of investigative Dermatology* 125: pp. 1149–1155.

Hijnen, DJ, Knol, EF, Gent, YY, *et al.* (2013). "CD8+ T cells in the lesional skin of atopic dermatitis and psoriasis patients are an important source of IFN-, IL-13, IL-17, and IL-22". *Journal of Investigative Dermatology* 133: pp. 973–979.

Hirschhorn, Joel N. (2005). "Genetic Approaches to Studying Common Diseases and Complex Traits". *Pediatr Res* 57: pp. 74–77.

Hitchon, Carol A., Danning, Carol L., Illei, Gabor G., *et al.* (2002). "Gelatinase expression and activity in the synovium and skin of patients with erosive psoriatic arthritis". *The Journal of rheumatology* 29: pp. 107–117.

Hoffman, M. M., Ernst, J., Wilder, S. P., *et al.* (2013). "Integrative annotation of chromatin elements from ENCODE data". *Nucleic Acids Res* 41: pp. 827–41.

Homey, B., Alenius, H., Müller, A., *et al.* (2002). "CCL27–CCR10 interactions regulate T cell–mediated skin inflammation". *Nature medicine* 8: pp. 157–165.

Hon, G., Wang, W., and biology, Ren -B. computational (2009). "Discovery and annotation of functional chromatin signatures in the human genome". *PLoS computational biology*.

Howell, K. J., Kraiczy, J., Nayak, K. M., *et al.* (2018). "DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes . . . " *Gastroenterology* 154: pp. 585–598.

Howie, B. N., Donnelly, P., and Marchini -J. (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies". *PLoS genetics*.

Hsieh, I. N., Liou, J. P., Lee, H. Y., *et al.* (2014). "Preclinical anti-arthritic study and pharmacokinetic properties of a potent histone deacetylase inhibitor MPT0G009". *Cell death & disease* 5: e1166.

Hu, Stephen, Yu, Hsin-Su, Yen, Feng-Lin, *et al.* (2016). "Neutrophil extracellular trap formation is increased in psoriasis and induces human -defensin-2 production in epidermal keratinocytes". *Scientific Reports* 6: p. 31119.

Hu, Xiaoyu and Ivashkiv, Lionel B (2009). "Cross-regulation of signaling pathways by interferon-gamma: implications for immune responses and autoimmune diseases." *Immunity* 31: pp. 539–50.

Huber, M., Brüstle A. and . . ., Reinhard -K., Guralnik, A., *et al.* (2008). "IRF4 is essential for IL-21-mediated induction, amplification, and stabilization of the Th17 phenotype". *Proc Natl Acad Sci U S A* 105: p. 2084620851.

Hudak, S., Hagen, M., Liu, Y., *et al.* (2002). "Immune surveillance and effector functions of CCR10+ skin homing T cells". *The Journal of Immunology* 169: pp. 1189–1196.

Hundhausen, C., Bertoni, A., Mak, R. K., *et al.* (2012). "Allele-Specific Cytokine Responses at the HLA-C Locus: Implications for Psoriasis". *Journal of Investigative Dermatology* 132: pp. 635–641.

Hunt, KA, Mistry, V, Bockett, NA, *et al.* (2013). "Negligible impact of rare autoimmune-locus coding-region variants on missing heritability". *Nature* 498: pp. 232–235.

Husted, J. A., Thavaneswaran -A., *et al.* (2011). "Cardiovascular and other comorbidities in patients with psoriatic arthritis: a comparison with patients with psoriasis". *Arthritis care & research* 63: pp. 1729–1735.

Hüffmeier, U., Uebe, S., Ekici, A. B., *et al.* (2010). "Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis". *Nature* . . . 42: pp. 996–999.

Ibrahim, G., Waxman, R., and Helliwell, P. S. (2009). "The prevalence of psoriatic arthritis in people with psoriasis". *Arthritis Rheum* 61: pp. 1373–8.

Icen, M., Crowson, CS., McEvoy, MT., *et al.* (2009). "Trends in incidence of adults-onset psoriasis over three decades: a population study". *J Am Acad Dermatol* 60: 394–401.

Idel, S., Dansky, H. M., and L., Breslow -J. (2003). "A20, a regulator of NFB, maps to an atherosclerosis locus and differs between parental sensitive C57BL/6J and resistant FVB/N strains". *Proc Natl Acad Sci U S A* 1000: pp. 14235–14240.

Itoh, N, Itoh, Y, Tassoni, A, *et al.* (2018). "Cell-specific and region-specific transcriptomics in the multiple sclerosis model: Focus on astrocytes". *Proc Natl Acad Sci U S A* 115: e302–309.

Jabbari, Ali, Suárez-Fariñas, Mayte, Dewell, Scott, *et al.* (2012). "Transcriptional profiling of psoriasis using RNA-seq reveals previously unidentified differentially expressed genes". *The Journal of investigative dermatology* 132: pp. 246–249.

Jacobson, Christine C., Kumar, Sandeep, and Kimball, Alexa B. (2011). "Latitude and psoriasis prevalence". *Journal of the American Academy of Dermatology* 65: pp. 870–873.

Jaitin, D. A., Kenigsberg, E., and Keren-Shaul -H. (2014). "Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types". *Science* 343: pp. 776–779.

Jansen, R., Hottenga, J. J., and G., Nivard -M. (2017). "Conditional eQTL analysis reveals allelic heterogeneity of gene expression". *Human molecular genetics* 26: pp. 1444–1451.

Javierre, BM, Burren, OS, Wilder, SP, *et al.* (2016). "Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters". *Cell* 16: pp. 1369–1384.

Jenuwein, T. and Allis, CD. (2001). "Translating the histone code". *Science* 293: pp. 1074–1080.

Jiang, X, Tian, H, Fan, Y, *et al.* (2012). "Expression of tumor necrosis factor alpha-induced protein 3 mRNA in peripheral blood mononuclear cells negatively correlates with disease severity in psoriasis . . . " *Clinical and Vaccin Immunology*.

Johansen, C., Kragballe, K., and Rasmussen -M. (2004). "Activator protein 1 DNA binding activity is decreased in lesional psoriatic skin compared with nonlesional psoriatic skin". *British Journal of Dermatology* 151: pp. 600–607.

John, S., Sabo, P. J., Canfield, T. K., *et al.* (2013). "Genome-scale mapping of DNase I hypersensitivity". *Current protocols in molecular biology* 27.

Johnson, D. S., Mortazavi, A., Myers, R. M., *et al.* (2007). "Genome-wide mapping of in vivo protein-DNA interactions". *Science* 316: pp. 1497–1502.

Johnson-Huang, LM., McNutt, NS., and Krueger, JG. (2009). "Cytokine-producing dendritic cells in the pathogenesis of inflammatory skin diseases". *Journal of clinical immunology* 29: pp. 247–256.

Jones, SM., Armas, JB., and Cohen, MG (1994). "Psoriatic arthritis: outcome of disease subsets and relationship of joint disease to nail and skin disease." *Journal of Rheumatology* 33: 834–9.

Jones, S.M., Dixey, J., Hall, N.D., *et al.* (1997). "Expression of the cutaneous lymphocyte antigen and its counter-receptor E-selectin in the skin and joints of patients with psoriatic arthritis". *Br J Rheumatol* 36: 748–57.

Jongbloed, Sarah L., Lebre, M. C., Fraser, Alasdair R., *et al.* (2006). "Enumeration and phenotypical analysis of distinct dendritic cell subsets in psoriatic arthritis and rheumatoid arthritis". *Arthritis research & therapy* 8: R15.

Jonker, M. J., Leeuw, W. C. de, and Marinković -M. (2014). "Absence/presence calling in microarray-based CGH experiments with non-model organisms". *Nucleic acids* 42: e94.

Joost, S, Zeisel, A, Jacob, T, *et al.* (2016). "Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity". *Cell systems* 3: pp. 221–237.

Jordan, C. T., Cao, L., Roberson, E. D. O., *et al.* (2012). "Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis". *The American Journal of Human Genetics* 90: pp. 796–808.

Jostins, Luke, Ripke, Stephan, Weersma, Rinse K, *et al.* (2012). "Host microbe interactions have shaped the genetic architecture of inflammatory bowel disease". *Nature* 491: p. 119.

Jung, M., Sabat, R., and Krätzschmar -J. (2004). "Expression profiling of IL-10-regulated genes in human monocytes and peripheral blood mononuclear cells from psoriatic patients during IL-10 therapy". *European journal of immunology* 34: pp. 481–493.

Junta, C. M., Sandrin-Garcia, P., Fachin-Saltoratto, AL., *et al.* (2009). "Differential gene expression of peripheral blood mononuclear cells from rheumatoid arthritis patients may discriminate immunogenetic, pathogenic and treatment features". *Immunology* 127: pp. 365–372.

Kagami, Shinji, Rizzo, Heather L, Lee, Jennifer J, *et al.* (2010). "Circulating Th17, Th22, and Th1 cells are increased in psoriasis". 130: pp. 1373–1383.

Kaminsky, Z. A., Tang, T., Wang, S. C., *et al.* (2009). "DNA methylation profiles in monozygotic and dizygotic twins". *Nature Genetics* 41: pp. 240–245.

Kanehisa, M. and Goto, S. (2000). "KEGG: kyoto encyclopedia of genes and genomes". *Nucleic acids research* 28: pp. 27–30.

Karason, A., Gudjonsson, J. E., and Upmanyu -R. (2003). "A susceptibility gene for psoriatic arthritis maps to chromosome 16q: evidence for imprinting". *The American Journal of Human Genetics* 72: pp. 125–131.

Karason, A., Thorvardur, JL., and Bjorn, G. (2009). "A strong heritability of psoriatic arthritis over four generations—the Reykjavik Psoriatic Arthritis Study". *Rheumatology* 48: 1424–1428.

