



# **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

This is my abstract...

# **Acknowledgements**

Thank you, thank you, thank you.

## 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. 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 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 and validate some of the tools and approaches. The peak filtering strategy for ATAC using IDR was proposed and implemented by Dr Gabriele Migliorini and I conducted additional analysis to validate it. 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. All the resources for fine-mapping analysis using genotyping or summary statistics data were provided by Dr Adrián Cortés. General advice on analysis were provided by Dr Silvia Salatino, Dr Hai Fang, Dr Katie Burnham, Dr Félicie Constantino, Dr Gabriele Migliorini and Dr Adrián Cortés and Enrique Váquez de Luis. Single-cell advice and scripts for some of the analysis was provided 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.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Declarations</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>Abbreviations</b>	<b>xvi</b>
<b>1 Introduction</b>	<b>1</b>
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 . . . . .	11
1.2.4 Therapeutic intervention . . . . .	15
1.3 Genetics of psoriasis and psoriatic arthritis . . . . .	17
1.3.1 Heritability . . . . .	17
1.3.2 Genome-wide association studies . . . . .	17
1.3.3 Relevance of non-coding variants in disease susceptibility .	22
1.3.4 The role of GWAS in highlighting immune-relevant cell types and pathways . . . . .	23
1.3.5 Limitations of GWAS . . . . .	27
1.4 Functional interpretation of GWAS in complex diseases . . . . .	29
1.4.1 Overcoming GWAS limitations . . . . .	29
1.4.2 Understanding the epigenetic landscape in complex diseases	29
1.4.3 The chromatin landscape . . . . .	32
1.4.4 Transcriptional profiles in disease . . . . .	37
1.4.5 Transcriptional regulation in complex diseases . . . . .	40
1.4.6 The use of fine-mapping to prioritise functional causal variants . . . . .	41
1.4.7 Generation, integration and interpretation of genomic data	42
1.4.8 Aims and objectives . . . . .	44

## CONTENTS

---

<b>2 Material and Methods</b>	<b>47</b>
2.1 Ethical approval and recruitment of study participants . . . . .	47
2.1.1 Psoriasis patient recruitment . . . . .	47
2.1.2 PsA patient recruitment . . . . .	48
2.1.3 Healthy volunteer recruitment . . . . .	48
2.2 Sample processing . . . . .	49
2.2.1 PBMCs and synovial fluid cells isolation . . . . .	49
2.2.2 Skin biopsies processing and adherence assays . . . . .	50
2.2.3 Fixation, cryopreservation and cell culture . . . . .	50
2.2.4 Primary cell isolation using magnetic-activated cell sorting	51
2.2.5 Primary cell isolation using fluorescence-activated cell sorting . . . . .	52
2.3 Experimental protocols . . . . .	53
2.3.1 ATAC - Chromatin Accessibility . . . . .	53
2.3.2 Chromatin immunoprecipitation with sequencing library preparation by Tn5 transposase . . . . .	56
2.3.3 RNA extraction and gene expression quantification . . . . .	57
2.3.4 DNA extraction and rs4672405 genotyping . . . . .	60
2.3.5 Mass cytometry using cytometry by time of flight (CyTOF)	60
2.4 Computational and statistical analysis . . . . .	62
2.4.1 ATAC data analysis . . . . .	62
2.4.2 ChIPm data analysis . . . . .	65
2.4.3 Gene expression analysis . . . . .	67
2.4.4 Genomic region annotation, enrichment analysis and pathway visualisation . . . . .	69
2.4.5 Statistical fine-mapping . . . . .	71
2.4.6 Mass cytometry data analysis . . . . .	73
<b>3 Establishing laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples</b>	<b>75</b>
3.1 Introduction . . . . .	75
3.1.1 Principle of ATAC-seq and compatibility with clinical samples . . . . .	75
3.1.2 ATAC-seq limitations and advances in optimisation . . . . .	76
3.1.3 Challenges of ATAC data analysis . . . . .	77
3.1.4 The challenge of working with clinical samples . . . . .	80
3.1.5 Aims . . . . .	81
3.2 Results . . . . .	82
3.2.1 Establishment of an ATAC-seq data analysis pipeline . . . . .	82
3.2.2 Assessment of ATAC-seq transposition times in relevant cell types . . . . .	96
3.2.3 Comparison of ATAC-seq with Fast-ATAC protocol . . . . .	97
3.2.4 Limitations of ATAC-seq and Fast-ATAC to assess chromatin accessibility in keratinocytes . . . . .	99
3.2.5 Effect of cryopreservation and fixation in the chromatin landscape of immune primary cells . . . . .	105
3.3 Discussion . . . . .	116

## CONTENTS

---

3.3.1	ATAC: methodological aspects and pipeline establishment	117
3.3.2	The challenges of performing differential chromatin accessibility analysis . . . . .	119
3.3.3	Studying the chromatin landscape from psoriasis skin biopsies . . . . .	121
3.3.4	Characterisation of the effect of preservative techniques in the chromatin landscape . . . . .	122
3.3.5	Limitations . . . . .	123
3.4	Conclusions . . . . .	124
<b>4</b>	<b>Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis</b>	<b>125</b>
4.1	Introduction . . . . .	125
4.1.1	The systemic and skin-specific manifestations of psoriasis .	125
4.1.2	The personalised epigenome in disease . . . . .	126
4.1.3	Transcriptional profiles in psoriasis . . . . .	127
4.1.4	Chromatin accessibility, gene expression and genetic variability . . . . .	131
4.1.5	Fine-mapping using summary statistics . . . . .	132
4.2	Aims . . . . .	133
4.3	Results . . . . .	134
4.3.1	Psoriasis and healthy controls: cohort description and datasets . . . . .	134
4.3.2	Investigation of psoriasis-specific changes in the enhancer mark H3K27ac in peripheral blood immune cell populations	137
4.3.3	Interrogation of chromatin accessibility changes in psoriasis peripheral blood immune cells . . . . .	143
4.3.4	Differential gene expression analysis in circulating immune cells in psoriasis . . . . .	149
4.3.5	RNA-seq in epidermis from psoriasis patients . . . . .	161
4.3.6	Comparison of systemic and tissue-specific gene expression signatures in psoriasis . . . . .	169
4.3.7	Integration of chromatin accessibility and expression data for peripheral blood immune cells in psoriasis . . . . .	170
4.3.8	Fine-mapping of psoriasis GWAS loci and functional interpretation . . . . .	173
4.3.9	Allele-specific differences in chromatin accessibility at the GWAS locus 2p15 . . . . .	178
4.4	Discussion . . . . .	184
4.4.1	Chromatin accessibility and H3K27ac landscape in psoriasis immune cells . . . . .	184
4.4.2	Dysregulation of gene expression in psoriasis circulating immune cells . . . . .	186
4.4.3	Integration of chromatin accessibility and gene expression data . . . . .	189
4.4.4	Tranciptomic profiles in lesional and uninvolved psoriatic epidermis . . . . .	189

## CONTENTS

---

4.4.5	LncRNAs in psoriasis . . . . .	192
4.4.6	Differences in transcriptional dysregulation in peripheral blood and skin . . . . .	194
4.4.7	Fine-mapping using summary statistics and integration with epigenetic data . . . . .	195
4.4.8	Limitations in the approach . . . . .	198
4.5	Conclusions . . . . .	199
<b>5</b>	<b>Cross-tissue comparison of chromatin accessibility, gene expression signature and cytokine production in PsA</b>	<b>200</b>
5.1	Introduction . . . . .	200
5.1.1	The relevance of cell type and tissue specificity in the study of PsA . . . . .	200
5.1.2	Bulk transcriptomic studies in PsA and their limitations . .	201
5.1.3	Transcriptomics and proteomics at the single cell resolution	202
5.1.4	Multi-omics approach in the study of complex diseases . .	203
5.1.5	Integration of fine-mapping GWAS SNPs and functional data in PsA . . . . .	204
5.2	Aims . . . . .	206
5.3	Results . . . . .	207
5.3.1	PsA patients cohort description and datasets . . . . .	207
5.3.2	Immune cellular composition of blood and synovial fluid in the PsA cohort . . . . .	209
5.3.3	Differential chromatin accessibility analysis in immune cells reveals differences between synovial fluid and peripheral blood . . . . .	211
5.3.4	Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility . . . . .	219
5.3.5	Differential gene expression analysis in paired circulating and synovial immune cells . . . . .	221
5.3.6	Characterisation of CD14 <sup>+</sup> monocyte heterogeneity in PsA using scRNA-seq . . . . .	232
5.3.7	Identification of additional pathways dysregulated between synovial fluid and peripheral blood CD14 <sup>+</sup> monocytes using scRNA-seq data . . . . .	241
5.3.8	Integration of mass cytometry differences in protein expression with chromatin accessibility and gene expression data in PsA CD14 <sup>+</sup> monocytes. . . . .	243
5.3.9	Prioritisation and interpretation of PsA GWAS SNPs . . . . .	248
5.4	Discussion . . . . .	260
5.4.1	Characterising chromatin accessibility in PsA samples . . .	260
5.4.2	Bulk qPCR expression profiling and integration with paired chromatin accessibility data . . . . .	262
5.4.3	The relevance of monocytes in the PsA synovial fluid . . . .	265
5.4.4	Charcaterisation of synovial and peripheral blood CD14 <sup>+</sup> monocytes by scRNA-seq: subpopulations and additional relevant pathways . . . . .	266

## CONTENTS

---

5.4.5	Mass cytometry data supports changes in chromatin accessibility and gene expression of pro-inflammatory mediators in PsA CD14 <sup>+</sup> monocytes . . . . .	270
5.4.6	The use of PsA functional data to inform fine-mapping GWAS loci . . . . .	272
5.4.7	Limitations of the study . . . . .	274
5.5	Conclusions . . . . .	276
<b>6</b>	<b>General discussion</b>	<b>278</b>
6.1	The importance of cell type and context specificity in disease . . . . .	278
6.2	The epigenetic landscape in psoriasis and PsA clinical samples: systemic vs affected tissue . . . . .	280
6.3	Cell type and tissue specific transcriptional profiles in clinical cohorts . . . . .	281
6.4	Data integration in multi-omic studies . . . . .	284
6.5	GWAS fine-mapping and integration with patients-derived epigenetic data . . . . .	285
6.6	Clinical relevance and translation . . . . .	286
6.7	Future work . . . . .	287
6.8	Concluding remarks . . . . .	291
<b>Appendices</b>		<b>293</b>
<b>A</b>	<b>Tables</b>	<b>296</b>
A.1	Chapter 3 Tables . . . . .	296
A.2	Chapter 4 Tables . . . . .	296
A.3	Chapter 5 Tables . . . . .	300
<b>B</b>	<b>Additional figures</b>	<b>303</b>
B.1	Chapter 3 Figures . . . . .	303
B.2	Chapter 4 Figures . . . . .	305
B.3	Chapter 5 Figures . . . . .	309
<b>Bibliography</b>		<b>353</b>

# List of Figures

1.1	Environmental triggers and genetic predisposition leading to psoriasis and PsA (adapted from Nestle <i>et al.</i> , 2009). . . . .	13
1.2	The epigenetic landscape and its role in regulation of gene expression. . . . .	31
3.1	Measurements for quality control assessment in ATAC-seq samples	84
3.2	FRIP and peak calling at different sequencing depths in ATAC-seq libraries. . . . .	88
3.3	Peak calling filtering using IDR analysis in ATAC-seq samples. . .	89
3.4	Work flow illustrating the strategy to account for ATAC background noise prior to differential analysis. . . . .	92
3.5	Normalisation and differential chromatin accessibility analysis for different cut-offs using quantile normalisation limma voom and DESeq2. . . . .	94
3.6	Comparison of the DARs identified by differential analysis using limma voom or DESeq for the consensus peak list filtered at the optimal cut-off 80%. . . . .	95
3.7	Enrichment analysis for the significant DARs identified by DESeq2 between CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> cells. . . . .	96
3.8	Assessment of the effect of transposition times on the ATAC-seq measurements . . . . .	98
3.9	Differences in mitochondrial DNA abundance and TSS enrichment between ATAC-seq and Fast-ATAC protocols. . . . .	100
3.10	Quality control assessment of different ATAC protocols in keratinocytes from a psoriatic lesion and NHEKs. . . . .	102
3.11	Quality control assessment of Fast-ATAC and Omni-ATAC in cultured NHEK . . . . .	104
3.12	Comparison of 4 different ATAC protocols applied to psoriasis and/or healthy keratinocytes and NHEKs cells with published ENCODE DNase-seq data. . . . .	106
3.13	Experimental design to assess the impact of cryopreservation and fixation in the chromatin accessibility of immune primary cells. . .	108
3.14	Total number of ATAC-seq reads for the fresh, frozen and fixed CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> samples for 3 volunteers (CTL1-3). .	109
3.15	Fragment size density distribution for ATAC-seq fresh, fixed and frozen in CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> cells. . . . .	110
3.16	ATAC-seq enrichment of nucleosome-free and di-nucleosome fragments at the TSS and surroundings in CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> samples for the three conditions. . . . .	111
3.17	Genomic features annotation for the ATAC-seq peaks called in each of the fresh,frozen and fixed samples from CD14 <sup>+</sup> monocytes and total CD4 <sup>+</sup> . . . . .	112

## LIST OF FIGURES

---

3.18	PCA analysis based on the ATAC-seq chromatin accessibility landscape in fresh, fixed and frozen samples. . . . .	113
3.19	Comparison of the $\log_2$ normalised ATAC-seq counts at the consensus list of peaks in fresh, fixed and frozen conditions. . . . .	114
3.20	Differential chromatin accessibility at the <i>TNFSF14</i> gene between ATAC-seq fresh and ATAC-seq frozen in CD14 <sup>+</sup> monocytes. . . . .	116
4.1	Quality control evaluation of the H3K27ac ChIPm libraries in immune cells isolated from psoriasis and control samples. . . . .	139
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. . . . .	140
4.3	Differential H3K27ac modification at a putative intergenic enhancer region in circulating CD14 <sup>+</sup> monocytes between psoriasis patients and healthy controls. . . . .	142
4.4	Quality control assessment of the ATAC libraries generated from circulating immune cells in psoriasis and control samples. . . . .	143
4.5	Clustered heatmap and conserved TF binding sites enrichment analysis in the consensus list of called ATAC peaks in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells from the patients and controls cohort. . . . .	145
4.6	Epigenetic landscape at two ATAC differential accessible regions between patients and controls in CD8 <sup>+</sup> cells. . . . .	148
4.7	Epigenetic landscape at <i>DTD1</i> locus showing differential H3K27ac and chromatin accessibility between psoriasis patients and controls in CD8 <sup>+</sup> cells. . . . .	150
4.8	PCA illustrating the sample variability based on the gene expression profiles for all 72 samples. . . . .	151
4.9	Magnitude and significance of the gene expression changes between psoriasis patients and healthy controls in four immune cell types. . . . .	153
4.10	RNA-seq expression levels of the lncRNA <i>HOTAIRM1</i> and its experimentally validated target <i>UPF1</i> in psoriasis and healthy controls CD14 <sup>+</sup> monocytes. . . . .	155
4.11	Differential expression between psoriasis and healthy controls of <i>MAPK</i> and <i>DUSP</i> genes contributing to MAPK signalling enrichment in CD14 <sup>+</sup> monocytes and CD8 <sup>+</sup> T cells. . . . .	158
4.12	Mapping of the DEGs identified in CD8 <sup>+</sup> cells between psoriasis patients and healthy controls onto the TNF- $\alpha$ and the chemokine signalling pathways. . . . .	160
4.13	PCA for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. . . . .	161
4.14	Magnitude and significance of the gene expression changes between matched lesional and uninvolved epidermal biopsies from three psoriasis patients. . . . .	162

## LIST OF FIGURES

---

4.15 Overlap of the significantly differentially expressed genes between lesional and uninvolved epidermal sheets, split epidermis and whole skin biopsies. . . . .	164
4.16 Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the HIF-I signalling pathway. . . . .	167
4.17 Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the NOD-like signalling pathway. . . . .	168
4.18 Differential gene expression and chromatin accessibility landscape for <i>TIAM1</i> and <i>TRAT1</i> genes in CD8 <sup>+</sup> cells. . . . .	172
4.19 Epigenetic landscape at the location of the SNPs in the 4 SNPs in the 90% credible set for the <i>SLC45A1/TNFRSF9</i> psoriasis GWAS locus. . . . .	179
4.20 Epigenetic landscape at the location of the 95% credible set of AS fine-mapped SNPs for the chr2p15. . . . .	181
4.21 rs4672505 genotype and chromatin accessibility at chr2:62,559,749-62,561,442 in CD8 <sup>+</sup> cells. . . . .	182
4.22 Potential role of rs4672505 in regulating <i>B3GNT2</i> gene expression. . . . .	183
5.1 Comparative percentages of peripheral blood and synovial fluid immune cellular composition from the PsA cohort. . . . .	210
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. . . . .	212
5.3 PCA based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and synovial fluid. . . . .	213
5.4 Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states. . . . .	215
5.5 Enrichment of PsA DARs for the FANTOM5 eRNA dataset. . . . .	217
5.6 Differentially accessible regions located within gene bodies in CD14 <sup>+</sup> monocytes and NK cells from PsA patients. . . . .	218
5.7 Functional pathways significantly enriched and unique for either synovial fluid open or peripheral blood open DARs in CD14 <sup>+</sup> monocytes, CD4m <sup>+</sup> ,CD8m <sup>+</sup> and NK cells. . . . .	221
5.8 Heatmap illustrating fold changes in gene expression between synovial fluid and peripheral blood across CD14 <sup>+</sup> monocytes and mCD4 <sup>+</sup> and mCD8 <sup>+</sup> T cells. . . . .	223
5.9 Gene expression changes in immune-relevant genes between synovial fluid and peripheral blood in CD14 <sup>+</sup> monocytes, mCD4 <sup>+</sup> and mCD8 <sup>+</sup> cells. . . . .	224
5.10 Comparison of immune-relevant gene expression modulation across PsA tissues (synovial fluid vs peripheral blood) and in PsA patients verus healthy controls. . . . .	226
5.11 Chromatin accessibility landscape at the qPCR differentially expressed <i>FN1</i> gene in CD14 <sup>+</sup> monocytes. . . . .	229
5.12 Visualisation using UMAP dimensional reduction of the synovial fluid and peripheral blood CD14 <sup>+</sup> monocytes. . . . .	234

## LIST OF FIGURES

---

5.13 Identification and characterisation of CD14 <sup>+</sup> monocytes subpopulations in synovial fluid and peripheral blood. . . . .	235
5.14 Heatmap and hierarchical clustering for the top eight marker genes of the CD14 <sup>+</sup> monocyte subpopulations identified in synovial fluid and peripheral blood from three PsA patients. . . . .	237
5.15 Functional characterisation of the MC-1 and MC-4 CD14 <sup>+</sup> PsA monocyte subpopulations. . . . .	240
5.16 Biaxial representation of CD14 <sup>+</sup> monocytes expressing TNF- $\alpha$ , MCP-1, osteopontin and IP-10 in synovial fluid and peripheral blood before and after protein transport blockade with BFA using mass cytometry. . . . .	244
5.17 Differences in the percentage of CD14 <sup>+</sup> monocytes producing TNF- $\alpha$ , osteopontin, MCP-1 and IP-10 in synovial fluid and peripheral blood. . . . .	246
5.18 Chromatin landscape in CD14 <sup>+</sup> monocytes in the proximity to <i>CCL2</i> and <i>CXCL10</i> . . . . .	247
5.19 Enrichment of PsA GWAS fine-mapped SNPs for DARs and ATAC peaks in CD14 <sup>+</sup> monocytes, mCD4 <sup>+</sup> , mCD8 <sup>+</sup> and NK cells. . . . .	252
5.20 Association plot for the Bayesian fine-mapping results of the 5q31 ( <i>CSF2/P4HA2</i> ) and <i>RUNX3/SYF3</i> PsA GWAS signal. . . . .	254
5.21 Epigenetic landscape at the genomic location of the fine-mapped SNP rs11249213 at the <i>RUNX3</i> PsA GWAS signal. . . . .	256
5.22 Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal. . . . .	259
B.1 Distribution of the background read counts from all the master list peaks absent in each sample. . . . .	303
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. . . . .	303
B.3 Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples. . . . .	304
B.4 Fast-ATAC and Omni-ATAC NHEK pre-sequencing profiles of relative abundance of DNA library fragment sizes. . . . .	304
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. . . . .	305
B.6 Percentage of MT reads and called peaks after IDR p-value filtering in the ATAClibraries generated in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> isolated from psoriasis patients and healthy controls. . . . .	305
B.7 PCA analysis illustrating batch effect in ATAC and RNA-seq samples. . . . .	306
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. . . . .	306
B.9 Mapping quality control for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. . . . .	307

## LIST OF FIGURES

---

B.10 Distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles for all 72 samples. . . . .	307
B.11 Correlation in gene expression between the miRs <i>MIR31HG</i> and the target gene <i>HIF-1A</i> in lesional and uninvolved skin. . . . .	308
B.12 Assessment of the percentage of mitochondrial 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. . . . .	309
B.13 Enrichment of eQTL SNPs for the consensus list of ATAC peaks used to perform differential chromatin accessibility analysis in each cell type. . . . .	309
B.14 Permutation analysis SF vs PB in CD14 <sup>+</sup> monocytes,CD4m <sup>+</sup> ,CD8m <sup>+</sup> and NK. . . . .	310
B.15 Enrichment of DARs in the 15 chromatin states of the cell type specific Roadmap Epigenomic Project segmentation maps. . . . .	311
B.16 Identification of the CD14 <sup>+</sup> monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. . . . .	311
B.17 Expression levels of antigen presentation and phagocytosis characteristic genes in synovial fluid and peripheral blood CD14 <sup>+</sup> monocytes. . . . .	312
B.18 Differential gene expression between synovial fluid and peripheral blood PsA CD14 <sup>+</sup> monocytes and correlation with qPCR and chromatin accessibility . . . . .	313
B.19 Enrichment of PsA GWAS fine-mapped SNPs for publicly available eQTL datasets. . . . .	314

# 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 . . . . .	20
2.1	Summary table of samples/cohorts recruited in this thesis for generation of various datasets. . . . .	49
2.2	Antibody panel used for FACS separation of primary cell populations in Chapter 3 controls and Chapter 5 PsA samples. . .	53
2.3	Molecules targeted by the mass cytometry antibody panel for peripheral blood and synovial fluid samples. . . . .	62
3.1	Summary table of ATAC analysis methodology for peak calling, filtering and differential analysis. . . . .	78
3.2	ATAC-seq percentage of mitochondrial reads and fraction of reads in called peaks (FRiP). . . . .	86
3.3	Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies. . .	103
3.4	Summary results from the differential chromatin accessibility analysis comparing ATAC-seq frozen or fixed chromatin landscape to the reference ATAC-seq fresh. . . . .	115
4.1	Summary table of the most comprehensive transcriptional studies in psoriasis skin and blood. . . . .	128
4.2	Description and metadata of the psoriasis patients cohort. . . . .	135
4.3	Description of the healthy control cohort. . . . .	137
4.4	Summary results from the differential H3K27ac analysis between psoriasis patients and healthy controls in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells. . . . .	141
4.5	Summary results from the differential chromatin accessibility analysis between psoriasis patients and healthy controls in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells. . . . .	147
4.6	Summary results from the differential gene expression analysis between psoriasis patients and healthy controls in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells. . . . .	152
4.7	Overlap between putative psoriasis GWAS genes and the reported significantly DEGs in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells. . . . .	154
4.8	Pathways enriched for DEGs between psoriasis patients and healthy controls in CD14 <sup>+</sup> monocytes and CD8 <sup>+</sup> cells. . . . .	156
4.9	Summary results of the DGE analysis between uninvolved and lesional psoriatic epidermal biopsies. . . . .	163

---

## LIST OF TABLES

---

4.10	Most relevant pathways enriched for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	166
4.11	Overlap between the DEGs in the four circulating immune cell types (psoriasis patients versus controls) and the DEGs in psoriasis patients skin biopsies (lesional versus uninvolved).	170
4.12	Summary results for the fine-mapping analysis of 26 psoriasis Immunochip GWAS loci.	174
4.13	SNPs from the 90% credible set of the successfully fine-mapped psoriasis loci overlapping ATAC accessible chromatin in four cell types.	177
5.1	Description and metadata of the PsA patients cohort.	208
5.2	Datasets generated for each sample in the PsA cohort.	209
5.3	Summary results of the differential chromatin accessibility analysis between synovial fluid and peripheral blood in PsA samples.	214
5.4	Characterisation of the DARs located within genes in each of the four cell types from PsA samples.	217
5.5	Immune genes with significant modulated expression in synovial fluid and proximal to a DAR in ATAC.	228
5.6	Pathway enrichment analysis for the modulated genes between synovial fluid and peripheral blood in CD14 <sup>+</sup> and mCD4 <sup>+</sup> .	231
5.7	Summary table of the PsA GWAS loci presenting log <sub>10</sub> ABF≥3 for the fine-mapping lead SNP.	249
5.8	PsA fine-mapped SNPs from the 90% credible sets overlapping accessible chromatin identified by ATAC in four cell types.	253
5.9	Publicly available <i>cis</i> -eQTL datasets reporting an effect for the PsA 5q31 GWAS locus fine-mapped SNPs (90% credible set) overlapping ATAC accessible regions.	257
A.1	Enrichment of ATAC-seq reads across the TSS for the CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> samples fresh, frozen and fixed.	296
A.2	Size of the master lists generated by DiffBind for the H3K27ac differential analysis between psoriasis patients and healthy controls in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells.	297
A.3	Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChiPm assay.	297
A.4	Functional interactions and overlap with another study for the differentially expressed lncRNAs in each cell type.	298
A.5	Additional enriched pathways for DEGs between psoriasis and healthy controls in CD14 <sup>+</sup> monocytes and CD8 <sup>+</sup> cells.	298
A.6	Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	299
A.7	Enriched pathways for the scRNA-seq DEGs between synovial fluid PsA CD14 <sup>+</sup> monocytes.	300
A.8	PsA GWAS Immunochip loci presenting log <sub>10</sub> ABF<3 for the fine-mapping lead SNP in the association analysis.	301

---

# Abbreviations

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

## LIST OF TABLES

---

<b>NSAID</b>	nonsteroidal antiinflammatory drug
<b>Omni-ATAC</b>	Omni- assay for transposase-accessible chromatin
<b>OR</b>	Odds ratio
<b>PBC</b>	PCR bottle-necking coefficient
<b>PBMC</b>	Peripheral blood mononuclear cells
<b>PCA</b>	Principal component analysis
<b>PCA</b>	Principal component analysis
<b>PI</b>	Protein inhibitor
<b>PP</b>	Posterior probability
<b>PRC</b>	Polycomb repressor complex
<b>PsA</b>	Psoriatic arthritis
<b>PTM</b>	Post-translational modification
<b>qPCR</b>	quantitative polymerase chain reaction
<b>RA</b>	Rehumatoid arthritis
<b>RNA-seq</b>	RNA sequencing
<b>ROS</b>	Reactive oxygen species
<b>SDS</b>	Sodium dodecyl sulfate
<b>SF</b>	synovial fluid
<b>SLE</b>	Systemic lupus erythematosus
<b>SpA</b>	Spondyloarthritis
<b>T1D</b>	Type 1 diabetes
<b>T2D</b>	Type 2 diabetes
<b>TAD</b>	Topological associating domain
<b>Th</b>	T-helper
<b>TLR</b>	Toll-like receptor
<b>TNF</b>	Tumour necrosis factor
<b>UCSC</b>	University of California Santa Cruz
<b>WGS</b>	Whole genome sequencing
<b>WHG</b>	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 as two distinct 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,

## **Introduction**

---

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). On the other hand, cases of PsA in the general population varies between 0.04 and 1.2% occurring in 10 to 30% of psoriasis patients, evidencing the strong association between the two diseases (Gelfand et al. 2005; Reich et al. 2015; Perera et al. 2012). Overall, data suggest a steady increase in both psoriasis and PsA prevalence over the last 30 years (Springate et al. 2017; Organization 2016).

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

## **Introduction**

---

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 (Sampogna et al. 2012). Overall, psoriasis and PsA represent a significant burden for the economy due to treatment costs and associated morbidity. Treatment and management-associated costs per psoriasis patient per year in 2015 in the UK accounted for £4,000 to £14,000, before and after requirements of biological therapy, respectively, and were further increased in PsA (Burgos-Pol and Martnez-Sesmero 2016; Poole et al. 2010).

### **1.1.2 Psoriasis and inflammatory dermatoses**

The skin is the biggest organ in the human body constituting 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, 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

## **Introduction**

---

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/sacroiliac or peripheral) and the presence of extra-articular features (Rudwaleit et al. 2009).

Major histocompatibility (MHC) class I molecules present intracellular peptides (self or from infectious agents) to T cells, encoded by the human leukocyte antigen (HLA) A, B and C genes. HLA-B27 is the strongest genetic association for the SpA family. Studies in human families and rat models with HLA-B27 positive status have shown manifestation of different SpA, 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 (Baeten et al. 2013). Interestingly, the axial and peripheral classification of SpA may be supported by true immunopathological differences between the two (Porcher et al. 2005; Appel et al. 2011; Vandooren et al. 2004). Nevertheless, the knowledge of cellular and molecular processes contributing to pathophysiology is still limited and further research will impact on the classification and understanding of SpA family.

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), a distinctive feature of PsA (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 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 in 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 (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 and scalp

## Introduction

---

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 (CASPAN) 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

## Introduction

---

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
Patients global health assessment	Evaluation of the patient's health by the patient (scale 0 to 5)
Physician global health assessment	Evaluation 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 (Figure 1.1). The origins of both pathologies, as well as the connection between skin and joint inflammation, still remain controversial.

## **Introduction**

---

### **Environmental factors and disease**

A variety of exposures are proposed as risk factors for the development and worsening of psoriasis and PsA. A wide range of drugs including anti-depressants, anti-hypertensives and anti-cytokine therapies 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 human immunodeficiency 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 international 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).

### **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

## **Introduction**

---

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 remodeling lead to osteolysis, bone resorption and erosion at the affected joints (Mensah et al. 2008). Bone erosion is also the main histopathological process driving dactylitis, where bone lysis resolves in shortening of the digits (Gladman et al. 2005). 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 bony spurs formation along the insertion sites, similar to RA, causing structural debilitation of the joints (Benjamin and McGonagle 2009; Finzel et al. 2014).

## **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)- $\alpha$  and  $\gamma$  are

## **Introduction**

---

innate immune cytokines involved in disease initiation and are produced mainly by circulating plasmacytoid dendritic cells (pDCs) and myeloid DC (mDCs), respectively, as well as by T cells in lesional skin (Leanne2009; Perera et al. 2012; Hijnen et al. 2013). Increased mRNA levels for both IFNs have been detected in skin plaques and shown to contribute to lymphocyte recruitment and maintenance of DC activation (Schmid et al. 1994). Tumour necrosis factor (TNF)- $\alpha$  is another key cytokine involved in the dysregulated innate immune response observed in psoriasis and PsA. TNF- $\alpha$  is produced by activated keratinocytes, mast cells, natural killer (NK) cells and also adaptive immune cell types, including the T helper (Th)/CD4 $^{+}$  activated cell subsets Th-1 and Th-17, that infiltrate skin lesions and inflamed joints (Perera et al. 2012; Lizzul et al. 2005). TNF- $\alpha$  causes activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- $\kappa$ B), 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- $\alpha$  has a prominent role in bone turnover and bone remodeling, 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 the mDCs and macrophages in inflamed skin and to a lesser degree by psoriatic keratinocytes (Lee et al. 2004; Li et al. 2018). IL-23 binds to the IL-23 receptor (IL23R), highly expressed by the lesion-resident DCs and T cells and also by circulating CD4 $^{+}$  (Tonel et al. 2010). In psoriasis, IL-23 mediates the pathogenic loop between activated keratinocytes and T cells, where activation of the IL-23 pathway leads to Th-17 cell differentiation and increased IL-17 cytokine levels as a result of NF- $\kappa$ B activation (McGeachy et al. 2009). IL-17 signalling maintains the Th-17 immune

## **Introduction**

---

mediated response through recruitment and activation of neutrophils, induction of pro-inflammatory cytokines, including interleukin-1 $\beta$  (IL-1 $\beta$ ) 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).

### **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 T cytotoxic(Tc)/CD8 $^{+}$  cells (Black and ArdernJones 2007). Upon damage by environmental triggers keratinocytes release cationic AMP 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-1, IL-6 and TNF- $\alpha$ , 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

## **Introduction**

---

keratinocyte-specific genes from the late cornified envelope (LCE) family (Tsoi et al. 2012).

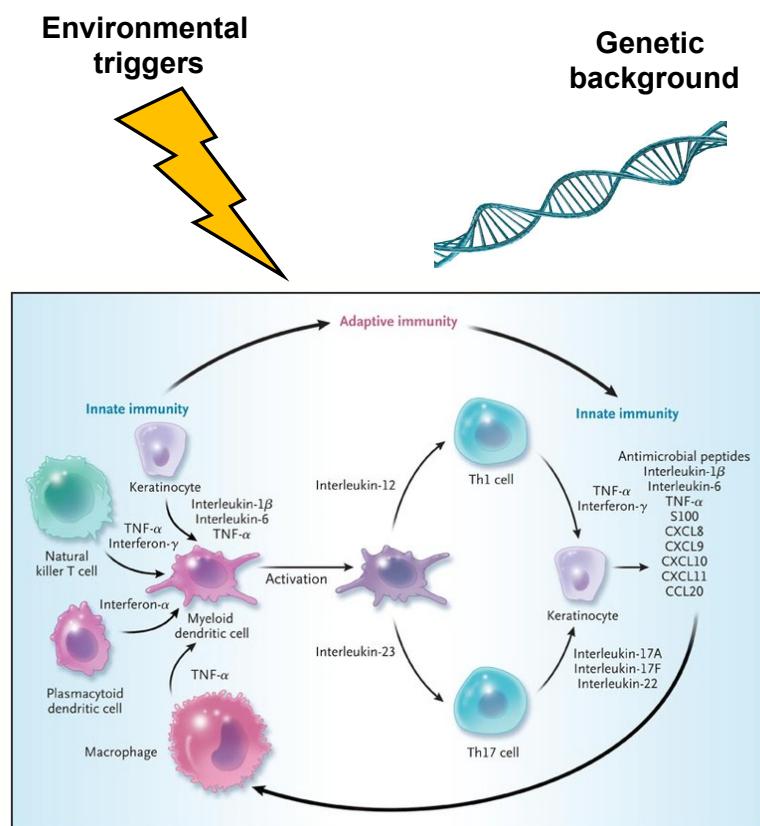
**T cells.** T lymphocytes have been considered the most relevant cell types in the initiation and maintenance of psoriasis and PsA. Skin-resident memory T cells have been demonstrated to have a key role in psoriatic lesion development in mice models (Boyman et al. 2004). In human case reports, bone 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<sup>+</sup> and not CD8<sup>+</sup> cells (Wrone-Smith and Nickoloff 1996). Nevertheless, CD8<sup>+</sup> 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-1 $\beta$  and IL-6, induce activation and differentiation of naïve CD4<sup>+</sup> and CD8<sup>+</sup> into effector pathogenic Th-17 and Tc17 cells producing IL-17 (Weaver et al. 2007). IL-17<sup>+</sup> CD8<sup>+</sup> cells have been found in 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). Likewise, Th-17 infiltrated cells have been found in the epidermis of psoriatic lesions (Lowes et al. 2008; Pène et al. 2008).

**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 on activation through Toll-like receptor (TLR)7/9 by the keratinocytes self-DNA and LL-37 complex infiltrate into the lesional and uninvolved dermis of psoriasis patients (Nestle et

## Introduction

---

al. 2005; Lande et al. 2007). In contrast, quiescent mDCs are epidermal resident cells that undergo maturation in presence of the IFN- $\alpha$  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 production (Lee et al. 2004).



**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.

**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 $\alpha$  production (Perera et al. 2012; Mahil et al. 2016). Similarly, mouse models for chronic psoriasiform skin inflammation have demonstrated macrophage migration into affected skin and how production of TNF- $\alpha$  contributes to maintenance of skin lesions (Stratis et al. 2006; Wang et

## **Introduction**

---

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 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 metalloproteinases responsible for bone erosion through differentiation into osteoclasts have been found to be similar to those found 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 driven by high content of perforin and granzymes (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- $\gamma$ , promoting Th-1 expansion and the adaptive immune response (Martin-Fontecha et al. 2004; Ferlazzo et al. 2004). In psoriasis, significant increase of cells expressing NK markers have been found in lesional compared to uninvolved skin (Cameron and Kirby 2003; Ottaviani et al. 2006). Expansion of NK  $CD3^-$   $CD56^{bright}$  cells in inflamed joints was observed in a cohort including RA, PsA and AS patients (Dalbeth and Callan 2002). Moreover, NK cells in RA have shown to trigger osteoclastogenesis and bone destruction *in vitro* and in mouse models (Soderstrom et al. 2010) . Amongst the target cell receptors regulating NK cells function, the killer immunoglobulin-like receptor (KIR) family includes activating and inhibitory members. The inhibitory receptor KIR2DL1 and the activatory receptor KIR2DS1 recognise HLA-Cw\*06:02, strongly associated

## **Introduction**

---

with psoriasis and PsA (Tobin et al. 2011). Interestingly, gene based studies have shown genetic variability in *KIR2DS1*, associated with psoriasis and PsA susceptibility, and also AS and RA (uszczek 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<sup>+</sup> 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.** The role of B cells in the pathophysiology of psoriasis and PsA remains unclear. B cells are mainly involved in the humoral adaptive immune response through antibody production. However, they also act as APCs, regulate CD4<sup>+</sup> 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). Recent studies on the imiquimod-induced psoriasis mice model have demonstrated a more severe inflammation in CD19<sup>-/-</sup> knock-out mice lacking of 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).