Kasela, Silva, Kisand, Kai, Tserel, Liina, *et al.* (2017). "Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells". *Plos Genet* 13: e1006643.

Katschke, K. J., Rottman, J. B., Ruth, J. H., *et al.* (2001). "Differential expression of chemokine receptors on peripheral blood, synovial fluid, and synovial tissue monocytes/macrophages in rheumatoid arthritis". *Arthritis & Rheumatism* 44: pp. 1022–1032.

Kavanaugh, Arthur F. and Ritchlin, Christopher T. (2006). "Systematic review of treatments for psoriatic arthritis: an evidence based approach and basis for treatment guidelines". *The Journal of Rheumatology* 33: pp. 1417–1421.

Keating, GM. (2017). "Apremilast: a review in psoriasis and psoriatic arthritis". *Drugs* 77: pp. 459–472.

Keermann, M., Kõks, S., and Reimann -E. (2015). "Transcriptional landscape of psoriasis identifies the involvement of IL36 and IL36RN". *BMC Genomics* 16.

Kent, James W., Sugnet, Charles W., Furey, Terrence S., *et al.* (2002). "The Human Genome Browser at UCSC". *Genome Research* 12: pp. 996–1006.

Kent, L. (2009). "Freezing and thawing human embryonic stem cells". *Journal of visualized experiments: JoVE* 34: p. 1555.

Kent, W. J., Zweig, A. S., Barber, G., *et al.* (2010). "BigWig and BigBed: enabling browsing of large distributed datasets". *Bioinformatics* 26: pp. 2204–2207.

Khera, AV., Chaffin, M., Aragam, KG., *et al.* (2018). "Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations". *Nature Genetics* 50: 1219–1224.

Kichaev, G. and Pasaniuc, B. (2015). "Leveraging functional-annotation data in trans-ethnic fine-mapping studies". *The American Journal of Human Genetics* 97: pp. 260–27.

Kim, G. K., of, clinical, and Q., Del J. (2010). "Drug-provoked psoriasis: is it drug induced or drug aggravated?: understanding pathophysiology and clinical relevance". *The Journal of clinical and aesthetic dermatology* 3: pp. 32–38.

Kim, SH, Han, SY, Azam, T, *et al.* (2005). "Interleukin-32: a cytokine and inducer of TNF". *Immunity* 22: pp. 131–142.

Kimball, AB, Kawamura, T, Tejura, K, *et al.* (2002). "Clinical and immunologic assessment of patients with psoriasis in a randomized, double-blind, placebo-controlled trial using recombinant human interleukin …" *Archives of dermatology* 138: pp. 1341–1346.

Kino, Tomoshige, Hurt, Darrell E., Ichijo, Takamasa, *et al.* (2010). "Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor". *Science signaling* 3.

Kiselev, VY, Andrews, TS, and M, Hemberg (2019). "Challenges in unsupervised clustering of single-cell RNA-seq data". *Nature Reviews Genetics*.

Knight, J. C. (2014). "Approaches for establishing the function of regulatory genetic variants involved in disease". *Genome medicine* 6: p. 92.

Kondrack, Robyn M, Harbertson, Judith, Tan, Joyce T, *et al.* (2003). "Interleukin 7 regulates the survival and generation of memory CD4 cells". 198: pp. 1797–1806.

Kopydlowski, KM, Salkowski, CA, and MJ, Cody (1999). "Regulation of macrophage chemokine expression by lipopolysaccharide in vitro and in vivo". *The Journal of Immunology* 163: pp. 1537–1544.

Kraan, C. T. M., Baggen, J. M., Bodegraven, van A. A., *et al.* (2012). "Defective IL-1A expression in patients with Crohns disease is related to attenuated MAP3K4 signaling". *Human immunology* 73: pp. 912–919.

Ku, C. S., Loy, E. Y., Salim, A., *et al.* (2010). "The discovery of human genetic variations and their use as disease markers: past, present and future". *Journal of human genetics* 55: pp. 403–415.

Kuijk, AWR. van, Reinders-Blankert, P., Smeets, TJM., *et al.* (2006). "Detailed analysis of the cell infiltrate and the expression of mediators of synovial inflammation and joint destruction in the synovium of patients with psoriatic arthritis . . . " *Annals of the Rheumatic Diseases* 65: pp. 1551–1557.

Kumasaka, N, Knights, AJ, and DJ, Gaffney (2018). "High-resolution genetic mapping of putative causal interactions between regions of open chromatin". *Nature genetics* 51: pp. 128–137.

Kumasaka, Natsuhiko, Knights, Andrew, and Gaffney, Daniel (2015). "Fine-mapping cellular QTLs with RASQUAL and ATAC-seq". *Biorxiv*: p. 018788.

Kundaje, A., Meuleman, W., Ernst, J., *et al.* (2015). "Integrative analysis of 111 reference human epigenomes". *Nature* 518: pp. 317–330.

Kurdi, A. T., Bassil, R., Olah, M., *et al.* (2016). "Tiam1/Rac1 complex controls Il17a transcription and autoimmunity". *Nature Communications* 7.

Kusaba, Hitoshi, Ghosh, Paritosh, Derin, Rachel, *et al.* (2005). "Interleukin-12-induced interferon- production by human peripheral blood T cells is regulated by mammalian target of rapamycin (mTOR)". 280: pp. 1037–1043.

Lande, R., Gregorio, J., Facchinetti, V., *et al.* (2007). "Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide". *Nature* 449.

Landt, Stephen G., Marinov, Georgi K., Kundaje, Anshul, *et al.* (2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia". *Genome research* 22: pp. 1813–1831.

Langewouters, A. M. G., Erp, P. E. J., Jong, Emgj, *et al.* (2008). "Lymphocyte subsets in peripheral blood of patients with moderate-to-severe versus mild plaque psoriasis". *Archives of dermatological research* 300: pp. 107–113.

Langmead, B. and Salzberg, S. L. (2012). "Fast gapped-read alignment with Bowtie 2". *Nature methods*.

Lassen, KG, CI, McKenzie, Mari, M, *et al.* (2016). "Genetic coding variant in GPR65 alters lysosomal pH and links lysosomal dysfunction with colitis risk". *Immunity* 44: pp. 1392–1405.

Lee, Donghyung, Bigdeli, Bernard T., Riley, Brien P., *et al.* (2013). "DIST: direct imputation of summary statistics for unmeasured SNPs". *Bioinformatics* 29: pp. 2925–2927.

Lee, Edmund, Trepicchio, William L., Oestreicher, Judith L., *et al.* (2004). "Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris". *Journal of Experimental Medicine* 199: pp. 125–130.

Lee, Mark N., Ye, Chun, Villani, Alexandra-Chloé C., *et al.* (2014). "Common genetic variants modulate pathogen-sensing responses in human dendritic cells". *Science (New York, N.Y.)* 343: pp. 1246980–9.

Lee, S. K., Jeon, E. K., Kim, Y. J., *et al.* (2009). "A global gene expression analysis of the peripheral blood mononuclear cells reveals the gene expression signature in psoriasis". *Annals of dermatology* 21: pp. 237–242.

Lee, Yoontae, Jeon, Kipyoung, Lee, Jun, *et al.* (2002). "MicroRNA maturation: stepwise processing and subcellular localization". *The EMBO journal* 21: pp. 4663–4670.

Lei, W., Luo, Y., Lei, W., *et al.* (2009). "Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis". *Scandinavian journal journal of rheumatology* 38: pp. 369–374.

Leipe, Jan, Grunke, Mathias, Dechant, Claudia, *et al.* (2010). "Role of Th17 cells in human autoimmune arthritis". 62: pp. 2876–2885.

Lerman, Galya, Avivi, Camila, Mardoukh, Corine, *et al.* (2011). "MiRNA expression in psoriatic skin: reciprocal regulation of hsa-miR-99a and IGF-1R". *PloS one* 6.

Lewis, B. P., Burge, C. B., and P., Bartel -D. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets". *Cell* 120: pp. 15–20.

Li, B., Tsoi, L., Swindell, W., *et al.* (2014). "Transcriptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into disease mechanisms". *J Invest Dermatol* 134: pp. 1828–1838.

Li, C., Li, H., Huang, L., *et al.* (2017). "H19 lncRNA regulates keratinocyte differentiation by targeting miR-130b-3p". *Cell death & disease* 8: e3174.

Li, H., Handsaker, B., Wysoker, A., *et al.* (2009). "The sequence alignment/map format and SAMtools". *Bioinformatics* 25: pp. 2078–2079.

Li, H, Yao, Q, Mariscal, AG, *et al.* (2018). "Epigenetic control of IL-23 expression in keratinocytes is important for chronic skin inflammation". *Nature Communications* 9.

Li, X., AW., O'Regan, and Berman, JS. (2003). "IFN- Induction of Osteopontin Expression in Human Monocytoid Cells". *Journalof interferon & cytokine research* 23: pp. 259–265.

Li, Y., and Kellis, M. (2016). "Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases". *Nucleic acids research* 44: e144.

Lin, Andrew M., Rubin, Cory J., Khandpur, Ritika, *et al.* (2011). "Mast Cells and Neutrophils Release IL-17 through Extracellular Trap Formation in Psoriasis". *The Journal of Immunology* 187: pp. 490–500.

Lin, S.H, Ho, J., Li, S.C, *et al.* (2019). "MiR-146a-5p Expression in Peripheral CD14+ Monocytes from Patients with Psoriatic Arthritis Induces Osteoclast Activation, Bone Resorption, and Correlates with Clinical Response". *Journal of Clinical Medicine* 8.

Lin, Y., Liu, L., Sheng, Y., *et al.* (2018). "A catalog of potential putative functional variants in psoriasis genome-wide association regions". *PloS one*.

Linton, PJ, Bautista, B, and Biederman, E. (2003). "Costimulation via OX40L expressed by B cells is sufficient to determine the extent of primary CD4 cell expansion and Th2 cytokine secretion in vivo". *Journal of experimental medicine* 197: pp. 875–83.

Liontos, Larissa M., Dissanayake, Dilan, Ohashi, Pamela S., *et al.* (2011). "The Src-like adaptor protein regulates GM-CSFR signaling and monocytic dendritic cell maturation". *The Journal of Immunology*: p. 903292.

Litherland, SA, Xie, TX, and KM, Grebe (2005). "Signal transduction activator of transcription 5 (STAT5) dysfunction in autoimmune monocytes and macrophages". *Journal of autoimmunity* 24: pp. 297–310.

Liu, Ting, Zhang, Lingyun, Joo, Donghyun, *et al.* (2017). "NF-B signaling in inflammation". *Signal Transduction and Targeted Therapy* 2: p. 17023.

Liu, Y., Aryee, M. J., and Padyukov L. and, Fallin -M. D. (2013). "Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis". *Nature Biotechnology* 31: pp. 142–147.

Liu, Y, Beyer, A, and R, Aebersold (2016). "On the dependency of cellular protein levels on mRNA abundance". *Cell* 165: pp. 535–550.

Liu, Y, Chang, JC, Hon, CC, *et al.* (2018). "Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis". *bioRxiv*.

Liu, Ying, Helms, Cynthia, Liao, Wilson, *et al.* (2008). "A Genome-Wide Association Study of Psoriasis and Psoriatic Arthritis Identifies New Disease Loci". *PLoS Genetics* 4: pp. 1–14.

Lizzul, P. F., Aphale, A., Malaviya, R., *et al.* (2005). "Differential expression of phosphorylated NF-B/RelA in normal and psoriatic epidermis and downregulation of NF-B in response to treatment with . . . " *Journal of Investigative dermatology* 124: pp. 1275–1283.

Lonnberg, AS., Skov, L., Skytthe, A., *et al.* (2013). "Heritability of psoriasis in a large twin sample". *British Journal of Dermatology* 169: pp. 412–416.

Lonsdale, J., Thomas, J., Salvatore, M., *et al.* (2013). "The genotype-tissue expression (GTEx) project". *Nature Genetics* 45: pp. 580–585.

Lories, RJU, Derese, I, and FP, Luyten (2008). "Activation of nuclear factor kappa B and mitogen activated protein kinases in psoriatic arthritis before and after etanercept treatment". *Clinical and and experimental rheumatology* 26: pp. 96–102.

Lotem, J, Levanon, D, Negreanu, V, *et al.* (2013). "Runx3-mediated transcriptional program in cytotoxic lymphocytes". *PloS one*.

Love, M. I., Huber, W., and Anders -S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome biology* 15: p. 550.

Lowes, MA, Kikuchi, T, and J, Fuentes-Duculan - (2008). "Psoriasis vulgaris lesions contain discrete populations of Th1 and Th17 T cells". *Journal of Investigative dermatology* 128: pp. 1207–1211.

Lu, J, Ding, Y, Yi, X, *et al.* (2016). "CD19+ B cell subsets in the peripheral blood and skin lesions of psoriasis patients and their correlations with disease severity". *Brazilian Journal of Medical and Biological Research* 49: e5374.

Luger, K., Mäder, A. W., Richmond, R. K., *et al.* (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution". *Nature*.

Lundström, W, Highfill, S, Walsh, ST., *et al.* (2013). "Soluble IL7R potentiates IL-7 bioactivity and promotes autoimmunity". *Proc Natl Acad Sci U S A* 110: e1761–1770.

Luszczek, W., Mańczak, M., Cisło, M., *et al.* (2004). "Gene for the activating natural killer cell receptor, KIR2DS1, is associated with susceptibility to psoriasis vulgaris". *Human immunology* 65: pp. 758–766.

M, Bar-Eli, Gallily, R, and HA, Cohen (1979). "Monocyte function in psoriasis." *Journal of Investigative dermatology* 73: pp. 147–149.

Mahil, Satveer K., Capon, Francesca, and Barker, Jonathan N. (2016). "Update on psoriasis immunopathogenesis and targeted immunotherapy". *Seminars in Immunopathology* 38: pp. 11–27.

Majumdar, I., Ahuja, V., and Paul, J. (2017). "Altered expression of Tumor Necrosis Factor Alpha -Induced Protein 3 correlates with disease severity in Ulcerative Colitis". *Scientific Reports* 7.

Malek, T. R. and Immunity, Castro -I. (2010). "Interleukin-2 receptor signaling: at the interface between tolerance and immunity". *Immunity*.

Maller, J. B., McVean, G., Byrnes, J., *et al.* (2012). "Bayesian refinement of association signals for 14 loci in 3 common diseases". *Nat Genet* 44: pp. 1294–301.

Mandal, A. and Viswanathan, C. (2015). "Natural killer cells: in health and disease". *Hematology/oncology and stem cell therapy* 8: pp. 47–55.

Mao, Liming, Kitani, Atsushi, Similuk, Morgan, *et al.* (2018). "Loss-of-function CARD8 mutation causes NLRP3 inflammasome activation and Crohn's disease". *Journal of Clinical Investigation* 128: pp. 1793–1806.

Marlow, GJ, Gent, D van, and LR, Ferguson (2013). "Why interleukin-10 supplementation does not work in Crohn's disease patients". *World journal of gastroenterology* 19: pp. 3931–3941.

Marouli, E., Graff, M., Medina-Gomez, C., *et al.* (2017). "Rare and low-frequency coding variants alter human adult height". *Nature* 542: pp. 186–190.

Marrakchi, S., Guigue, P., and R., Renshaw -B. (2011). "Interleukin-36–receptor antagonist deficiency and generalized pustular psoriasis". *The New England Journal of Medicine* 365: pp. 620–628.

Marshall, C. R., Howrigan, D. P., and Merico -D. (2017). "Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects". *Nature Genetics* 49: pp. 27–35.

Martin, Marcel (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". *EMBnet. journal*.

Martin-Fontecha, Alfonso, Thomsen, Lindy L, Brett, Sara, *et al.* (2004). "Induced recruitment of NK cells to lymph nodes provides IFN- for T H 1 priming". 5: p. 1260.

Maurano, M. T., Humbert, R., Rynes, E., *et al.* (2012). "Systematic localization of common disease-associated variation in regulatory DNA". *Science* 337: pp. 1190–1195.

McCormack, WJ., Parker, AE., and O'Neill, LA. (2009). "Toll-like receptors and NOD-like receptors in rheumatic diseases". *Arthritis research and therapy* 11: p. 243.

McGeachy, Mandy J., Chen, Yi, Tato, Cristina M., *et al.* (2009). "The interleukin 23 receptor is essential for the terminal differentiation of interleukin 17-producing effector T helper cells in vivo." *Nat. Immunol.* 10: pp. 314–24.

McGonagle, D., Tan, AL., and Benjamin, M. (2007). "The biomechanical link between skin and joint disease in psoriasis and psoriatic arthritis: what every dermatologist needs to know". *Annals of the Rheumatic Diseases* 67: pp. 1–4.

McGonagle, D., Tan, Ai Lyn, Watad, Abdulla, *et al.* (2019). "Pathophysiology, assessment and treatment of psoriatic dactylitis". *Nature Reviews Rheumatology* 15: pp. 113–122.

McGonagle, Dennis, Ash, Zoe, Dickie, Laura, *et al.* (2011). "The early phase of psoriatic arthritis". *Annals of the Rheumatic Diseases* 70.

Mease, Philip J. (2011). "Measures of psoriatic arthritis: Tender and Swollen Joint Assessment, Psoriasis Area and Severity Index (PASI), Nail Psoriasis Severity Index (NAPSI), Modified Nail Psoriasis Severity Index (mNAPSI), Mander/Newcastle Enthesitis Index (MEI), Leeds Enthesitis Index (LEI), Spondyloarthritis Research Consortium of Canada (SPARCC), Maastricht Ankylosing Spondylitis Enthesis Score (MASES), Leeds Dactylitis Index (LDI), Patient Global for Psoriatic Arthritis, Dermatology Life Quality Index (DLQI), Psoriatic Arthritis Quality of Life (PsAQOL), Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F), Psoriatic Arthritis Response Criteria (PsARC), Psoriatic Arthritis Joint Activity Index (PsAJAI), Disease Activity in Psoriatic Arthritis (DAPSA), and Composite Psoriatic Disease Activity Index (CPDAI)". *Arthritis Care & Research* 63.

Meglio, Di P, Cesare, Di A, Laggner, U, *et al.* (2011). "The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans". *PloS one*.

Meglio, Paola, Villanova, Federica, and Nestle, Frank O. (2014). "Psoriasis". *Cold Spring Harbor Perspectives in Medicine* 4.

Mehta, N.N., Yu, Y., Saboury, B., *et al.* (20119). "Systemic and vascular inflammation in patients with moderate to severe psoriasis as measured by [18F]-fluorodeoxyglucose positron emission tomography-computed tomography (FDG-PET/CT): a pilot study". *Arch Dermatol.* 147: pp. 1031–9.

Meijerink, Marjolein, Ulluwishewa, Dulantha, Anderson, Rachel C, *et al.* (2011). "Cryopreservation of monocytes or differentiated immature DCs leads to an altered cytokine response to TLR agonists and microbial stimulation". 373: pp. 136–142.

Menon, B., Gullick, N. J., and J., Walter -G. (2014). "Interleukin-17+ CD8+ T cells are enriched in the joints of patients with psoriatic arthritis and correlate with disease activity and joint damage progression". *Arthritis & Rheumatology* 66: pp. 1272–1281.