### **1.2.4 Therapeutic intervention**

Psoriasis and PsA are currently incurable diseases, with available treatments focused on alleviating symptoms. For instance, topical therapies

## **Introduction**

---

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 Coroller 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; M. 2017; Polachek et al. 2017). Biologic systemic agents represent the most specific treatment option for severe psoriasis and PsA notably TNF- $\alpha$  inhibitors (TNFi). Three TNFi have been approved for the treatment of psoriasis: etanercept, infliximab and adalimumab (**Ahil2016**). 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 (Branisteanu et al. 2015). 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. The concordance of psoriasis is greater in monozygotic (33-55%) compared to dizygotic twins (13-21%), giving a heritability estimate of 80%, while no difference in concordance is reported for PsA, probably due to lack of statistical power and appropriate diagnosis (Farber et al. 1974; Duffy et al. 1993; Pedersen et al. 2008). In the general population, approximately 40% of patients with psoriasis or PsA have a family history in first degree relatives (Gladman et al. 1986). Interestingly, the recurrence rate in first-degree relatives has been shown to be greater in PsA (40%) compared to psoriasis (8%) in a study in the Icelandic population (Chandran et al. 2009). Altogether, this may suggest differences in the heritability between the two phenotypes and a stronger genetic contribution in PsA.

#### 1.3.2 Genome-wide association studies

Linkage analysis of psoriasis and PsA in family pedigrees presenting an autosomal dominant condition yielded nine psoriasis susceptibility loci (PSORS1-9) with PSORS1 showing the strongest genetic association (Consortium2003; Capon 2017). Nevertheless, the inability of independent studies to reproduce these results for regions other than PSOR1, 2 and 4, highlighted the limitations of linkage studies 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

## **Introduction**

---

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 solely used 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 and PsA GWAS were published in 2007, with a total of 63 independent genetic associations identified at genome-wide significance ( $p\text{-value}\leq 5\times 10^{-8}$ ) to date (Table 1.3), explaining 28% of the 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 (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). HLA-C has been consistently identified as the most significant locus with the greatest effect size. Additional MHC-I and MHC-II associations with disease risk have been identified for HLA-A, HLA-B and HLA-DQA1 through step-wise conditional analysis(Okada et al. 2014).

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, also included meta-analysis with the largest available psoriasis cohorts at the time (Tsoi et al. 2012). This meta-analysis has since been further expanded, yielding 16

## **Introduction**

---

additional associations and reinforcing the importance of NF $\kappa$ B and cytotoxicity pathways in disease pathophysiology (Tsoi et al. 2015; Tsoi et al. 2017). Meta-analysis of GWAS across Caucasian and Chinese populations revealed 4 new associations, as well as population-specific effect or allelic heterogeneity in 11 loci, including MHC-I genes, demonstrating the value of trans-ethnic approach, to further understand the heterogeneous genetic susceptibility to psoriasis in different populations (Yin et al. 2015).

Conducting independent GWAS for psoriasis and PsA has shown differences in HLA-C and HLA-B allele frequencies. Interestingly higher association with HLA-B has been found in PsA individuals compared to psoriasis patients without joint inflammation (Winchester et al. 2012; Okada et al. 2014). GWAS associations for previously identified psoriasis loci such as *IFNLR1*, *IFIH1* and *NFKBIA* were also found as genome-wide significant when using PsA cases, and PsA-specific independent signals for *IL23R* and *TNFAIP3* demonstrated stronger associations compared to psoriasis patients only with cutaneous affection (Ellinghaus et al. 2010; Stuart et al. 2015).

Furthermore, PsA GWAS using Immunochip has also revealed a specific association in chromosome 5q31 not reported previously (Bowes et al. 2015).

Overall, GWAS have demonstrated shared and distinct genetic architectures for psoriasis and PsA. It is important to take into account that these results affected by imprecise phenotyping of cases, one of the many challenges in the systematic comparison between the two diseases.

## Introduction

---

**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 Ellinghaus *et al.*, 2010 and the Stuart *et al.*, 2015 studies included stratified association analysis of psoriasis 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 <i>IL-12B</i>
(Nair et al. 2009)	Eur		1,409/1,436	Psoriasis, PsA	<i>IL-23A</i> , <i>IL-23R</i> , <i>IL-12B</i> , <i>TNIP1</i> , <i>TNFIP3</i> , <i>IL-4</i> and <i>IL-13</i>
(Stuart et al. 2010) (Ellinghaus et al. 2010)	WA, Eur German		1,831/2,546 472/1,146	Psoriasis, PsA Psoriasis	<i>NOS2</i> , <i>FBXL19</i> , <i>PSMA6</i> - <i>NFKBIA</i> <i>TRAF3IP2</i>
(Strange et al. 2010)	Eur		2,622/5,667	Psoriasis, PsA	<i>LCE3D</i> (PSOR2), <i>IL28RA</i> , <i>REL</i> , <i>IFIH1</i> , <i>ERAP1</i> , <i>TYK2</i> and <i>HLA-C/ERAP1</i>
(Zhang et al. 2008)	Chinese		1,139/1,132	Psoriasis	(type epistasis epistasis epistasis)
(Sun et al. 2010)	Chinese		8,312/12,919	Psoriasis, PsA I)	<i>PTTG1</i> , <i>CSMD1</i> , <i>GJB2</i> ,
(Tsoi et al. 2012)*	WA, Eur		10,588/22,806	Psoriasis, PsA	<i>SERPINB8</i> , <i>ZNF816A</i> <i>CARD14</i> (PSOR4), <i>RUNX3</i> , <i>B3GNT2</i> , <i>ELMO1</i> , <i>STAT3</i>

## Introduction

---

(Tsoi et al. 2015)*	WA, Eur	15,000/27,000	Psoriasis, PsA	1q31.1, 5p13.1, <i>PLCL2</i> , <i>NFKBIZ</i> ,
(Bowes et al. 2015)	British, WA and Eur	Irish, 1,962/8,923	PsA	<i>CAMK2G</i> 5q31 PsA-specific
(Stuart et al. 2015)	Australians	1,430/1,417	Psoriasis, PsA	PsA-specific secondary signals (main text), new psoriasis-specific ( <i>CDKAL1</i> and <i>CAMK2G</i> ), stronger psoriasis <i>LCE</i> association
(Yin et al. 2015)	WA, Eur, Asian	15,369/19,517	Psoriasis, PsA	<i>LOC144817</i> , <i>COG6</i> , <i>RUNX1</i> and <i>TP63</i> ; signals with ethnic heterogeneity
(Tsoi et al. 2017)*	WA, Eur	19,032/39,498	Psoriasis, PsA	<i>CHUK</i> , <i>IKBKE</i> , <i>FASLG</i> , <i>KLRK1</i> , <i>PTEN</i>

## **Introduction**

---

### **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 (Nica2014; Fairfax et al. 2012; Fairfax et al. 2012; Naranbhai et al. 2015). 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

## **Introduction**

---

SNPs located 14kb upstream (Small et al. 2011). Moreover, this haplotype also showed association with genes in *trans*, highlighting 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 pathophysiological relevant cell types and molecular pathways associated with disease. 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 control (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 candidate genes are selected by proximity to the associated variants and fail to include 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

## **Introduction**

---

opposite directions, with AS, Crohns 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- IL-23 and anti- 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 Graves 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 (Costantino2015; Hanson et al. 2018).

## **Skin barrier**

**GWAS have highlighted keratinocyte-specific genes such as *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 association with disease**

## **Introduction**

---

is driven by a deletion in two of *LCE3B* and *LCE3C* (*LCE3C\_LCE3B\_del*) (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- $\kappa$ B as well as overexpression of psoriasis pathophysiological relevant genes, including *IL-6* and *TNFA*, in keratinocyte cell lines (Jordan et al. 2012).

### **NF- $\kappa$ B and TNF pathways**

The NF- $\kappa$ B pathway is involved in the regulation of the innate and adaptive immune response and dysregulation of NF- $\kappa$ B contributes to the development of many chronic inflammatory diseases (Liu et al. 2017). In fact, elevated levels of NF- $\kappa$ B 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- $\kappa$ B and TNF signalling pathways including *TNIP1*, *TNFAIP3*, *NFKBIA*, *REL*, *TRAF3IP2*, *CHUK*, *IKBKE* and *FASLG* (Nair et al. 2008; Ellinghaus et al. 2010; Hffmeier 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- $\kappa$ B activation and the IL-17/IL-23 axis (Hffmeier 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- $\kappa$ B and MAP kinases activation in endothelial cells (Dand et al. 2017; Wang et al. 2014).

## **Introduction**

---

### **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 diseases 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 psoriasis 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). Interestingly, an *IL-23* signal secondary to the one reported by Tsoi *et al.*, 2012 has been specifically associated with PsA (Bowes et al. 2015). Regarding the genetics of the Th-17 pathway, its relevance is

## **Introduction**

---

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 at 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 the chr1p36.23, chr2p15 and chr9q31.2 in psoriasis and PsA. 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 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

## **Introduction**

---

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 compared to 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). 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 putative causal variants for each GWAS association is required previous to functional validation of their putative pathogenic effect through molecular and cellular assays and *in vivo* models. Statistical fine-mapping partially overcomes some of the GWAS limitations by further refining 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 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.

### **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

## **Introduction**

---

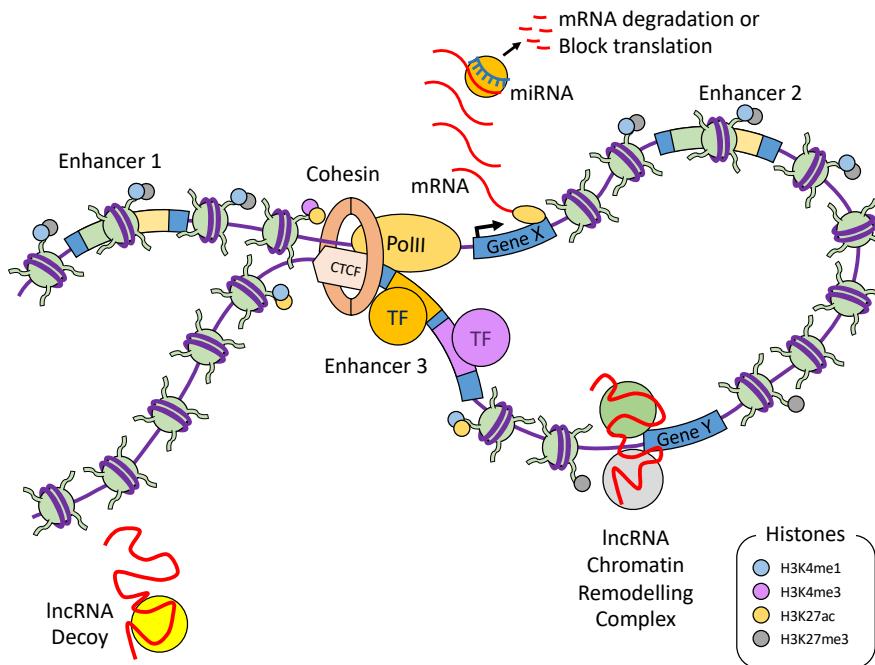
proteins which serve as scaffold for the DNA compaction into chromatin, as well as changes in DNA methylation and regulatory functions of non-coding RNAs (Figure 1.2). Environmental and intrinsic factors can trigger changes in the epigenome that result in dysregulation of gene expression and function.

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 et al. 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 demonstrated in QTL studies, where 50 to 90% the genetic variants rare 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; Naranhai 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).

## Introduction

---



**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.

The epigenetic landscape and its role in regulation of gene expression. Changes to chromatin conformation brings promoters into proximity with distal regulatory elements allowing the recruited TF complexes to regulate gene expression. 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 by altering mRNA stability. Similarly, miRNA can regulate the levels of mRNA by either preventing translation or transcript degradation.

## **Introduction**

---

### **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 octameric 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 nucleosome as well as their interaction across nucleosomes in the vicinity (Polach et al. 2000; Pepenella et al. 2014). Additionally, chromatin structure can also be influenced by adenosine 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-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 sites sequencing (DNase-seq), which uses the non-specific double strand endonuclease DNase I to preferentially cut in nucleosome-free regions known as DNase hypersensitive

## Introduction

---

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

## **Introduction**

---

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<sub>2</sub> 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

## **Introduction**

---

et al. 2009; Creyghton and the 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 Nature 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 iuses the

## **Introduction**

---

Tn5 transposase, accelerating library preparation and increasing the sensitivity of the results.

### **DNA methylation**

DNA methylation involves the transferal 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), psoriasis and PsA (Lei et al. 2009; Liu et al. 2013; Zhang et al. 2010). For example, regulation of TNF- $\alpha$  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

## **Introduction**

---

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; Batliwala 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

## **Introduction**

---

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' polyadenylation (Derrien et al. 2012). lncRNAs can positively and negatively regulate transcription through different mechanisms including guidance of chromatin modifiers such as DMTs 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 psoriasis (Müller et al. 2014; Shi et al. 2014; Li et al. 2014; Ahn et al. 2016).

## **Introduction**

---

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

## **Introduction**

---

to 88% of the genes with an associated *cis*-eQTL (Yan et al. 2002; Pickrell et al. 2010).

Additionally, 5' end RNA-sequencing 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 MS results in a soluble isoform of the *TNFRSF1A* protein with TNF antagonistic function (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. For instance, in thyroid autoimmunity, the risk allele of an intronic SNP in the thyroid stimulating hormone receptor (*TSHR*) gene reduces *TSHR* protein expression in IFN- $\alpha$  stimulated thyroid cells (Stefan et al. 2014). Alterations in TF binding sites can also affect looping and long-range chromatin interactions between enhancers and promoters. For instance, in prostate cancer this phenomenon causes up-regulated expression of the oncogene *SOX9* due

## **Introduction**

---

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  $\alpha$ -thalassaemia, where this phenomenon leads to dysregulated downstream activation of all  $\alpha$ -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 PPAR $\gamma$  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

## **Introduction**

---

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 (**Wakefield2007**). 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**

## **Introduction**

---

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).

## **Introduction**

---

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 functional landscape **or** **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 protocol(s) appropriate for cells, tissues and clinical samples of interest, with optimisation of the required methodologies and analytical tools including** (Chapter 3). Appropriate establishment of the methodological and analytical tools to study chromatin accessibility in

## **Introduction**

---

clinical samples in the group was required at the start of the project. This work will investigate different aspects of ATAC protocols, with 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 epigenetic landscape and gene expression in psoriasis patients.** Identifying differentially accessible or histone modified regions together with changes in gene expression 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,

## **Introduction**

---

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 included 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 psoriasis (fulfilling the Psoriasis Area and Severity Index (PASI) classification, Table 1.1) and in a flare (active disease state). Recruited patients were required

## **Material and Methods**

---

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 according to the PsA Response Criteria (PsARC) (Table 1.2) and a physician global assessment questionnaire, with concomitant psoriasis and in a flare. Patients had to present an oligoarticular phenotype, 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, preferably British or European, without family history of psoriasis, PsA, RA or SpA and not having suffered from an infection in the two weeks prior to

## **Material and Methods**

---

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.

<b>Chapter</b>	<b>Samples/cohorts</b>	<b>Description</b>
Chapter 3	Healthy controls	CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> cells from peripheral blood (fresh, frozen and fixed)
	Skin biopsies	Healthy and psoriatic skin
Chapter 4	Cohort 1A and 1B	Psoriasis and healthy control cells isolated from peripheral blood
Chapter 5	PsA samples	Cells isolated from synovial fluid and peripheral blood

**Table 2.1: Summary table of samples/cohorts recruited in this thesis for generation of various datasets.** For each of the chapters, the samples or name of the cohorts (when applicable) recruited and used to generate different types of data together with a short description are included. Detailed explanation is included in each of the chapters where the results for the processed samples are presented.

## **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 Hanks 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

## **Material and Methods**

---

(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-Owsiaak and colleagues (GutowskaOwsiaak and Schaupp 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 KGM-2 BulletKit (Lonza) supplemented with 0.06mM Ca<sup>2+</sup> and cultured in a collagen IV coated 96-well plate at 37°C 5% CO<sub>2</sub> 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**

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<sup>6</sup> PBMCs was carried out using a modified version of the protocol described by Kent and colleagues(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,

## **Material and Methods**

---

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 and rested for 30 min at 37°C, 5% CO<sub>2</sub> 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  $5 \times 10^6$  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 37C, 5% CO<sub>2</sub> 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 were separated using magnetic-activated cell sorting (MACS, Miltenyi) following the manufacturer's instructions. Consecutive positive selection was performed using Miltenyi beads for CD19<sup>+</sup>, CD8<sup>+</sup>, CD14<sup>+</sup>

## **Material and Methods**

---

monocytes and CD4<sup>+</sup> 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.

### **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<sup>6</sup> 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, cells were resuspended in FACS buffer prior to separation.

From the healthy control samples of Chapter 3 (Table 2.1), CD14<sup>+</sup> monocytes and CD3<sup>+</sup> CD14<sup>-</sup> CD4<sup>+</sup> 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<sup>+</sup> B cells, memory T cells (CD3<sup>+</sup> CD14<sup>-</sup> CD4<sup>+</sup> CD45RA<sup>-</sup> and CD3<sup>+</sup> CD14<sup>-</sup> CD8<sup>+</sup> CD45RA<sup>-</sup>), CD14<sup>+</sup> monocytes and CD56<sup>+</sup> 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.

## Material and Methods

---

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 for assessment of chromatin accessibility in different primary cells, including CD14<sup>+</sup> monocytes, CD4<sup>+</sup> and CD8<sup>+</sup> T cells, CD19<sup>+</sup> B cells and CD56<sup>+</sup> 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.

## **Material and Methods**

---

### **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

## **Material and Methods**

---

hematopoietic cells, the protocol combined cell lysis and transposition into a single step. In this thesis, Fast-ATAC was performed in skin biopsies (Chapter ?? 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 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 fragmentation 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 30

## **Material and Methods**

---

million paired-end reads. 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 ?? 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 Schmidl and colleagues for immunoprecipitation, tagmentation of bed-bound chromatin, de-crosslinking and DNA amplification (Schmidl et al. 2015). Immunoprecipitation was carried out with 1 $\mu$ g of the Diagenode antibody (C15410196). For each

## **Material and Methods**

---

sample, an aliquot of chromatin was processed in parallel without incubation with the anti-H3K27ac antibody (control input). 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  $3 \times 10^6$  cells were resuspended in 350 $\mu$ L of RNAProtect (Qiagen) 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,

## **Material and Methods**

---

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 50 million total reads per sample to maintain an appropriate level of sensitivity for subsequent expression analysis.

## **Material and Methods**

---

### **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 UCB. RT2 Profiler PCR 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<sup>+</sup> monocytes, memory CD4<sup>+</sup> and CD8<sup>+</sup> 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 manufacturers 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

## **Material and Methods**

---

and size selection. Afterward, appropriate sequencing Illumina indexes were incorporated into the samples through library preparation. Sequencing was performed using PE HiSeq4000 with 26bp for read 1 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'-CGTGTG GCCAGGATAGTCT-3') primers annealing up and down stream 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

## **Material and Methods**

---

functional markers simultaneously assayed and minimises the spillover of signal, two of the main limitations in flow cytometry.

Mass cytometry using CyTOF was performed by Dr Nicole Yager in collaboration with UCB for all the PsA samples (Chapter 5 in Table 2.1 and Table 5.2). 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.

## **Material and Methods**

---

### **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- $\gamma$ , CD-11b  
CD45RO, CD56, HLA-DR, IL-13, CD117, CD4, IL4, IL-2, TNF $\alpha$ ,  
IL-21, FceR, CD3, CD161, MCP-1, 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 used in the CyTOF staining panel are listed. Note the panel includes antibodies targeting intra-cellular molecules and also surface markers, to identify the cell populations of interest for further analysis of the intracellular cytokine and chemokine production.

## **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 (M. 2011) before general QC assessment using fastqc (S. 2010). Trimmed reads were aligned to the reference

## **Material and Methods**

---

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 pval cut off for filtering called peaks was determined for each cell type using Irreproducibility Discovery Rate (IDR) analysis (as further detailed in Chapter ??ch:Results1). 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

## **Material and Methods**

---

range of pvals (from 0 to  $10^{-45}$ ) and IDR analysis done for the resulting pairs of filtered peak sets. For each of the filtering pvals, the percentage of peaks sharing IDR rank between the two pseudoreplicates was determined, and the optimal pval 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  $\pm 250\text{bp}$  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  $\pm 1,000\text{bp}$  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 260340bp 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. 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

## **Material and Methods**

---

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 ( $\leq 5\text{Kb}$ ) 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 (see Section ATAC-seq, Fast-ATAC and Omni-ATAC data analysis) 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

## **Material and Methods**

---

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 (FCs) for each gene was conducted by UCB collaborators. When comparing synovial fluid and peripheral blood isolated cells from patients, FC was calculated for each patient as the ratio between the  $2^{-dCt}$  in each of the tissues. Therefore, in a particular cell type three FCs (one per individual) were provided for each gene. When comparing expression in cells isolated from peripheral blood of PsA patients versus healthy controls, the FC for each gene and each individual 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 FCs, 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  $\log_2\text{FC}$  were enabled when using DESeq2.

## **Material and Methods**

---

Differentially expressed transcripts were identified based on False discovery rate (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>. The 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 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

## **Material and Methods**

---

on the CD14<sup>+</sup> 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 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.bbnan`). Cluster identification was performed using the Leiden algorithm with default resolution 1 (`scanpy.leiden`) (Traag et al. 2018) followed by dimensional reduction and visualisation using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) algorithm (`scanpy.umap`). To identify marker genes of a particular cluster, differential analysis vs the rest of the cells was conducted using `scanpy.rank_genes_groups` function with p-values calculation using Wilcoxon-signed rank test and FDR with Benjamini-Hochberg test. The same function was used to conduct the differential gene expression analysis between CD14<sup>+</sup> monocytes of synovial fluid vs peripheral blood.

### **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](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html)).

Genomic feature enrichment analysis was performed using XGR built-in data (including eQTLs, FANTOM enhancers, histone marks and TFBS, amongst others) and the function `xGRviaGenomicAnno` (Fang et al. 2016).

## **Material and Methods**

---

When annotating regions (ATAC or ChIPm peaks) the summit of the peaks 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. Input data used were genes annotating differentially accessible ATAC peaks or differentially expressed genes, and background was defined as all the annotated ATAC peaks from the differential analysis master list or all the gencode hg19 detected genes, depending on the 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<sup>+</sup> 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 to compute enrichment of the psoriasis and PsA GWA LD blocks for an annotation feature, in this case differentially accessible ATAC peaks, 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. A number of 20,000 permutations were conducted. Similarly, enrichment analysis for PsA GWAS fine-mapped SNPs and differentially accessible regions or consensus

## Material and Methods

---

ATAC peaks was conducted using 1,000 backgrounds of SNPs generated by SNPsnap ([https://data.broadinstitute.org/mpg/snpsnap/match\\_snps.html](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 co-localised 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 FC 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, association testing and calculation of PP and were implemented in collaboration with Dr Adrián Cortés.

## Material and Methods

---

### Psoriasis fine-mapping using Immunochip summary statistics

Fine mapping was performed for X of the risk loci reported by the psoriasis Immunochip GWAS study from Tsoi and colleagues, for which only the summary statistics of the GPC Immunochip cohort (2,997 cases and 9,183 controls) were available (Tsoi et al. 2012). The summary statistics file included the pval, 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 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 ( $>0.8$ ). Association analysis and 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 ( $\sigma^2$ ); (2) the greater the variance the bigger the size effects obtained 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 ( $\beta$ ) and standard error (SE). SE is calculated as  $\sqrt{\sigma^2}$  and  $\beta$  is determined using the z-scores of each of the interrogated SNPs as:  $\beta = \text{z-score} * \text{SE}$ . It is important to note that step-wise conditional analysis is not performed when using summary statistics imputation. Similarly to the genotyping fine-mapping approach, PP for the SNPs in a particular window (signal) were calculated and ranked to set the threshold of the 50 and 90 % credible set of SNPs.

## **Material and Methods**

---

### **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. 2015). Access to the data post-quality control was kindly provided by Dr Anne Barton (The University of Manchester) PCA analysis was performed using only pruned SNPs with flashpca (Abraham and Inouye 2014) and the 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 SNPTTEST and including the previously calculated PCs as covariates (Burton 2007). Approximated Bayes Factors (ABF) were calculated for the lead signal and step-wise conditional analysis was performed if  $\log_{10}ABF > 3$ . 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 were identified using manual annotation. For the CD14<sup>+</sup> monocyte population identified within

## **Material and Methods**

---

each tissue at 0h and 6h, the percentage of TNF- $\alpha$ , MCP-1 and osteopontin positive staining cells were calculated. 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 the 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  $\leq$ 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 mitochondrial 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  $10 \times 10^6$  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).

**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 Turner *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 <i>et al.</i> , 2016	MACS2 (-nomodel), summit extension +/−250bp, rank summits by p-value	Maximally significant overlapping peaks.	non-Quantile unsupervised hierarchical clustering.
ENCODE*	MACS2 -nomodel, pairwise IDR analysis, filtering IDR<10%	Choosing longest pairwise	NA
Turner <i>et al.</i> , 2018	MACS2 (-nomodel -q 0.01)	Merging all filtered called peaks from the different cell types.	De novo: DiffReps fragment size 50bp. with

Alasoo <i>et al.</i> , 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<0.01).
Qu <i>et al.</i> , 2017	ZINBA PP>0.99.	Merging of filtered peaks from each individual sample.	Quantile normalisation and peak based in house Pearson correlation method.
Rendeiro <i>et al.</i> , 2016	MACS2 (-nomodel -extsize 147)	Merge of peaks from all samples in an iterative process including permutations	Peak based: quantile normalisation and Fisher exact test (FDR<0.05).
Scharer <i>et al.</i> , 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).

## **Establishment of methods to assess genome-wide chromatin accessibility**

---

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).

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, used for ChIP data, which 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 protocol(s) 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<sup>+</sup> monocytes and total CD4<sup>+</sup> (CD4<sup>+</sup>) 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 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.

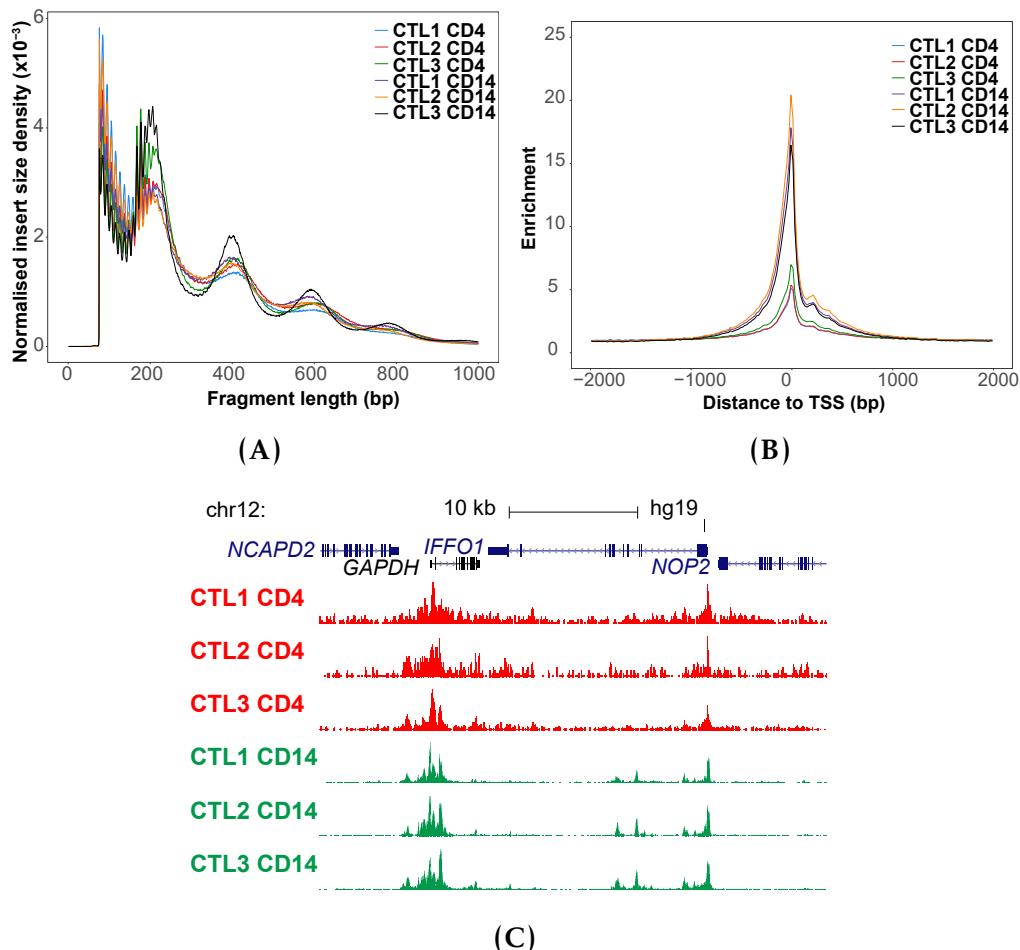
The most appropriate criteria and parameters to implement were investigated using ATAC-seq data generated for paired CD14<sup>+</sup> monocytes and CD4<sup>+</sup> T cells from three healthy individuals (referred as CTL1-3). This data corresponds to the fresh samples generated to test the effect of cryopreservation 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) ( $\leq 147$ bp, 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 that was investigated and implemented 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<sup>+</sup> samples, and were much higher (17 to 20) in the CD14<sup>+</sup> samples. The lower sample quality of the CD4<sup>+</sup> compared to CD14<sup>+</sup> 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*)



**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 shown are (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.  $CD14^+$  monocytes and  $CD4^+$  tracks are colour-coded in green and red, respectively.

## **Establishment of methods to assess genome-wide chromatin accessibility**

---

and the NOP2 Nucleolar Protein (*NOP2*) gene, showing more background reads and lower signal for the CD4<sup>+</sup> 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. 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<sup>+</sup> than in CD4<sup>+</sup> cells and was not directly related to any of the other quality control measurements. 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.

Both TSS and FRiP are appropriate signal-to-noise measures, with recommended threshold values from ENCODE and Alasoo and colleagues (Alasoo *et al.*, 2018 and previously published in bioRxiv in 2017) being FRiP between 10-20% and TSS between 6-10. 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<sup>+</sup> CTL1 and CTL2, being borderline for the 6 fold-enrichment TSS 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

<b>Sample</b>	<b>% mitochondrial reads</b>	<b>Fraction of reads in peaks</b>
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).

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.

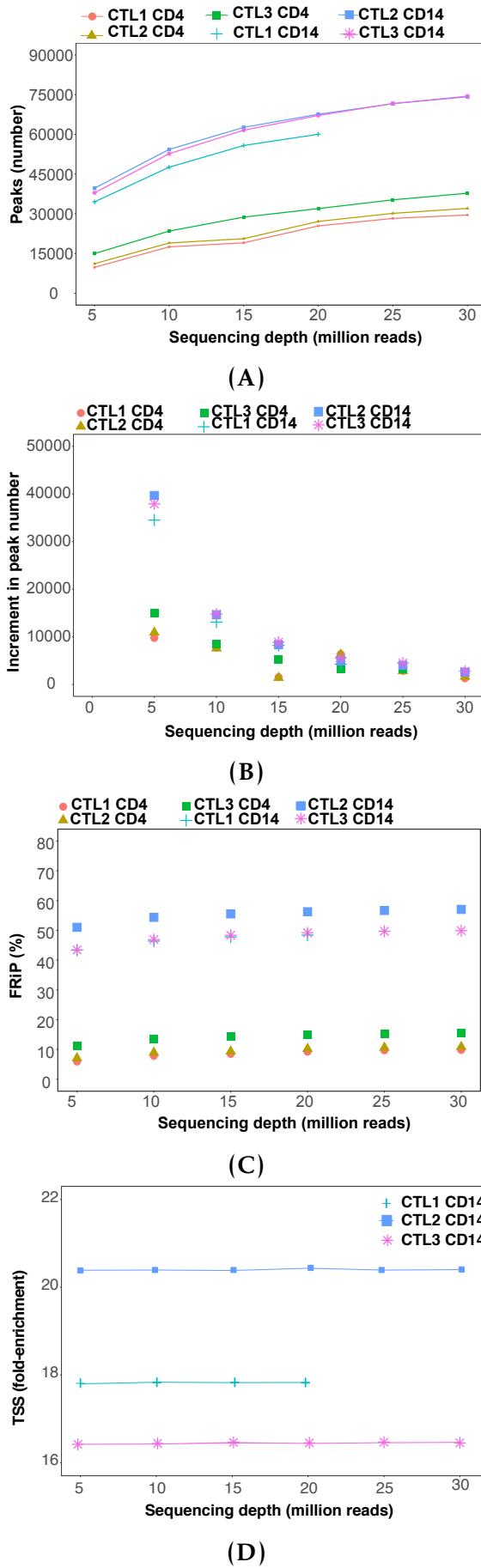
A systematic analysis of the effect of sequencing depth and the sample quality on peak calling was conducted to better understand the effect of both variables on the downstream analysis. For each of the six samples, random subsampling of reads was performed in every 5 million increment, ranging from 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 ??A), beginning to plateau at approximately 25 million reads (Figure 3.2B). Moreover, lower number of peaks were detected in CD4<sup>+</sup> samples compared to CD14<sup>+</sup> monocytes when

## **Establishment of methods to assess genome-wide chromatin accessibility**

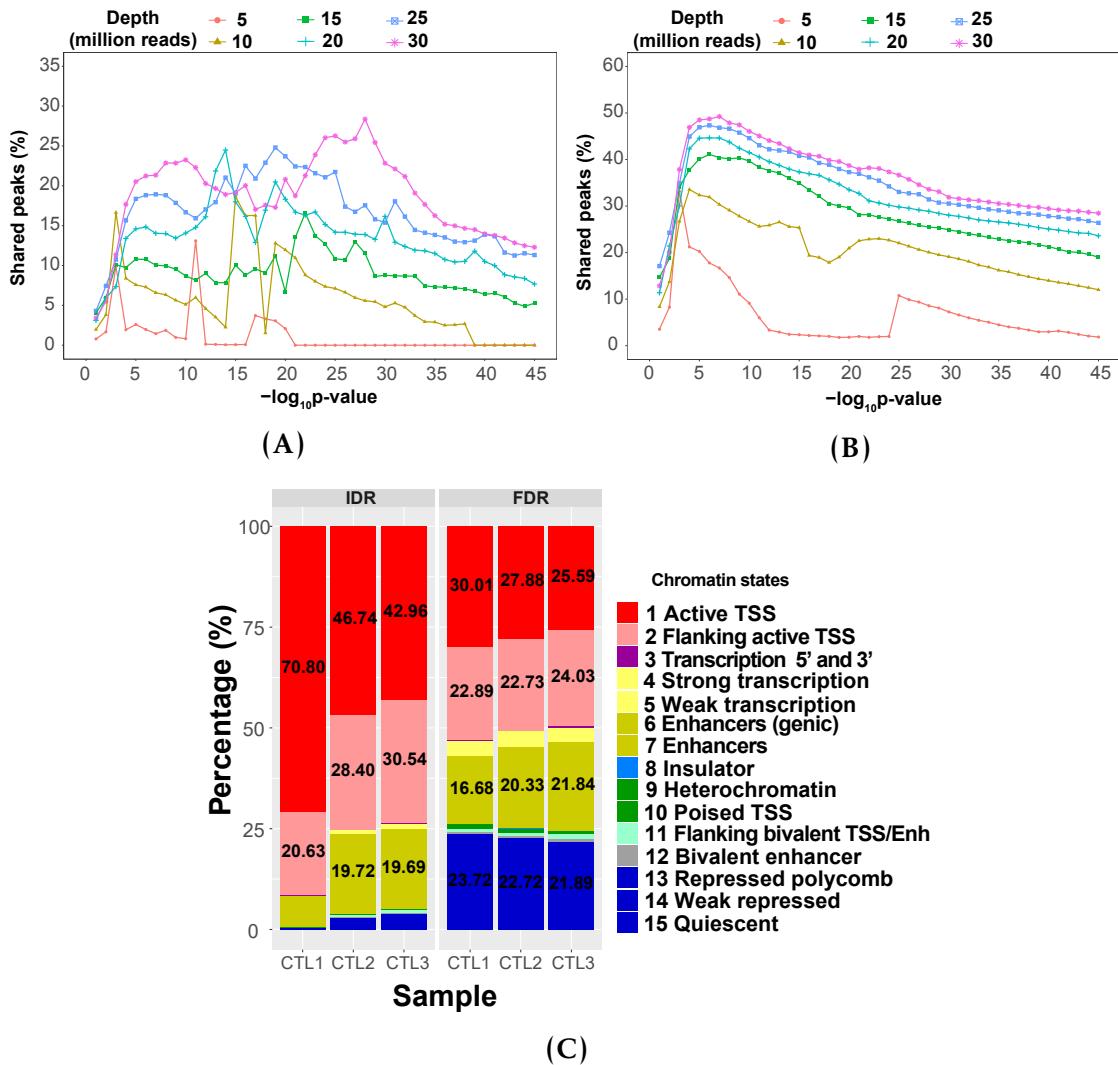
---

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.