Mensah, K. A., Schwarz, E. M., and T., Ritchlin -C. (2008). "Altered bone remodeling in psoriatic arthritis". *Current rheumatology reports* 10: pp. 311–317.

Menter, Alan, Korman, Neil J., Elmets, Craig A., *et al.* (2009). "Guidelines of care for the management of psoriasis and psoriatic arthritis Section 3. Guidelines of care for the management and treatment of psoriasis with topical therapies". *Journal of the American Academy of Dermatology* 60: pp. 643–659.

Mesko, B., Poliskal, S., and Szegedi, A. (2010). "Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets". *BMC medical genomics* 3: p. 15.

Messeguer, X., Escudero, R., Farré, D., *et al.* (2002). "PROMO: detection of known transcription regulatory elements using species-tailored searches". . . .

Miao, Y. L., Xiao, Y. L., Du, Y., *et al.* (2013). "Gene expression profiles in peripheral blood mononuclear cells of ulcerative colitis patients". *World Journal of gastroenterology* 19: pp. 3339–3346.

Michel, T, Poli, A, Cuapio, A, *et al.* (2016). "Human CD56bright NK cells: an update". *The Journal of immunology* 196: pp. 2923–2931.

Milani, P., Escalante-Chong, R., and reports, Shelley -B. C. (2016). "Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells". *Scientific reports* 6: p. 25474.

Mino, T., Murakawa, Y., Fukao, A., *et al.* (2015). "Regnase-1 and roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms". *Cell* 161: pp. 1058–1073.

Mizoguchi, Fumitaka, Slowikowski, Kamil, Wei, Kevin, *et al.* (2018). "Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis". *Nature communications* 9: p. 789.

Mo, Angela, Marigorta, Urko M, Arafat, Dalia, *et al.* (2018). "Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease". *Genome Medicine* 10.

Mo, JX., Shi, SJ., Zhang, Q., *et al.* (2011). "Synthesis, transport and mechanism of a type I prodrug: L-carnitine ester of prednisolone". *Mol Pharm.* 3: pp. 1629–40.

Moll, J. M. H., in, arthritis, and rheumatism, Wright -V. (1973). "Psoriatic arthritis". *Seminars in arthritis and rheumatism*.

Morimoto, Junko, Kon, Shigeyuki, Matsui, Yutaka, *et al.* (2010). "Osteopontin; as a target molecule for the treatment of inflammatory diseases". 11: pp. 494–505.

Mortazavi, Ali, Williams, Brian A., McCue, Kenneth, *et al.* (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq". *Nature methods* 5: pp. 621–628.

Moser, B, Desai, DD, Downie, MP, *et al.* (2007). "Receptor for advanced glycation end products expression on T cells contributes to antigen-specific cellular expansion in vivo". *The Journal of Immunology* 179: pp. 8051–8058.

Mueller, C., Edmiston, H., Carpenter, C., *et al.* (2013). "One-Step Cell and Tissue Preservative for Morphologic and Molecular Analysis". *PLoS One*.

Mukherjee, R, Barman, PK, Thatoi, PK, *et al.* (2015). "Non-classical monocytes display inflammatory features: validation in sepsis and systemic lupus erythematous". *Scientific reports* 3886.

Müller, Nike, Döring, Frank, Klapper, Maja, *et al.* (2014). "Interleukin-6 and tumour necrosis factor- differentially regulate lincRNA transcripts in cells of the innate immune system in vivo in human subjects with rheumatoid arthritis". *Cytokine* 68: pp. 65–68.

Muntyanu, Anastasiya, Abji, Fatima, Pollock, Remy, *et al.* (2017). "Abstract 156. Gene Expression in Cellular Subsets in Psoriatic Disease".

Myers, A., Kay, LJ., Lynch, SA., *et al.* (2005). "Recurrence risk for psoriasis and psoriatic arthritis within sibships". *Rheumatology* 44: pp. 773–776.

Nagano, T., Lubling, Y., Stevens, T. J., *et al.* (2013). "Single-cell Hi-C reveals cell-to-cell variability in chromosome structure". *Nature* 502: pp. 59–64.

Nair, R. P., Ruether, A., Stuart, P. E., *et al.* (2008). "Polymorphisms of the IL12B and IL23R genes are associated with psoriasis". *J Invest Dermatol* 128: pp. 1653–61.

Nair, R. P., Duffin, K. C., Helms, C., *et al.* (2009). "Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways". *Nat Genet* 41: pp. 199–204.

Naranbhai, Vivek, Fairfax, Benjamin P, Makino, Seiko, *et al.* (2015). "Genomic modulators of gene expression in human neutrophils". *Nat Commun* 6: p. 7545.

Natarajan, R, Salloum, FN, and BJ, Fisher (2007). "Activation of hypoxia-inducible factor-1 via prolyl-4 hydoxylase-2 gene silencing attenuates acute inflammatory responses in postischemic myocardium". *American Journal of physiology, heart and circulatory physiology* 293: pp. 1571–1580.

Nestle, Frank O., Conrad, Curdin, Tun-Kyi, Adrian, *et al.* (2005). "Plasmacytoid predendritic cells initiate psoriasis through interferon- production". *The Journal of Experimental Medicine* 202: pp. 135–143.

Nestle, Frank O., Kaplan, Daniel H., and Barker, Jonathan (2009). "Psoriasis". *New England Journal of Medicine* 361: pp. 496–509.

Nica, Parts, Glass, *et al.* (2011). "The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study". *The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study* 7: pp. 1–9.

Nickoloff, B. J. and Wrone-Smith, T. (1999). "Injection of pre-psoriatic skin with CD4+ T cells induces psoriasis". *Am J Pathol* 155: pp. 145–58.

Nickoloff, B. J., Qin, J. Z., and Nestle, F. O. (2007). "Immunopathogenesis of psoriasis". *Clin Rev Allergy Immunol* 33: pp. 45–56.

Nora, Elphège P. P., Goloborodko, Anton, Valton, Anne-Laure L., *et al.* (2017). "Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization". *Cell* 169: pp. 930–1073741824.

O'Garra, Anne, Vieira, Pedro L., Vieira, Paulo, *et al.* (2004). "IL-10-producing and naturally occurring CD4+ Tregs: limiting collateral damage." *J. Clin. Invest.* 114: pp. 1372–8.

Okada, Y., Han, B., Tsoi, L. C., *et al.* (2014). "Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes". *The American Journal of Human Genetics* 95: pp. 162–172.

Oliveira, Mfsp, Rocha, B. O., and V., Duarte -G. (2015). "Psoriasis: classical and emerging comorbidities". *Anais brasileiros de dermatologia* 90: pp. 9–20.

Organization, World Health (2016). *Global report on psoriasis*. Report.

Ortega, C, S, Fernández-A, and JM, Carrillo (2009). "IL-17-producing CD8+ T lymphocytes from psoriasis skin plaques are cytotoxic effector cells that secrete Th17-related cytokines". *Journal of leukocyte biology* 86: pp. 435–443.

Ottaviani, C, Nasorri, F, and Bedini, C (2006). "CD56brightCD16– NK cells accumulate in psoriatic skin in response to CXCL10 and CCL5 and exacerbate skin inflammation". *European journal of immunology* 36: pp. 118–128.

Oudelaar, A. M., Davies, J. O. J., and acids …, Downes -D. J. (2017). "Robust detection of chromosomal interactions from small numbers of cells using low-input Capture-C". *Nucleic acids research* 45: e184.

Ouyang, Jing, Zhu, Xiaomei, Chen, Yuhai, *et al.* (2014). "NRAV, a long noncoding RNA, modulates antiviral responses through suppression of interferon-stimulated gene transcription". *Cell host & microbe* 16: pp. 616–626.

Palau, Nuria, Julià, Antonio, Ferrándiz, Carlos, *et al.* (2013). "Genome-wide transcriptional analysis of T cell activation reveals differential gene expression associated with psoriasis". *BMC genomics* 14: p. 825.

Pan, G., Tian, S., Nie, J., *et al.* (2007). "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells". *Cell stem cell* 1: pp. 299–312.

Pandey, R. R., Mondal, T., Mohammad, F., *et al.* (2008). "Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation". *Molecular cell* 32: pp. 232–246.

Pang, Ken C., Dinger, Marcel E., Mercer, Tim R., *et al.* (2009). "Genome-wide identification of long noncoding RNAs in CD8+ T cells". *Journal of immunology (Baltimore, Md. : 1950)* 182: pp. 7738–7748.

Papagrigoraki, Anastasia, Maurelli, Martina, Giglio, Micol, *et al.* (2017). "Advanced Glycation End Products in the Pathogenesis of Psoriasis". 18: p. 2471.

Parameswaran, N and S, Patial (2010). "Tumor necrosis factor- signaling in macrophages". *Critical Reviews in Eukaryotic Gene Expression* 20: pp. 87–103.

Park, SY, Lee, SW, Kim, HY, *et al.* (2016). "SIRT1 inhibits differentiation of monocytes to macrophages: amelioration of synovial inflammation in rheumatoid arthritis". *Journal of Molecular Medicine* 94: pp. 921–31.

Partsch, G, Steiner, G, Leeb, BF, *et al.* (1997). "Highly increased levels of tumor necrosis factor-alpha and other proinflammatory cytokines in psoriatic arthritis synovial fluid." *The Journal of dermatology* 24: pp. 518–523.

Pasaniuc, B and Price, A.L (2017). "Dissecting the genetics of complex traits using summary association statistics". *Nature Reviews Genetics* 18: pp. 117–127.

Patrick, MT., Stuart, PE., and Raja, K. (2018). "Genetic signature to provide robust risk assessment of psoriatic arthritis development in psoriasis patients". *Nature Communications*.

Pedersen, OB, Svendsen, AJ, and oEjstrup, L (2008). "On the heritability of psoriatic arthritis. Disease concordance among monozygotic and dizygotic twins". *Annals of the rheumatic diseases* 67: pp. 1417–1421.

Peeters, J. G. C., Vervoort, S. J., Tan, S. C., *et al.* (2015). "Inhibition of super-enhancer activity in autoinflammatory site-derived T cells reduces disease-associated gene expression". *Cell reports* 12: pp. 1986–1996.

Pelikan, Richard C, Kelly, Jennifer A, Fu, Yao, *et al.* (2018). "Enhancer histone-QTLs are enriched on autoimmune risk haplotypes and influence gene expression within chromatin networks." *Nat Commun* 9: p. 2905.

Pène, J, Chevalier, S, and L, Preisser (2008). "Chronically inflamed human tissues are infiltrated by highly differentiated Th17 lymphocytes". *The Journal of immunology* 180: pp. 7423–7430.

Pepenella, S., Murphy, K. J., and J., Hayes -J. (2014). "Intra-and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure". *Chromosoma* 123: pp. 3–13.

Perera, G. K., Di Meglio, P., and Nestle, F. O. (2012). "Psoriasis". *Annu Rev Pathol* 7: pp. 385–422.

Peters, J. E., Lyons, P. A., Lee, J. C., *et al.* (2016). "Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease". *PLoS One*.

Petersen, C. P., Bordeleau, M. E., Pelletier, J., *et al.* (2006). "Short RNAs repress translation after initiation in mammalian cells". *Molecular cell*.

Petri, M., Singh, S., Tesfasyone, *et al.* (2009). "Longitudinal expression of type I interferon responsive genes in systemic lupus erythematosus." *Lupus* 18: 980–9.

Petronis, Arturas (2010). "Epigenetics as a unifying principle in the aetiology of complex traits and diseases". *Nature* 465: p. 721.

Picelli, S., Faridani, O. R., Björklund, ÅK, *et al.* (2014). "Full-length RNA-seq from single cells using Smart-seq2". *Nature protocols* 9: pp. 171–181.

Pickrell, J. K., Marioni, J. C., Pai, A. A., *et al.* (2010). "Understanding mechanisms underlying human gene expression variation with RNA sequencing". *Nature* 464: pp. 768–772.

Pitzalis, C., Cauli, A., Pipitone, N., *et al.* (1996). "Cutaneous lymphocyte antigen-positive T lymphocytes preferentially migrate to the skin but not to the joint in psoriatic arthritis". *Arthritis Rheum* 39: 137–45.

Pivarcsi, A., Ståhle, M., and Sonkoly -E. (2014). "Genetic polymorphisms altering microRNA activity in psoriasis–a key to solve the puzzle of missing heritability?" *Experimental dermatology* 23: pp. 620–624.

Pogribny, I. P. and Tryndyak, V. P. Bagnyukova -T. V. (2009). "Hepatic epigenetic phenotype predetermines individual susceptibility to hepatic steatosis in mice fed a lipogenic methyl-deficient diet". *Journal of hepatology* 51: pp. 176–186.

Polach, K. J., Lowary, P. T., and Widom -J. (2000). "Effects of core histone tail domains on the equilibrium constants for dynamic dna site accessibility in nucleosomes1". *Journal of molecular biology* 298: pp. 211–223.

Polachek, A., Li, S., care, *et al.* (2017). "Clinical Enthesitis in a Prospective Longitudinal Psoriatic Arthritis Cohort: Incidence, Prevalence, Characteristics, and Outcome". *Arthritis care & research* 69: pp. 1685–1691.

Ponts, N., Harris, E. Y., Prudhomme, J., *et al.* (2010). "Nucleosome landscape and control of transcription in the human malaria parasite". *Genome Research* 20: pp. 228–238.

Porcher, R., Said-Nahal, R., and A., D'Agostino -M. (2005). "Two major spondylarthropathy phenotypes are distinguished by pattern analysis in multiplex families". *Arthritis Care & Research* 53: pp. 263–271.

Présumey, J, Courties, G, and P, Louis-Plence - (2013). "Nicotinamide phosphoribosyltransferase/visfatin expression by inflammatory monocytes mediates arthritis pathogenesis". *Annals of the Rheumatic Diseases* 72: pp. 1717–1724.

Proksch, E., Brandner, J. M., and Jensen, J. M. (2008). "The skin: an indispensable barrier". *Exp Dermatol* 17: pp. 1063–72.

Qian, F., Deng, J., Cheng, N., *et al.* (2009). "A non-redundant role for MKP5 in limiting ROS production and preventing LPS-induced vascular injury". *The EMBO journal* 28: pp. 2896–2907.

Qu, C, NS, Brinck-Jensen, Zang, M, *et al.* (2014). "Monocyte-derived dendritic cells: targets as potent antigen-presenting cells for the design of vaccines against infectious diseases". *International Journal of infectious diseases* 19: pp. 1–5.

Qu, K., Zaba, L. C., Satpathy, A. T., *et al.* (2017). "Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors". *Cancer Cell* 32: pp. 27–41.

Qu, Kun, Zaba, Lisa C., Giresi, Paul G., *et al.* (2015). "Individuality and variation of personal regulomes in primary human T cells". *Cell systems* 1: pp. 51–61.

Quinlan, A. R. and Hall, I. M. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". *Bioinformatics*.

Raj, Towfique, Rothamel, Katie, Mostafavi, Sara, *et al.* (2014). "Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes". *Science* 344: pp. 519–523.

Rajasekhar, M., Olsson, AM., and Steel, KJA (2017). "MicroRNA-155 contributes to enhanced resistance to apoptosis in monocytes from patients with rheumatoid arthritis". *Journal of autoimmunity* 79: pp. 53–62.

Rao, D. A., Gurish, M. F., and Marshall, J. L. Slowikowski -K. (2017). "Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis". *Nature* 542: pp. 110–114.

Rauch, I, Müller, M, and T, Decker (2013). "The regulation of inflammation by interferons and their STATs". *Jak-Stat* 2: e23820.

Raurell-Vila, H., Ramos-Rodriguez, M., and Pasquali, L. (2018). "Assay for Transposase Accessible Chromatin (ATAC-Seq) to Chart the Open Chromatin Landscape of Human Pancreatic Islets". *Methods in Molecular Biology* 1766: pp. 197–208.

Ray-Jones, H. F., McGovern, A., Warren, P., *et al.* (2017). "[Abstract] Capture Hi-C identifies chromatin interactions between psoriasis-associated genetic loci and disease candidate genes." *UNPUBLISHED*.

Reich, K. (2012). "The concept of psoriasis as a systemic inflammation: implications for disease management". *Journal of the European Academy of Dermatology ...* 26: pp. 3–11.

Reich, K., Krüger, K., and Mössner -R. (2009). "Epidemiology and clinical pattern of psoriatic arthritis in Germany: a prospective interdisciplinary epidemiological study of 1511 patients with plaque-type psoriasis". *The British Journal of Dermatology* 160: pp. 1040–1047.

Reich, Kristian, Papp, Kim A., Matheson, Robert T., *et al.* (2015). "Evidence that a neutrophil–keratinocyte crosstalk is an early target of IL-17A inhibition in psoriasis". *Experimental Dermatology* 24: pp. 529–535.

Rendeiro, André F., Schmidl, Christian, Strefford, Jonathan C., *et al.* (2016). "Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks". *Nature communications* 7: p. 11938.

Reveille, J. D., Sims, A. M., Danoy, P., *et al.* (2010). "Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci". *Nature Genetics* 42: pp. 123–127.

Ritchlin, C., Haas-Smith, S. A., and Hicks -D. (1998). "Patterns of cytokine production in psoriatic synovium". *The Journal of rheumatology* 25: pp. 1544–1552.

Rittling, SR. and Singh, R. (2015). "Osteopontin in immune-mediated diseases". *Journal of dental research* 94: pp. 1638–1645.

Rivollier, A, He, J, Kole, A, *et al.* (2012). "Inflammation switches the differentiation program of Ly6Chi monocytes from antiinflammatory macrophages to inflammatory dendritic cells in the colon". *Journal of Experimental Medicine* 209: pp. 139–155.

Rizova, E. and Corroller, M. (2001). "Topical calcitriolstudies on local tolerance and systemic safety". *British Journal of Dermatology* 144: pp. 3–10.

Romanowska, M, Yacoub, Al N, and H, Seidel (2008). "PPAR enhances keratinocyte proliferation in psoriasis and induces heparin-binding EGF-like growth factor". *Journal of Investigative Dermatology* 128: pp. 110–124.

Rosenberger, C., Solovan, C., and D., Rosenberger -A. (2007). "Upregulation of hypoxia-inducible factors in normal and psoriatic skin". *Journal of Investigative Dermatology* 127: pp. 2445–2452.

Ross, E. L., D'Cruz, D., and Morrow, W. J. (2000). "Localized monocyte chemotactic protein-1 production correlates with T cell infiltration of synovium in patients with psoriatic arthritis". *The Journal of rheumatology* 27: pp. 2432–2443.

Rossetto, Cyprian C. and Pari, Gregory (2012). "KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome". *PLoS pathogens* 8.

Rotem, A., Ram, O., Shoresh, N., *et al.* (2015). "Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state". *Nature . . .*

Rousseeuw, Peter J (1987). "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis". 20: pp. 53–65.

Ruchusatsawat, K, Wongpiyabovorn, J, and P, Protjaroen (2011). "Parakeratosis in skin is associated with loss of inhibitor of differentiation 4 via promoter methylation". *Human pathology* 42: pp. 1878–887.

Rudwaleit, M., Heijde, Van D., and Landewé -R. (2009). "The development of Assessment of Spondyloarthritis International Society (ASAS) classification criteria for axial spondyloarthritis (part II): validation and final selection". *Annals of the rheumatic diseases* 68: pp. 777–783.