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<sup>+</sup> T cells, which had lower TSS values compared to CD14<sup>+</sup> monocytes (Figure 3.3A and B). The shape of the curves were also influenced by the sample quality. For appropriate sequencing depth (15-20 million reads), the CD14<sup>+</sup> 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<sup>+</sup> samples reached two local maxima (Figure 3.3A).



**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 depth.



**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) CD14<sup>+</sup> monocytes and (B) CD4<sup>+</sup>, two representative samples differing in quality for this analysis. (C) Annotation (as percentage of total) of the CTL1, CTL2 and CTL3 CD4<sup>+</sup> ATAC-seq peaks filtered for FDR<0.01 or optimal p-value from the IDR analysis ( $p\text{-value}=10^{-11}$ ) with the corresponding cell type specific Roadmap Epigenomics Project chromatin segmentation map.

Filtering the CD4<sup>+</sup> 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 3.3 c). 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**

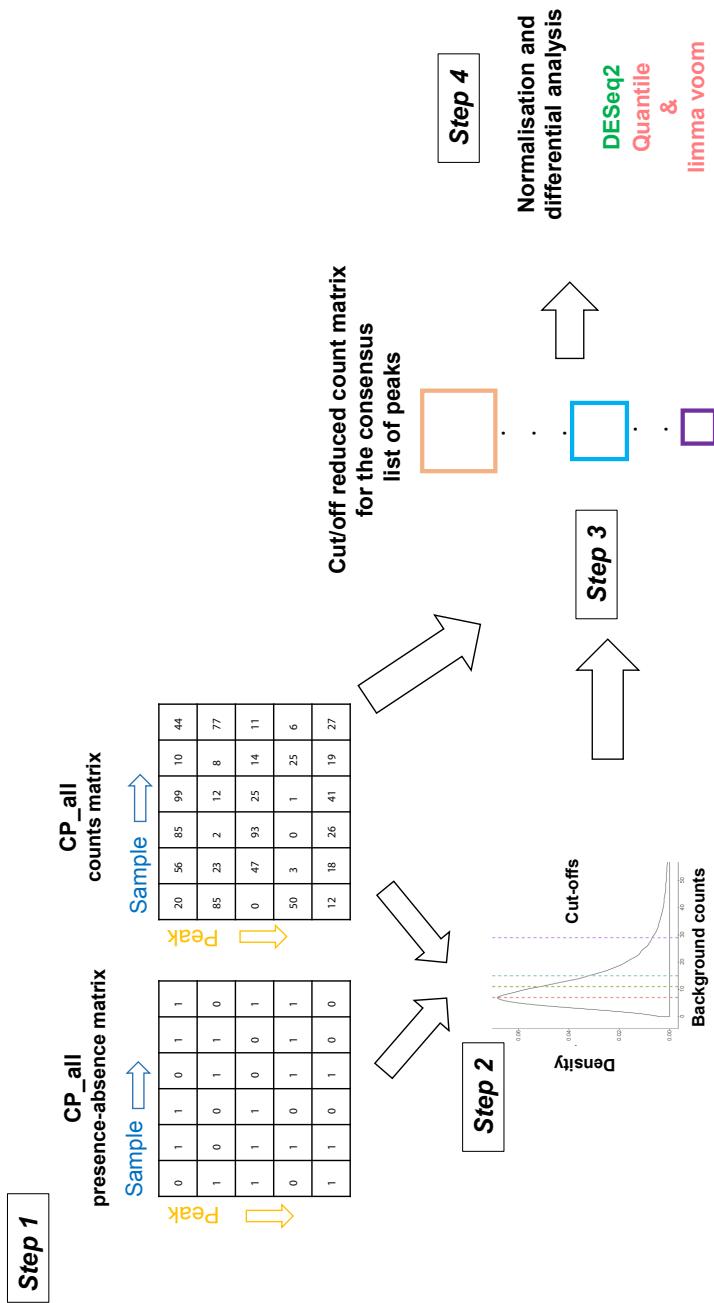
A peak-based approach using the number of reads overlapping the peaks included in the consensus list of ATAC peaks (CP\_all) was implemented to perform differential analysis. One of the main limitations of the ATAC-seq and Fast-ATAC protocols (discussed in 3.2.2 and 3.2.3) is the background signal. Therefore, an empirical cut-off was identified to minimise the impact 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 the use of limma voom to perform differential chromatin accessibility analysis. The second strategy used DESeq2 for both, normalisation and differential analysis.**

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)

## **Establishment of methods to assess genome-wide chromatin accessibility**

---

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 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 are retained. The resulting reduced matrix of low-noise peaks for each cut-off value was then normalised using quantile or DESeq2 (library size and variance stabilization (Love et al. 2014)) and used to conduct differential analysis with limma voom or DESeq2 respectively. The results from this were used to determine an appropriate threshold for filtering background reads, as well as to compare the two analytical methods.

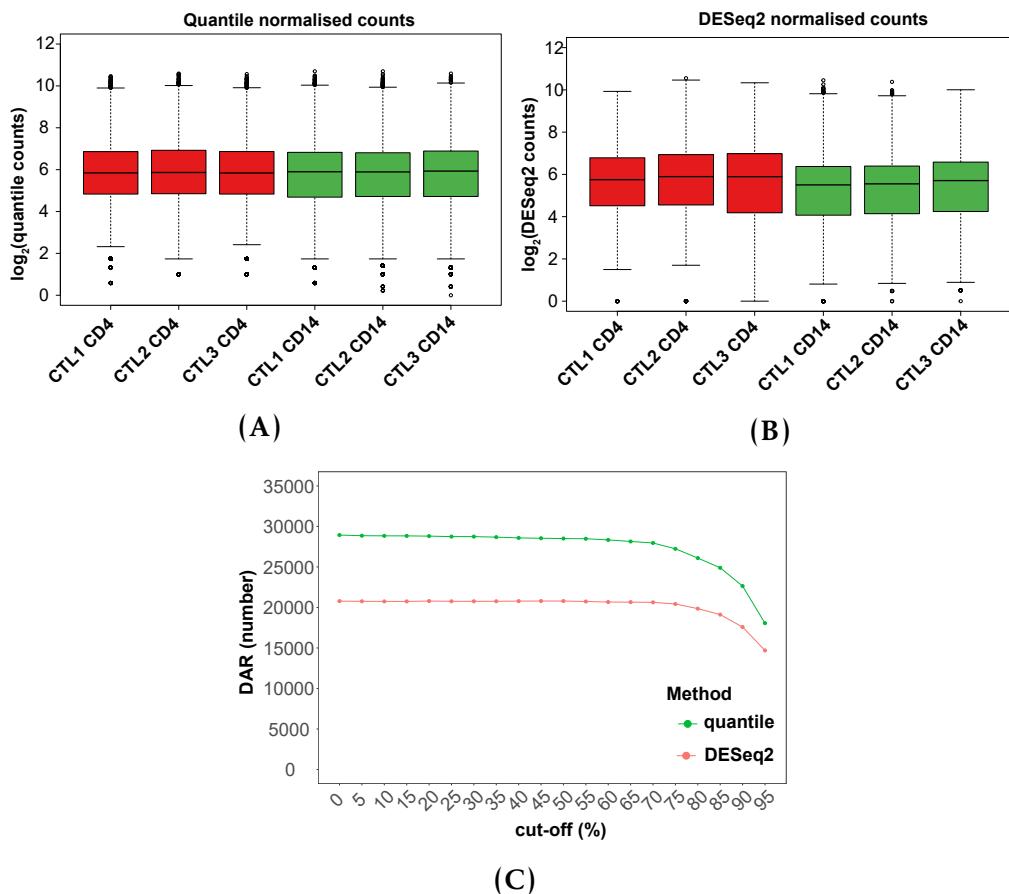


**Figure 3.4: Work flow illustrating the strategy to account for ATAC background noise prior to differential analysis.** A combined consensus list of ATAC peaks (CP\_all) was built as explained in Chapter 2, including all the significant peaks that were shared by at least 30% of the samples included 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).

Both normalisation methods performed appropriately across all filter values and samples, with the quantile normalisation showing slightly greater consistency across the two groups (Figure 3.5A and B). Differential chromatin accessibility analysis using limma voom showed a greater number of significant (FDR  $\leq 0.01$  and FC  $\geq 1.5$ ) 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 ??C). This suggested a reduction in the number of false positive hits resulting from peaks with 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.

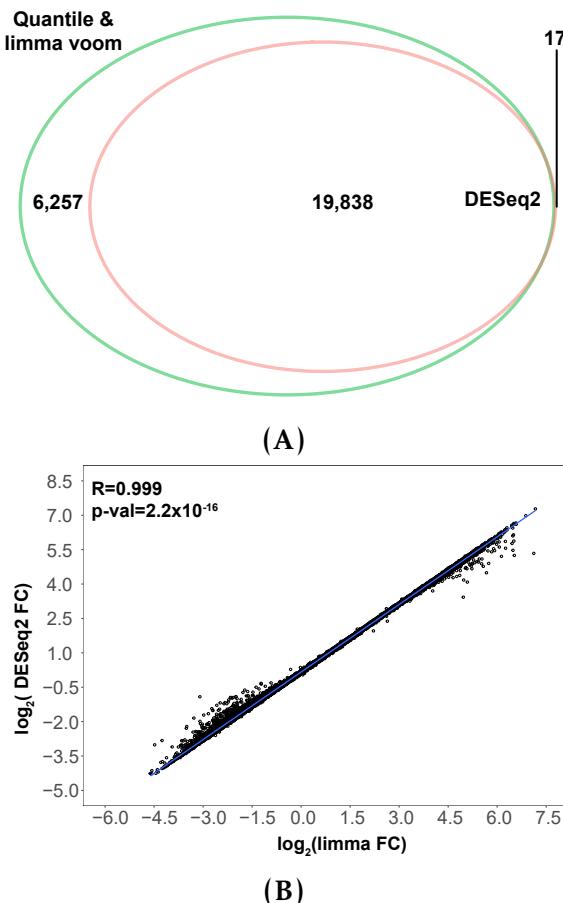
The vast majority of the 19,855 DARs called as significant using the more conservative method (DESeq2) were recapitulated by limma voom at the same significance threshold (FDR $<0.01$  and fold change $>1.5$ ) (Figure 3.6 a). FDR rank revealed that out of the first 19,855 limma voom DARs 18,768 matched 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 ( $R=0.999$ ,  $p\text{-value}=2.2^{-16}$ ) (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 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 $^{+}$

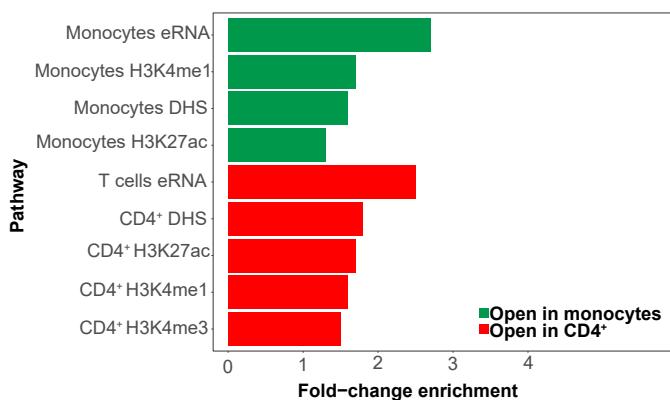


**Figure 3.5: Normalisation and differential chromatin accessibility analysis for different cut-offs using quantile normalisation& limma voom and DESeq2.** Boxplots representing the log<sub>2</sub> read counts for each of the peaks from the unfiltered consensus list of peaks normalised by (A) quantile or (B) DESeq2 in the three CD14<sup>+</sup> monocytes and total CD4<sup>+</sup> healthy control paired samples. (C) Representation of the number of significant DARs (FDR<0.01 and FC>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.

DHSs (Figure 3.7). Conversely, the top enriched features for CD14<sup>+</sup> 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<sup>+</sup> monocytes and CD4<sup>+</sup> 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 cut-off 80%.** (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 consensus list of ATAC peaks for the 80% optimal cut-off using limma voom or DESeq2. (B) Representation of the correlation between limma voom and DESeq2 log<sub>2</sub> 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<sup>+</sup> monocytes and CD4<sup>+</sup> cells.** Barplot representing the fold change for the top significantly enriched (FDR<0.01) FANTOM5 eRNAs and, 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<sup>+</sup> monocytes (open in monocytes) or CD4<sup>+</sup> (open in CD4<sup>+</sup>).

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

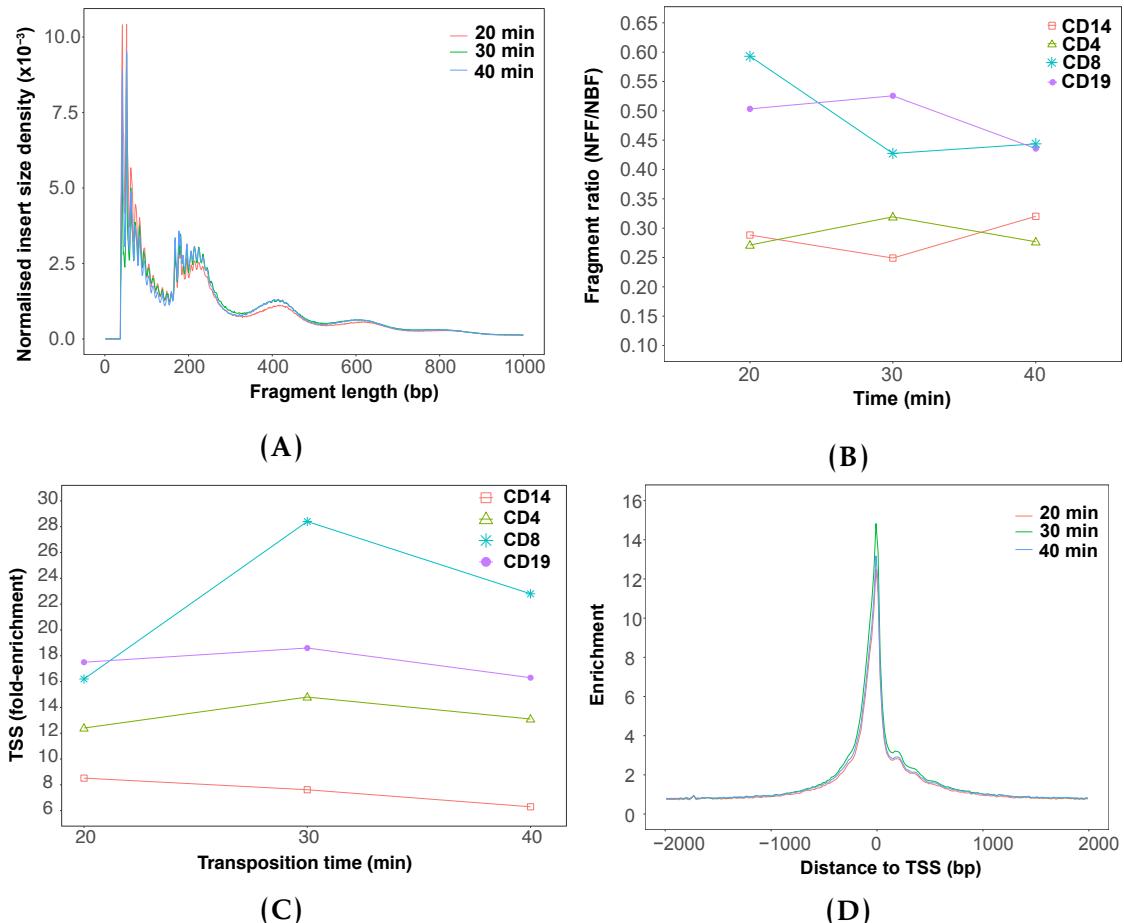
The duration of the transposition reaction was optimised for the main immune cell types of interest for this thesis. ATAC-seq was performed for three different transposition times (20, 30 and 40 min) in CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, total CD8<sup>+</sup> (CD8<sup>+</sup>) and CD19<sup>+</sup> 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 ATAC related readouts was explored. The duration of the transposition time is known to have an effect on the proportion of nucleosome-free and nucleosome-bound (mono-nucleosomes and beyond) regions tagged by the adapters. Ideally, the transposition reaction should maximise NFF ( $\leq 150\text{bp}$ ), where TF and other proteins bind, without loosing mono- and di-nucleosome fragments characteristic of the ATAC libraries. All three transposition times produced appropriate fragment size distributions which recapitulated the ATAC characteristic nucleosome periodicity pattern (Figure 3.8A). The variation of the ratio between NFF and nucleosome-bound fragments (NBF) ( $>150\text{bp}$ ) with

the transposition times was heterogenous across cell types (Figure 3.8B). For example, CD8<sup>+</sup> presented the greatest NFF/NBF ratio for 20 min of transposition whereas the NFF/NBF ratio reached a maximum at 40 min for the CD14<sup>+</sup> monocytes. Nevertheless, the change in this ratio across transposition times was very moderate in all cell types but for 20 min in CD8<sup>+</sup> cells. In all four cell types and transposition times, the ATAC libraries showed NFF/NBF ratios lower than 7, considered the length ratio cut-off to designate ATAC libraries lacking appropriate fragment size distribution according to Alasoo and colleagues (Alasoo et al. 2018).

Transposition times 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). Differences between TSS fold-enrichments at 30 and 40 min were very modest and did not differ in more than 4 units in CD8<sup>+</sup> 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.

### **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



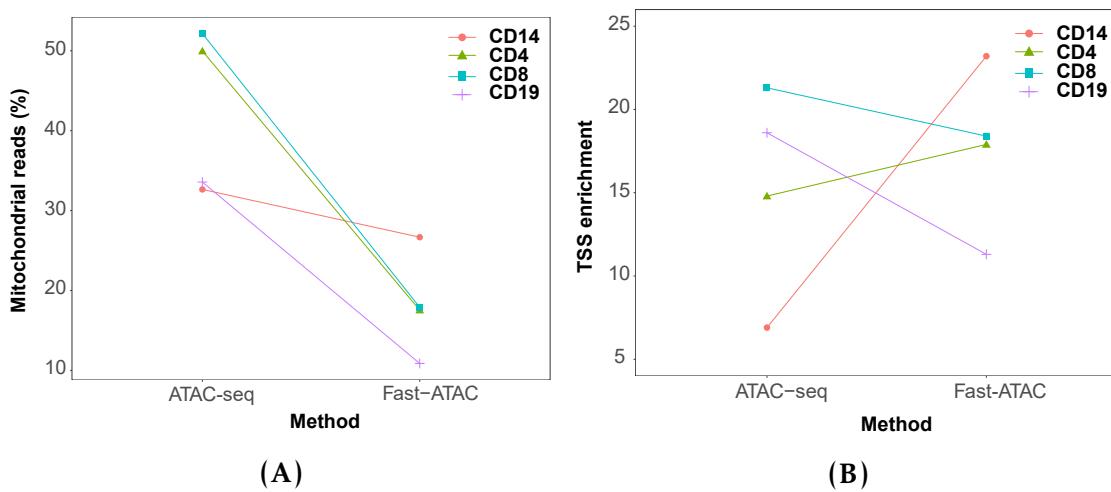
**Figure 3.8: Assessment of the effect of transposition times on the ATAC-seq measurements.** (A) Representative plot of the ATAC-seq fragment sizes density distribution following 20, 30 and 40 min of transposition in CD8<sup>+</sup> cells. Changes across different transposition times in CD14<sup>+</sup>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

advantages of Fast-ATAC, namely reduction of mitochondrial reads and signal enhancement, before implementing it as the replacement for the ATAC-seq protocol, which had been used for processing 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 3.2.2 transposed for 30 min in the same four cell types.

The percentage of mitochondrial reads in the Fast-ATAC libraries was lower for all cell types analysed compared to ATAC-seq (Figure 3.9A). Importantly, the mitochondrial percentage in CD4<sup>+</sup> and CD8<sup>+</sup> cells showed a reduction from 50% to less than 20%. With regard to the background signal, differing trends were observed across cell types (Figure 3.9B) with no change in TSS enrichment for CD14<sup>+</sup> monocytes or CD4<sup>+</sup> T cells while CD8<sup>+</sup> and CD19<sup>+</sup> cells showed a reduction with Fast-ATAC when compared to ATAC-seq. 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 patient recruitments (see Chapter ??).

### **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 keratinocytes, one of the most relevant cell types in psoriasis 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), the epidermis was isolated from a psoriatic lesional skin biopsy, digested with trypsin and filtered through a cell strainer to ensure a single-cell suspension,



**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<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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 3.2.2 transposed for 30 min.

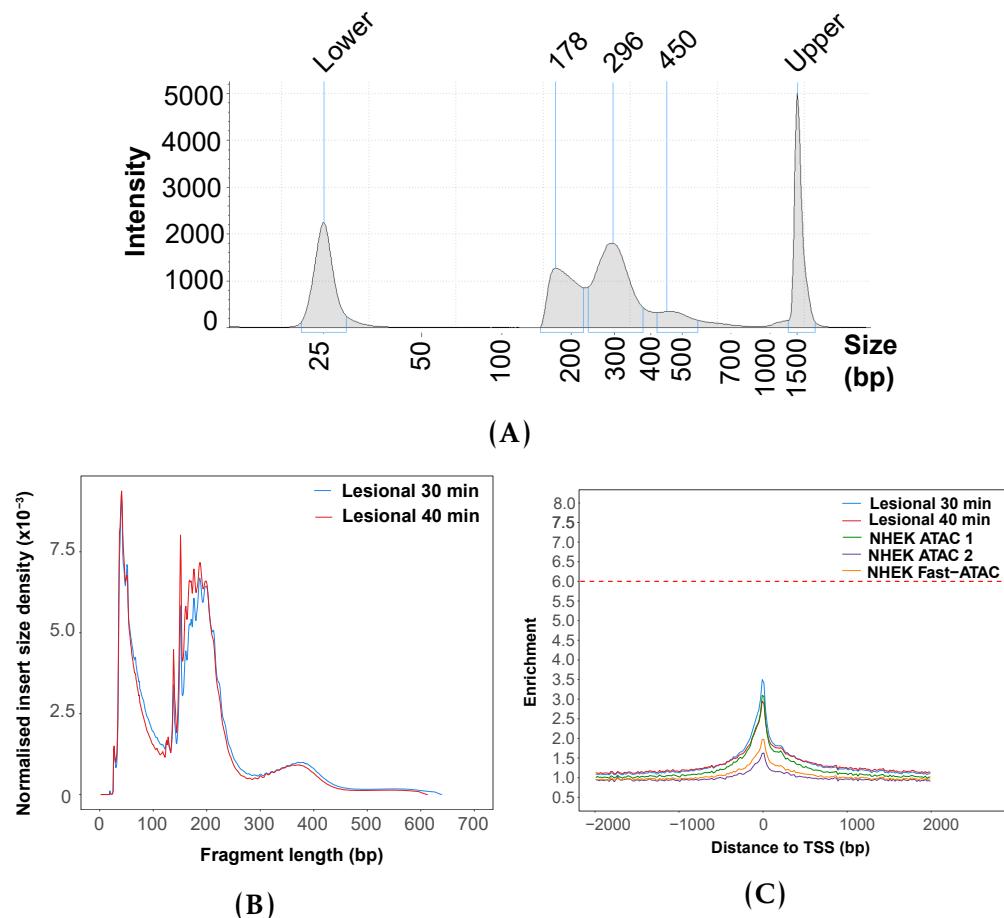
as detailed in Chapter 2. Approximately 50,000 cells from the suspension were counted and ATAC 1 was performed for two different transposition times (30 and 40 min). Since biopsy handling and lesional epidermal keratinocytes are particularly challenging, this was considered the best system to test the performance of the standard protocol in the clinical setting of interest. Library quality control based on pre-sequencing relative abundance of DNA fragment sizes recapitulated the characteristic ATAC nucleosome pattern every ~200bp generated by transposition of nucleosome-free and nucleosome-bound DNA in both samples (Figures 3.10A and B.2). This was consistent with the fragment size distribution from the NGS data, presenting NFF and NBF (mono- and di-nucleosomes only) for both transposition times (Figure 3.10B). However the relative abundance of the mono-nucleosome fragments appeared to be approximately equal to the NFF, which is not observed in higher quality libraries. Libraries for both transposition times presented TSS fold enrichment below

the acceptable cut-off threshold of 6, showing slightly better signal (3.5 fold-enrichment) in the 30 min library (Figure 3.10C).

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). However, dead cells and free-DNA released by apoptotic cells throughout processing could contribute to the increased background noise observed above.

Following Buenrostro and colleagues' ATAC-seq protocol, a modified version for keratinocytes by Bao and colleagues (named ATAC 2 in this subsection) and the Fast-ATAC protocols were released (Bao et al. 2015; Corces et al. 2016). Interestingly, Bao's protocol was applied directly on the cell culture plate containing adherent NHEKs, avoiding a trypsinisation step that could increase cell death. In line with Bao and colleagues protocol, to prevent background noise due to presence of dead cells affecting the assessment of the different protocols, two systems were implemented. First, all the cells obtained from the epidermis of control skin biopsies were cultured for 3h in a 96-well plate and washed afterward to minimise the presence of apoptotic cells. This procedure, known as adherence assay, allows the isolation of viable undifferentiated keratinocytes. Second, cultured NHEKs (50,000 cells) were also incorporated as a control to test the performance of the different ATAC protocols. The three ATAC protocols (ATAC 1, ATAC 2 and Fast-ATAC using C1 conditions, Table 3.3) were performed directly on the plate adherent cells with no cell detachment step.

For all three protocols, the library size distribution of sequenced fragments showed the presence of NFF and a poorly defined nucleosome pattern, particularly for the ATAC 2 protocol in both NHEKs and adherent keratinocytes from skin biopsies (Figure 3.11A). TSS enrichments were also very low,



**Figure 3.10: Quality control assessment of different ATAC protocols in keratinocytes from a psoriatic lesion and NHEKs.** (A) Pre-sequencing quantification of DNA fragment sizes in the ATAC libraries and (B) the density distribution of sequenced fragments for ATAC 1 libraries generated in 50,000 keratinocytes in suspension isolated from lesional skin from a psoriasis patient. Two transposition times (30 and 40 min) were tested in the same sample keratinocytes suspension and only the pre-sequencing profiles of relative abundance of DNA library fragment sizes at 30 min transposition has been included. The pre-sequencing profiles of relative abundance of DNA library fragment sizes for 40 min transposition is in Figure B.2. (C) Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from the ATAC 1 psoriasis lesional keratinocytes libraries (previously mentioned in (A) and (B) and NHEK libraries generated with ATAC 1, ATAC 2 and Fast-ATAC protocols performed directly on the 96-well plate adherent cells.

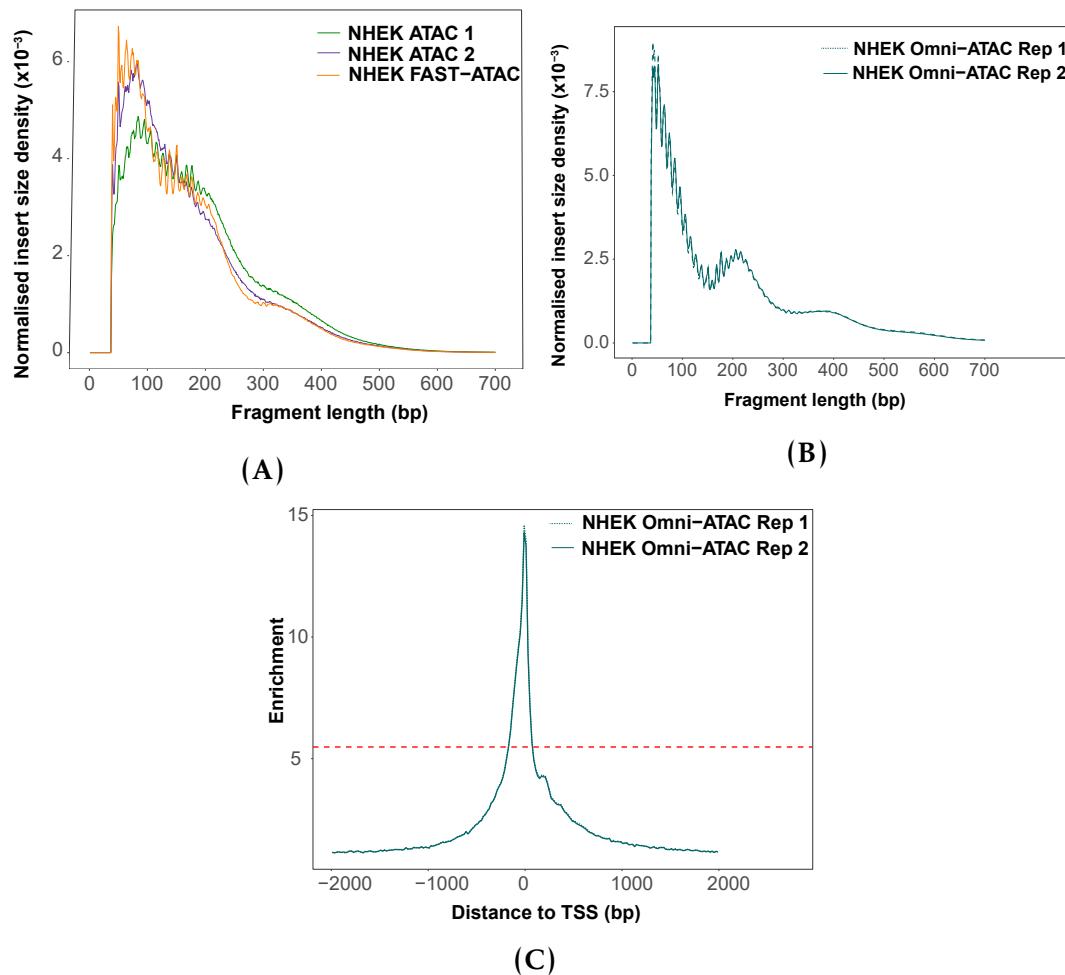
Protocol	Lysis and transposition	Key parameters
ATAC 1 (Buenrostro et al. 2013)	Two steps	0.1% NP-40 and 2.5µL Tn5
ATAC 2 (Bao et al. 2015)	Two steps	0.05% NP-40 and 5µL Tn5
Fast-ATAC (Corces et al. 2016)	One step	C1*: 0.01% digitonin, 2.5µL Tn5 C2: 0.01% digitonin, 0.5µL Tn5 C3: 0.025% digitonin, 0.5µL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

particularly for the ATAC 2 protocol, and in the two cell systems under the acceptable cut-off of 6 (Figures 3.10C and B.3).

Further optimisation of the Fast-ATAC protocol was performed by modifying the original concentration of NP-40 detergent and the Tn5 enzyme (Table 3.3 C1-4). Library quality control using pre-sequencing profiles of relative abundance of DNA library fragment sizes to assess the DNA fragment size distributions failed to show the nucleosome pattern profile expected in ATAC, and therefore NGS was not performed (Figure B.4A, B and C).

Towards the end of experimental work for 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 3.10B). Moreover, high TSS enrichment values (approximately 20 fold) were observed for the two replicates (Figure 3.10 c). When performing overlap



**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) ATAC1, ATAC2 and Fast-ATAC or (B) Omni-ATAC protocols. (C) Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from two Omni-ATAC technical replicates.

between the Omni-ATAC sample peaks filtered based on low stringency p-value (0.01) and the ENCODE DHSs from 125 cell types, the highest percentage (approximately 55%) was reported for NHEKs (Figure 3.12A).

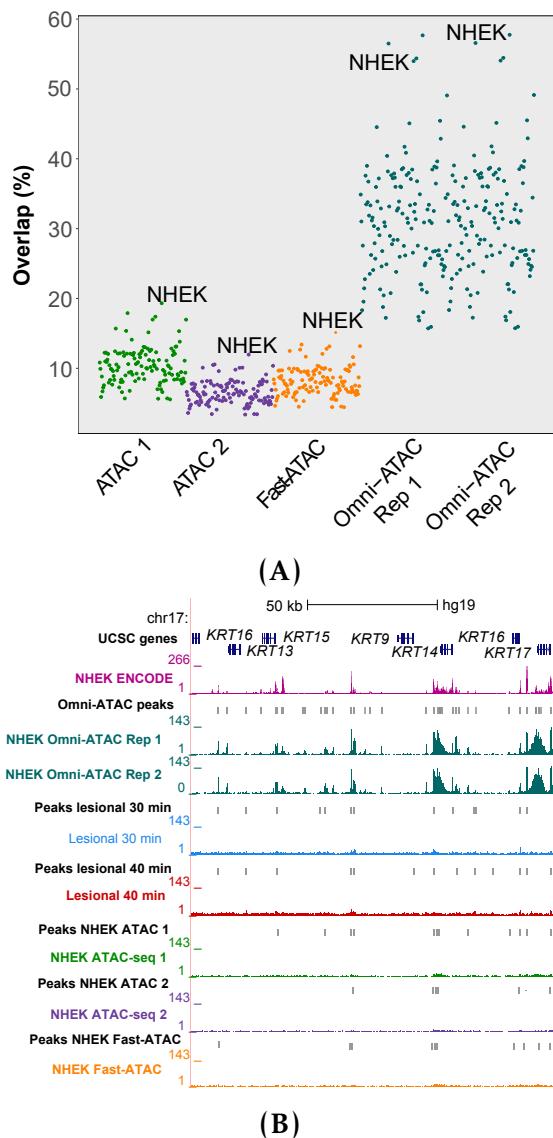
In contrast, the same analysis for the ATAC 1, ATAC 2 and Fast-ATAC peaks only showed 20% or less overlap with ENCODE NHEKs, supporting the higher quality and specificity of Omni-ATAC accessible regions in keratinocytes, even at a very low p-value filtering. The differences in the quality of ATAC signal 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).

### **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 and a collaborative project to investigate these was established with High-Throughput Genomics 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 3 healthy volunteers (Chapter 3 in 2.1) matched for sex and age (3 females, mean

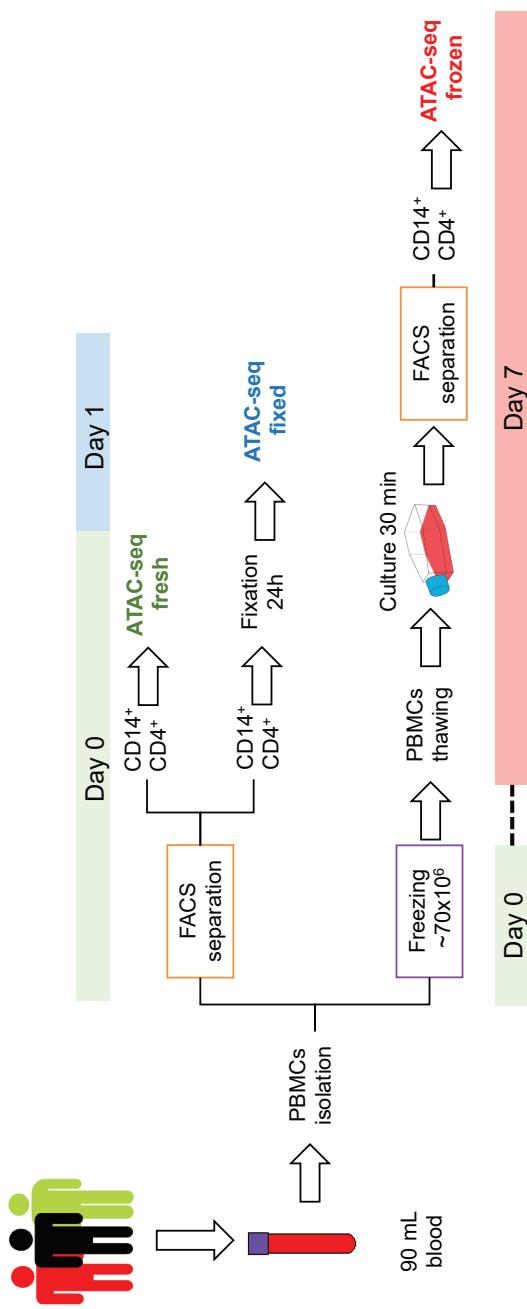


**Figure 3.12: Comparison of 4 different ATAC protocols applied to psoriasis and/or healthy keratinocytes and NHEKs cells with published ENCODE DNase-seq data.** (A) Enrichment (% of overlap) of observed peaks called in each ATAC sample (using low stringent p-value) with open DHS chromatin regions in 125 ENCODE cell types. (B) UCSC Genome Browser track showing open chromatin at the chr17 keratin (KRT) family gene locus for different ATAC protocols with the normalised ATAC read density (y-axis) shown.