S, Shen-Orr, Shai and Furman, David (2013). "Variability in the immune system: of vaccine responses and immune states." *Curr. Opin. Immunol.* 25: pp. 542–7.

Sahlol, N.Y., Mostafa, M.S., Madkour, L.A., *et al.* (2019). "Low TNFAIP3 expression in psoriatic skin promotes disease susceptibility and severity". *PLOS ONE* 14.

Sahmatova L. and, Sügis-E., Šunina, M., Hermann, H., *et al.* (2017). "Signs of innate immune activation and premature immunosenescence in psoriasis patients". *Scientific Reports* 7.

Said-Nahal, R., Miceli-Richard, C., D'Agostino, MA., *et al.* (2001). "Phenotypic diversity is not determined by independent genetic factors in familial spondylarthropathy". *Arthritis Rheumatology* 45: pp. 478–484.

Sampogna, Francesca, Tabolli, Stefano, and Abeni, Damiano (2012). "Living with psoriasis: prevalence of shame, anger, worry, and problems in daily activities and social life". *Acta dermato-venereologica* 92: pp. 299–303.

Saxena, H., Deshpande, DA., and Tiegs, BC. (2012). "The GPCR OGR1 (GPR68) mediates diverse signalling and contraction of airway smooth muscle in response to small reductions in extracellular pH". *British journal of pharmacology* 16: pp. 981–990.

Schafflick, D, Cole, M, Hartlehnert, M, *et al.* (2018). "Single-cell transcriptomics identifies drivers of local inflammation in multiple sclerosis". *bioRxiv*.

Scharer, Christopher D., Blalock, Emily L., Barwick, Benjamin G., *et al.* (2016). "ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells". *Scientific reports* 6: p. 27030.

Schauer, C, Janko, C, Munoz, LE, *et al.* (2014). "Aggregated neutrophil extracellular traps limit inflammation by degrading cytokines and chemokines". *Nature medicine* 20: pp. 511–517.

Schett, G., Coates, L. C., and R., Ash -Z. (2011). "Structural damage in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: traditional views, novel insights gained from TNF blockade, and . . . " *Arthritis research and therapy* 13.

Schmid, P. and Itin P. and, Cox -D. (1994). "The type I interferon system is locally activated in psoriatic lesions". *Journal of interferon research* 14: pp. 229–234.

Schmidl, Christian, Rendeiro, André F., Sheffield, Nathan C., *et al.* (2015). "ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors". *Nature methods* 12: p. 963.

Schmitt, J. and Rosumeck, S. (2014). "Efficacy and safety of systemic treatments for moderate-to-severe psoriasis: meta-analysis of randomized controlled trials". *British Journal of Dermatology* 170: pp. 274–303.

Schmitt, J. and Wozel, G. (2005). "The psoriasis area and severity index is the adequate criterion to define severity in chronic plaque-type psoriasis". *Dermatology* 210: pp. 194–199.

Schork, N. J., Murray, S. S., Frazer, K. A., *et al.* (2009). "Common vs. rare allele hypotheses for complex diseases". *Curr Opin Genet Dev* 19: pp. 212–9.

Shadidi, KR, Aarvak, T, and JE, Henriksen (2003). "The chemokines CCL5, CCL2 and CXCL12 play significant roles in the migration of Th1 cells into rheumatoid synovial tissue". *Scandinavian journal of immunology* 57: pp. 192–198.

Shapiro, J., Cohen, A. D., David, M., *et al.* (2007). "The association between psoriasis, diabetes mellitus, and atherosclerosis in Israel: a case-control study". *Journal of the American Accademy of Dermatology* 56: pp. 629–634.

Shen, L., Shao, N. Y., Liu, X., *et al.* (2013). "diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates". *PloS one*.

Shi, Lihua, Zhang, Zhe, Yu, Angela M., *et al.* (2014). "The SLE transcriptome exhibits evidence of chronic endotoxin exposure and has widespread dysregulation of non-coding and coding RNAs". *PloS one* 9.

Shi, LZ, Wang, R, Huang, G, *et al.* (2011). "HIF1–dependent glycolytic pathway orchestrates a metabolic checkpoint for the differentiation of TH17 and Treg cells". *Journal of experimental medicine* 208: p. 1367.

Shu, Jin, Li, Ling, Zhou, Lan-Bo, *et al.* (2017). "IRF5 is elevated in childhood-onset SLE and regulated by histone acetyltransferase and histone deacetylase inhibitors". *Oncotarget* 8: p. 47184.

Singh, Y, Chen, H, Zhou, Y, *et al.* (2015). "Differential effect of DJ-1/PARK7 on development of natural and induced regulatory T cells". *Scientific reports*.

Small, K. S., Hedman, ÅK, Grundberg, E., *et al.* (2011). "Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes". *Nature Genetics* 43: pp. 561–564.

Smallwood, SA, Lee, HJ, Angermueller, C, *et al.* (2014). "Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity". *Nature Methods* 11: pp. 817–820.

Smemo, S., Tena, J. J., Kim, K. H., *et al.* (2014). "Obesity-associated variants within FTO form long-range functional connections with IRX3". *Nature* 507: pp. 371–375.

Smith, Emily M., Lajoie, Bryan R., Jain, Gaurav, *et al.* (2016). "Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus". *American journal of human genetics* 98: pp. 185–201.

Smith, Judith A., Barnes, Michael D., Hong, Dihua, *et al.* (2008). "Gene expression analysis of macrophages derived from ankylosing spondylitis patients reveals interferon dysregulation". *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 58: pp. 1640–1649.

Soccio, Raymond E., Chen, Eric R., Rajapurkar, Satyajit R., *et al.* (2015). "Genetic variation determines PPAR function and anti-diabetic drug response in vivo". *Cell* 162: pp. 33–44.

Soderstrom, K., Stein, E., Colmenero, P., *et al.* (2010). "Natural killer cells trigger osteoclastogenesis and bone destruction in arthritis". *Proc Natl Acad Sci U S A* 107: pp. 13028–13033.

Solomon, M. J., Larsen, P. L., and Cell, Varshavsky -A. (1988). "Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene". *Cell* 53: pp. 937–947.

Sos, Brandon, Fung, Ho-Lim, Gao, Derek, *et al.* (2016). "Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay". *Genome Biology* 17: p. 20.

Spadaro, A., Scrivo, R., Moretti, T., *et al.* (2004). "Natural killer cells and gamma/delta T cells in synovial fluid and in peripheral blood of patients with psoriatic arthritis". *Clinical and experimental rheumatology* 22: pp. 389–394.

Spain, S. L. and Barrett, J. C. (2015). "Strategies for fine-mapping complex traits". *Human molecular genetics* 24R1: pp. 111–119.

Springate, D.A., Parisi, R., and Kontopantelis, E. (2017). "Incidence, prevalence and mortality of patients with psoriasis: a UK population-based cohort study". *British Journal of Dermatology* 176: pp. 650–658.

Srivastava, Shivani, Macaubas, Claudia, Deshpande, Chetan, *et al.* (2010). "Monocytes are resistant to apoptosis in systemic juvenile idiopathic arthritis". 136: pp. 257–268.

Stefan, M., Wei, C., Lombardi, A., *et al.* (2014). "Genetic-epigenetic dysregulation of thymic TSH receptor gene expression triggers thyroid autoimmunity". *Proc Natl Acad Sci U S A* 111: pp. 12562–7.

Stoeckman, A. K., Baechler, E. C., and A., Ortmann -W. (2006). "A distinct inflammatory gene expression profile in patients with psoriatic arthritis". *Genes and Immunity* 7: pp. 583–591.

Strange, A., Capon, F., Spencer, C. C., *et al.* (2010). "A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1". *Nat Genet* 42: pp. 985–90.

Stratis, Athanasios, Pasparakis, Manolis, Rupec, Rudolf A., *et al.* (2006). "Pathogenic role for skin macrophages in a mouse model of keratinocyte-induced psoriasis-like skin inflammation". *Journal of Clinical Investigation* 116: pp. 2094–2104.

Stuart, Philip E., Nair, Rajan P., Ellinghaus, Eva, *et al.* (2010). "Genome-wide association analysis identifies three psoriasis susceptibility loci". *Nature genetics* 42: pp. 1000–1004.

Stuart, Philip E., Nair, Rajan P., Tsoi, Lam C., *et al.* (2015). "Genome-wide association analysis of psoriatic arthritis and cutaneous psoriasis reveals differences in their genetic architecture". *The American Journal of Human Genetics* 97.

Sugiyama, H, Gyulai, R, Toichi, E, *et al.* (2005). "Dysfunctional blood and target tissue CD4+ CD25high regulatory T cells in psoriasis: mechanism underlying unrestrained pathogenic effector T cell proliferation". *The Journal of Immunology* 174: pp. 64–73.

Sullivan, KE, Reddy, ABM, and K, Dietzmann (2007). "Epigenetic regulation of tumor necrosis factor alpha". *Molecular and cellular biology* 27: pp. 5147–5160.

Sun, Lian-Dang, Cheng, Hui, Xue-Qin, Yang, *et al.* (2010). "Association analyses identify six new psoriasis susceptibility loci in the Chinese population". *Nature Genetics* 42.

Swindell, W. R., Remmer, H. A., Sarkar, MK., *et al.* (2015). "Proteogenomic analysis of psoriasis reveals discordant and concordant changes in mRNA and protein abundance". *Genome Medicine* 7: p. 86.

Swindell, W. R., Sarkar, M. K., Liang, Y., *et al.* (2017). "RNA-seq identifies a diminished differentiation gene signature in primary monolayer keratinocytes grown from lesional and uninvolved psoriatic skin". *Scientific reports*.

Taams, Leonie S (2018). "Inflammation and immune resolution". *Clinical & Experimental Immunology* 193: pp. 1–2.

Taganov, K. D., Boldin, M. P., J., Chang -K., *et al.* (2006). "NF-B-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses". *Proc Natl Acad Sci U S A* 103: pp. 12481–12486.

Takagi, Rie, Higashi, Takehiro, Hashimoto, Kumiko, *et al.* (2008). "B cell chemoattractant CXCL13 is preferentially expressed by human Th17 cell clones". 181: pp. 186–189.

Tan, NS, Michalik, L, Noy, N, *et al.* (2001). "Critical roles of PPAR/ in keratinocyte response to inflammation". *Genes & Development* 15: pp. 3263–3277.

Tang, F., Barbacioru, C., Wang, Y., *et al.* (2009). "mRNA-Seq whole-transcriptome analysis of a single cell". *Nature methods* 6: p. 377.

Tang, Fuchou, Barbacioru, Catalin, Nordman, Ellen, *et al.* (2010). "RNA-Seq analysis to capture the transcriptome landscape of a single cell". *Nature protocols* 5: pp. 516–535.

Tang, H., Jin, X., Li, Y., *et al.* (2014). "A large-scale screen for coding variants predisposing to psoriasis". *Nature Genetics* 45: pp. 45–50.

Tassiulas, I., Duncan, S.R., Centola, M., *et al.* (1999). "Clonal characteristics of T cell infiltrates in skin and synovium of patients with psoriatic arthritis". *Hum Immunol* 60: 479–91.

Taylor, W, Gladman, D, Helliwell, P., *et al.* (2006). "Classification criteria for psoriatic arthritis: development of new criteria from a large international study". *Arthritis & Rheumatism* 54: pp. 2665–2673.

Telfer, NR, Chalmers, RJ, Whale, K, *et al.* (1992). "The role of streptococcal infection in the initiation of guttate psoriasis." *Arch Dermatol* 128: pp. 39–42.

Tervaniemi, Mari H., Katayama, Shintaro, Skoog, Tiina, *et al.* (2016). "NOD-like receptor signaling and inflammasome-related pathways are highlighted in psoriatic epidermis". *Scientific reports* 6: p. 22745.

Thrastardottir, T and Love, TJ. (2018). "Infections and the risk of psoriatic arthritis among psoriasis patients: a systematic review". *Rheumatology international* 38: pp. 1385–1397.

Thurner, M., Bunt, M. van de, Torres, J. M., *et al.* (2018). "Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci". *Elife*.

Toberer, Ferdinand, Sykora, Jaromir, Göttel, Daniel, *et al.* (2011). "Tissue microarray analysis of RANKL in cutaneous lupus erythematosus and psoriasis". *Experimental dermatology* 20: pp. 600–602.

Tobin, AM, Lynch, L, Kirby, B, *et al.* (2011). "Natural killer cells in psoriasis". *Journal of innate immunity* 3: pp. 403–410.

Togayachi, Akira, Kozono, Yuko, Kuno, Atsushi, *et al.* (2010). "3GnT2 (B3GNT2), a major polylactosamine synthase: analysis of B3GNT2-deficient mice". 479: pp. 185–204.

Tonel, Giulia, Conrad, Curdin, Laggner, Ute, *et al.* (2010). "Cutting edge: A critical functional role for IL-23 in psoriasis." *J. Immunol.* 185: pp. 5688–91.

Traag, V, Waltman, L, and arXiv:1810.08473, Eck NJ, van preprint (2018). "From Louvain to Leiden: guaranteeing well-connected communities". *arXiv preprint arXiv:1810.08473*.

Trynka, G. and Raychaudhuri, S. (2013a). "Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases". *Current opinion in genetics & development* 23: pp. 635–641.

Trynka, G., Hunt, K. A., Bockett, N. A., *et al.* (2011). "Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease". *Nature Genetics* 43: pp. 1183–1201.

Trynka, Gosia and Raychaudhuri, Soumya (2013b). "Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases". *Curr Opin Genetics Dev* 23: pp. 635–641.

Tsoi, L. C., Spain, S. L., Knight, J., *et al.* (2012). "Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity". *Nat Genet* 44: pp. 1341–8.

Tsoi, L. C., Stuart, P. E., Tian, C., *et al.* (2017). "Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants". *Nature Genetics* 8: p. 15382.

Tsoi, Lam C, Spain, Sarah L, Ellinghaus, Eva, *et al.* (2015a). "Enhanced meta-analysis and replication studies identify five new psoriasis susceptibility loci". *Nature Communications* 6: p. 7001.

Tsoi, L.C., Iyer, M.K., Stuart, P.E., *et al.* (2015b). "Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin". *Genome Biology* 16: p. 24.

Tsou, Chia-Lin, Peters, Wendy, Si, Yue, *et al.* (2007). "Critical roles for CCR2 and MCP-3 in monocyte mobilization from bone marrow and recruitment to inflammatory sites". *The Journal of Clinical Investigation* 117: pp. 902–909.

Ueno-Shuto, K., Kato, K, Tasaki, Y, *et al.* (2014). "Lipopolysaccharide decreases single immunoglobulin interleukin-1 receptor-related molecule (SIGIRR) expression by suppressing specificity protein 1 (Sp1) via the TLR4-p38 pathway in monocytes and neutrophils". *Journal of Biological Chemistry*.

Valdimarsson, H., Thorleifsdottir, RH., Sigurdardottir, SL., *et al.* (2009). "Psoriasis–as an autoimmune disease caused by molecular mimicry". *Trends in Immunology* 30: pp. 494–501.

Valk, E., Leung, R., Kang, H., *et al.* (2006). "T cell receptor-interacting molecule acts as a chaperone to modulate surface expression of the CTLA-4 coreceptor". *Immunity* 25: pp. 807–821.

Vandooren, B, Kruithof, E, Yu, DTY, *et al.* (2004). "Involvement of matrix metalloproteinases and their inhibitors in peripheral synovitis and down-regulation by tumor necrosis factor blockade in spondylarthropathy". 50: pp. 2942–3953.

Vence, L., Schmitt, A., and E., Meadows -C. (2015). "Recognizing guttate psoriasis and initiating appropriate treatment". *West Virginia Medical Journal* 111: pp. 26–28.

Vereecke, L, Beyaert, R, and Loo, G van (2011). "Genetic relationships between A20/TNFAIP3, chronic inflammation and autoimmune disease".

Vernes, Sonja C., Spiteri, Elizabeth, Nicod, Jérôme, *et al.* (2007). "High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders". *American journal of human genetics* 81: pp. 1232–1250.

Vierstra, Jeff, Wang, Hao, John, Sam, *et al.* (2014). "Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH". *Nature Methods* 11: pp. 66–72.

Villani, AC, Satija, R, Reynolds, G, *et al.* (2017). "Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors". . . .

Villanova, Federica, Meglio, Paola, and Nestle, Frank O. (2013). "Biomarkers in psoriasis and psoriatic arthritis". *Annals of the rheumatic diseases* 72.

Visscher, PM, Wray, NR, Zhang, Q, *et al.* (2017). "10 years of GWAS discovery: biology, function, and translation". *The American Journal of Human Genetics* 101: pp. 5–22.

Wakefield, J. (2007). "A Bayesian measure of the probability of false discovery in genetic epidemiology studies". *The American Journal of Human Genetics*.

Wallace, Chris, Cutler, Antony J., Pontikos, Nikolas, *et al.* (2015). "Dissection of a Complex Disease Susceptibility Region Using a Bayesian Stochastic Search Approach to Fine Mapping". *PLoS genetics* 11.

Wang, H., Kess, D., Lindqvist, A. K. B., *et al.* (2008). "A 9-centimorgan interval of chromosome 10 controls the T cell-dependent psoriasiform skin disease and arthritis in a murine psoriasis model". *The Journal of Immunology* 180: pp. 5520–5529.

Wang, Honglin, Peters, Thorsten, Kess, Daniel, *et al.* (2006). "Activated macrophages are essential in a murine model for T cell–mediated chronic psoriasiform skin inflammation". *Journal of Clinical Investigation* 116: pp. 2105–2114.

Wang, Jie, Zibetti, Cristina, Shang, Peng, *et al.* (2018). "ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration". 9: p. 1364.

Wang, Jun, Al-Lamki, Rafia S., Zhu, Xinwang, *et al.* (2014). "TL1-A can engage death receptor-3 and activate NF-kappa B in endothelial cells". *BMC nephrology* 15: p. 178.

Ward, L. D. and Kellis, M. (2012). "Interpreting noncoding genetic variation in complex traits and human disease". *Nature biotechnology*.

Weaver, CT, Hatton, RD, and PR, Mangan (2007). "IL-17 family cytokines and the expanding diversity of effector T cell lineages". *Annual Reviews* 25: pp. 821–852.

Weiss, G., Shemer, A., and Academy, Trau -H. of the (2002). "The Koebner phenomenon: review of the literature". *Journal of the European Academy . . .*

Weitzmann, MN, Cenci, S, Rifas, L, *et al.* (2000). "Interleukin-7 stimulates osteoclast formation by up-regulating the T-cell production of soluble osteoclastogenic cytokines". *Blood* 96: pp. 1873–1878.

Welter, D., MacArthur, J., Morales, J., *et al.* (2013). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations". *Nucleic acids . . .*

Westra, HJ, Peters, MJ, Esko, T, *et al.* (2013). "Systematic identification of trans eQTLs as putative drivers of known disease associations". *Nature . . .*

Whitney, Adeline R., Diehn, Maximilian, Popper, Stephen J., *et al.* (2003). "Individuality and variation in gene expression patterns in human blood". *Proc Natl Acad Sci U S A* 100: pp. 1896–1901.