## **Establishment of methods to assess genome-wide chromatin accessibility**

---

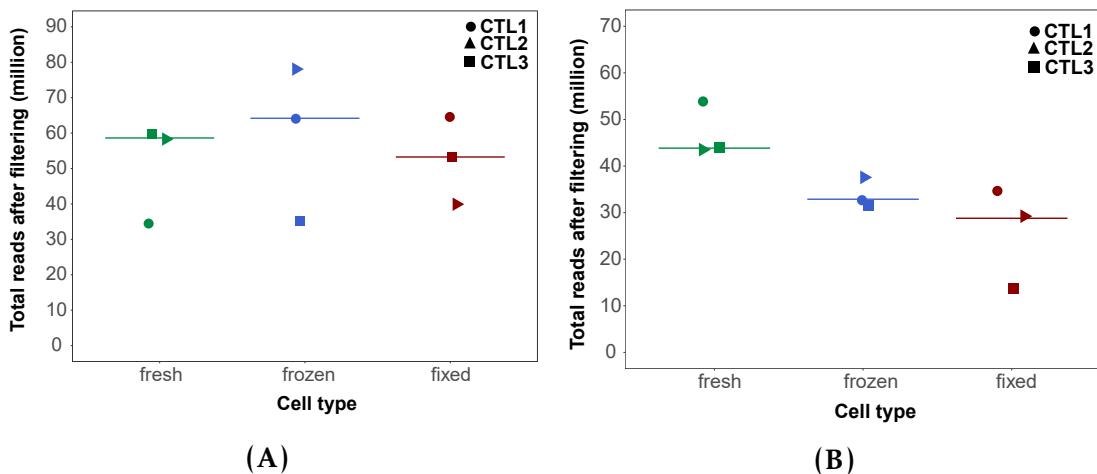
age 25.6 years old) was processed on different days to simulate the experimental design when using patient samples (experimental design summarised in Figure 3.13). PBMCs were prepared from 90mL blood using a Ficoll gradient and CD14<sup>+</sup> monocytes and CD4<sup>+</sup> T cells were isolated by FACS, as detailed in Chapter 2. ATAC-seq was performed on 50,000 CD14<sup>+</sup> monocytes and CD4<sup>+</sup> 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). To investigate the utility of cryopreservation, 70x10<sup>6</sup> million PBMCs were cryopreserved 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<sup>+</sup> monocytes and CD4<sup>+</sup> 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, 3 matched ATAC-seq libraries were generated in two different cell populations: ATAC-seq fresh, ATAC-seq fixed and ATAC-seq frozen.



**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<sup>+</sup> monocytes or CD4<sup>+</sup> 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).

### Chromatin accessibility in the different experimental conditions

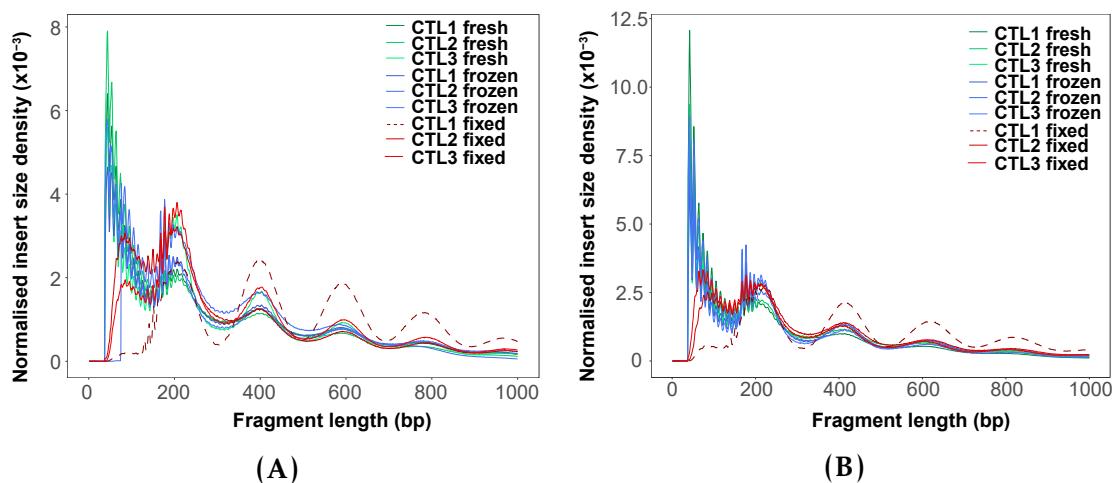
All samples from each of the two cell types had more than 15 million reads, which have previously been shown as the minimum for successful ATAC-seq analysis and peak calling (Figure 3.14). The median number of reads across the fresh, frozen and fixed were more similar for CD14<sup>+</sup> monocytes (58.6, 64.2 and 39.6 million reads, respectively) than in the CD4<sup>+</sup> 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).



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

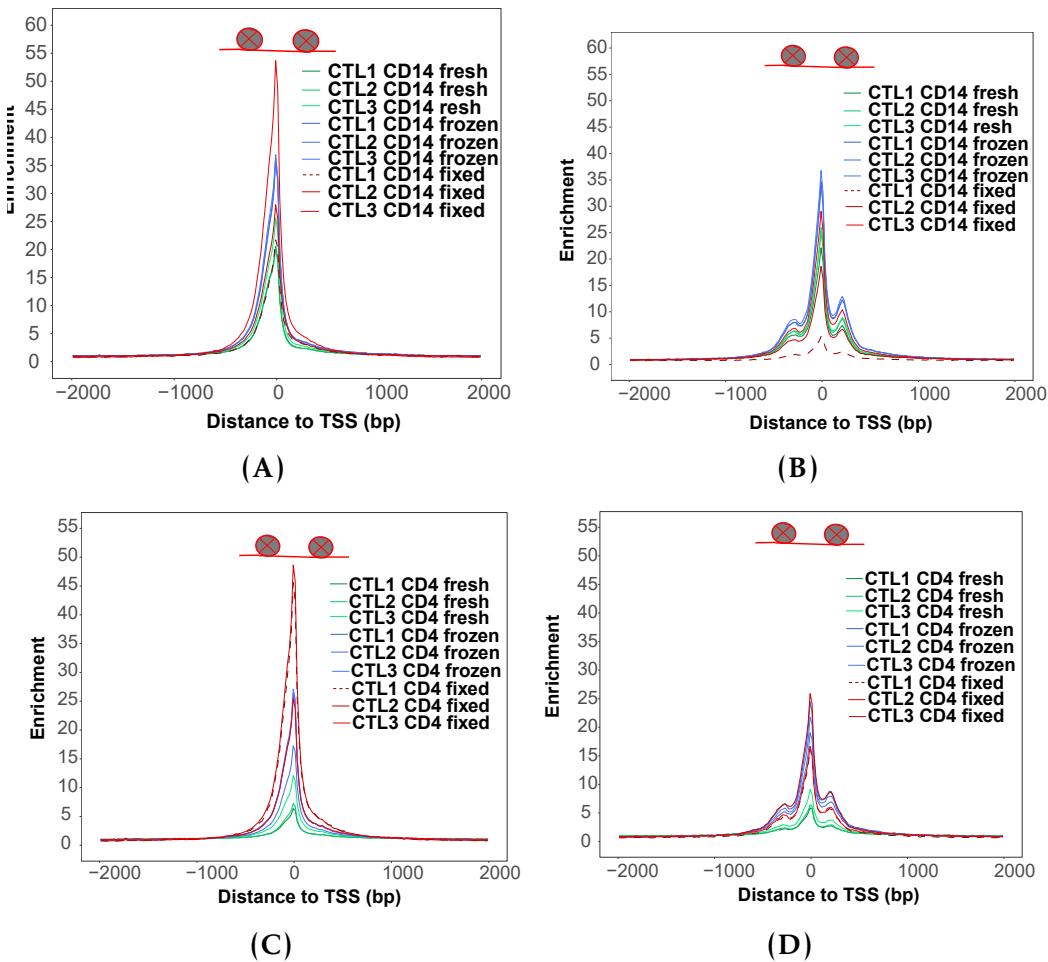
The ATAC-seq signal-to-noise ratios across the TSS showed a similar median for the fresh and fixed CD14<sup>+</sup> monocytes libraries (17.4 and 16.5 fold-enrichment, respectively) and was higher for the frozen samples (26.3 fold-enrichment)(Table A.1). The median TSS enrichments in the frozen and fixed CD4<sup>+</sup> 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).

In terms of the fragment size distribution, the profiles of all the samples, except fixed CD14<sup>+</sup> monocytes and CD4<sup>+</sup> from CTL1, were similar, showing NFF <150bp and fragments corresponding to mono-, di-, tri- and tetra-nucleosomes (Figure 3.15 a and b). The fixed samples presented a lower density of NFF when compared to fresh and frozen, which had very similar distributions in both cell types. In particular, fixed CD14<sup>+</sup> monocytes and CD4<sup>+</sup> in CTL1 had extremely low abundance of NFF (Figure 3.15 red-dashed line in A and B), consistent with the very low TSS enrichment for CTL1 CD14<sup>+</sup> ATAC-seq fixed cells.



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

Chromatin structure across and within the TSS was then investigated following Scharer and colleagues (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<sup>+</sup> fresh samples presenting the lowest 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 a characteristic periodicity in the TSS surroundings, with two peaks of enrichment mapping at the up-stream and down-stream positioned nucleosomes (Figure 3.16B and D).

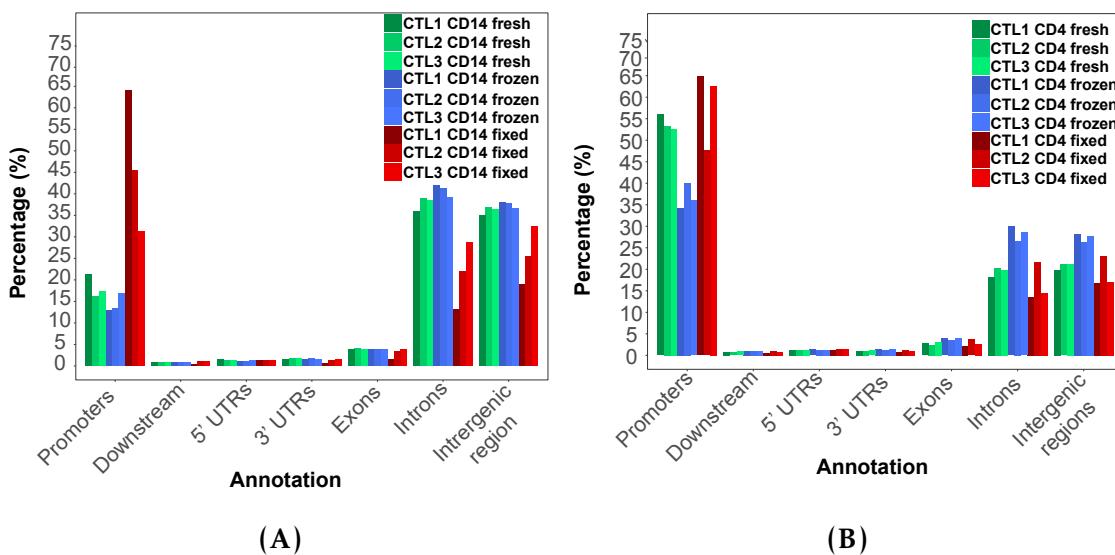


**Figure 3.16: ATAC-seq enrichment of nucleosome-free and di-nucleosome fragments at the TSS and surroundings in CD14<sup>+</sup> monocytes and CD4<sup>+</sup> samples for the three conditions.** Nucleosome-free fragments (<150bp) and di-nucleosome (between 260 and 340bp) were selected *in silico* and enrichment analysis was carried out +/-1Kb across all the Ensembl annotated TSS.

Although fixation in CTL1 CD14<sup>+</sup> reduced the abundance of nucleosome-free tagged fragments, the ATAC-seq signal at those regions was clearly enriched when compared to the background. In contrast, DSP in CTL1 CD14<sup>+</sup> appeared to increase the efficiency of ATAC-seq in tagging NBF (di-, tri- and tetra-nucleosomes) but the loss of chromatin structure around the TSS indicated that those fragments are likely to have been displaced from their original location. Conversely, the fresh CD4<sup>+</sup> samples showed low enrichment for NFF in the TSS, in line with their overall TSS enrichment (Table A.1) and indicating weakly the two nucleosome positioned in the TSS surroundings. In summary, freezing and,

importantly fixing (with exception of the two cell types from CTL1) appeared to maintain the overall chromatin structure.

Annotation of the significant peaks from each sample (filtered 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). The higher percentage of peaks annotated as promoters for some ATAC-seq libraries revealed the preferential location of high quality and robust peaks at this feature when compared to other annotations. Overall, the genomic annotation of the peaks across the three conditions revealed accessible ATAC-seq regions at meaningful genomic features.

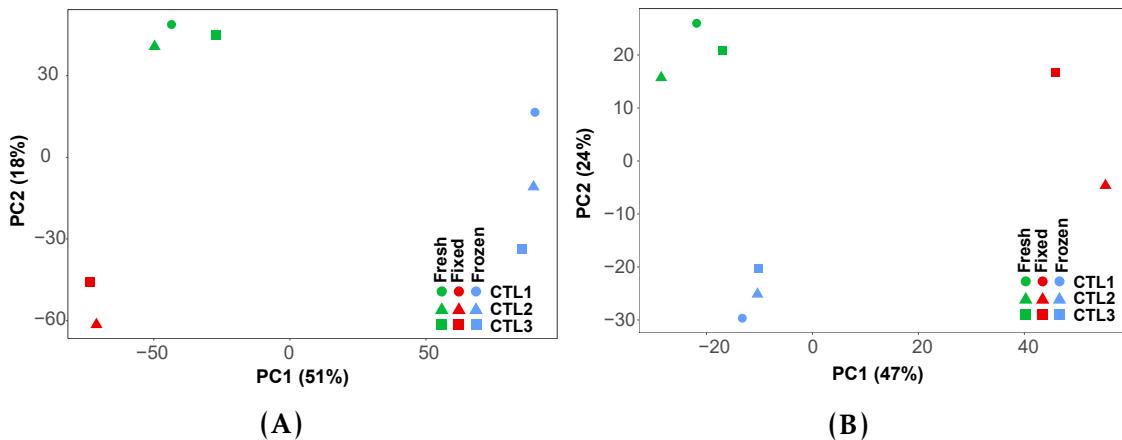


**Figure 3.17: Genomic features annotation for the ATAC-seq peaks called in each of the fresh,frozen and fixed samples from CD14<sup>+</sup> monocytes and CD4<sup>+</sup>. Overlap was performed between the genomic features and the list of (A)CD14<sup>+</sup> monocytes and (B) CD4<sup>+</sup> peaks filtered for FDR<0.01 in each sample from each of the three conditions (fresh=green, frozen=blue and fixed=red).**

### Differential analysis demonstrates discrete significant changes in chromatin accessibility across conditions

In order to investigate genome wide differences between ATAC-seq fresh (biological reference) and ATAC-seq frozen and fixed, read counts were retrieved

at the location of each peak including in the CD14<sup>+</sup> monocytes and CD4<sup>+</sup> consensus list of ATAC regions built including all the samples from the three conditions (CP\_CD14\_all\_cond and CP\_CD4\_all\_cond, respectively). For this analysis the data relating to CTL1 fixed samples from CD14<sup>+</sup> and CD4<sup>+</sup> cells were removed given the low quality metrics previously described.



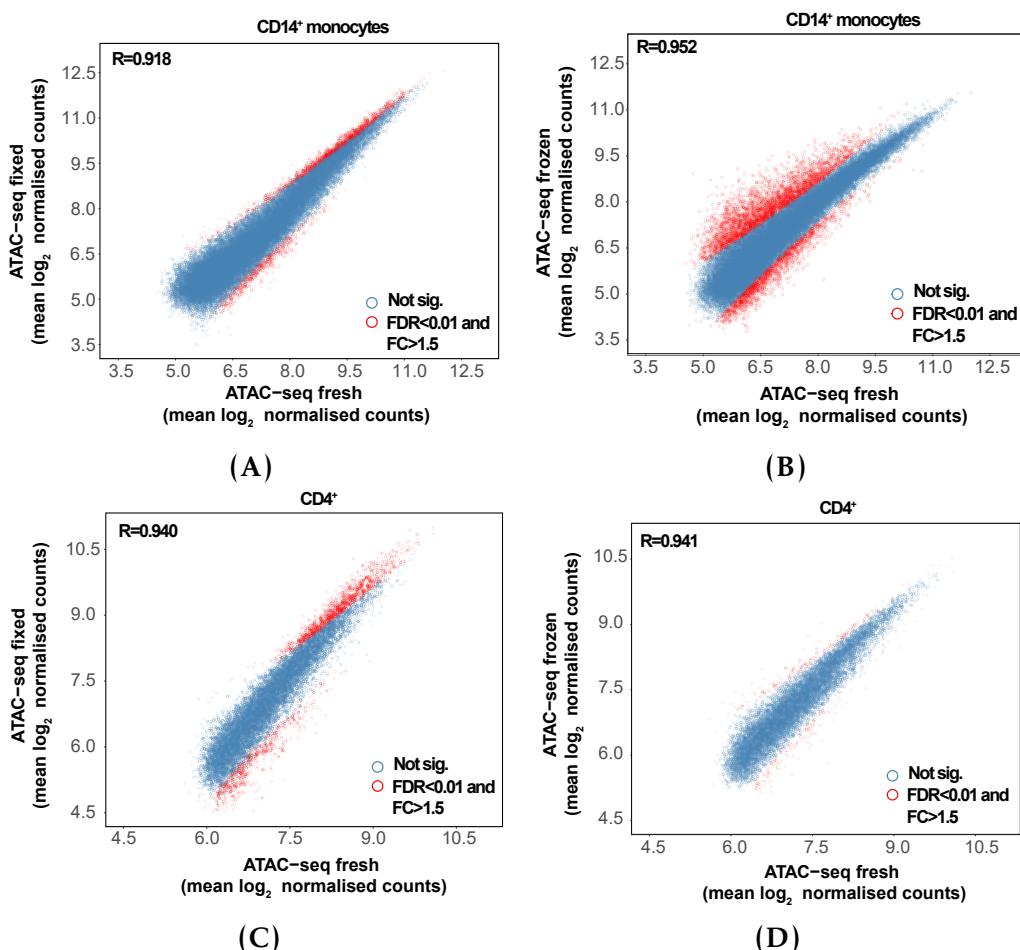
**Figure 3.18: PCA analysis based on the ATAC-seq chromatin accessibility landscape in fresh, fixed and frozen samples.** PCA analysis was performed using the normalised counts across the consensus list of ATAC peaks from the combined fresh, fixed and frozen samples (CP\_CD14\_all\_cond and CP\_CD4\_all\_cond) in (A) CD14<sup>+</sup> monocytes or (B) CD4<sup>+</sup> 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.

Principal component analysis (PCA) performed using DESeq2 normalised counts for each of the two consensus list of peaks demonstrated differences across the three ATAC-seq conditions in the two cell types. Plotting the first two principal components (PCs) showed sample clustering based on condition, which explained the largest variability within the two cell types. The first PC explaining 51% of the variance in the chromatin accessibility landscape across samples, separated ATAC-seq fresh and fixed from the ATAC-seq frozen samples in CD14<sup>+</sup> monocytes (Figure 3.18A). The second PC (16% of variance) showed more moderate changes between fresh and fixed libraries. In contrast, PCA

## Establishment of methods to assess genome-wide chromatin accessibility

analysis in CD4<sup>+</sup> showed fixed to present the largest differences versus fresh in the genome-wide accessibility (Figure 3.18B).

To further compare chromatin accessibility across the three conditions, the normalised read counts at each consensus list of peaks were used to contrast fresh vs fixed or frozen. The majority of the regions showed highly correlated ATAC-seq normalised counts between fresh and frozen or fixed, with the lowest correlation ( $R=0.918$ ) found between fresh and frozen CD14<sup>+</sup> monocytes (Figure 3.19A, B, C and D).



**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 (A) and (C) fresh versus fixed or (B) and (D) fresh versus frozen samples. Pearson correlation coefficient ( $R$ ) is indicated.

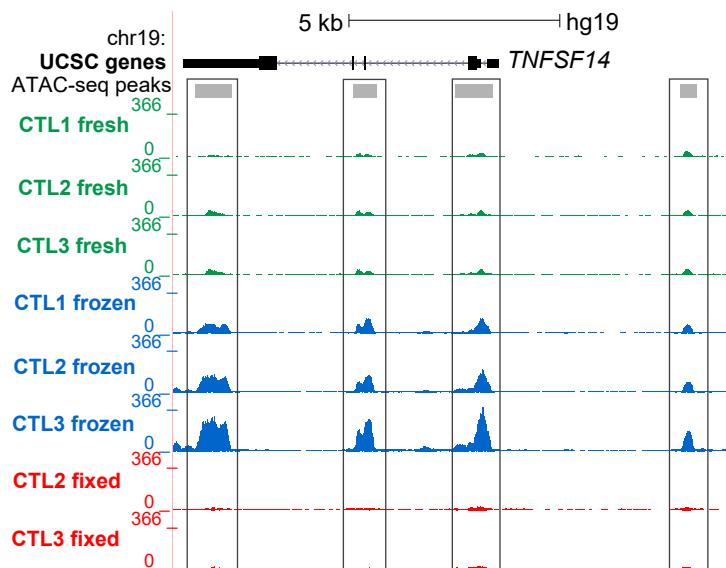
The regions showing lower correlation in mean counts were significant DARs (FDR<0.01 and fold change>1.5) when performing differential chromatin accessibility analysis between ATAC-seq fresh and ATAC-seq fixed or frozen using DESeq2. The number of significant DARs (FDR<0.01 and fold change>1.5) reported in each of the comparisons mirrored the PCA analysis results (Table 3.4). In CD14<sup>+</sup> monocytes, the largest number of DARs versus fresh were reported for the ATAC-seq frozen (5,269 regions in Figure 3.19B). Conversely, in CD4<sup>+</sup> the greatest differences in chromatin accessibility were found between fresh and fixed ATAC-seq samples (1,564 DARs in Figure 3.19D). In the performance of the differential analysis, the limited samples size (only two ATAC-seq fixed libraries) and the borderline quality of the ATAC-seq CD4<sup>+</sup> samples partially skewed the normalisation process, as can be observed by the location of the significant DARs in Figure 3.19D. Altogether, the number of DARs identified in each comparison did not involve more than 14.2% of the total ATAC-seq regions included in the analysis, representing a discrete proportion of the genome-wide accessible regions studied in this two cell types.

Cell type	Fresh vs Frozen	Fresh vs Fixed
CD14 <sup>+</sup> monocytes	5,269 (12.9%)	1,838 (4.5%)
CD4 <sup>+</sup>	282 (2.1%)	1,564 (14.2%)

**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.** Number of significant DARs (FDR<0.01 and fold change>1.5) identified using CP\_CD14\_all\_cond and CP\_CD4\_all\_cond filtered for background counts using 80% cut-off (optimal cut-off identified for this analysis, data not shown). In brackets are shown the percentage of DARs over the total number of regions included in the differential analysis.

An example of differences in chromatin accessibility between fresh and frozen ATAC-seq libraries in CD14<sup>+</sup> included four regions within and downstream of the *TNFSF14* gene (Figure 3.20. *TNFSF14* is the ligand for a receptor from the TNF-receptor superfamily. *TNFSF14* is involved in T cell activation, induction of apoptosis and also in bone destruction mediated by

monocytes and synovial cells interactions in RA. Three of the DARs were located at the promoter, an exon and the 3'UTR of the gene (respectively) and one was found at approximately 5Kb upstream the gene. All DARs were more open in ATAC-seq frozen libraries when compared to fresh and did not show any changes between fresh and fixed conditions.



**Figure 3.20: Differential chromatin accessibility at the TNFSF14 gene between ATAC-seq fresh and ATAC-seq frozen in CD14<sup>+</sup> 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<sup>+</sup> monocytes. The four DARs were more accessible in ATAC-seq frozen when compared to ATAC-seq fresh. ATAC-seq fixed was similar to ATAC-seq fresh at these four locations. Tracks are colour-coded by condition(green=fresh, blue=frozen and red=fixed).

### 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 experiment presented in this chapter. This enabled the establishment of a

pipeline and approach to be implemented for investigation of 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.

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. Successful profiles of DNA relative abundance for the ATAC libraries, showing 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. 2017), 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 of the effect of transposition times, one of the most critical variables in ATAC that can be cell type specific, beyond the conditions from Buenrostro's publication. 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). Assessment of three different transposition times in the ATAC-seq protocol showed heterogenous impact across cell types on the ratio of NFF/NBF and overall no major changes in the signal-to noise ratios.

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 $<4.1\times10^{-5}$ ) (Figure B.5A). Regarding signal intensity of the libraries, a trend for improved signal-to-noise when using Fast-ATAC was only found in CD14<sup>+</sup> monocytes and CD<sup>+</sup> 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 fold-enrichment in CD4<sup>+</sup> cells (unpaired Wilcoxon signed-ranked test p-values=0.024) and not a significant improvement for any of the remaining three haematopoietic cell types (Figure B.5B). In fact, Corces and colleagues Fast-ATAC publication only showed data demonstrating improved TSS by Fast-ATAC for CD4<sup>+</sup> 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<sup>+</sup> 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 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 list of peaks 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 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 the tissue represented a great opportunity to investigate the chromatin accessibility landscape at the main site of inflammation using ATAC. Variations of the ATAC-seq protocol (with different NP40 and Tn5 conditions) ATAC 1, ATAC 2 and Fast-ATAC were found to perform very poorly in keratinocytes isolated from skin biopsies and also in cultured NHEKs. The fact that similar results were obtained in keratinocytes isolated through different systems as well as in NHEKS indicated that the main reason for poor performance of these protocols was intrinsic to the cell type and not driven by compromising cell viability through the system used to isolate the cells. As differentiation progresses, keratinocytes synthesis an insoluble protein structure that progressively replaces the plasma membrane, which may have been impairing appropriate cell permeabilisation and efficient transposition. Interestingly, using increased Tn5 concentration to perform appropriately in NHEKS did not appear to improve ATAC quality libraries in the data presented. Similarly, the additional optimisation of the Fast-ATAC protocol modifying the concentration of detergent and Tn5 failed to improve the quality of the data.

Testing of the Omni-ATAC protocol in keratinocytes demonstrated substantial improvement in performance vs ATAC-seq and Fast-ATAC. The data was consistent with Corces *et al.*, 2017, where successful results in NHEKs were shown, and it encourages future testing of Omni-ATAC in keratinocytes from psoriasis patients biopsies processed through adherent assay to minimise the presence of dead cells. This opens an avenue to explore the chromatin landscape in lesional and uninvolved psoriasis biopsies that was not compatible with the time frame of this thesis and will be conducted as future work.

### **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. It was therefore of interest to test the ability of this fixative to perform well in ATAC (Attar et al. 2018). Alternatively cryopreservation of PBMCs had historically been used but formal assessment of the effect of this process in the chromatin landscape of different cell types had not yet been conducted.

DSP fixed samples presented overall lower abundance of NFF, when compared to the fresh and frozen libraries. DSP performed very poorly in the two cell types from CTL1, which presented extremely low abundance of NFF and predominance of NBF. Interestingly, despite the abundance of di-nucleosome fragments in these two samples, the chromatin structure across the TSS failed to reproduce the position of the TSS flanking nucleosome, which could be due to nucleosome displacement, as DSP does not cross-link DNA to proteins. Since this effect was only observed for CTL1, problems relating to inappropriate performance of the fixation protocol on that particular day may be the cause.

After removing the CTL1 samples, consideration of the chromatin accessibility landscape in all the remaining samples clearly showed that the differences by condition were greater than the differences between individuals. This was confirmed by performing differential analysis, which revealed a moderate number of DARs when comparing fixed or frozen to the fresh reference ATAC-seq. As expected, CD14<sup>+</sup>monocytes appeared to be more sensitive to cryopreservation than CD4<sup>+</sup> cells and present greater differences in the chromatin accessibility landscape when compared to fresh samples. This has been previously reported in the context of cytokine response, where

cryopreservation of monocytes and immature DCs have shown a skewed profile compared to fresh counterparts (Meijerink et al. 2011).

The results presented here could be limited by the borderline quality of some of the libraries, as previously explained, which was due to the issues of consistency intrinsic to the ATAC-seq protocol and could be distorting the findings to some extent. Nevertheless, this study still provides some useful information regarding the effect of using DSP in sorted cell populations or cryopreserved PBMCs. Depending on the sample size foreseen in each particular study, if the aim is to perform paired ATAC and scRNA-seq based on microfluidics methods in the same samples, and the final biological question, these two preservative protocols could be considered for implementation. Additionally, as the ATAC-seq protocol has been improved, ATAC has also been successfully conducted in frozen tissues and formaldehyde samples and can be considered when establishing the experimental design (Corces et al. 2017; Chen et al. 2016). In this thesis, 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 cells were used to generate the results presented in the following two chapters.

### **3.3.5 Limitations**

Several limitations have been described throughout this chapter. The main factor involves that many of the results have been generated in a limited number of samples, with low number of replicates, limited power to identify differences and, in some instances, lacking of 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 was first conducted in one replicate of each method and a comparison in a larger cohort with statistical evidence for the results could only be performed at the end of the thesis. Interestingly, the trends observed in the pilot comparison were confirmed when using the entire cohort of samples from the thesis, and showed not to agree with the claim made in the Fast-ATAC protocol publication, in which the protocol was supposed to improve the signal in all the hematopoietic cell types. This also highlights the importance of conducting experiments to reproduce published findings in a particular 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 (Leanne2009). A number of studies have identified systemic components of psoriasis, including an increase of circulating Th-17, Th-1 and Th-22 cells in patients' blood and an impaired inhibitory function of circulating Tregs (Kagami et al. 2010; Sugiyama et al. 2005). Activated T cells isolated from patients' blood have the ability to induce skin lesions in xenotransplantation models of psoriasis (Wrone-Smith and Nickoloff 1996; Nickoloff and Wrone-Smith 1999). Psoriasis patients also present increased risk for PsA following skin lesions as well as other co-morbidities, such as CVD (Ibrahim et al. 2009; Shapiro et al. 2007). Increased levels of cytokines including TNF- $\alpha$ , IFN- $\gamma$  and IL-2, amongst others, have also been reported in psoriasis sera in a large case control

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

---

study (Bai et al. 2018). 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 retina 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 is of

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

---

particular interest here. 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 SF mCD4<sup>+</sup> 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).

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

**Table 4.1: Summary table of the most comprehensive transcriptional studies in psoriasis skin and blood.** SB= whole 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 Technology discrepancies
(Li et al. 2014)	SB (L=92, C=82)	microarray RNA-seq	and Technology discrepancies and lncRNAs targets
(Keermann et al. 2015)	SB (L=12, U=12, C=12)	microarray RNA-seq	co-regulation Dormant psoriasis signature and <i>IL36</i> expression in psoriasis skin
(Tsoi et al. 2015) (Swindell and Remmer 2015)	SB (L=97, U=29, C=90) SB (L=14, U=14)	RNA-seq RNA-seq and mass-spectrometry RNA-seq	Psoriatic skin-specific new lncRNAs 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)	RNA-seq	NOD-like and inflammasome pathways
(Coda et al. 2012)	PBMCs (PS=6, C=5) and SB	Microarray	Partial overlap between PBMCs and skin DEGs
(Lee et al. 2009) (Mesko et al. 2010)	(L=5, U=5) PBMCs (PS=5, C=8) PBMCs (PS=15, IBD=12, RA=12, C=18)	Microarray TaqMan customised array (96 genes)	202 DEGs, circulating gene expression signature 6 psoriasis-specific DEGs

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

---

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

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

---

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. 2015).

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 2011; Li et al. 2014; Tsoi et al. 2015). LncRNAs expression has also proven a role in psoriasis pathophysiology, with approximately 1,000 species differentially expressed between lesional and uninvolved skin (Tsoi et al. 2015). 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 and Remmer 2015).

The majority of the transcriptional studies have been performed in whole skin biopsies containing a mix of tissues from the epidermis, dermis, basal layer, 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 whole skin biopsies (Swindell et al. 2017; Tervaniemi et al. 2016). These results reinforce the importance of using homogenous tissue and cell type samples to better dissect the altered biological processes contributing to the development of psoriasis at the site of inflammation.

### **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

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

---

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 versus 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 ( $\alpha$ ,  $\beta$  and  $\gamma$ ) 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  $\alpha$ - compared to  $\beta$  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

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

---

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 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

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

---

array (Pasaniuc and Genetics 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 isolate four relevant immune cells types (CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup>) 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 .

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

---

**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)	—	—	—	—	—	—

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

---

The mean age of the cohort was 55 years old and the mean disease duration 331.5 months. All patients presented 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 $\geq$ 10 (Woolacott et al. 2006; Finlay 2005). In this cohort, six out of ten patients had PASI $\geq$ 10, and so were categorised as having severe psoriasis and two had PASI<7 showing a mild 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. Patients PS1011, PS2015, PS2001 and PS2314 presented nail pitting, which has been defined as one of the markers for increased risk of developing joint affection 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 10 sex and age-matched healthy individuals (Table 4.3).

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

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

Sample ID	Sex	Age (years)	ATAC	ChIPm	RNA-seq
<b>Cohort 1A</b>					
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
<b>Cohort 1B</b>					
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.

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 whole skin biopsies, with a more heterogeneous cell type composition compared to epidermis, which could not be expanded due to time constrains.

### 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 ???. 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 non-redundant fraction and PCR

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

---

bottlenecking coefficients PBC1 and PBC2 (see 2.4.2), most of the libraries passed these filtering criteria (Table A.3). Only CD8<sup>+</sup> CTL7 and the CD19<sup>+</sup> PS2000 and PS2314 libraries failed the recommended complexity according to non-redundant fraction 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 measured by cross correlation analysis NSC and RSC coefficients (above 1.05 and 0.8, respectively), following ENCODE standards (Landt et al. 2012), with CD8<sup>+</sup> CTL7 and the CD19<sup>+</sup> PS2000 and PS2314 libraries being in the lower based on the RSC threshold.

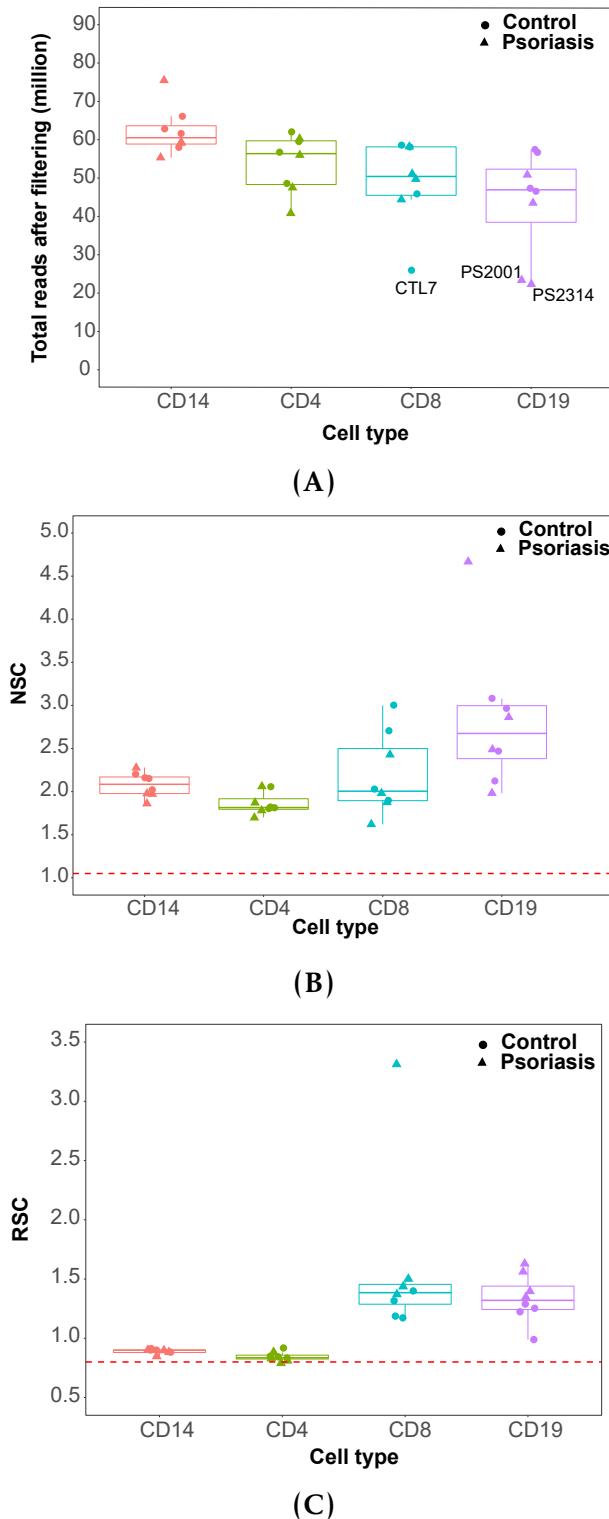
Due to the small sample size and the potential effect of borderline quality control samples in the differential analysis, CD8<sup>+</sup> CTL7 and the CD19<sup>+</sup> PS2000 and PS2314 libraries were removed from downstream analysis based on low library complexity (PBC1) and borderline signal-to-noise (RSC) (Figure 4.1B and C) and Table A.3).

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). PCA per cell type identified PS2314 CD8<sup>+</sup> as an outlier not removed based on the previous quality control analysis. This sample was not included in the downstream differential analysis to avoid the impact that outliers can have when working with small sample size.