Wikramanayake, T. C., Stojadinovic, O., and Tomic-Canic, M. (2014). "Epidermal Differentiation in Barrier Maintenance and Wound Healing". *Adv Wound Care (New Rochelle)*. Vol. 3: pp. 272–280.

Williams, Fionnuala, Meenagh, Ashley, Sleator, Carole, *et al.* (2005). "Activating Killer Cell Immunoglobulin-Like Receptor Gene KIR2DS1 Is Associated With Psoriatic Arthritis". *Human Immunology* 66: pp. 836–841.

Wirasinha, RC, Vijayan, D, *et al.* (2018). "GPR65 inhibits experimental autoimmune encephalomyelitis through CD4+ T cell independent mechanisms that include effects on iNKT cells". *Immunology and cell . . .*

Wolf, Alexander F, Angerer, Philipp, and Theis, Fabian J (2018). "SCANPY: large-scale single-cell gene expression data analysis". 19: p. 15.

Wolk, K, Witte, E, Wallace, E, *et al.* (2006). "IL-22 regulates the expression of genes responsible for antimicrobial defense, cellular differentiation, and mobility in keratinocytes: a potential role in psoriasis". *European journal of immunology* 36: pp. 1309–1323.

Wong, Brian R, Rho, Jaerang, Arron, Joseph, *et al.* (1997). "TRANCE is a novel ligand of the tumor necrosis factor receptor family that activates c-Jun N-terminal kinase in T cells". 272: pp. 25190–25194.

Wong, W. F., Kohu, K., Chiba, T., *et al.* (2011). "Interplay of transcription factors in T-cell differentiation and function: the role of Runx". *Immunology* 132: pp. 157–164.

Woolacott, N., Hawkins, N., Kainth, A., *et al.* (2006). "Etanercept and Efalizumab for the Treatment of Moderate to Severe Psoriasis". *Etanercept and Efalizumab for the Treatment of Moderate to Severe Psoriasis* 10: pp. 1–233.

Woolf, E, Xiao, C, Fainaru, O, *et al.* (2003). "Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis". *Proc Natl Acad Sci U S A*.

Wray, Naomi R. (2005). "Allele frequencies and the r 2 measure of linkage disequilibrium: impact on design and interpretation of association studies". *Twin Research and Human Genetics* 8: pp. 87–94.

Wrone-Smith, T. and Nickoloff, B. J. (1996). "Dermal injection of immunocytes induces psoriasis". *The Journal of clinical investigation* 98: pp. 1878–1887.

Wu, Q, Jin, H, Yang, Z, *et al.* (2010). "MiR-150 promotes gastric cancer proliferation by negatively regulating the pro-apoptotic gene EGR2". *Biochemical and Biophysical Research Communications* 392: pp. 340–345.

Xia, P., Fang, X., Zhang, Z.H, *et al.* (2012). "Dysregulation of miRNA146a versus IRAK1 induces IL-17 persistence in the psoriatic skin lesions". *Immunol Lett* 148: pp. 151–162.

Xia, Yu-Ping, Li, Baosheng, Hylton, Donna, *et al.* (2003). "Transgenic delivery of VEGF to mouse skin leads to an inflammatory condition resembling human psoriasis". 102: pp. 161–168.

Xinmin, L., Kim, J., Zhou, J., *et al.* (2005). "Use of signal thresholds to determine significant changes in microarray data analyses". *Genetics and Molecular Biology* 28.

Yagil, C., Hubner, N., Monti, J., *et al.* (2005). "Identification of Hypertension-Related Genes Through an Integrated Genomic-Transcriptomic Approach". *Circulation Research* 96: 617–625.

Yamazaki, H., Hiramatsu, N., and Hayakawa -K. (2009). "Activation of the Akt-NF-B pathway by subtilase cytotoxin through the ATF6 branch of the unfolded protein response". *The Journal of immunology* 183: pp. 1480–1487.

Yan, Di, Issa, Naiem, Afifi, Ladan, *et al.* (2017). "The Role of the Skin and Gut Microbiome in Psoriatic Disease". *Current Dermatology Reports* 6: pp. 94–103.

Yan, Hai, Yuan, Weishi, Velculescu, Victor E., *et al.* (2002). "Allelic variation in human gene expression". *Science* 297: pp. 1143–1143.

Yanaba, K, Kamata, M, Ishiura, N., *et al.* (2013). "Regulatory B cells suppress imiquimod-induced, psoriasis-like skin inflammation". *Journal of leukocyte biology* 94: pp. 563–573.

Yang, J., Benyamin, B., McEvoy, B. P., *et al.* (2010). "Common SNPs explain a large proportion of the heritability for human height". *Nat Genet* 42: pp. 565–9.

Yang, J., Hu, S., Zhao, L., *et al.* (2016a). "Selective programming of CCR10(+) innate lymphoid cells in skin-draining lymph nodes for cutaneous homeostatic regulation". *Nature Immunology* 17: pp. 48–56.

Yang, Jun, Diaz, Norma, Adelsberger, Joseph, *et al.* (2016b). "The effects of storage temperature on PBMC gene expression". 17: p. 6.

Yao, P and Fox, PL. (2013). "Aminoacyl-tRNA synthetases in medicine and disease". *EMBO molecular medicine* 5: pp. 332–343.

Yao, Yi, Liu, Rebecca, Shin, Min, *et al.* (2014). "CyTOF supports efficient detection of immune cell subsets from small samples". *Journal of immunological methods* 415: pp. 1–5.

Yen, J. H., Moore, B. E., Nakajima, T., *et al.* (2001). "Major histocompatibility complex class I–recognizing receptors are disease risk genes in rheumatoid arthritis". *Journal of experimental medicine* 193: pp. 1159–1167.

Yin, Xianyong, Low, Hui Q., Wang, Ling, *et al.* (2015). "Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility". *Nature communications* 6: pp. 1–11.

Yogo, T, Nagamiya, H, Seto, M, *et al.* (2016). "Structure-Based Design and Synthesis of 3-Amino-1,5-dihydro-4H-pyrazolopyridin-4-one Derivatives as Tyrosine Kinase 2 Inhibitors". *Journal of medicinal chemistry* 59: pp. 733–749.

Yoon, Bo, Yoo, Su-Jin, Choi, Yeon, *et al.* (2014). "Functional phenotype of synovial monocytes modulating inflammatory T-cell responses in rheumatoid arthritis (RA)". 9: e109775.

Yosef, N and Regev, A (2016). "Writ large: Genomic dissection of the effect of cellular environment on immune response". 354: pp. 64–68.

Zaba, LC, Cardinale, I, and P, Gilleaudeau (2007). "Amelioration of epidermal hyperplasia by TNF inhibition is associated with reduced Th17 responses". *Journal of experimental medicine* 204: pp. 3183–3194.

Zeeuwen, Patrick L, Jongh, Gys J de, Diana, Rodijk-Olthuis, *et al.* (2008). "Genetically programmed differences in epidermal host defense between psoriasis and atopic dermatitis patients." *PLoS ONE* 3: e2301.

Zgraggen, S., Huggenberger, R., Kerl, K., *et al.* (2014). "An important role of the SDF-1/CXCR4 axis in chronic skin inflammation". *PLoS One*.

Zhang, Chen, Wang, Chen, Jia, Zhenyu, *et al.* (2017). "Differentially expressed mRNAs, lncRNAs, and miRNAs with associated co-expression and ceRNA networks in ankylosing spondylitis". *Oncotarget* 8: p. 113543.

Zhang, F., Wu, L., Qian, J., *et al.* (2016). "Identification of the long noncoding RNA NEAT1 as a novel inflammatory regulator acting through MAPK pathway in human lupus". *Journal of autoimmunity* 75: pp. 96–104.

Zhang, F., Wei, K., Slowikowski, K., *et al.* (2018). "Defining Inflammatory Cell States in Rheumatoid Arthritis Joint Synovial Tissues by Integrating Single-cell Transcriptomics and Mass Cytometry". *bioRxiv*.

Zhang, P., Su, Y., Chen, H., *et al.* (2010). "Abnormal DNA methylation in skin lesions and PBMCs of patients with psoriasis vulgaris". *Journal of dermatological science* 60: pp. 40–42.

Zhang, X., Cowper-Sal, R., Bailey, S. D., *et al.* (2012). "Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24. 3 prostate cancer risk locus". *Genome Research* 22: pp. 1437–1446.

Zhang, X. J., Huang, W., Yang, S., *et al.* (2009). "Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21". *Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21*.

Zhang, Y., Liu, T., and A., Meyer -C. (2008). "Model-based analysis of ChIP-Seq (MACS)". *Genome Biology* 9: R137.

Zhao, C, Zhang, H, Wong, WC, *et al.* (2009). "Identification of novel functional differences in monocyte subsets using proteomic and transcriptomic methods". *Journal of proteome research* 8: pp. 4028–4038.

Zhao, S., Fung-Leung, W. P., Bittner, A., *et al.* (2014). "Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells". *PloS one*.

Zhong, Y, Kinio, A, and immunology, Saleh M in (2013). "Functions of NOD-like receptors in human diseases". *Frontiers in immunology*.

Zhu, K. J., Zhu, C. Y., Shi, G., *et al.* (2012). "Association of IL23R polymorphisms with psoriasis and psoriatic arthritis: a meta-analysis". *Inflammation Research* 61: pp. 1149–1154.

Ziegenhain, C, Vieth, B, Parekh, S, *et al.* (2017). "Comparative analysis of single-cell RNA sequencing methods". *Molecular cell* 65: pp. 631–643.

Zuk, O., Hechter, E., and R., Sunyaev -S. (2012). "The mystery of missing heritability: Genetic interactions create phantom heritability". *Proc Natl Acad Sci U S A* 4.

Zuo, Xianbo, Sun, Liangdan, Yin, Xianyong, *et al.* (2015). "Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis". *Nature communications* 6: p. 6793.