### **H3K27ac differential analysis**

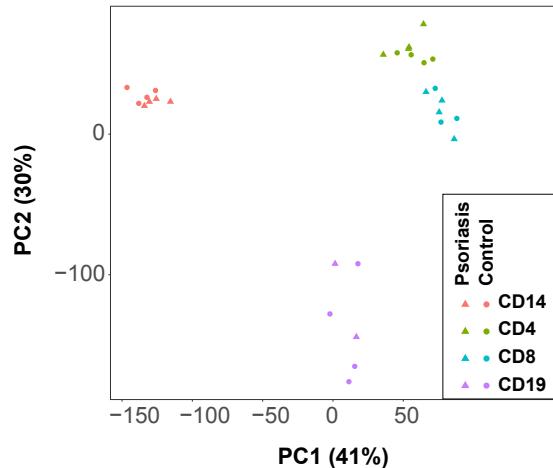
An exploratory analysis was conducted to find differences in H3K27ac patterns between patients and controls using DiffBind in each cell type. DiffBind

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

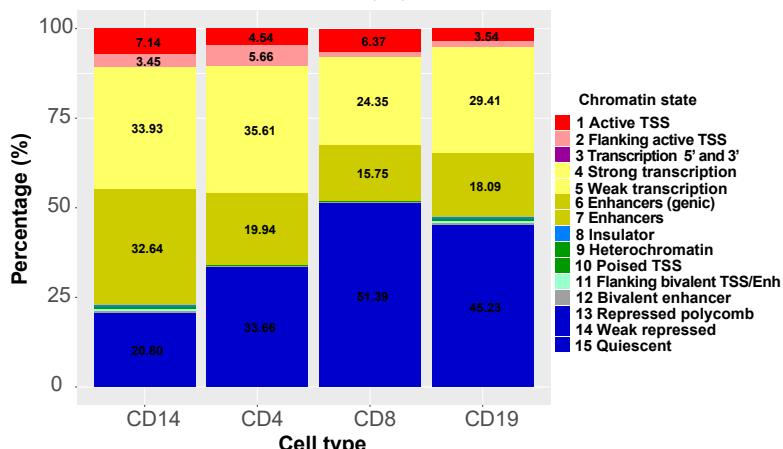


**Figure 4.1: Quality control evaluation of the H3K27ac ChIPm libraries in immune cells isolated from psoriasis and control samples.** For each of the cell types boxplots representing (A) million of reads after filtering, (B) normalised strand cross-correlation coefficient (NSC) and (C) relative strand 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 (C) and d) 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).

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



(A)



(B)

**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<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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.

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

---

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 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<sup>+</sup> monocytes to 51.39% in CD8<sup>+</sup> 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<sup>+</sup> monocytes had the greatest number of differentially modified enhancers (8 significant sites), followed by CD4<sup>+</sup> (4) and CD8<sup>+</sup> (1) (Table 4.4).

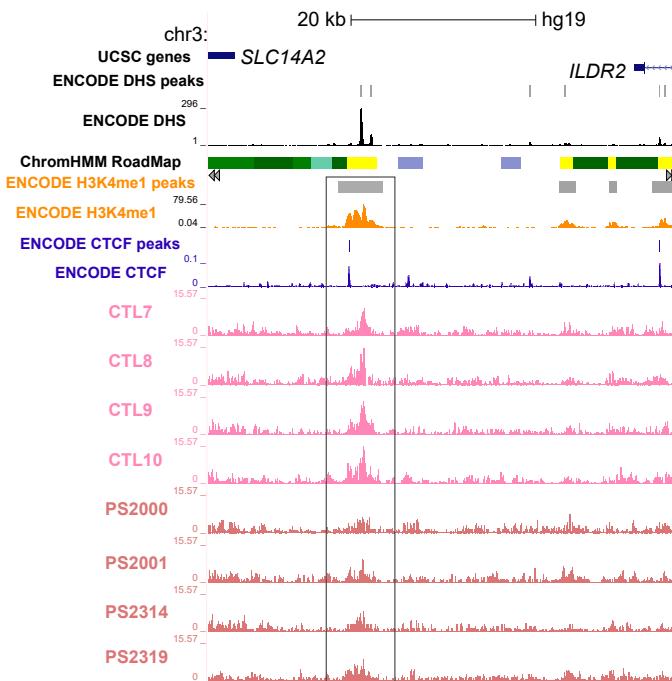
Cell type	Differential H3K27ac genome-wide analysis	Differential H3K27ac enhancers analysis
CD14 <sup>+</sup>	15	8
CD4 <sup>+</sup>	0	4
CD8 <sup>+</sup>	8	1
CD19 <sup>+</sup>	12	0

**Table 4.4: Summary results from the differential H3K27ac analysis between psoriasis patients and healthy controls in CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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<sup>+</sup> monocytes and CD8<sup>+</sup> 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 versus controls, several H3K27ac differentially modified regions were proximal to genes relevant in chronic inflammation. For example, a differential H3K27ac region in CD14<sup>+</sup> 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 enhancer by the Roadmap Epigenomics Project chromatin

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

segmentation map. Additionally this site was overlapping a DHS and H3K4me1 (enhancer mark) modification and a CTCF-binding site in K562 cells.



**Figure 4.3: Differential H3K27ac modification at a putative intergenic enhancer region in circulating CD14<sup>+</sup> 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<sup>+</sup> monocytes (lower H3K27ac enrichment in psoriasis patients compared to healthy controls). CD14<sup>+</sup> 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<sup>+</sup> monocytes and CD8<sup>+</sup> 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

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

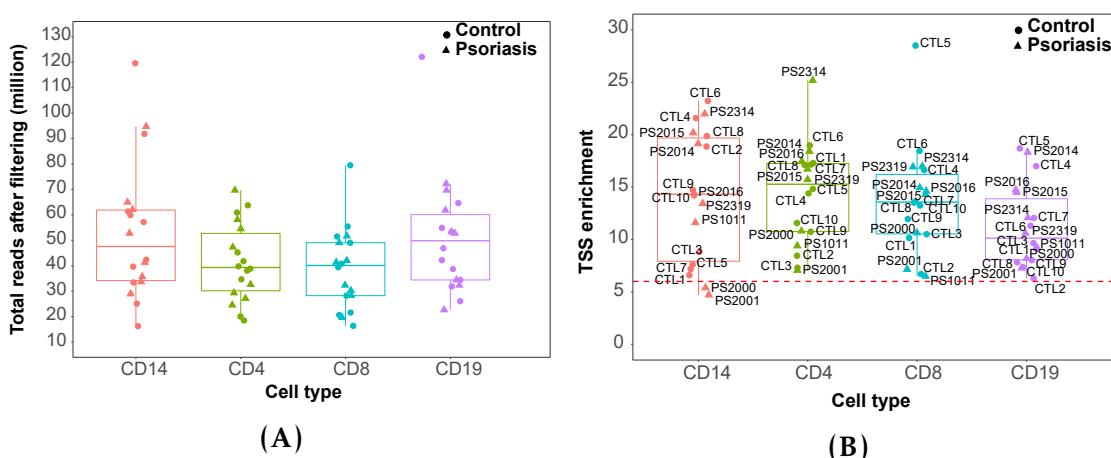
cell types. Altogether, the results presented in this pilot cohort did not show relevant global epigenetic changes in H3K27ac sites between psoriasis patients and controls for these cell types and sample size.

#### 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 (Table ??) 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).



**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.

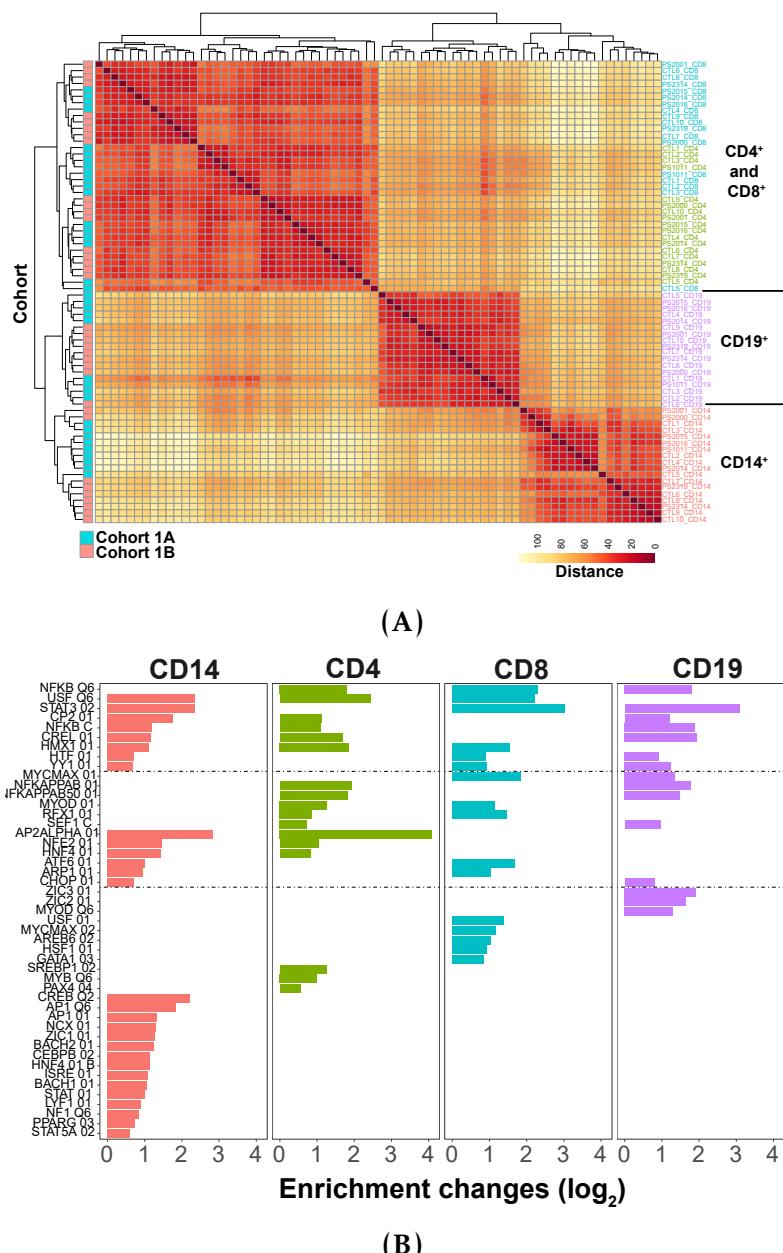
## **Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis**

---

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<sup>+</sup> PS2000 and PS2001 monocytes showed TSS<6 and were removed from downstream analysis together with CD19<sup>+</sup> CTL2, which TSS was borderline. Regarding the number of called peaks, most samples showed between 10,000 and 35,000 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<sup>+</sup> 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 heatmap illustrating sample distance using 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<sup>+</sup> monocytes, CD19<sup>+</sup> and CD4<sup>+</sup>/CD8<sup>+</sup> T cells (Figure 4.5A). Within each cell type cluster, samples did not separate based on disease condition, suggesting an absence of major global differences in the chromatin accessibility landscape between psoriasis patients and healthy controls. Within the cell types ,some grouping of ATAC samples by batch was observed (Figure 4.5A).

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



**Figure 4.5: Clustered heatmap and conserved TF binding sites enrichment analysis in the consensus list of called ATAC peaks in CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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 4 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.

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

---

PCA analysis revealed batch effect between cohort 1A and cohort 1B samples, likely driven by the use of ATAC-seq and Fast-ATAC protocols, respectively (representative example in CD8<sup>+</sup> cells Figure B.7A). This analysis also revealed CTL5 as a cohort 1A outlier sample for all the cell types and therefore it 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 4 cell types (named here as CP\_CD14, CP\_CD4, CP\_CD8, and CP\_CD19) (detailed in Chapter 2). Each cell type consensus list of called peaks was significantly enriched ( $FDR < 0.01$ ) for conserved TF binding sites (Figure 4.5B). For example, enrichment of conserved NF $\kappa$ B binding motifs was identified across each cell type 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 et al. 2013). Similarly, peaks from each cell type consensus list also showed enrichment for relevant cell type *cis*-eQTLs. For example, eQTLs from unstimulated and stimulated (LPS or IFN- $\gamma$ ) monocytes were the most significantly enriched ( $FDR < 0.01$ ) in the CP\_CD14 (unstimulated fold-enrichment 5.1, LPS 2h fold-enrichment 4.7 and IFN- $\gamma$  fold-enrichment 5.0 from Fairfax et al., 2014 data) when compared to other eQTL datasets. In summary, this observed enrichment of eQTL SNPs and conserved TF binding sites demonstrated functional relevance of the accessible chromatin regions further investigated for differential analysis.

The differential analysis between patients and controls was performed on the ATAC normalised read counts of each cell type consensus list using DESeq2

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

---

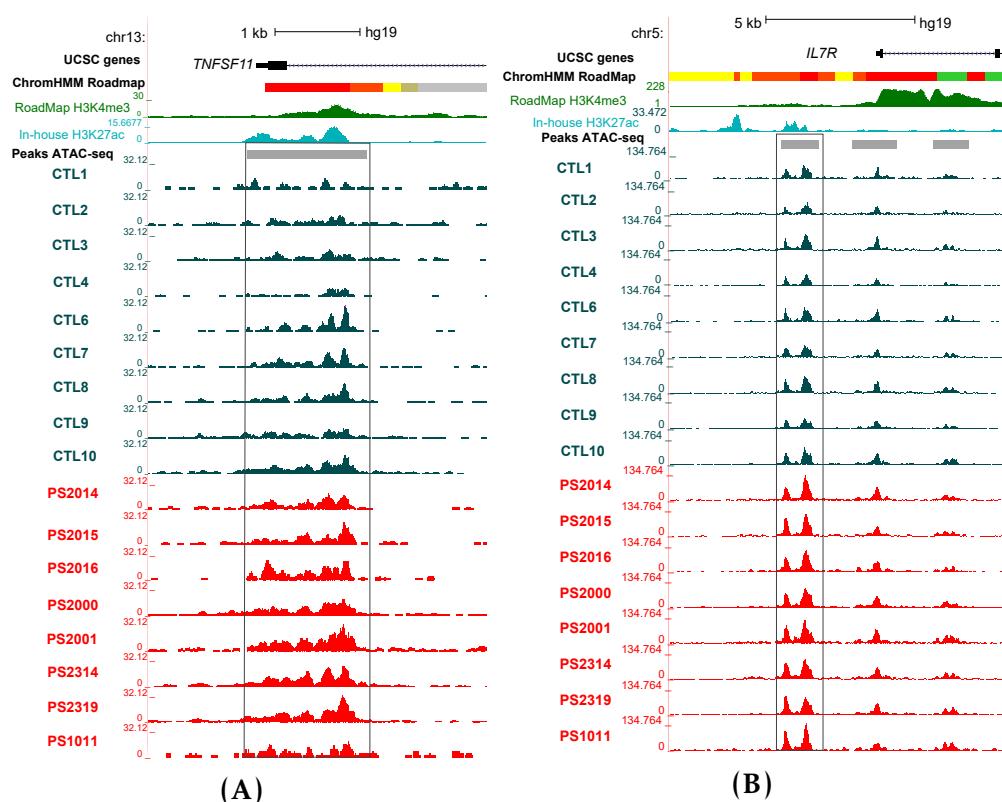
and including the ATAC protocol as a covariate to account for the aforementioned batch effect. Genome-wide differential chromatin accessibility analysis in CD8<sup>+</sup> cells revealed 55 significant (FDR<0.05) differentially accessible regions (DARs) between psoriasis patients and healthy controls (Table 4.5), of which 17 showed FDR<0.01. In contrast, CD14<sup>+</sup> monocytes, CD4<sup>+</sup> and CD19<sup>+</sup> cells only showed one or no DARs.

Cell type	Number of DARs FDR<0.05
CD14 <sup>+</sup>	1
CD4 <sup>+</sup>	0
CD8 <sup>+</sup>	55
CD19 <sup>+</sup>	1

**Table 4.5: Summary results from the differential chromatin accessibility analysis between psoriasis patients and healthy controls in CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> T cell eRNA data from the FANTOM5 project. Of the CD8<sup>+</sup> DARs, 8 overlapped significantly expressed eRNAs in the same cell type. These included a region at the TSS of the *TNSF11* gene and another upstream of the *IL7R* promoter, which were more accessible in the 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.

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



**Figure 4.6: Epigenetic landscape at two ATAC differential accessible regions between patients and controls in CD8<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> 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.

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

---

Other interesting CD8<sup>+</sup> DARs were found nearby *MAP3K7CL* and *NFKB1* 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 co-localisation for one region in CD8<sup>+</sup> 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 in the Roadmap Epigenomics Project segmentation map and did not interact with the promoter of any gene in CD8<sup>+</sup> cells according to Hi-C and promoter Hi-C (Javierre et al. 2016). Interestingly, SNPs within this region are reported to be eQTL for *DTD1* in whole blood (<https://gtexportal.org/home/eqtls>).

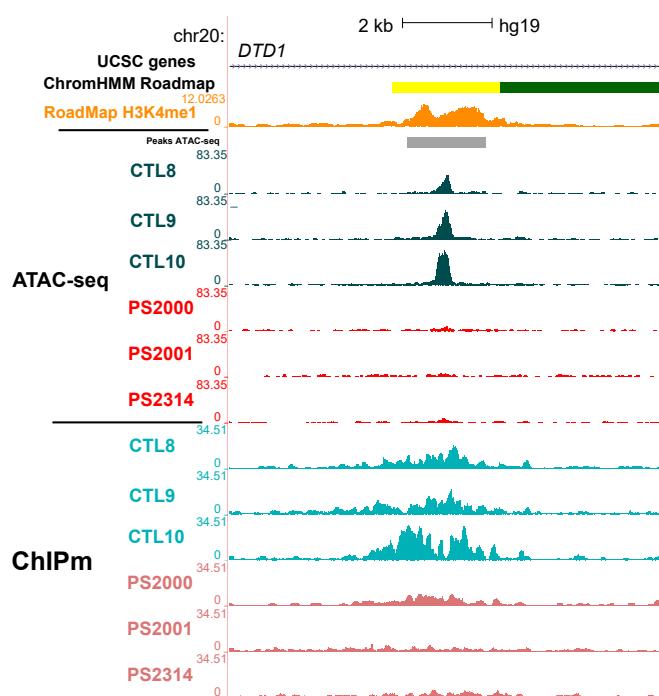
#### **4.3.4 Differential gene expression analysis in circulating immune cells in psoriasis**

##### **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. 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

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

---

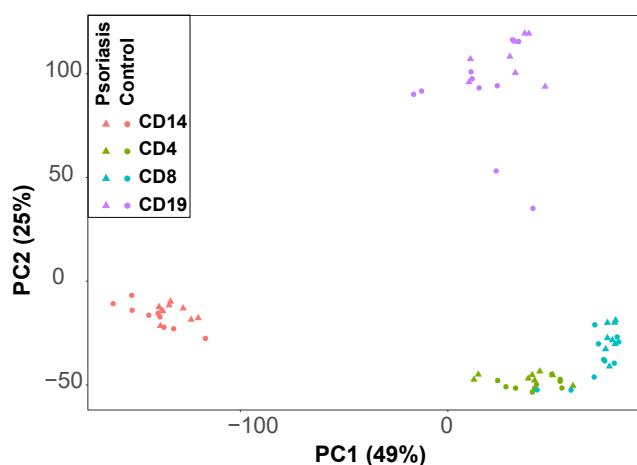


**Figure 4.7: Epigenetic landscape at *DTD1* locus showing differential H3K27ac and chromatin accessibility between psoriasis patients and controls in CD8<sup>+</sup> cells.** UCSC Genome Browser view illustrating the normalised ATAC read density and H3K27ac normalised fold-enrichment (y-axis). Tracks 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<sup>+</sup> cells are also shown.

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

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<sup>+</sup> monocytes, CD4<sup>+</sup> and CD8<sup>+</sup> T cells, and CD19<sup>+</sup> 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).



**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.

## 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 as a covariate to account for the aforementioned batch effect. For each cell type, differentially expressed mRNAs were identified using two thresholds: FDR <0.05 and FDR <0.01 (Table 4.6).

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

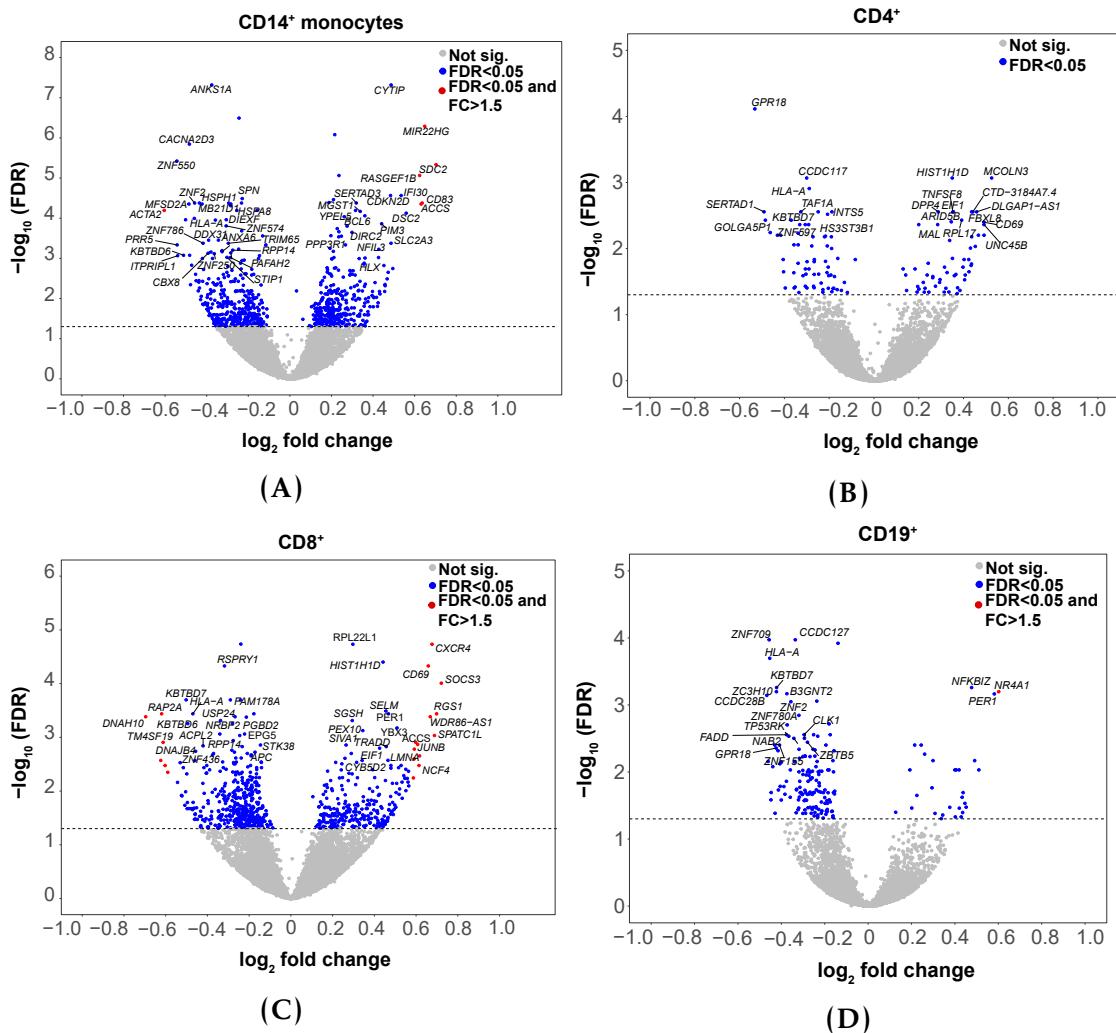
Cell type	mRNA FDR<0.05/0.01	lncRNA FDR<0.05/0.01	Up-regulated FDR<0.05/0.01	Down-regulated FDR<0.05/0.01
CD14 <sup>+</sup>	671/229	28/8	331/112	368/125
CD4 <sup>+</sup>	108/40	12/4	56/20	64/24
CD8 <sup>+</sup>	651/175	31/5	269/67	418/113
CD19 <sup>+</sup>	167/71	6/2	29/13	144/60

**Table 4.6: Summary results from the differential gene expression analysis between psoriasis patients and healthy controls in CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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 thresholds the number of up- and down-regulated genes is included.

CD14<sup>+</sup> monocytes and CD8<sup>+</sup> showed the largest number of differentially expressed mRNAs between psoriasis patients and controls with the highest magnitude of fold change (Figure 4.9). **CD14<sup>+</sup> monocytes and CD4<sup>+</sup> T cells showed similar numbers of genes up-regulated and down-regulated in psoriasis patients when compared to the healthy controls. In contrast, in CD8<sup>+</sup> and CD19<sup>+</sup> cells a larger number of modulated genes were down-regulated in patients compared to controls.** In CD14<sup>+</sup> monocytes, the most significant up-regulated genes (fold change>1.5) included *SDC2*, *CD83* and *MIR22G* (Figure 4.9A). In CD8<sup>+</sup>, the top differentially expressed genes with fold change>1.5 included *SOCS3* and *CXCR4* (Figure 4.9C), whereas in CD4<sup>+</sup> 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 patients vs controls in four analysed cell types was moderate.

Enrichment of significant DEGs (FDR<0.05) amongst genes associated with psoriasis from GWAS was analysed using the NHGRI-EBI catalog (<https://www.ebi.ac.uk/gwas>) (Table 4.7). CD8<sup>+</sup> had the largest number of DEGs (7 hits) overlapping putative GWAS genes, followed by CD14<sup>+</sup> monocytes and CD4<sup>+</sup> (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

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



**Figure 4.9: Magnitude and significance of the gene expression changes between psoriasis patients and healthy controls in four immune cell types.** Volcano plots illustrating the results of the differential gene expression analysis in (A) CD14<sup>+</sup> monocytes, (B) CD4<sup>+</sup>, (C) CD8<sup>+</sup> and (D) CD19<sup>+</sup> cells. For each gene, the log<sub>2</sub>fold change represents the change in expression for that 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.

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

---

*NFKBIZ*. In CD4<sup>+</sup> and CD8<sup>+</sup> T cells, these overlaps represented a statistically significant enrichment of GWAS genes amongst the differentially expressed genes in patients when compared to controls (Fisher exact test, p-values 2.6x10<sup>-3</sup> and 9.36x10<sup>-4</sup>, respectively).

Cell type	Number of GWAS overlaps	Up-regulated genes	Down-regulated genes
CD14 <sup>+</sup>	3	<i>NFKBIA</i>	<i>IL23A, FASLG</i>
CD4 <sup>+</sup>	3	<i>TNFAIP3, NFKBIZ</i>	<i>FASLG</i>
CD8 <sup>+</sup>	7	<i>TNFAIP3, NFKBIA, ETS1, SOCS1, NFKBIZ</i>	<i>B3GNT2, FASLG</i>
CD19 <sup>+</sup>	2	<i>NFKBIZ</i>	<i>B3GNT2</i>

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

### The role of lncRNAs in psoriasis circulating immune cells

In addition to protein coding genes, some of the DEGs identified were classified as lncRNAs. CD8<sup>+</sup> and CD14<sup>+</sup> monocytes were the two cell types showing the largest number of dysregulated lncRNAs between psoriasis patients and controls (Table 4.6).

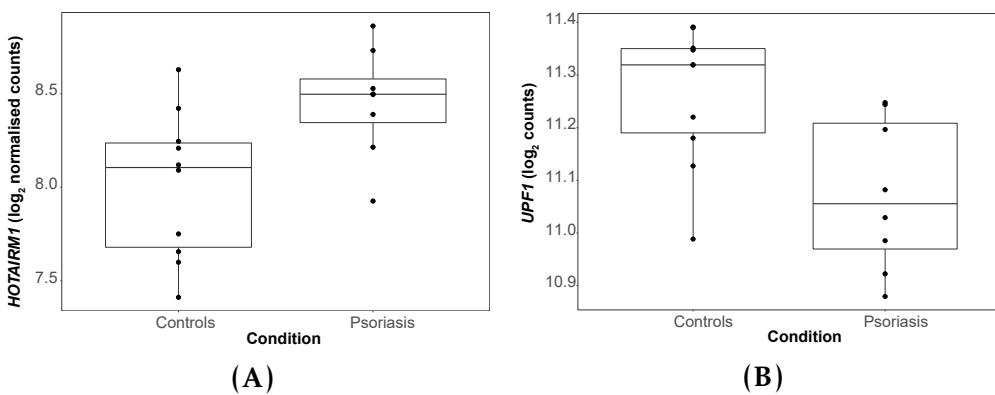
The majority of the differentially expressed lncRNAs (FDR<0.05) were found to have a functional interacting partner based on NPInter database (Table A.4), which retrieves functional interactions between non-coding RNAs and biomolecules (proteins, RNAs and DNAs) which have been published (Hao et al. 2016). Some of the lncRNAs were also found to be differentially expressed in PBMCs when comparing PsA patients versus healthy controls in a study conducted by Dolcino and colleagues (Dolcino et al. 2018).

Examples of cell type-specific differentially expressed lncRNAs between psoriasis patients and healthy controls included *DYNLL1-AS1* (or *NAV*), *HOTAIRM1* and *NEAT1* in CD14<sup>+</sup> monocytes, *DLGAP1-AS1* in CD4<sup>+</sup>, *UBXN8*

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

and *MIR146A* in CD8<sup>+</sup> and *LINC00324* in CD19<sup>+</sup>. Only one lncRNA, *RP11-218M22.1* appeared to be dysregulated between psoriasis and healthy controls in all the cell types.

Amongst the cell type-specific dysregulated lncRNAs, *DYNLL1-AS1* has been shown to affect the histone modifications of some critical IFN-stimulated genes (ISGs) including *IFITM3* and *MxA*, leading to down-regulation of their expression (Ouyang et al. 2014). *DYNLL1-AS1* was down-regulated in CD14<sup>+</sup> monocytes from psoriasis patients when compared to controls but no up-regulation of *IFITM3* and *MxA* was found. Conversely, *HOTAIRM1*, another CD14<sup>+</sup> monocytes specific lncRNA was up-regulated in psoriasis patients (Figure 4.10A). The experimentally validated target for *HOTAIRM1* reported by NPInter database was the RNA helicase and ATPase *UPF1* (Hao et al. 2016), which was found to be down-regulated in CD14<sup>+</sup> monocytes in psoriasis versus control samples in this data (Figure 4.10B). Lastly, *NEAT1*, was also up-regulated in psoriasis patients compared to controls in CD14<sup>+</sup> monocytes in line with a study in SLE CD14<sup>+</sup> monocytes (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<sup>+</sup> monocytes.** Expression is illustrated as the log<sub>2</sub> 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.

For CD8<sup>+</sup> cells, the miR *MIR146A* was the most interesting non-coding RNAs, showing down-regulation in psoriasis patients compared to controls and

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

previously reported to have a role in negative regulation of the innate immunity, inflammatory response and antiviral pathways (Taganov et al. 2006).

### Pathway enrichment analysis for the DEGs

To investigate the biological role of the significantly modulated genes, pathway enrichment analysis was performed for each cell type using DEG with FDR<0.05 and no fold change cut-off. Biologically relevant pathways were significantly enriched (FDR<0.01) for CD14<sup>+</sup> monocytes and CD8<sup>+</sup> cells (Table 4.8 and A.5). In CD19<sup>+</sup> cells, only one pathway (generic transcription, p-value=8.2x10<sup>-12</sup>, fold change=6.26) was significantly enriched, while in CD4<sup>+</sup> cells no enrichment was seen for any pathways.

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

**Table 4.8: Pathways enriched for DEGs between psoriasis patients and healthy controls in CD14<sup>+</sup> monocytes and CD8<sup>+</sup> 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.

Two of the significant enriched pathways, MAPK signalling and IL-12 mediated signalling, were found to be enriched in both CD14<sup>+</sup> monocytes and CD8<sup>+</sup> cells (Table 4.8). MAPK signalling was of particular interest as DEGs contributing to the enrichment of this pathway in both cell types included

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

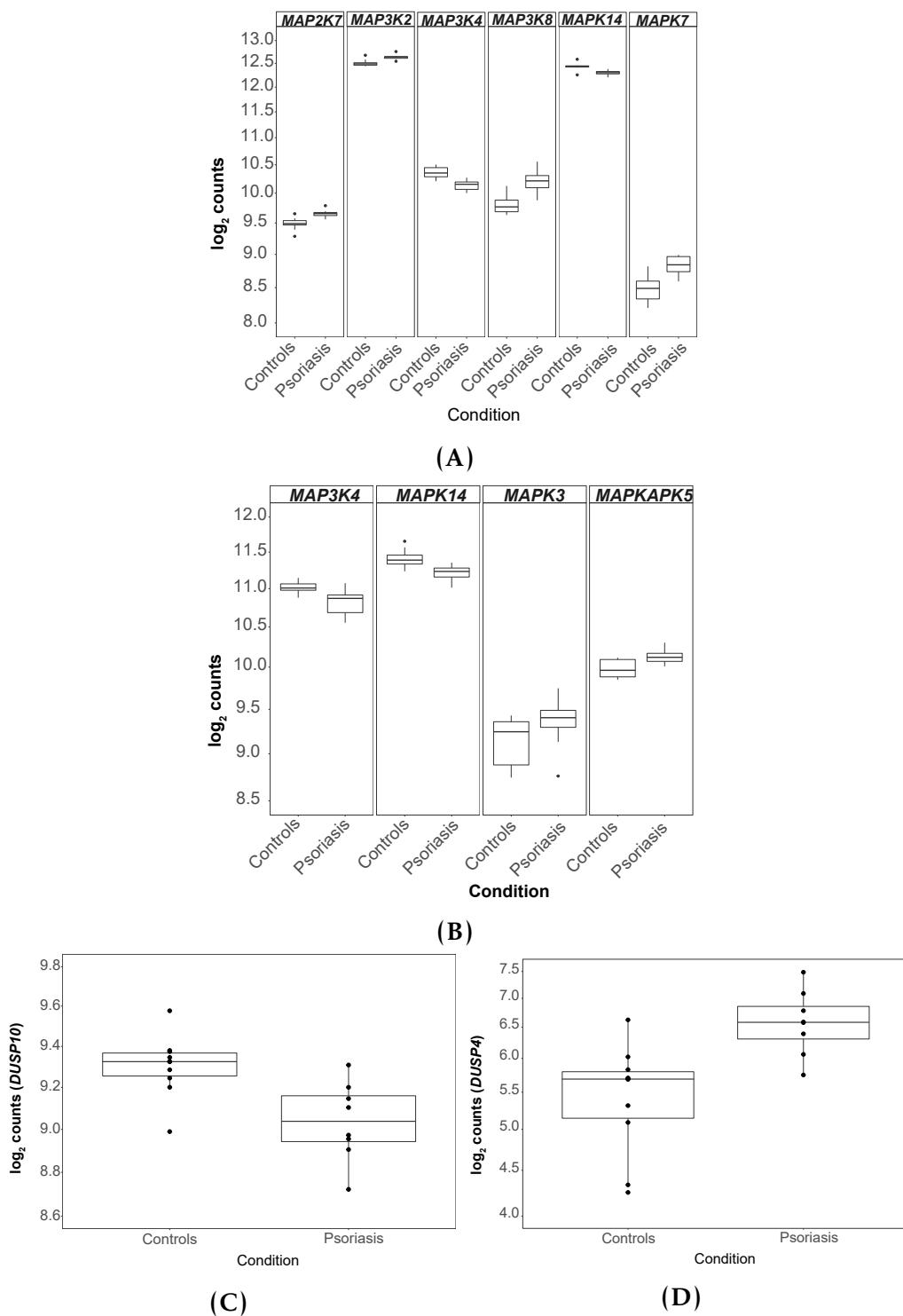
---

a number of MAPK gene members differentially expressed between psoriasis patients and healthy controls (Figure 4.11). Two of those MAPK genes, *MAP3K4* and *MAPK14*, were down-regulated in psoriasis vs controls in CD14<sup>+</sup> monocytes and CD8<sup>+</sup> 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 Crohn's 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 the immune response (Qian et al. 2009), also contributed to the enrichment of the MAPK pathway in CD14<sup>+</sup> monocytes and CD8<sup>+</sup> cells. *DUSP10* was down-regulated in psoriasis only in CD14<sup>+</sup> monocytes (Figure 4.11C) and knock-out in mice has been associated with enhanced inflammation (Qian et al. 2009). Conversely, *DUSP4* was up-regulated in psoriasis CD8<sup>+</sup> cells vs healthy controls (Figure 4.11D) and has been demonstrated to have a pro-inflammatory role in a sepsis mice model (Cornell et al. 2010).

Regarding enrichment of the IL-12 signalling, CD14<sup>+</sup> monocytes from psoriasis showed down-regulation of *STAT4* and *STAT5A* in patients compared to controls. Neither *STAT4* or *STAT5A* were differentially expressed in CD8<sup>+</sup> cells. By contrast, *IFNG* expression in psoriasis patients was lower than in healthy controls in CD8<sup>+</sup> cells but showed no differences in CD14<sup>+</sup> monocytes. Also, *IL2RA* only showed up-regulation in CD8<sup>+</sup> T cells from psoriasis patients which may have an effect in the formation of the IL2-R $\alpha$  and the signalling by this cytokine involved in effector and regulatory T cell differentiation (Malek and Immunity 2010).

The platelet-derived growth factor (PDGF- $\beta$ ) signalling pathway was only enriched in CD14<sup>+</sup> monocytes (Table 4.8). Within this pathway the *SLA* gene was down-regulated in psoriasis patients. A *SLA* knock-out mice model has shown

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



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

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

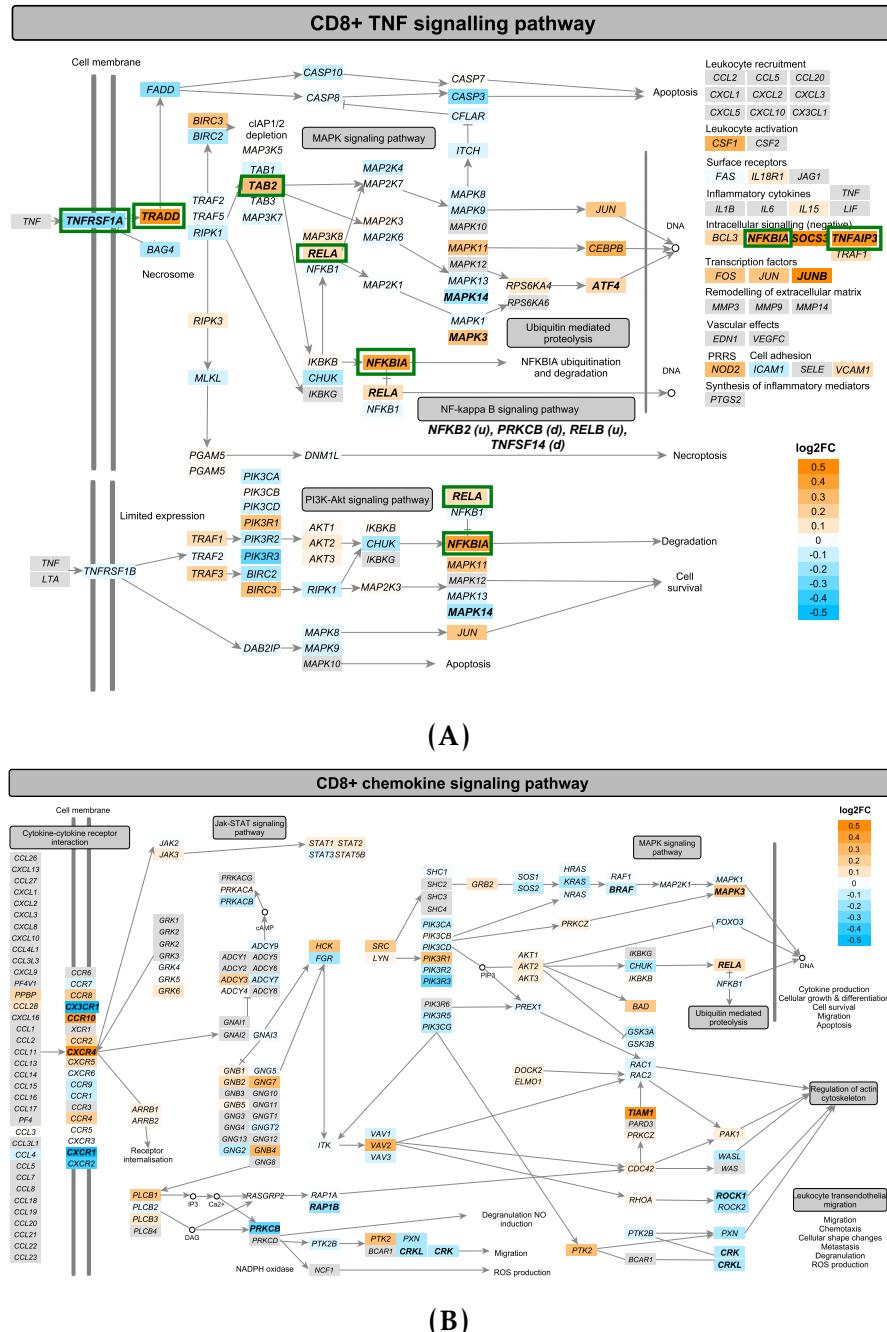
---

impaired IL-12 and TNF- $\alpha$  and failure of T cell stimulation by GM-CSG treated bone marrow-derived DCs (Liontos et al. 2011).

A number of very relevant inflammatory pathways in psoriasis were enriched only in CD8 $^{+}$  cells. These included TNF, NF- $\kappa$ B and chemokine signalling (Table 4.8), with some of the DEGs contributing to the enrichment of more than one of the pathways. NF- $\kappa$ B inhibitor A (*NFKBIA*) and the TNF- $\alpha$  induced protein 3 (*TNFAIP3*) were unexpectedly up-regulated in CD8 $^{+}$  cells from psoriasis compared to healthy controls. *NFKBIA* up-regulation contributed to the enrichment of all three pathways (Figure 4.12A in green box and B) and *TNFAIP3* was a member of the TNF and NF- $\kappa$ B pathways (Figure 4.12A in green box). Moreover, *NFKBIA* and *TNFAIP3* are two of the psoriasis GWAS associated genes and were also up-regulated in psoriasis CD14 $^{+}$  monocytes and CD4 $^{+}$  cells, respectively (Table 4.7). The enrichment of these three pathways in CD8 $^{+}$  cells was also driven by a mix of up-regulation (*JUNB*, *RELA*, *RELB* and *NFKB2*) and down-regulation (*ATF2* and *ATF4*) of pro-inflammatory genes, which effect on the overall immune response is difficult to predict.

Up-regulation and down-regulation of specific chemokines was seen in CD8 $^{+}$  cells from psoriasis patients when compared to healthy controls (Figure 4.12B), including up-regulation of *CCR10*, the receptor for the chemotactic skin-associated chemokine CCL27. Other up-regulated chemokine receptors included *CXCR4* gene, the receptor for the chemokine SDF-1, highly expressed in skin (Zgraggen et al. 2014). Of note, none of the genes coding for well-known psoriasis drug target genes, including TNF- $\alpha$ , IL-17 and IL-6, were up-regulated in any of the 4 cells types from psoriasis patients compared to healthy controls.

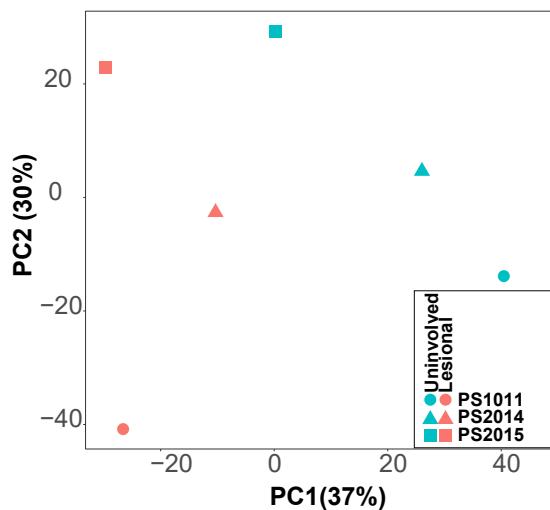
# Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis



#### 4.3.5 RNA-seq in epidermis from psoriasis patients

##### Data processing and quality control

For the three paired uninvolved-lesional samples (Table ??), all had a mapping rate greater than 80% (Figure B.9A). The number of reads after filtering mapped to Ensembl genes ranged between 29.5 and 33.2 million (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).



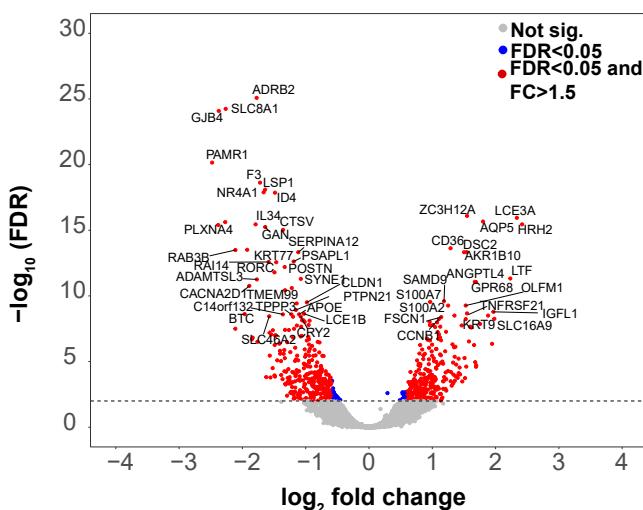
**Figure 4.13: PCA for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients.** First and second component of the PCA analysis performed on the normalised number of reads mapping to the Ensembl list of mRNAs and lncRNAs detected in lesional and uninvolved epidermis isolated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

PCA using the normalised number of reads mapping to the genes after filtering revealed separation of the lesional samples from the uninvolved by the first PC, which explained 37% of the variance (Figure ??). The second PC explained 30% of the variance and correlated with the patient ID. Overall, PCA analysis revealed substantial variation between the lesional and uninvolved samples and biological variability across individuals, for which the paired design in the differential gene expression analysis accounted.

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

### Summary of the differential gene expression results

Differential gene expression analysis revealed a total of 1,227 ( $FDR < 0.05$ ) and 702 ( $FDR < 0.01$ ) genes differentially expressed between uninvolved and lesional epidermis skin biopsies, including mRNAs and lncRNAs (Table ??). Amongst the 1,227 DEGs, a similar proportion of genes up- (559 genes) and down-regulated (629) in lesional skin when compared to uninvolved were identified (Figure 4.14) and 46 were annotated as lncRNAs (Table ??). The magnitude of changes in gene expression between lesional and uninvolved skin were notably larger when compared to the changes in expression from analysis in peripheral blood, with 874 out of 1,227 genes showing fold change  $> 1.5$ .



**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.

Amongst the DEGs between the uninvolved and lesional skin, five genes ( $FDR < 0.05$ ) overlapped with putative GWAS genes (Table ??). *IFIH1*, *NOS2*, *LCE3D* and *STAT3* were up-regulated in lesional compared to uninvolved skin, whereas *TNFAIP3* was down-regulated. Only *IFIH1* failed to show differential expression at an  $FDR < 0.01$ .

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

---

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

**Table 4.9: Summary results of the differential gene expression analysis between uninvolved and lesional psoriatic epidermal biopsies.** Number of differentially expressed mRNAs and lncRNAs are reported for two threshold 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 also specified.

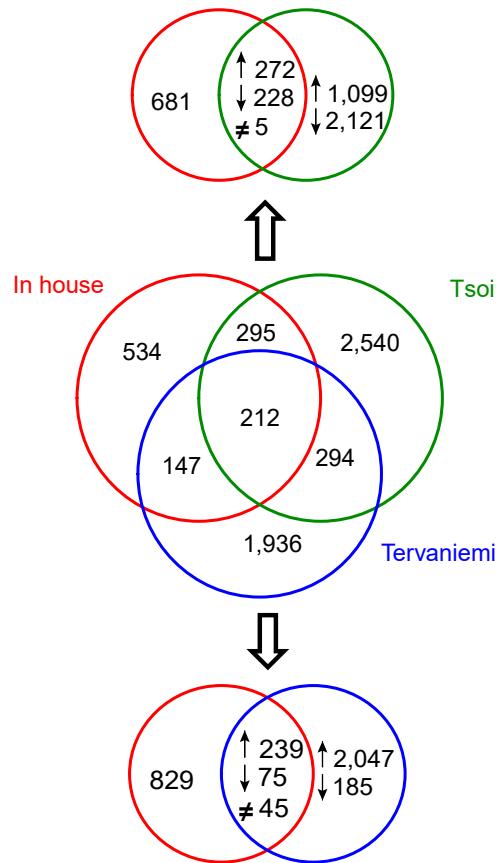
### Overall comparison with other skin transcriptomic studies

As detailed in Chapter 2, the approach to study differential gene expression in skin is different from previously published studies using whole punch biopsies to compare lesional and uninvolved skin from psoriasis patients. During the course of this project a study was published by Tervaniemi and colleagues that also aimed to characterise the transcriptional profiles of the epidermis from psoriasis patients lesional and uninvolved skin (Tervaniemi et al. 2016). In order to explore the similarities between the two studies, a comparison for the DEGs identified between lesional and uninvolved matched samples was conducted.

**Tervaniemi study showed a total of 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 in this thesis, and reported a larger number of up-regulated than down-regulated hits (Figure 4.15 bottom panel). A total of 359 out of the 1,227 DEGs (29.25%) identified by the in-house study 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**

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

studies but showed opposite direction. For example *SERPINB2* was down-regulated in the in-house data and up-regulated in the Tervaniemi results.



**Figure 4.15: Overlap of the significantly differentially expressed genes between lesional and unininvolved epidermal sheets, split epidermis and full-thickness skin biopsies.** The central venn diagram illustrated the DEGs overlapping between this study (in house), Tervaniemi *et al.*, 2016 split epidermis biopsies and Tsoi *et al.*, 2015 full thickness 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 the in-house data and the Tsoi *et al.* (top) or Tervaniemi *et al.* (bottom).

In addition to the Tervaniemi study, our results were further contrasted to one of the most recent comprehensive RNA-seq studies comparing lesional and unininvolved full thickness skin biopsies from psoriasis patients (Tsoi *et al.* 2015). Out of the 3,725 DEGs reported by Tsoi and colleagues, 507 genes were shared between the two studies (41% of the in-house DEGs) and 24 corresponded to dysregulated lncRNAs. Out of the 507 commonly dysregulated genes in the

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

---

two datasets, 272 were up-regulated, 228 down-regulated and 7 showed opposite direction of change (Figure 4.15 top panel).

Overlap across the three studies only identified 212 DEGs shared by the three datasets. Despite having a the larger sample size, the Tsoi and colleagues study did not capture all the DEGs from the in-house or Tervaniemi (506 overlapping genes, Figure 4.15 middle panel) data.

### **Dysregulated lncRNAs in the psoriatic lesional skin**

In addition to protein coding genes, a total of 46 lncRNAs were also significantly (FDR<0.05) differentially expressed between uninvolved and lesional skin in the three psoriasis patients from this study. Out of the 46 differentially regulated lncRNAs, 37 had a functional experimental partner functionally validated according to NPInter database (Hao et al. 2016). An interesting example was *H19* which was significantly down-regulated in the lesional skin when compared to uninvolved. *H19* has been described to directly bind miR-130b-3p, which down-regulates Desmoglein 1 (*DSG1*), a gene promoting keratinocyte differentiation (Li et al. 2017). Nevertheless, *DSG1* did not appear as one of the DEGs between lesional and uninvolved skin.

Interestingly, four miRNAs (*MIR146A*, *MIR22HG*, *MIR31HG* and *MIR205HG*) were also captured with the standard library preparation for mRNAs and lncRNAs implemented in our project. The relevance of miR-146a has been already noted in the differential analysis from circulating immune cells. In lesional skin *MIR146A* was up-regulated when compared to uninvolved skin, consistently with other studies (Lerman2014; Tsoi et al. 2015).

Another relevant finding was the up-regulation of *MIR31HG* in lesional skin, which has also been reported by (Tsoi et al. 2015). In a study in head and neck carcinoma, *MIR31HG* expression was identified to target *HIF-1A*, inducing up-regulation by an unknown mechanisms (Wang et al. 2018). In our data, *HIF-*

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

---

*1A* showed up-regulation in lesional skin compared to uninvolved, and a trend for positive correlation ( $R=0.690$ ,  $p\text{-value}=0.12$ ) between normalised counts of this gene and the putative regulator *MIR31HG* was also observed (Figure B.11).

### Pathway enrichment analysis

In order to better understand the functional role of the DEGs ( $\text{FDR}<0.05$ ) between lesional and uninvolved epidermis from psoriasis patient skin biopsies, pathways enrichment analysis was performed. A number of pathways were significantly enriched ( $\text{FDR}<0.005$  and fold change $>2$ ) for DEGs found in our analysis (Tables 4.10 and A.6).

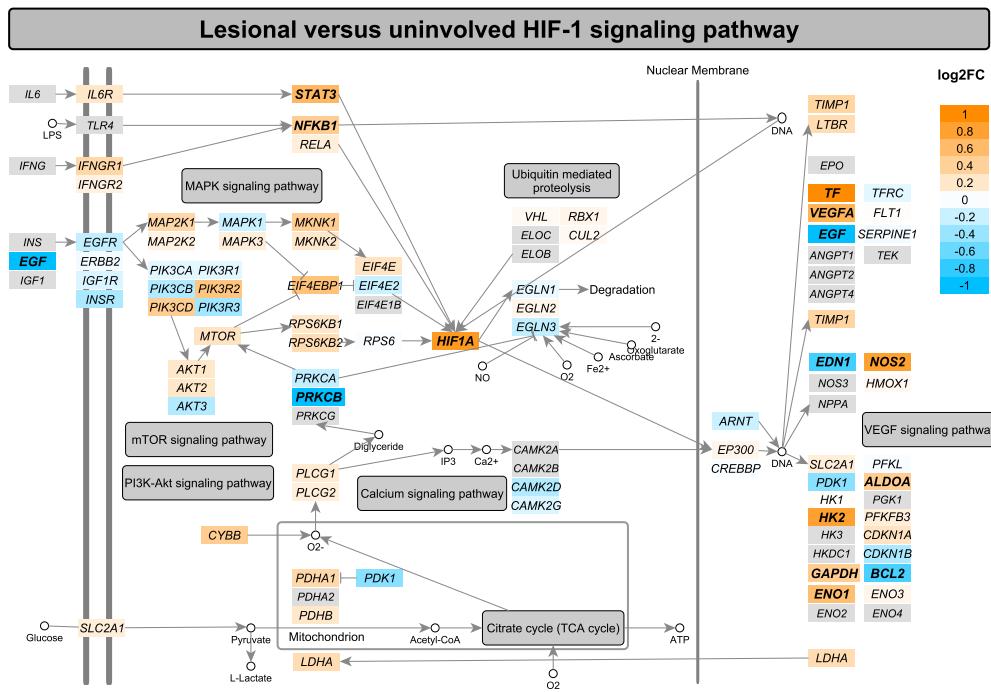
Pathway	FDR	Fold change
IFN- $\alpha/\beta$ /signalling	$5.4\times 10^{-5}$	3.65
Peroxisome proliferator-activated receptors (PPAR)	$1.5\times 10^{-4}$	3.42
NOD-like receptor signalling	$3.1\times 10^{-4}$	2.37
IL-17 signalling	$1.4\times 10^{-3}$	2.72
IL2-mediated signalling	$3.1\times 10^{-3}$	2.64
Hypoxia-inducible factor 1 (HIF-1) signalling	$5.0\times 10^{-3}$	2.31
Glycolysis/gluconeogenesis	$1.0\times 10^{-6}$	4.71
Cell cycle	$6.7\times 10^{-5}$	2.55
Apoptosis	$3.7\times 10^{-3}$	2.11
Arginine and proline metabolism	$7.5\times 10^{-6}$	5.28

**Table 4.10: Most relevant pathways enriched for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.** Significant pathways for  $\text{FDR}<0.005$  and fold change $>2$ . The analysis was performed using significantly DEGs ( $\text{FDR}<0.05$ ). Enriched pathways had a minimum of ten members overlapping with DEGs.

A number of pathways were related to alterations in cell cycle and metabolic processes, including hypoxia-inducible factor 1 (HIF-1) signalling, arginine and proline metabolism, glycolysis/gluconeogenesis and metabolism of carbohydrates. HIF-I signalling has been found to be up-regulated in psoriasis skin, likely through hypoxia caused by increased cell proliferation rates and epidermal thickening. In this data up-regulation of *HIF1A*, *VEGFA*, *ENO1* and

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

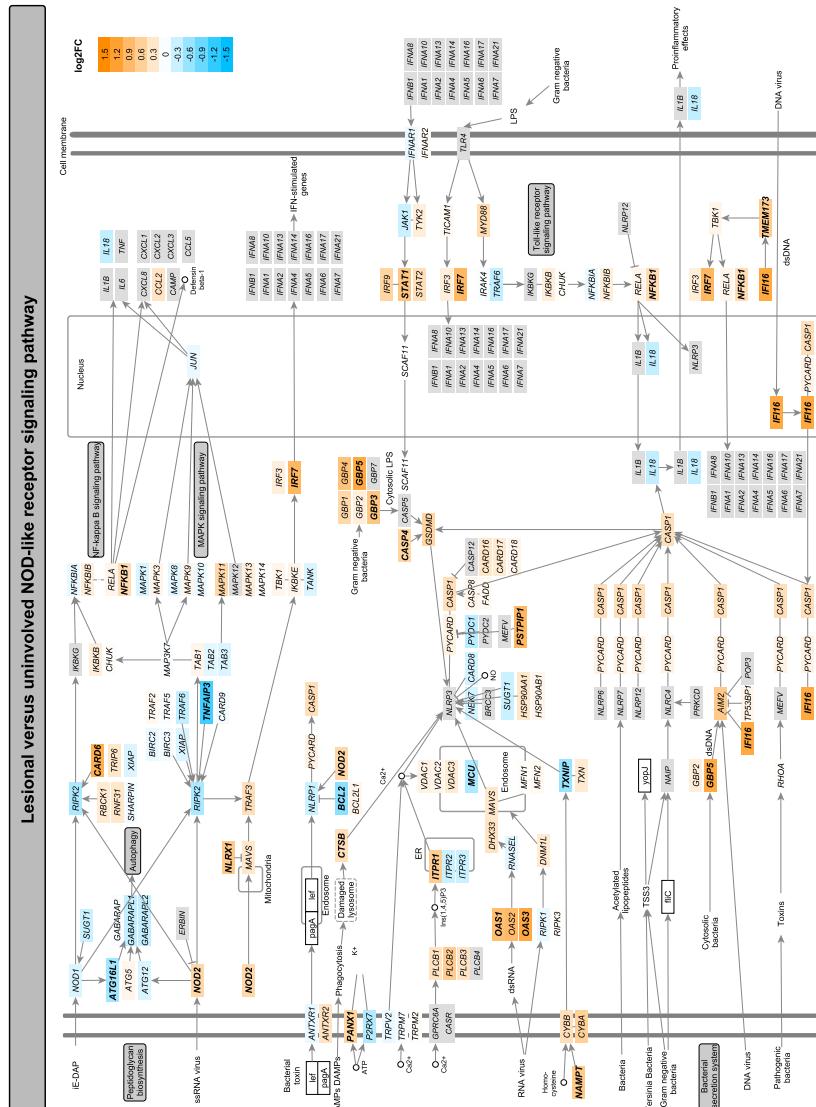
the GWAS gene *NOS2*, amongst others, contributed to the enrichment of this pathway (Figure 4.16).



**Figure 4.16: Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the HIF-1 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<sub>2</sub>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).

Immune relevant pathways including IFN, IL-17 and NOD-like receptor signalling were also identified in this analysis. The NOD-like receptor pathway responsible for detecting various pathogens and generating innate immune responses through NF- $\kappa$ B and MAPK activation, was enriched with 23 significantly DEGs , including *NOD2*, *CARD6* or *IFI16* (highly up-regulated) (Figure ?? in orange and bold) and *TNFAIP3* and *BCL-2* (down-regulated) (Figure ?? in blue and bold).

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



**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<sub>2</sub>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).

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

---

In addition to NOD-like signalling, IL-17 signalling was another enriched pathway well known to be relevant in the development of psoriasis. Enrichment of the IL-17 signalling pathway in our data is driven by up-regulation of the S100 protein family (*S100A7*, *S100A8* and *S100A9*) and chemokines such as *CCL20*, which binds the CCR6 receptor and is involved in DCs and T cell chemotaxis. *IL-17RE* which together with IL-17RA forms the receptor for IL-17C was down-regulated in lesional skin. Enrichment of DEGs between lesional and uninvolved skin for the peroxisome proliferator-activated receptor (PPAR) signalling highlighted the link between metabolic dysregulation (particularly lipids) and innate immunity. This pathway included up-regulation of the PPAR receptor  $\delta$  (*PPARD*), stearoyl-CoA desaturases such as *SCD* and *SCD5* involved in fatty acid synthesis and *CD36* which mediates fatty acid transport, also dysregulated in the Tsoi and/or 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 at the affected tissue (skin) and 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, although 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 and down-regulated in lesional epidermis when compared to uninvolved.

Early growth response genes were up-regulated in CD14 $^{+}$  monocytes (*EGR1* and *EGR3*) and CD4 $^{+}$  T cells (*EGR1*, *EGR2* and *EGR3*). In lesional skin,

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

---

*EGR2* and *EGR3* were also up-regulated with a much greater magnitude in the fold change ( $\log_2$ fold change -0.74 and -0.53, respectively) compared to the circulating immune cell types.

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

**Table 4.11:** Overlap between the DEGs in the four circulating immune cell types (psoriasis patients versus controls) and the DEGs in psoriasis patients skin biopsies (lesional versus 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. The pathways enriched for CD14<sup>+</sup> and CD8<sup>+</sup> DEGs were mostly immune-related pathways, including TCR , IL-12 , TNF and NF- $\kappa$ B signalling. In skin, the DEGs were not only enriched in immune-related pathways but also for pathways involved in metabolism, oxidative stress and cell cycle and genes contributing to the enrichment of immune-related pathways had a more pronounced pro-inflammatory signature, consistent with the skin being a site of more active inflammation compared to circulating immune cells in psoriasis.

### 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 in circulating immune cells from psoriasis patients has revealed a greater effect of disease status on gene expression than chromatin accessibility. To investigate whether an integrated approach combining evidence of differential

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

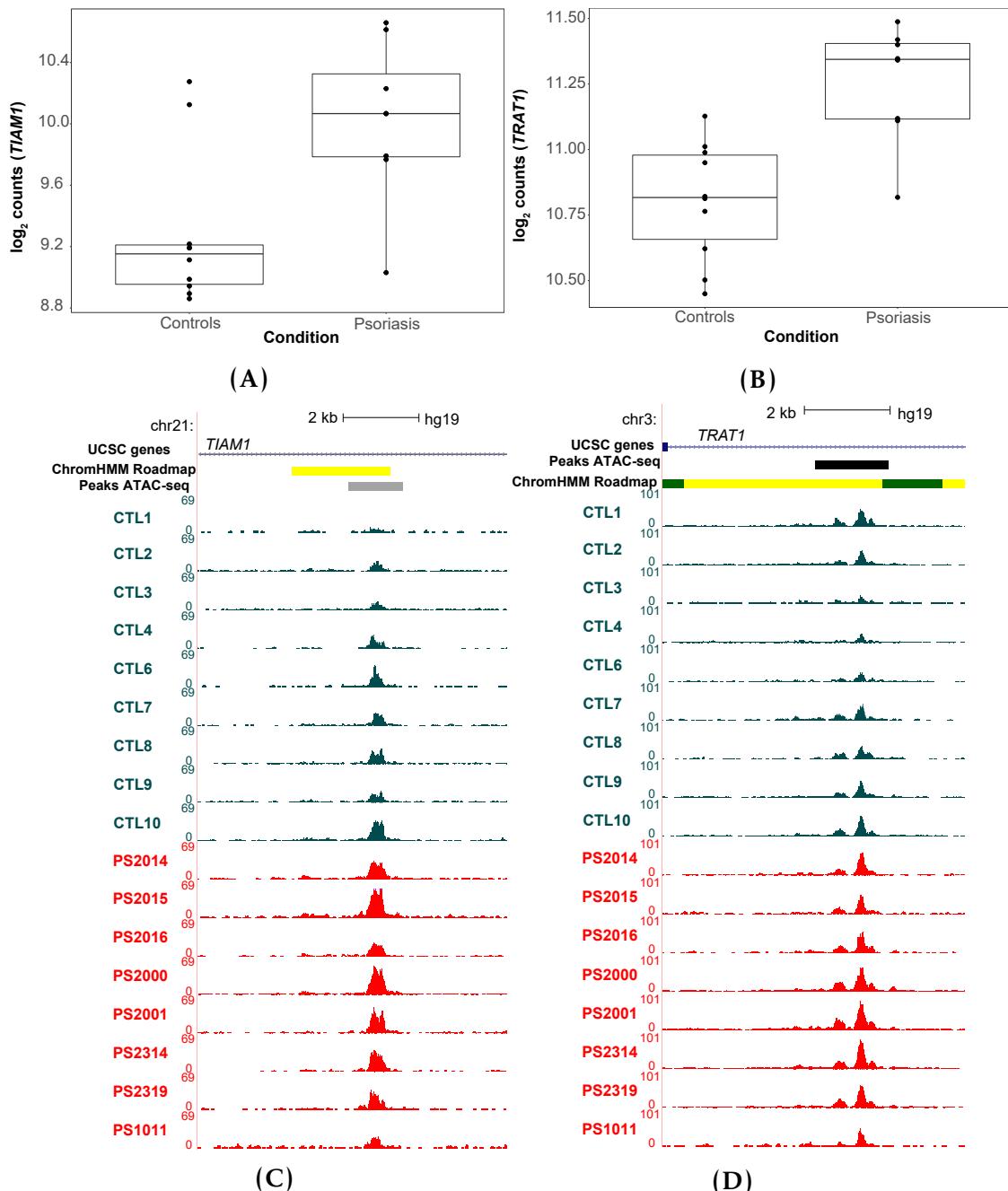
---

accessibility and expression, overlap between DEGs and the genes proximal to DARs ( $\leq 5\text{Kb}$ ) was investigated. Overlap was only found in CD8 $^{+}$  cells, where 6 out of the 53 DARs were annotated by proximity to an RNA-seq DEG in the same cell type (*ARL4A*, *ASCL2*, *ENTPD1*, *TIAM1*, *TRAT1* and *ZNF276*).

For example, T Cell lymphoma invasion and metastasis 1 (*TIAM1*), which activates IL-17 expression and T cell transendothelial migration during inflammation (Kurdi et al. 2016; Grard et al. 2009). This gene showed an increased expression (log<sub>2</sub>fold change 0.44) in psoriasis patients CD8 $^{+}$  cells (Figure 4.18A). Likewise, psoriasis CD8 $^{+}$  cells presented greater chromatin accessibility compared to healthy controls (log<sub>2</sub>fold change 0.41) in a region located at an intron of the *TIAM1* gene and annotated as an active enhancer according to the chromatin segmentation data in this cell type (Figure 4.18C). Common SNPs within this peak did not appear to be an eQTL regulating expression of any gene in CD8 $^{+}$  cells (Kasela et al. 2017) and chromatin conformation data did not reveal interaction of this particular region with the *TIAM1* promoter (Javiere2016), at least in unstimulated conditions, complicating the establishment of a mechanistic connection between chromatin accessibility and gene expression.

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

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



**Figure 4.18: Differential gene expression and chromatin accessibility landscape for *TIAM1* and *TRAT1* genes in CD8<sup>+</sup> cells.** Boxplots represent RNA-seq  $\log_2$  normalised counts for (A) *TIAM1* and (B) *TRAT1* genes in psoriasis and healthy controls CD8<sup>+</sup> cells. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) at an intron of the *TIAM1* (C) and *TRAT1* (x-axis) in CD8<sup>+</sup> cells. Both regions were identified as more accessible in psoriasis patients compared to healthy controls. 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<sup>+</sup> cells is included.

#### **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 GACP cohort (2,997 cases and 9,183 controls), one of the two included in the Immunochip psoriasis GWAS study from Tsoi *et al.*, 2012. Summary statistics of the second Immunochip cohort from Tsoi and colleagues (PAGE cohort) (Tsoi *et al.* 2012), were not publicly available through ImmunoBase at the time of this analysis. As explained in Chapter 2, fine-mapping from summary statistics with DIST used the statistics z-score of each of the genotyped SNPs from the GACP 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-genotypes SNPs, association analysis was performed and ABF, PP and credible set of SNPs were built for each of the signals (Tables 4.12).

Fine-mapping was performed for 25 of the Immunochip GWAS loci reported by Tsoi *et al.*, 2012, excluding the MHC and those loci which were in high LD with missense mutations showing experimentally proved or highly confident predicted damaging effects. Out of the 25 regions, 8 loci did not reach  $\log_{10}\text{ABF}>3$  (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 of the loci presenting  $\log_{10}\text{ABF}>3$ , the fine-mapping lead SNP was in low LD with the Tsoi *et al.*, 2012 GWAS lead SNP and/ or did not include it in the 90 or 50% credible set (Table 4.12 with \*). This is likely due to lack of power compared to Tsoi and colleagues Immunochip, as only summary statistics from the GACP cohort were available for this analysis.

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

**Table 4.12: Summary results for the fine-mapping analysis of 26 psoriasis Immunochip GWAS loci.** For the 17 Immunochip psoriasis GWAS loci showing  $\log_{10}ABF > 3$ , 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. The imputation status indicates whether the SNP was genotypes by the Immunochip array (No) or not (Yes). 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  $\log_{10}ABF \geq 3$ , 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  $\log_{10}ABF < 3$  are also included. FM=fine-mapping; RAF=reference panel allele frequency; OR=odds ratio; ABF=approximate Bayes factor; PP=posterior probability.

chr	Closer gene	FM lead SNP	RAF	Imputed	z	OR	GAPC	$\log_{10}ABF$ FM lead SNP	PP	90% credible set	Tsoi lead SNP
1	<i>IL28RA</i>	rs61774731*	0.10	No	-7.68	1.21	11.5	0.99	1		rs7552167
2	<i>FLJ16341/REL</i>	rs6714339*	0.14	No	-6.82	1.17	9.0	0.99	1		rs62149416
2	<i>IFIH1</i>	rs2111485*	0.59	No	5.94	1.27	6.7	0.50	2		rs17716942
5	<i>TNIP1</i>	rs17728338	0.06	No	7.45	1.59	10.6	0.40	6		rs2233278
5	<i>IL12B/ADRA1B</i>	rs12188300	0.05	No	7.71	1.58	11.2	0.18	9		rs12188300
6	<i>TNFAIP3</i>	rs1416173*	0.85	No	-6.36	1.23	7.7	0.15	10		rs582757
14	<i>NFKBIA</i>	rs74243591	0.21	No	-5.23	1.16	5.0	0.30	12		rs8016947
17	<i>NOS2</i>	rs117094752*	0.02	No	-6.53	1.22	7.3	0.94	1		rs28998802
1	<i>SLC45A1/TNFRSF9</i>	rs425371	0.25	Yes	5.52	1.13	5.6	0.14	22		rs11121129
1	<i>RUNX3</i>	rs61774731*	0.10	No	-7.68	1.13	11.5	0.99	1		rs7536201
2	<i>B3GNT2/TMEM17</i> (ch2p15)	rs9309343*	0.33	Yes	4.92	1.12	4.3	0.66	34		rs10865331
7	<i>ELMO1</i>	rs77840275*	0.10	No	-6.31	1.11	7.5	0.99	1		rs2700987
11	<i>ZC3H12C</i>	rs11213274	0.40	Yes	-4.78	1.14	4.0	0.05	69		rs4561177

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

$\log_{10}ABF < 3$										
chr	Closer gene	FM lead SNP	RAF	Imputed	z	OR	$\log_{10}ABF$ 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

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

---

The remaining 6 loci in high LD with the Tsoi *et al.*, 2012 GWAS lead SNP showed 90% credible set of SNPs ranging from 6 to 69 SNPs (Table 4.12). Of those 6 loci, *TNIP1*, *IL12B/ADRA1B* and *PTRF/STAT3* had 90% credible sets refined to less than 10 SNPs. *TNIP1* and *IL12B/ADRA1B* had been previously fine-mapped by Das and colleagues using dense genotyping with a customised array followed by association analysis (Das *et al.* 2014). Interestingly, 2 (rs75851973 and rs2233278) and 3 SNPs (rs918519, rs918518 and rs733589) from the *TNIP1* and *IL12B/ADRA1B* 90% credible sets, respectively, were amongst the set of significant variants and perfect near proxies ( $r^2 > 0.9$ ) reported by Das *et al.* for those two same loci. Interestingly, only 2 of the 6 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.

### **Integration with functional data**

A total of 144 unique SNPs formed the union of 90% credible sets from the 7 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 any of the DARs or differentially H3K27ac regions identified in CD14<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> cells. Conversely, overlap with the consensus master list ATAC peaks from each of the cell types revealed a total of 17 SNPs from 6 loci located at accessible chromatin in at least one cell type (*NFKBIA*(1 SNP), *PTRF/STAT3*(1 SNP), *SLC45A1/TNFRSF9*(4 SNPs), *TNIP1*(4 SNPs) and *ZC3H12C*(6 SNPs), *ETS1*(1 SNP)). However, no overlap was found for the 9 SNPs of the *IL12B/ADRA1B* 90% credible set. CD14<sup>+</sup> monocytes appeared as the cell type showing the largest proportion of accessible chromatin regions containing SNPs from the credible (2.3%), followed by CD19<sup>+</sup>, CD4<sup>+</sup> and CD8<sup>+</sup> cells (1.7, 1.6 and 1.05%, respectively) (Table 4.13). Altogether, integration of the SNPs from the credible set with ATAC accessible regions in four cell types allowed

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

---

to further refine the number of genetic variants with a putative functional role in psoriasis for 6 of the 7 analysed loci. Moreover, in the *PTRF/STAT3* and *SLC45A1/TNFRSF9* loci all the SNPs from the credible set overlapping accessible chromatin appeared to be cell type specific (Table 4.12 and 4.13). Of the 8 SNPs in the *PTRF/STAT3* credible set, only 1 overlapped accessible chromatin in CD14<sup>+</sup> monocytes. 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<sup>+</sup> monocytes-specific.

ATAC cell type master list	90% credible set overlapping SNPs (number)	Cell type specific overlap
CD14 <sup>+</sup> monocytes	13	<i>PTRF/STAT3</i> (1), <i>SLC45A1/TNFRSF9</i> (4)
CD4 <sup>+</sup>	6	None
CD8 <sup>+</sup>	4	<i>TNIP1</i> (2)
CD19 <sup>+</sup>	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 7 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 *SLC45A1/TNFRSF9* locus

As previously mentioned, integration of the *SLC45A1/TNFRSF9* 90% credible set of SNPs with ATAC data further refined the number of candidate functional SNPs from 22 to 4. *SLC45A1/TNFRSF9* was one of the new intergenic GWAS associations reported by Tsoi *et al.*, 2012 and is shared with UC and celiac disease (<https://www.immunobase.org/>). Amongst the 4 SNPs overlapping CD14<sup>+</sup> monocyte ATAC-specific peaks was the fine-mapping lead SNP rs425371, located at an intergenic region, approximately 269.3Kb downstream *TNFRSF9* gene (Figure 4.19 top panel). The other 3 SNPs overlapping CD14<sup>+</sup> monocytes

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

---

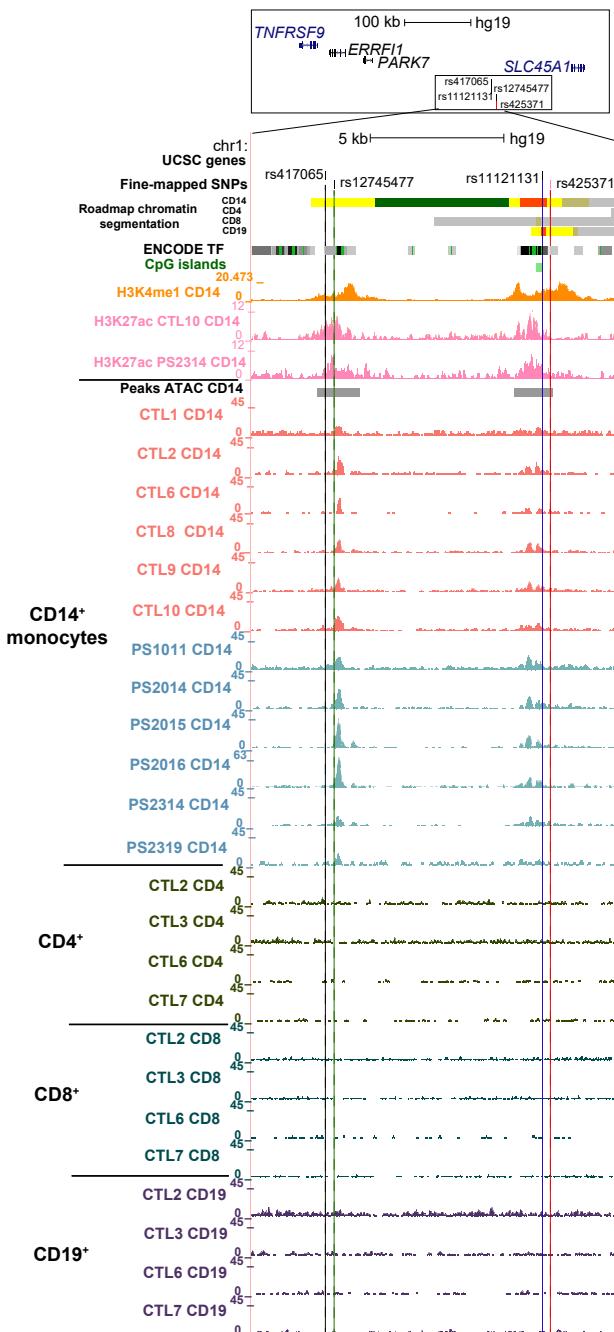
accessible chromatin included rs11121131, rs12745477 and rs417065 (Figure 4.19 top panel). Notably, the 2 ATAC peaks harbouring the 4 fine-mapped SNPs showed variable chromatin accessibility across individuals (Figure 4.19 bottom panel). Roadmap Epigenomics Project data showed enrichment for H3K4me1 in CD14<sup>+</sup> monocytes overlying the two ATAC peaks also overlapping with modest enrichment of in-house H3K27ac data in some of the samples. This was consistent with the designation of these two regions as enhancers by the CD14<sup>+</sup> monocytes chromatin segmentation map (Figure 4.19 bottom panel). No accessible chromatin was found at the location of the 4 fine-mapped SNPs in CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> cells, in line with the classification of these two regions as heterochromatin or repressed chromatin by the chromatin segmentation maps in the same 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 4.19 bottom panel), altogether reinforcing a putative role of these two regions in regulating gene expression.

rs12745477 and rs425371 (p-values 0.01 and 0.04) had an effect in regulating PARK7 expression levels in CD14<sup>+</sup> monocytes treated with LPS for 24h and in whole blood (Fairfax et al. 2012; Westra et al. 2013).

### **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) represents one of the GWAS associations located in an intergenic region was identified by the Immunochip study from Tsoi *et al.*, 2012. This locus is shared with other chronic inflammatory diseases including AS and CD (Cortes et al. 2013; Jostins et al. 2012). Although fine-mapping using summary statistics from the psoriasis GAPC Immunochip identified a signal with the lead showing

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



**Figure 4.19: Epigenetic landscape at the location of the SNPs in the 4 SNPs in the 90% credible set for the *SLC45A1/TNFRSF9* psoriasis GWAS locus.** The top panel illustrates the genomic location of the 4 SNPs in the *SLC45A1/TNFRSF9* fine-mapping 90% credible set overlapping ATAC accessible regions in CD14<sup>+</sup> monocytes and their distance from the *SLC45A1* and *TNFRSF9* genes. The bottom panel represents the UCSC Genome Browser view illustrating normalised ATAC read density and fold-enrichment of H3K27ac (both in-house dat(A) together with other publicly available epigenetic datasets (chromatin segmentation map, H3K4me1, ENCODE TF ChIP-seq and CpG islands) (y-axis) at the location of the 4 SNPs (rs425371, rs11121131, rs12745477 and rs417065) (x-axis) from the 90% credible set obtained in the fine-mapping analysis. ATAC and H3K27ac ChIPm data from some controls and psoriasis patients in the cohort are included for CD14<sup>+</sup> monocytes. Representative ATAC tracks from control samples are also shown for CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> cells to illustrate the absence of accessible chromatin at the location of the 4 SNPs.

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

---

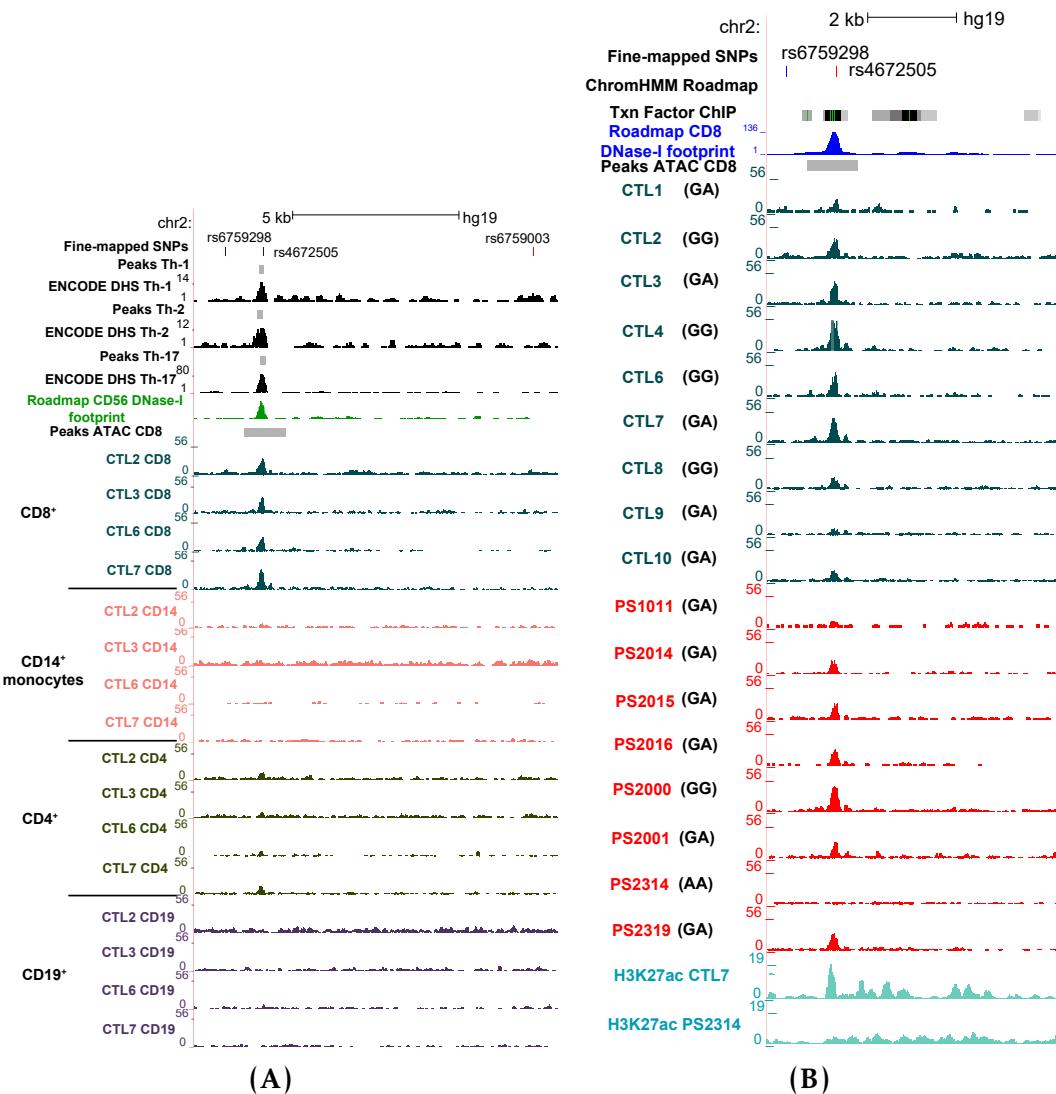
$\log_{10}\text{ABF}=4.3$ , it was not in LD with the psoriasis GWAS lead SNPs. Fine-mapping analysis in the AS UK Immunochip performed by Dr Anna Sanniti implicated rs4672505 (risk allele rs4672505\_A) as the SNP with the greatest PP (0.4) of the 95% credible set, together with another two SNPs (rs6759298 and rs6759003). Rs4672505 was also identified as the lead SNP in a multi-disease meta-analysis from Ellinghaus and colleagues (Ellinghaus et al. 2016).

Rs4672505 overlaps a CD8<sup>+</sup> T cell specific ATAC region, not present in CD14<sup>+</sup> monocytes, CD4<sup>+</sup> and CD19<sup>+</sup> cells (Figure 4.20A), that was not differentially accessible between patients and controls but did show marked variability across individuals (Figure 4.20B), 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<sup>+</sup> T cells, respectively (Figure 4.20A). Variability across individuals was also observed for H3K27ac enrichment was also observed according to the ChIPm data generated in cohort 1B (Figure 4.20B). ENCODE ChIP-seq experimental data from GM12878 showed binding for RUNX3, a psoriasis and AS GWAS associated gene and is involved in CD8<sup>+</sup> cell differentiation (Wong et al. 2011). 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 rs4672505. Altogether, integration of ATAC and publicly available epigenetic data indicated that rs4672505 was the most likely variant, amongst the three fine-mapped SNPs included in the 95% credible set from AS fine-mapping, to have a functional role which could explain the association of chr2p15 with psoriasis risk.

The genotype of each individual for rs4672505 was determined using Sanger sequencing. Amongst the eighteen samples (10 controls and 8 psoriasis patients), one (PS2314) was homozygous for the risk allele (A, MAF=0.43), 11

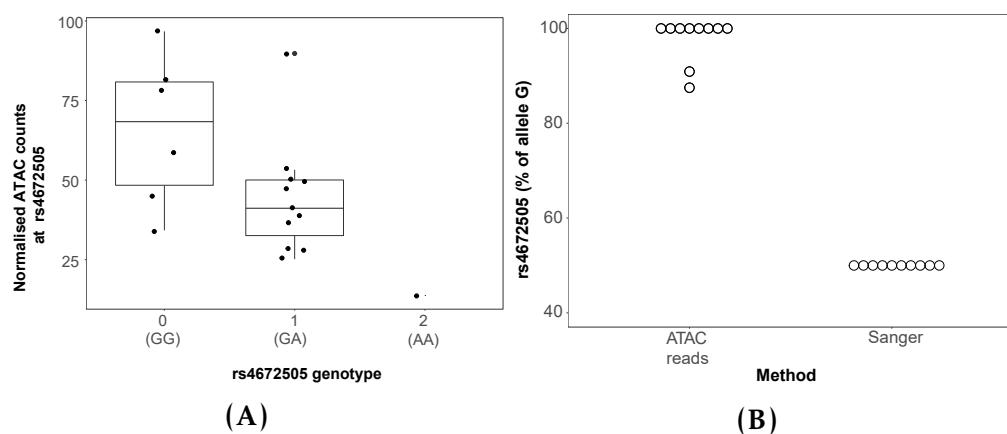
## Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis



**Figure 4.20: 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 3 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 4 controls in the cohort and the 4 cell types included in this study are shown. (B) UCSC Genome Browser view illustrating the normalised read density for CD8<sup>+</sup> ATAC (x-axis) generated in psoriasis patients and healthy controls, in-house H3K27ac ChIPm, ENCODE TF ChIP-seq and DNase-I footprint (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.

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

were heterozygous and 6 were homozygous for the protective allele (G) (Figure 4.20B). Interestingly, PS2314, the only homozygous individual for the risk allele, showed complete absence of the peak at rs4672505. To further investigate the role of rs4672505 genotype in the variability of chromatin accessibility across individuals, the 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.



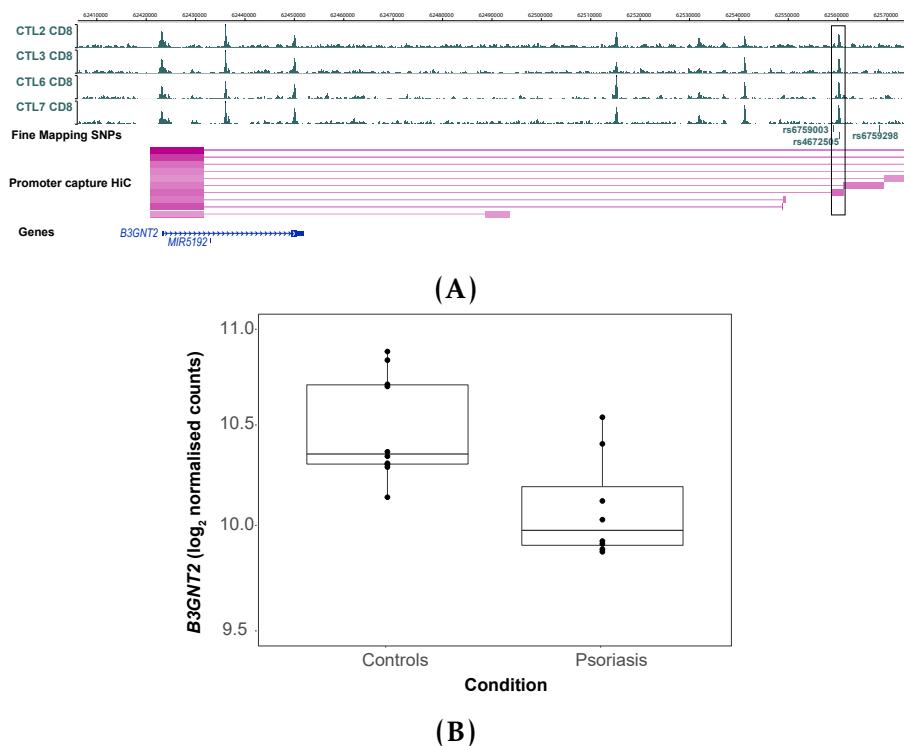
**Figure 4.21: rs4672505 genotype and chromatin accessibility at chr2:62,559,749-62,561,442 in CD8<sup>+</sup> cells.** (A) Boxplot illustrating the effect of the rs4672505 genotype on chromatin accessibility at the chr2:62,559,749-62,561,442 ATAC peak. Log<sub>2</sub> normalised ATAC counts adjusted for batch effect, also included as a covariate for the linear model, are plotted for each sample against the number of copies of the minor allele (G=0, AG=1, AA=2). (B) Representation of the percentage of ATAC reads overlapping rs4672505 and mapping to the major allele (G) in comparison to the Sanger genotype results for the eleven heterozygous individuals at this SNP.

Significant negative correlation ( $p\text{-value}=0.035$ ) was found, suggesting allele-dependent chromatin accessibility (Figure 4.21 a). Furthermore, allelic imbalance for the ATAC reads at rs4672505 position was investigated on those individuals identified as heterozygous by Sanger sequencing and 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 4.21 b). This finding was not driven by mapping bias, since A was the reference allele in the hg19 build used to map the ATAC data. Overall, these results showed evidence of greater chromatin

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

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 data from Javierre *et al.* 2016 in CD8<sup>+</sup> revealed a genome-wide significant interaction (CHiCAGO score=7.67) between a region containing rs4672505 and the promoter of the *B3GNT2* gene (Figure 4.22A).



**Figure 4.22: Potential role of rs4672505 in regulating *B3GNT2* gene expression.** (A) WASHU Genome Browser track showing in CD8<sup>+</sup> cells normalised ATAC read density in four of the healthy controls, the location of the three SNPs of the 95% credible set and the promoter capture HiC data depicting the regions interacting with the bait at the *B3GNT2* promoter. (B) Boxplot illustrating the *B3GNT2* log<sub>2</sub> normalised RNA-seq counts adjusted for batch effect in the psoriasis and healthy control groups.

This interaction was not found in any of the additional 16 human primary hematopoietic cell included in the study. No upstream interaction with *TMEM1* promoter was identified. Investigation of the publicly available T cell eQTL

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

---

dataset from Kasela *et al.* and Raj *et al.* did not show a significant eQTL for this SNPs 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 ( $\beta=-0.16$ ) for the minor allele (Jansen et al. 2017). In terms of gene expression based on disease state, significant (FDR<0.05) down-regulation of *B3GNT2* was observed in psoriasis patients when compared to controls in CD8 $^{+}$  cells (log<sub>2</sub>fold change=-0.317) (Figure 4.22B). These data provide evidence of allele-specific chromatin accessibility at chr2p15 in CD8 $^{+}$  T cells involving a GWAS SNP, with chromatin conformation capture and gene expression data suggesting that this modulates, at a distance, expression of *B3GNT2*, providing insights into the possible mechanistic basis of the disease association seen in psoriasis and other inflammatory diseases, including AS.

## **4.4 Discussion**

### **4.4.1 Chromatin accessibility and H3K27ac landscape in psoriasis immune cells**

Comparison of chromatin accessibility and H3K27ac histone modifications has revealed a small number of differential regions between patients and controls in the four cells types under study. For both epigenetic features, CD14 $^{+}$  monocytes and CD8 $^{+}$  cells had the largest number of discrete changes. In ATAC, greater accessibility in CD8 $^{+}$  cells from patients compared to controls was found at two regions proximal to *IL7R* and *TNFSF11*, respectively, that also overlap FANTOM eRNA in the same cell type. Both genes are well known for having a pro-inflammatory effect and be involved in chronic inflammatory diseases. For example, *TNFSF11* is downstream of the lead SNP for a CD risk locus, and

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

---

its protein product RANKL was found to be overexpressed in epidermis from psoriasis patients, highlighting the role of this gene in the pathophysiology of psoriasis (Toberer et al. 2011).

Integration of the ATAC and H3K27ac ChIPm differential analysis only found one overlapping region at an intron of the *DTD1*, a D-Tyr-tRNA deacylase 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). A role of aminoacylRNA synthetases, responsible for the loading of L-aminoacids into tRNAs, in modulation of inflammation and angiogenesis (Yao and medicine 2013). Although, intronic variants in *DTD1* have been associated with anxiety, onset of sleep irregularities and age-related variation of DNA methylation through GWAS (<https://www.ebi.ac.uk/gwas/>), no evidence of *DTD1* involvement in chronic inflammation and psoriasis has yet been reported. Altogether, the lack of overlap between DARs and differentially H3K27ac modified regions might be expected given that chromatin accessibility is driven by the interaction between a number of histone modifications, TFs, and structural proteins, such as CTCF.

The results in this chapter suggest that disease status does not involve global differences in chromatin accessibility and H3K27ac between patients and controls in immune cells isolated from peripheral blood. Recent similar studies performing ATAC in B cells from SLE patients have revealed larger differences in the chromatin accessibility landscape between patients and controls , while H3K27ac mapping in mCD4<sup>+</sup> cells isolated from juvenile idiopathic arthritis synovial fluid found approximately one thousand differential enhancers when compared to healthy control circulating cells (Scharer et al. 2016; Peeters et al. 2015). Conversely only small differences were found when comparing mCD4<sup>+</sup> from peripheral blood of patients and controls, highlighting the specificity of the

disease signature at the site of inflammation and 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 DARs or differential H3K27ac modifications. As for ATAC and ChIPm, CD14<sup>+</sup> monocytes and CD8<sup>+</sup> 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<sup>+</sup> compared to CD4<sup>+</sup> may suggest that, as in skin, CD8<sup>+</sup> are the main effector cells upon induced-activation by CD4<sup>+</sup> cells (Nickoloff and Wrone-Smith 1999). The importance of monocytes/macrophages in psoriasis has also been demonstrated by their presence in psoriatic skin where TNF- $\alpha$  production contributes towards maintenance of inflammation (Nickoloff 2000; Wang et al. 2006).

For example, *MAP3K4* down-regulation in LPS stimulated PBMCs has been identified as an immune-suppressive feature in CD leading to reduced expression of the cytokine IL-1 $\alpha$ . In the IL-12 signalling pathway which leads to T cell proliferation and IFN- $\gamma$  production through activation of TFs from the STAT family, CD14<sup>+</sup> cells presented down-regulation of *STAT4* and *STAT5A* in psoriasis versus controls. *STAT2*, other member of the STAT family, was found to be down-regulated in psoriasis PBMCs and in AS monocyte-derived macrophages when compared to controls 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 (Litherland et al. 2005). Further investigation to determine phosphorylated STAT4 and STAT5

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

---

protein abundance will be required to determine if the down-regulation at the transcript level observed in psoriasis CD14<sup>+</sup> monocytes is biologically relevant.

A very interesting observation in this data was the down-regulation of *IFNG* in psoriasis CD8<sup>+</sup> cells, a gene member of the IL-12 signalling pathway. 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 a SpA rat model (Smith et al. 2008; Fert et al. 2014). In AS monocytes and in the SpA rat model, *IFNG* down-regulation was accompanied by an overall inverse transcriptional response of IFN-regulated 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, either IL-23 or IL-17 were up-regulated in psoriasis CD8<sup>+</sup> T cells compared to healthy controls.

In CD8<sup>+</sup> specifically, DEGs showed significant enrichment for three very relevant pathophysiological pathways in psoriasis: NF-κB, TNF and chemokine signaling. Important cross-talk between the NF-κB and TNF signaling pathway was observed, with a number of dysregulated genes contributing to both. Interestingly, the enrichment of these pathways involved up-regulation of pro-inflammatory genes (e.g. *ATF2*, *ATF4*, *RELA*, *RELB*) but also increased expression of well-characterised immunoregulatory genes. These included *NFKBIA* and *TNFAIP3*, also up-regulated in CD14<sup>+</sup> monocytes and CD4<sup>+</sup> cells respectively, and both associated with psoriasis risk and a number of other chronic inflammatory diseases, including MS, RA, SLE and T1D (Vereecke et al. 2011). *NFKBIA* inhibits NF-κB by binding it and preventing translocation to the nucleus. Similarly, *TNFAIP3* undergoes upregulation 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

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

---

dysregulated in psoriasis PBMCs by Coda *et al.*, Lee *et al.* and Mesko *et al.* or in PBMCs from PsA patients versus controls (Dolcino et al. 2015). Interestingly, qPCR analysis in PBMCs from mild ( $\text{PASI} < 4.84$ ) and severe ( $\text{PASI} > 4.84$ ) psoriasis vulgaris revealed a significant negative correlation between *TNFAIP3* expression and disease severity (Jiang et al. 2012). This study also demonstrated *TNFAIP3* down-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 ( $\text{PASI} > 4.84$ ) according to Jian and colleagues (Jian2012). Altogether, the up-regulated expression of *TNFAIP3* and *NFKBIA* compared to healthy controls may be the result of a feedback loop reflecting a persistent inflammatory stimuli in psoriasis peripheral blood and a mechanism that limits the systemic inflammatory response to some extent (Idel et al. 2003).

Amongst the differential expression of chemokines, up-regulation of the chemokine receptor *CCR10* in  $\text{CD8}^+$  cells from psoriasis patients was of particular interest. In circulation, expression of *CCR10* is restricted to a subset of circulating mCD4 $^+$  and mCD8 $^+$  T cells expressing the cutaneous lymphocyte-associated antigen (CLA), and preferentially recruited to cutaneous sites of inflammation (Hudak et al. 2002). Indeed, an increase of  $\text{CCR10}^+$  infiltrated T lymphocytes in psoriatic skin, where keratinocytes express *CCR10* ligand *CCL27*, has been demonstrated (Homey et al. 2002). The up-regulation of *CCR10* in my data could potentially suggest an increase of mCD8 $^+$   $\text{CCR10}^+$  cells ready to migrate into the skin lesions. Overall, these data have revealed differential gene expression between psoriasis patients and controls for relevant immune genes showing pro- and anti-inflammatory effects in peripheral blood immune cells. Although down-regulation of pro-inflammatory genes and up-regulation of anti-inflammatory genes has been observed, understanding the overall effect of those interactions in the inflammatory response requires further investigation.

#### **4.4.3 Integration of chromatin accessibility and gene expression data**

In this chapter, greater changes in gene expression have been identified compared to chromatin accessibility. Strikingly, in CD8<sup>+</sup> cells, 687 transcripts were differentially expressed between psoriasis and healthy controls but only 55 regions showed differential chromatin accessibility when performing the same contrast and only six of the 687 were proximal to a DAR. This may relate to the complexity of gene regulation as part of a multi-component regulatory process manifesting in differential gene expression, where chromatin accessibility is only one of the events contributing to the entire process.

Two relevant DEG nearby a DAR were *TIAM1* and *TRAT1*, showing increased chromatin accessibility and gene expression in psoriasis CD8<sup>+</sup> cells compared to healthy controls. The T cell lymphoma invasion and metastasis 1 *TIAM1* is involved in IL-17 expression and cell migration into the inflamed tissue (Kurdi et al. 2016; Grard et al. 2009). However, no eQTL or chromatin conformation data in this cell type has been found to formally establish a link between the region harbouring this DAR and *TIAM1* expression. The TCR associated transmembrane adaptor 1 *TRAT1* gene is a positive regulator for TCR signalling and has been observed to be down-regulated in Tregs (Birzele).

#### **4.4.4 Transcriptomic profiles in lesional and uninvolved psoriatic epidermis**

Investigation of differences in the transcriptomic profile 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 fibroblast, adipocytes, keratinocytes from the epidermis and dermis and infiltrated immune cells.

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

---

A study from Ahn and colleagues demonstrated large differences in gene expression between whole biopsies and FACS-isolated keratinocytes, and the former may be masking keratinocyte-specific pathophysiological differences in many previous studies using psoriasis skin biopsies (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 the Tervaniemi *et al.* study contrasting gene expression between lesional and uninvolved epidermis split biopsies, mainly formed by epidermis, revealed an overlap of only 359 out of the 1,227 DEGs detected in my data (12.1% of Tervaniemi *et al.* DEGs). Interestingly, the overlap with the Tsoi *et al.* study using whole biopsies was similar (505, 13.1% of Tsoi *et al.* DEGs) and only 5 genes had opposite direction of change, in contrast with the 75 showing discrepancies with Tervaniemi's study. The similar percentage overlap with the Tsoi study despite the different source material could simply be the result of greater power in that study.

Genes consistently up-regulated across the three studies included genes from the *S100A* family. The *S100* family are located in the chr1p21 locus, which harbours genes involved in keratinocytes differentiation, act as calcium sensors and may also have a chemotactic effect (Eckert et al. 2004). In particular, *S100A9* and *S100A12* undergo up-regulation in psoriasis (Broome et al. 2003), with the latter involved in the T cell proliferative response and IFN- $\gamma$  and IL-2 production (Moser et al. 2007). *LCE3B*, also at the chr1p21 locus, was also upregulated in lesional skin compared to uninvolved in all three studies. *LCE3B/C\_del*, a psoriasis GWAS association, is found in approximately 60 to 70% of European psoriasis patients (Cid et al. 2009). As explained in Chapter 1, *LCE* gene expression is induced upon disruption of the skin barrier, and expression of *LCE3B* and *LCE3D* has been only detected in lesional but not uninvolved psoriatic skin of heterozygous individuals (Cid et al. 2009; Bergboer et al. 2011).

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

---

Pathway enrichment analysis for the DEGs between lesional and uninvolved skin revealed a number of relevant biological processes for psoriasis pathophysiology. These highlighted alterations in cell cycle and metabolic processes, including amino acid metabolism, glycolysis, and hypoxia (HIF-I signalling), which had been identified in other studies performing differential gene expression analysis between lesional and uninvolved skin or genome-wide pathway analysis (Coda et al. 2012; Gudjonsson et al. 2010; Aterido et al. 2016; Tervaniemi et al. 2016). Up-regulation of the hypoxia-inducible TFs HIF-1 $\alpha$  and HIF-2 $\alpha$  has been reported in lesional skin, correlating with an increase in *VEGF* transcript levels, a gene regulated by HIFs that mediates pathological angiogenesis also characteristic of psoriasis (Rosenberg 2007). No correlation was observed between *HIFA* and *VEGF* in the data presented here, likely due to the small sample size. Moreover, HIF-I signalling is also involved in regulating Th-17/Treg ratios and therefore in perpetuation and termination of the immune response (Dang et al. 2011).

Immune-related pathway enrichment were also found, including Th-17, IL-12 and NOD-like signalling. Interestingly, NOD-like signalling was found to be enriched in DEGs between lesional and uninvolved skin in a contemporary study by Tervaniemi and colleagues (Tervaniemi et al. 2016). Tervaniemi mainly attributed this novel pathway to the greater sensitivity of RNA-seq compared to microarrays to detect changes in gene expression for genes involved in this pathway. The fact that Tsoi *et al.* also used RNA-seq and did not show enrichment for NOD-like signalling is likely due to the type of biopsy, highlighting the value of studying epidermis instead of full thickness skin to uncover dysregulation of functional pathways in keratinocytes. NOD-like signalling involves signal transduction by NOD-like receptors, a type of pattern-recognition receptors, which can recruit and activate caspases into the inflammasomes or trigger inflammation through NF- $\kappa$ B and MAPK. Amongst the genes contributing to

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

---

this pathway *CARD6*, *IFI16*, *NOD2* and *NLRX1* overlapped with Tervaniemis data and showed up-regulation in both. Notably, polymorphisms in *NOD2* have been linked to inflammatory diseases such as CD, atopic eczema and arthritis and potentially with psoriasis and PsA (Zhong et al. 2013; Zhu et al. 2012).

Lastly, PPAR signaling appeared as one of the most significantly enriched pathways and represents a link between metabolic and innate immunity dysregulation in psoriasis. PPARs are TFs activated by fatty acid signaling with an anti-inflammatory role in the development of metabolic diseases and chronic inflammation such as RA (Straus and immunology 2007; Ji et al. 2001). Similarly to my study, *PPARD* was up-regulated in psoriatic lesional skin, 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).

### **4.4.5 LncRNAs in psoriasis**

The largest number of differentially expressed lncRNAs between psoriasis patients and controls were found in CD14<sup>+</sup> monocytes and CD8<sup>+</sup> T cells (28 and 31, respectively for FDR<0.05) with 46 lncRNAs differentially expressed between lesional and uninvolved skin (FDR<0.05). Several studies have contrasted lncRNAs in lesional compared to uninvolved or healthy skin (Li et al. 2014; Gupta et al. 2016; Ahn et al. 2016; Tsoi et al. 2015) but no study has been conducted to identify differentially lncRNAs in a cell type-specific manner in peripheral blood from psoriasis patients. The role of lncRNAs has been studied in RA, SLE, AS, and PsA (Muller 2014; Shi et al. 2014; Zhang et al. 2017; Dolcino et al. 2018).

Characterisation of lncRNA biological function is a developing field, which represents a limitation when interpreting these results (B et al. 2018). Some of the well-characterised dysregulated lncRNAs have a role in the immune response.

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

---

For example, in psoriasis CD14<sup>+</sup> monocytes the up-regulation of *HOTAIRM1* was found to be associated with down-regulation of the predicted target gene *USP1*. *UPF1* is involved in nonsense-mediated decay and in partnership with the monocyte chemotactic protein-1-induced protein-1 (*MCPIP1*) gene drives degradation of inflammation-related mRNAs to ensure maintenance of homeostasis (Mino et al. 2015). Down-regulation of *UPF1* in psoriasis CD14<sup>+</sup> monocytes may suggest impairment of this homeostatic mechanism, contributing to disease pathophysiology in this cell type. Another example of a relevant lncRNA up-regulated in CD14<sup>+</sup> monocytes was *NEAT1*, which has also shown up-regulation in SLE CD14<sup>+</sup> monocytes. Knock-down demonstrated impairment of TLR4 signalling and down-regulation of inflammatory genes including IL-6 and CXCL10 (Zhang et al. 2016).

*MIR146A* was found to be differentially expressed between lesional and uninvolved skin and also when comparing psoriasis CD8<sup>+</sup> T cells to healthy controls. Molecular studies have suggested a role for miR-146a as a negative regulator of the TLR4 pathway through inhibition of TNF associated factor 6 (*TRAF6*) and IL-1 receptor-associated kinase 1 (*IRAK1*) expression (Taganov et al. 2006). The up-regulation of *MIR146A* expression in lesional epidermis compared to uninvolved has been observed in other studies, and was also shown to be increased in lesional skin versus healthy biopsies (Lerman 2014; Tsoi et al. 2015; Li et al. 2014). Down-regulation of miR-146a levels in CD8<sup>+</sup> cells would support failure of one of the check-points controlling sustained inflammation and the subsequent pathophysiological implications, while the up-regulation in lesional skin would not be expected given the exacerbated inflammatory response in skin characteristic of psoriasis.

Other dysregulated non-coding RNAs in lesional epidermis relevant to psoriasis pathophysiology were *HG19* and *MIR31HG*, found to be down- and up-regulated, respectively. Silencing miR-31hg in keratinocyte immortal cell line

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

---

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 to identify putative targets for all the identified lncRNAs. Those interactions could then be used to identify relevant biological processes through network and pathway analysis using only those dysregulated lncRNAs matching dysregulated target genes, 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 Differences in transcriptional dysregulation in peripheral blood and skin**

Comparison of the dysregulated genes in circulating immune cells and in psoriatic skin revealed very limited overlap. CD14<sup>+</sup> monocytes and CD8<sup>+</sup> cells showed the greatest overlap of DEGs; however, almost half of the genes showed opposite direction in differential expression, consistent with a report by Coda comparing DEGs genes in psoriasis patients and controls PBMCs to genes dysregulated between lesional and uninvolved skin biopsies (Coda et al. 2012). Genes showing opposite changes in circulation and in skin included the GWAS gene *TNFAIP3*, *EGR2* and *EGR3*, amongst others. As previously mentioned, *TNFAIP3* down-regulation in lesional skin may reflect complete loss of an NF-κB pathway check-point to control and terminate the inflammatory response at the site of inflammation. Similarly, *EGR2* and *EGR3* are pivotal for control of inflammation and antigen-induced proliferation. In addition to its role in regulating inflammation, down-regulation of *EGR2* in skin may also increase keratinocyte proliferation as has been shown in certain types of cancer (Wu et al. 2010).

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

---

Differences are also observed in the distinct enriched pathways. For example, DEGs in skin not only showed enrichment for immune-related functions but also highlighted metabolic dysregulation that appears to be characteristic of this site of inflammation. Moreover, immune related pathway such as NOD-like signalling was only seen in skin. Likewise, up-regulation of genes from the *S100* family in lesional skin, such as *S100A7*, *S100A8*, *S100A9*, contributed to enrichment of IL-17 signalling and appeared to be a feature of dysregulated inflammation only in skin. Notably, these genes had also been reported as a specific hallmark of skin inflammation when compared to inflamed synovium from matched PsA patients, supporting the better outcomes for IL-17 antagonists in skin lesions compared to the inflammation of the synovium (Belasco et al. 2015).

### **4.4.7 Fine-mapping using summary statistics and integration with epigenetic data**

#### **Summary statistics fine-mapping and integration with functional data**

Fine-mapping of GWAS loci is one strategy to narrow down putative functionally relevant variants identified by GWAS studies. Using summary statistics from the psoriasis Immunochip GWAS GPC cohort, 26 of the genome-wide significant loci reported by Tosi and colleagues were fine-mapped (Tsoi et al. 2012). Successful fine-mapping based on ABF cut-off was achieved for 16 loci, 10 of them which association signal identified by the fine-mapping was not in LD with the GWAS signal reported by Tsoi and colleagues. This is probably due to a combined effect imputing summary statistics instead of genotypes and the reduced power for the association analysis due to smaller sample size compared to Tsoi and colleagues Immunochip (Tsoi et al. 2012). This is in line with the sensitivity of the association analysis to sample size previously shown in other studies (Bunt et al. 2015). In contrast, 6 of the loci

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

---

were successfully fine-mapped, with the fine-mapping lead SNP in high LD with the GWAS lead SNP. Fine-mapping successfully reduced the number of putative causal SNPs, with some loci showing less than 10 SNPs in the 90% credible set, such as *TNIP1*, *IL12B/ADRA1B* and *PTRF/STAT3*. Interestingly, the credible set of *TNIP1* included 6 SNPs, a lower number of putative causal SNPs compared to the 9 SNPs (which only included those in LD  $r^2 \geq 0.9$ ) reported by a study performing dense genotyping and association analysis for 8 psoriasis loci (Das et al. 2014). Integration of DARs and differentially H3K27ac modified regions with the SNPs from the credible set of the loci successfully fine-mapped did not reveal any overlap. In contrast, 16 SNPs out of the 126 forming the union of credible set of SNPs from 6 successfully fine-mapped loci, co-localised with accessible chromatin, and some of the with cell type specific ATAC peaks. Interestingly, integration with ATAC data further reduced the credible set of SNPs in the *SLC45A1/TNFRSF9* locus from 22 to 4, all of them overlapping CD14<sup>+</sup>-specific ATAC peaks. Integration of eQTL publicly available data showed regulation of *PARK7* expression by rs12745477 and rs425371 genotype. *PARK7* is a sensor for oxidative stress causing early-onset of Parkinson'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).

Similar approaches integrating 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) has been applied to perform fine-mapping for the Immunochip GWAS associated loci, including psoriasis, using summary statistics and incorporating in the model the 43 ENCODE and Roadmap Epigenomics Project annotation features most

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

---

enriched for psoriasis GWAS loci (Li et al. 2016). For example, Li and colleagues only reported two SNPs to explain the association of the *SLC45A1/TNFRSF9* locus to psoriasis but none was in LD with the GWAS signal from Tsoi *et al.*, 2012.

### **Allelic differences in chromatin accessibility at the chr2p15 locus**

One of the particularly interesting psoriasis GWAS associations is the chr2p15 locus, where the lead SNP is located in an intergenic region 140Kb and 150Kb away from *B3GNT2* and *TMEM1* respectively and is also associated with CD and AS (**Ostins2012**; Cortes et al. 2013). Fine-mapping at chr2p15 using psoriasis summary stats from the GAPC cohort failed to identify a signal in LD with the psoriasis GWAS from Tsoi and colleagues (Tsoi et al. 2012). The results published by Li and colleagues using RiVIERA also failed to fine-map this locus. Conversely, fine-mapping analysis in AS implicated rs4672505, which is in high LD with the psoriasis Immunochip chr2p15 SNP, and was also reported to be the lead SNP for this locus in the cross-disease association analysis including psoriasis, AS and CD (Ellinghaus et al. 2016). rs4672505 has also been shown as the 4<sup>th</sup> variant prioritised by the fine-mapping algorithm published by Farh and colleagues (**Farh2014**) and the integration with the in-house ATAC data demonstrated to be overlapping accessible chromatin at a CD8<sup>+</sup>-specific peak. Chromatin accessibility at this location was observed to vary between individuals, with complete loss of ATAC peak in some of them. Integration of genotyping data with ATAC revealed allele-dependent chromatin accessibility, with loss of chromatin accessibility correlating with the risk allele. Promoter capture-C data linked rs4672505 to the *B3GNT2* promoter only in CD8<sup>+</sup> cells, suggesting the accessible chromatin at rs4672505 may be highlighting an enhancer element interacting with *B3GNT2* promoter as a priming event (Javierre et al. 2016). The genotype effect of this SNP in chromatin accessibility

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

---

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.*B3GNT2* is a major polylactosamine synthase involved in the post-translational modifications 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 compared to wild-type (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*.

Up-regulation of *B3GNT2* in this context could be contributing to attenuation and modulation of CD8<sup>+</sup> activation. Under this scenario, the presence of the risk allele (A) at this stimulus-specific enhancer could increase risk of disease by reducing chromatin accessibility, in both homozygous and heterozygous individuals. 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 FAIRE-seq accessible chromatin, with the risk allele showing greater abundance at open chromatin and increased enhancer activity (Gaulton et al. 2010; Stefan et al. 2014). Nevertheless, establishing of a more comprehensive and accurate model to further explain the functional role of rs4672505 in psoriasis susceptibility will require additional work.

### **4.4.8 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.

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

---

Other limitations in this study include its relatively small sample size, lack of genotyping data and skin biopsies only being available for 3 patients in the cohort. These limitations are intrinsic to time and project budget constraints and will be addressed as the study continues. Moreover, at the time of sample recruitment this patients did not present self-reported joint affection 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, the analytical methods and strategies used in this chapter to perform data analysis, particularly incorporation of genotyping data into the differential analysis of chromatin accessibility and histones modifications. Further analysis could also be performed with the differentially expressed lncRNA integrating their experimentally validated targets to identify specific processes regulated by these RNA species in the context of disease.

## **4.5 Conclusions**

In this chapter, use of contemporary methods for epigenetic profiling (as established in the previous chapter) together with gene expression quantification has allowed characterisation of the regulatory landscape in relevant cell types isolated from psoriasis patients and healthy individuals. Minor 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. 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<sup>+</sup> and IL-22<sup>+</sup> CD4<sup>+</sup> 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

## **Cross-tissue comparison analysis in PsA**

---

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).

Nevertheless, 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 have been highlighted by differences in cell composition and cytokine production, amongst others, between peripheral blood and synovial fluid in PsA patients. For example, expansion mCD8<sup>+</sup> and elevated proportion of T cells expressing the cytokine receptors CCR6<sup>+</sup> and IL-23R<sup>+</sup> was observed in synovial fluid when compared to peripheral blood in PsA paired samples (Ross et al. 2000; Benham et al. 2013). Moreover, the elevated TNF- $\alpha$ , IL-1, IL-6 and IL-18 production in PsA synovial fluid was comparable to RA (**Kujik2006**).

### **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 and synovial fluid mononuclear cells (SFMCs) samples. 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

## **Cross-tissue comparison analysis in PsA**

---

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- $\alpha$  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 *in vitro* differentiated cells from naïve CD4 $^{+}$  isolated from peripheral blood and synovial fluid of PsA patients (Antoniv 2006; Leipe et al. 2010). The importance of investigating 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 (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 pre-selection 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 than other with less manipulation and variability than other scRNA-seq methods such as SmartSeq2 (Baran-Gale and functional 2017).

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 Abs recognising the molecular markers have been labelled with stable isotopes instead of fluorophores (Bandura et al. 2009). The use of isotopes enables incorporating up to 45 Abs 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

## **Cross-tissue comparison analysis in PsA**

---

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 cell populations 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 highly proliferative and active in pro-inflammatory cytokine secretion (Mizoguchi et al. 2018). Similarly, mass cytometry analysis performed in RA identified an expanded CD4<sup>+</sup> 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 reduce the putative number of causal SNPs accounting for a particular association in complex diseases. Fine-mapping using genotype level data incorporates a locus step-wise

## **Cross-tissue comparison analysis in PsA**

---

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 (Bowes et al. 2015). 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. 2015).

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, this data 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<sup>+</sup> monocytes, mCD4<sup>+</sup>, mCD8<sup>+</sup> 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<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> 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 (TJC66 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, in a scale of 1 to 5. These four measurements include joint and global assessment and form 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 immunosuppressive therapy.

## Cross-tissue comparison analysis in PsA

---

**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 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	TJC66/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)	— (±15.74)	44.3 (±59.6)	104	—	1.5(±0.5) /1.5(±0.5)	3 (±0.6)	3.2 (±0.8)	17.4 (±17.8)

## Cross-tissue comparison analysis in PsA

---

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

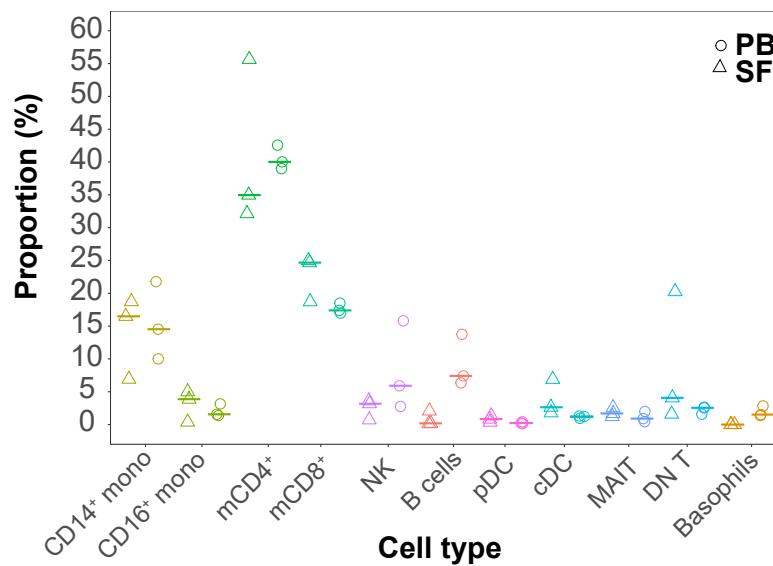
**Table 5.2: Datasets generated for each sample in the PsA cohort.** Four types of data were generated in paired synovial fluid and peripheral blood from the same individual. Fast-ATAC data was generated for CD14<sup>+</sup>, mCD4<sup>+</sup>, mCD8<sup>+</sup> and NK cells. RNA expression by PCR array was performed only for CD14<sup>+</sup>, mCD4<sup>+</sup> and mCD8<sup>+</sup> 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<sup>+</sup> (32.1-55.6%) constituted the most abundant cell type followed by mCD8<sup>+</sup> (16.9-24.9%) and CD14<sup>+</sup> "non-classical" monocytes (6.9-21.7%) (Figure 5.1). Consistent with previous studies, a trend of increased percentage of mCD8<sup>+</sup> 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

## Cross-tissue comparison analysis in PsA

cytotoxicity in PsA (Spadaro et al. 2004). Similarly, a tendency towards reduced proportions of B cells in synovial fluid compared to peripheral blood reinforced a lower contribution of the humoral immune response 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. Central DC=cDC, mucosal-associated invariant T=MAIT, DN=double negative.)

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

#### Data processing and quality control

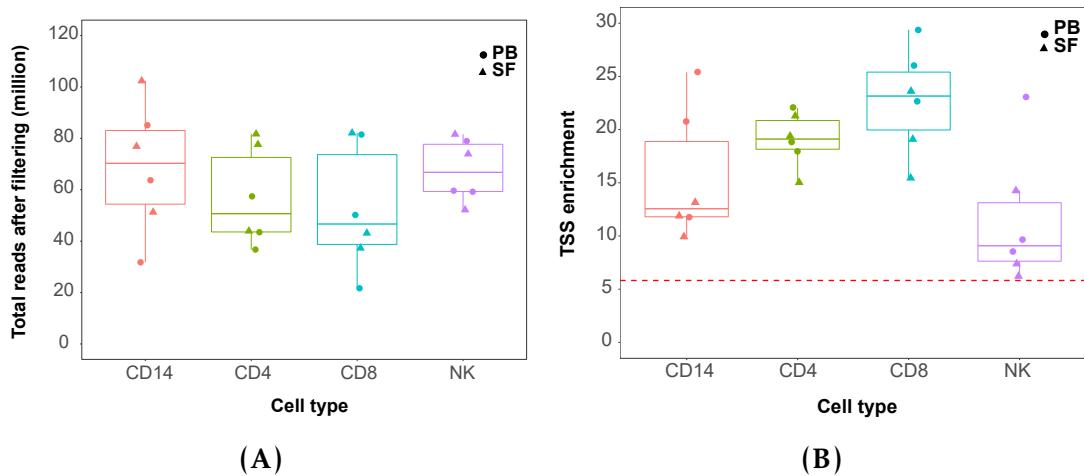
Chromatin accessibility was studied in the four most abundant cell types in PsA synovium fluid, CD14<sup>+</sup> monocytes, mCD4<sup>+</sup>, mCD8<sup>+</sup> and NK cells. The 24 ATAC samples (generated using Fast-ATAC protocol as detailed in Chapter 2) from four cell types were sequenced and processes using the in-house pipeline as previously detailed (Chapters 2 and 3). After filtering of duplicated and mitochondrial reads (Figure B.12A), 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<sup>+</sup> and mCD8<sup>+</sup> 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 threshold cut-off). Given the limited cohort size, these samples were not excluded, but it is worth noting that they could be contributing noise and thus reducing the power of the differential analysis. The number of accessible regions (peaks after filtering with p-values based on IDR analysis) ranged between 24x10<sup>3</sup> and 97x10<sup>3</sup> (Figure B.12B), and was also dependent of sample quality and cell type and no outliers were identified.

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 3) and PCA was conducted on the normalised counts mapping to each of those peaks.

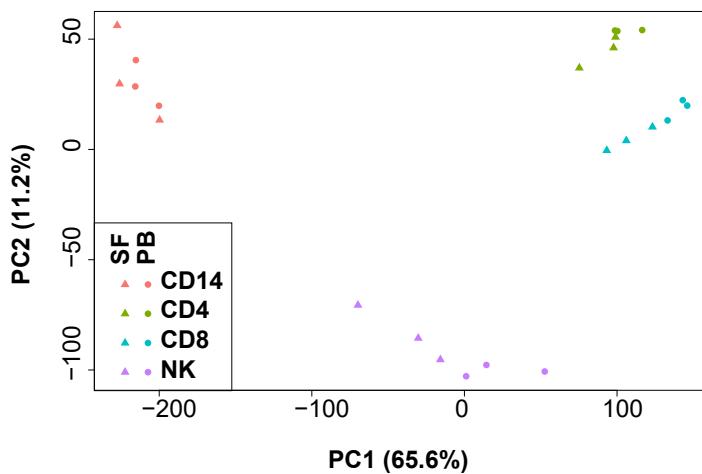
## Cross-tissue comparison analysis in PsA

---



**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 most distinct 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 (Figure 5.3).



**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 4 cell types ( $CD14^+$  monocytes, m $CD4^+$ , m $CD8^+$  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.

**CD14<sup>+</sup> monocytes present the greatest proportion of changes in chromatin accessibility**

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). Regions included in each consensus list of peaks were enriched for SNPs from the GTEx eQTL datasets, importantly whole blood, as well as for immune cells eQTL SNPs from publicly available data (Figure B.13), demonstrating the inclusion of functionally relevant regions in the downstream differential analysis.

Differential chromatin accessibility analysis between synovial fluid and peripheral blood using normalised counts retrieved for each of the cell type master lists using DESeq2 with a paired design (Table 5.3). Using an 80% cut-off for background noise (Chapter ??),  $CD14^+$  monocytes and NK cells presented

## Cross-tissue comparison analysis in PsA

---

the greatest total number (5,285 and 2,314, respectively) and proportion of DARs (23.3 and 8.9%, respectively).

Cell type	Total DARs	Proportion DARs (%)	Synovial fluid open DARs	Peripheral blood open DARs
CD14 <sup>+</sup>	5,285	23.3	3,779	1,506
CD4 <sup>+</sup>	1,329	4.3	621	708
CD8 <sup>+</sup>	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). SF=synovial fluid; PB=peripheral blood

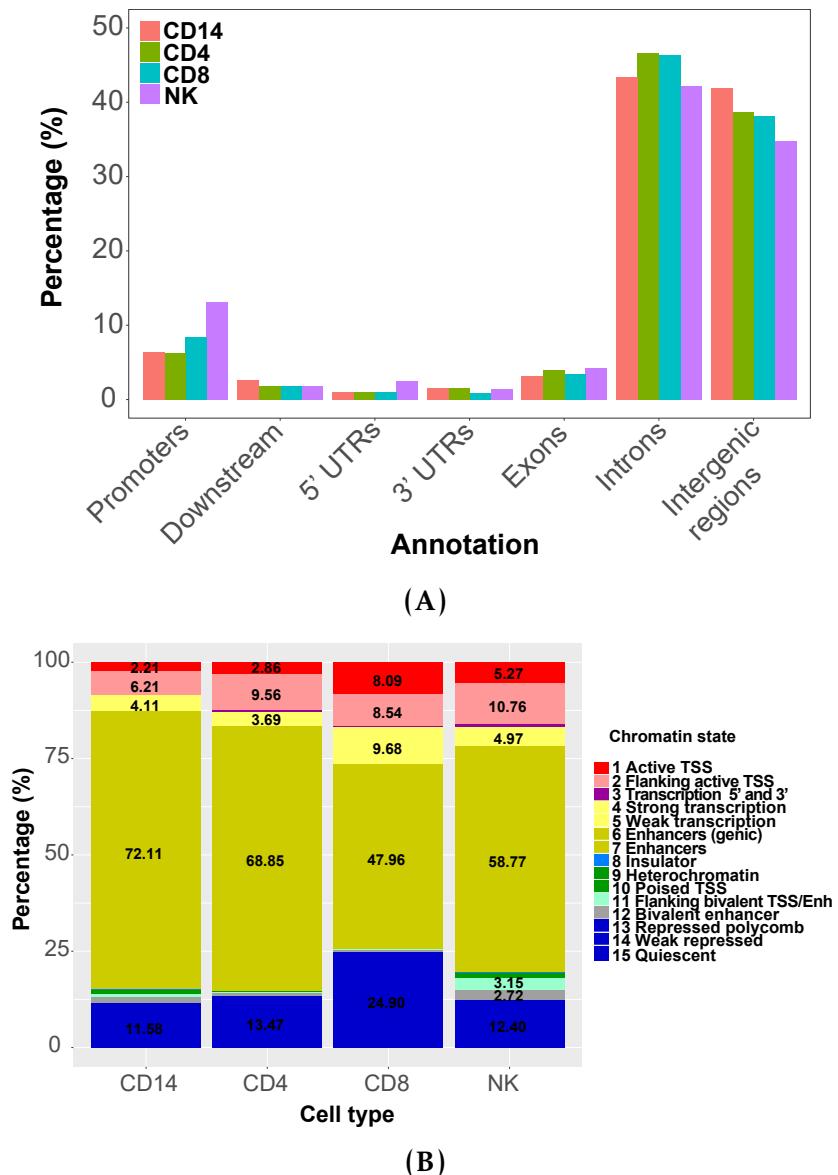
Considering the direction of the differential chromatin accessibility, DARs were divided in DARs more open in synovial fluid compared to peripheral blood (named here synovial fluid open DARs) and DARs less open in synovial fluid compared to peripheral blood (named here peripheral blood open DARs). In CD14<sup>+</sup> 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.

Permutation analysis using the unique ten possible permutations (given the small number of samples included in the analysis) did not show a greater number of DARs than the ones identified for the true groups, reinforcing the robustness of the differential analysis results (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).

## Cross-tissue comparison analysis in PsA

---



**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  $\leq 1,000\text{bp}$  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).

## **Cross-tissue comparison analysis in PsA**

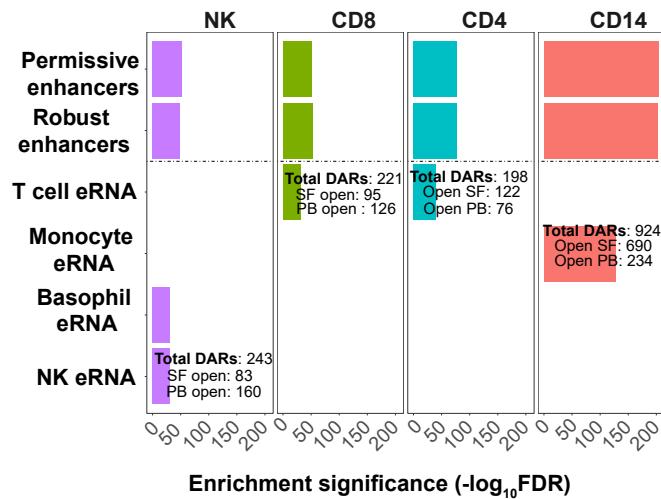
---

Universal promoter regions was 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 (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 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 significant enrichment for the corresponding cell type robust eRNA set. 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<sup>+</sup> T cells and 924 DARs (690 synovial fluid open and 234 open in peripheral blood) in CD14<sup>+</sup> monocytes.

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) annotated as weak or strong enhancers according to the cell type specific chromatin segmentation map previously illustrated.

## Cross-tissue comparison analysis in PsA

---



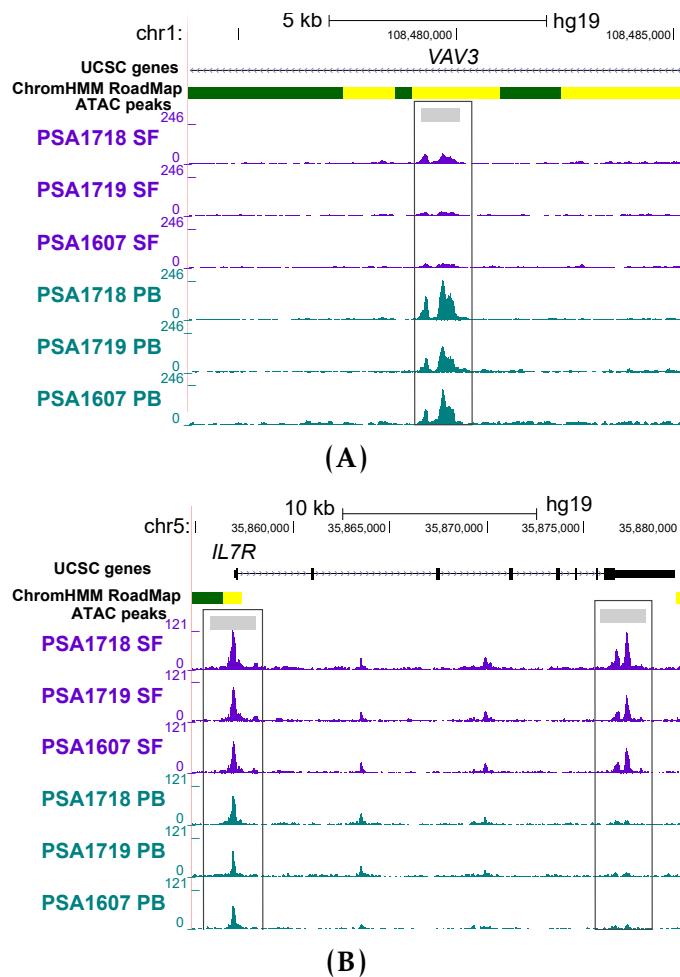
**Figure 5.5: Enrichment of PsA DOCs for the FANTOM5 eRNA dataset.** Robust enhancers have been defined as those detected at the genome-wide significant level in at least one primary cell type or tissue. Permissive enhancers are all detected eRNAs but not passing genome-wide filtering criteria (Andersson 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.

Cell type	DARs in gene body	Gene with more than one DAR	Enhancers	Introns
CD14 <sup>+</sup>	2,357	744	1,775	1,920
CD4 <sup>+</sup>	700	99	504	577
CD8 <sup>+</sup>	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 that 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.

## Cross-tissue comparison analysis in PsA

For example, differential chromatin accessibility analysis in NK 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 in CD14<sup>+</sup> monocytes of two DARs, located at the 5' and 3' UTRs of IL7R gene, were more accessible in synovial fluid compared to peripheral blood (Figure 5.6B).



**Figure 5.6: Differentially accessible regions located within gene bodies in CD14<sup>+</sup> 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<sup>+</sup> 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 FC>1.5. SF=synovial fluid; PB=peripheral blood.

When compared to all ATAC regions in each consensus list of peaks, DARs ionly showed to be significantly enriched for whole blood GTEx SNPs eQTLs (CD14<sup>+</sup> monocytes FDR=2.7x10<sup>-6</sup>, mCD4<sup>+</sup> FDR=7.9x10<sup>-3</sup>, mCD8<sup>+</sup> FDR=1.6x10<sup>-2</sup>, and NK FDR=1.9x10<sup>-3</sup>) but not for any of the other GTEx or immune cell types publicly available eQTL datasets. In all four cell types, overlap between putative psoriasis and PsA GWAS genes and proximal ( $\leq 5\text{Kb}$ ) DARs were found. CD14<sup>+</sup> monocytes presented the largest number of overlaps (13), followed by mCD8<sup>+</sup> (9), NK<sup>+</sup> (8) and mCD4<sup>+</sup> (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). This analysis using a total of 20,000 permutations only found significant enrichment of GWAS LD blocks for DARs between synovial fluid and peripheral blood in CD14<sup>+</sup> monocytes (empirical p-value=0.043).

### 5.3.4 Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility

To explore the potential function 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 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

## **Cross-tissue comparison analysis in PsA**

---

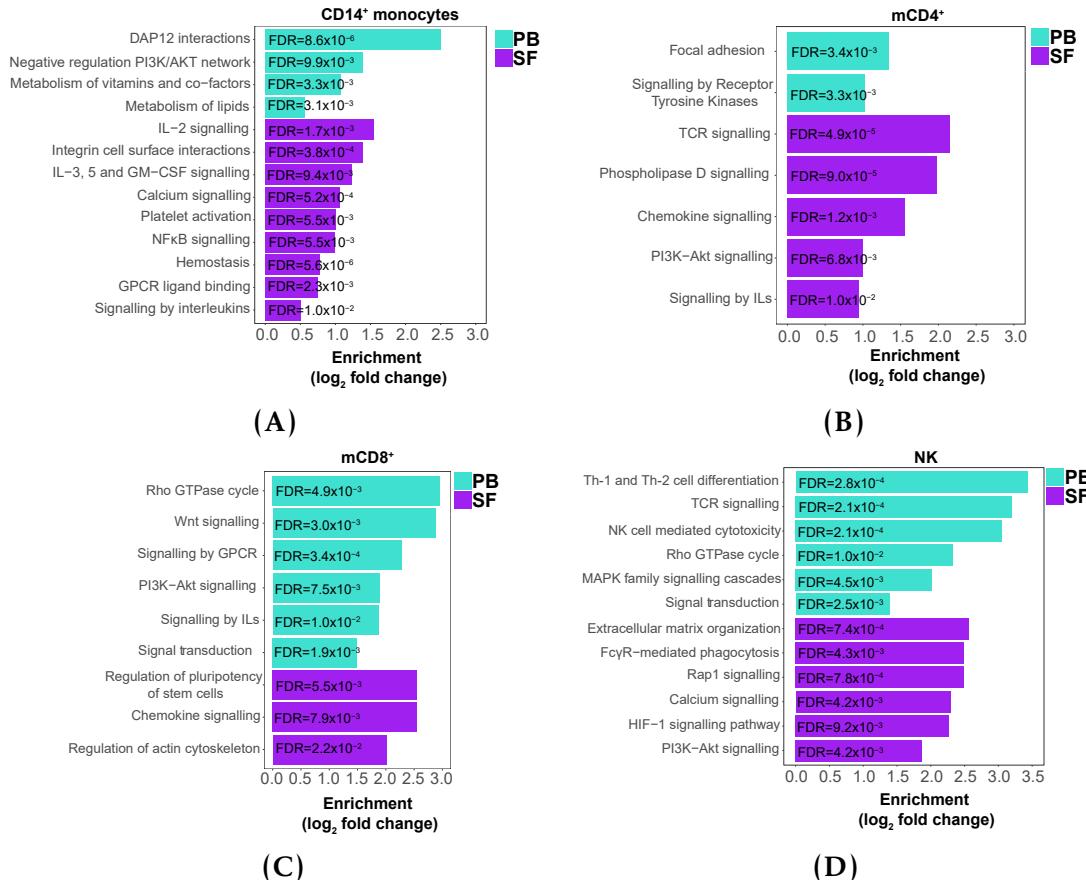
and vice versa (Figure 5.7). In CD14<sup>+</sup> monocytes, synovial fluid open DARs showed enrichment for pathways involved in hemostasis, integrin interactions, calcium signalling and regulation of immunity, inflammation and cell survival, importantly the NF- $\kappa$ B pathway (Figure 5.7A in purple). Synovial fluid open DARs in CD14<sup>+</sup> 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<sup>+</sup> monocytes were not enriched for any of these pathways, showing enrichment for DAP12 interactions and regulation of the PI3K/AKT network (Figure 5.7A in turquoise).

mCD4<sup>+</sup> 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- $\gamma$  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<sup>+</sup> 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<sup>+</sup> peripheral blood open DARs, as well as the Wnt signaling pathway (Figure 5.7C in turquoise). mCD8<sup>+</sup> synovial fluid open DARs showed enrichment for chemokine signalling and regulation of stem cell pluripotency (Figure 5.7C in purple).

NK synovial fluid open DARs showed enrichment for extracellular matrix organisation and RAP-1 signalling as well as for Fc $\gamma$  receptor (Fc $\gamma$ R)-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,

## Cross-tissue comparison analysis in PsA

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



**Figure 5.7: Functional pathways significantly enriched and unique for either synovial fluid open or peripheral blood open DARs in CD14<sup>+</sup> monocytes, CD4m<sup>+</sup>, CD8m<sup>+</sup> and NK cells.** Enrichment analysis was performed separately for genes annotating synovial fluid open DARs and peripheral blood open DARs. Log<sub>2</sub> 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<sup>+</sup> monocytes, (B) mCD4<sup>+</sup>, (C) mCD8<sup>+</sup> and (D) NK cells are shown. SF=synovial fluid; PB=peripheral blood.

### 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, qPCR gene expression analysis for 370 key genes in the inflammatory and autoimmune response was conducted in

## **Cross-tissue comparison analysis in PsA**

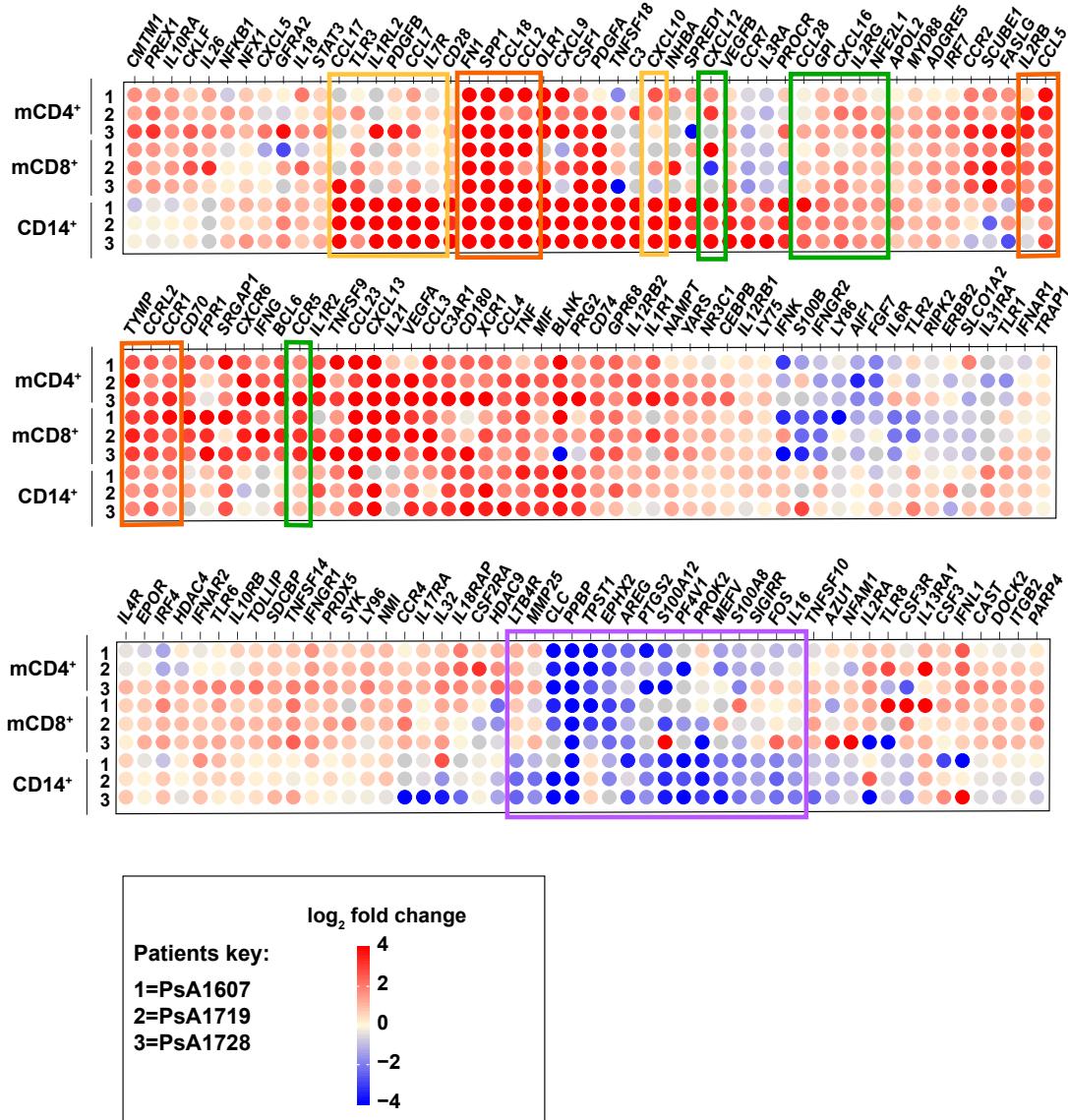
---

CD14<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> 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 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 between synovial fluid and peripheral blood showed differences in magnitude and reproducibility across the three cell types (Figure 5.8). Some of the modulated genes showed marked up-regulation (fold change>1.5) in synovial fluid compared to peripheral blood in the 3 cell types, for example *FN1*, *SPP1* or *CCL2*, amongst others (Figure 5.8 orange box). On the other hand, a number of genes presented reduced expression in synovial fluid (fold change<1.5) in at least one of the three cell types, including *FOS*, *IL16*, *PPBP* and *TPST1* (Figure 5.8 purple box). A number of genes demonstrated consistent change in the same direction in the three CD14<sup>+</sup> monocyte samples but not in T cells (Figure 5.8 dark yellow box), for example *CXCL10*. *CCR7* and *IL7R* were up-regulated in the synovial fluid of CD14<sup>+</sup> monocytes compared to peripheral blood; however changes between synovial fluid and peripheral blood were not consistent across patients in mCD4<sup>+</sup> and mCD8<sup>+</sup>. Moreover, differences in fold change were also observed for genes modulated in the same direction across the three cell types, for instance *VEGFB* and *CCL17* (Figure 5.8 green box).

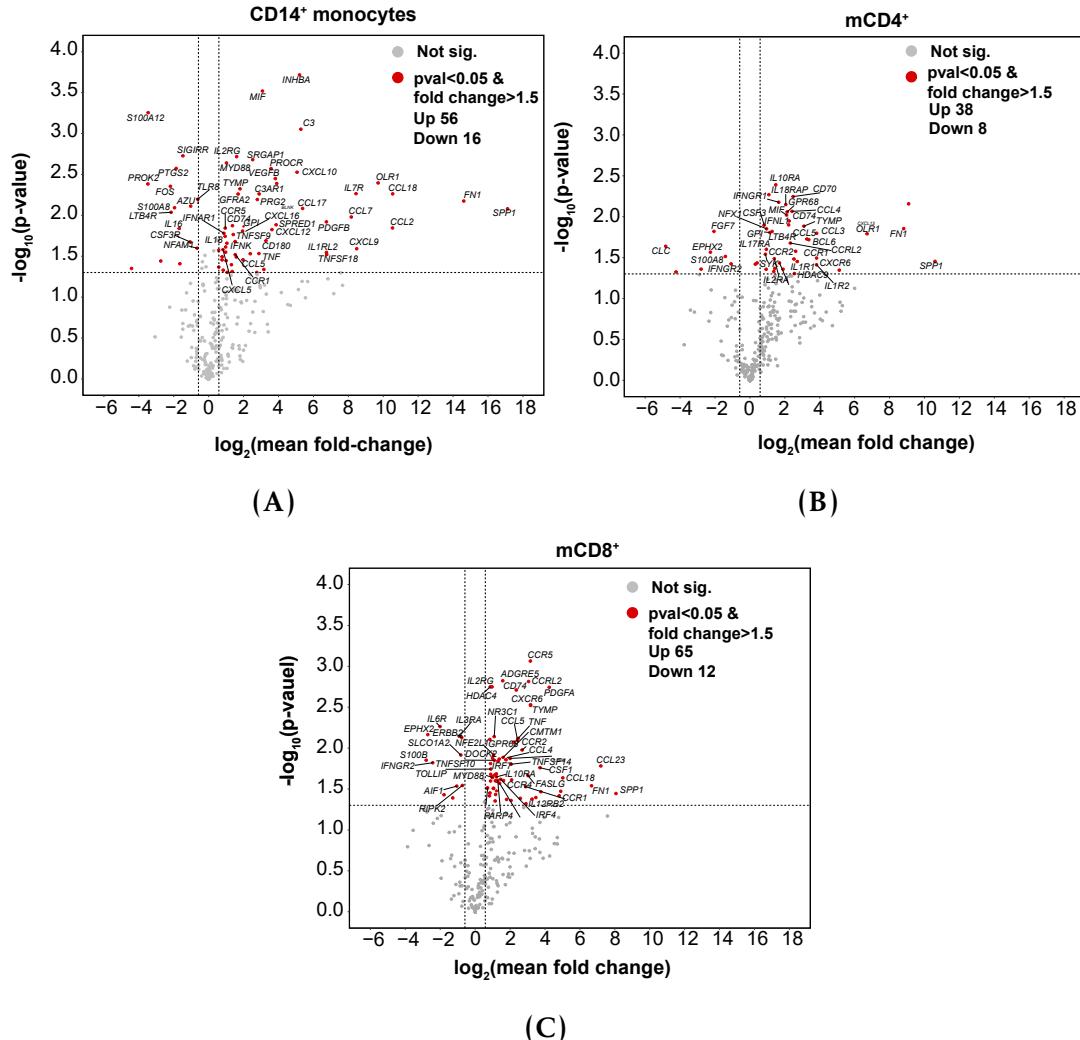
The majority of the modulated genes between synovial fluid and peripheral blood did not reach significance (FDR<0.05 or 0.1) after multiple testing correction, likely due to the small sample size. Following this, filtering using a p-value<0.05 as the statistical threshold for significance and a mean fold change>1.5, CD14<sup>+</sup> monocytes and mCD8<sup>+</sup> showed greater number of

## Cross-tissue comparison analysis in PsA



**Figure 5.8: Heatmap illustrating fold changes in gene expression between synovial fluid and peripheral blood across CD14<sup>+</sup> monocytes and mCD4<sup>+</sup> and mCD8<sup>+</sup> 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 ( $p\text{-value} < 0.05$ ) 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.

## Cross-tissue comparison analysis in PsA



**Figure 5.9: Gene expression changes in immune-relevant genes between synovial fluid and peripheral blood in CD14<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> cells.** Volcano plots showing differences in gene expression measured by qPCR array between synovial fluid and peripheral blood for (A) CD14<sup>+</sup> monocytes, (B) mCD4<sup>+</sup> and (C) mCD8<sup>+</sup> cells. The significance ( $\log_{10} p\text{-value}$ ) 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  $p\text{-value}<0.05$  and mean fold change  $>1.5$  in red, with the most significant genes labelled.

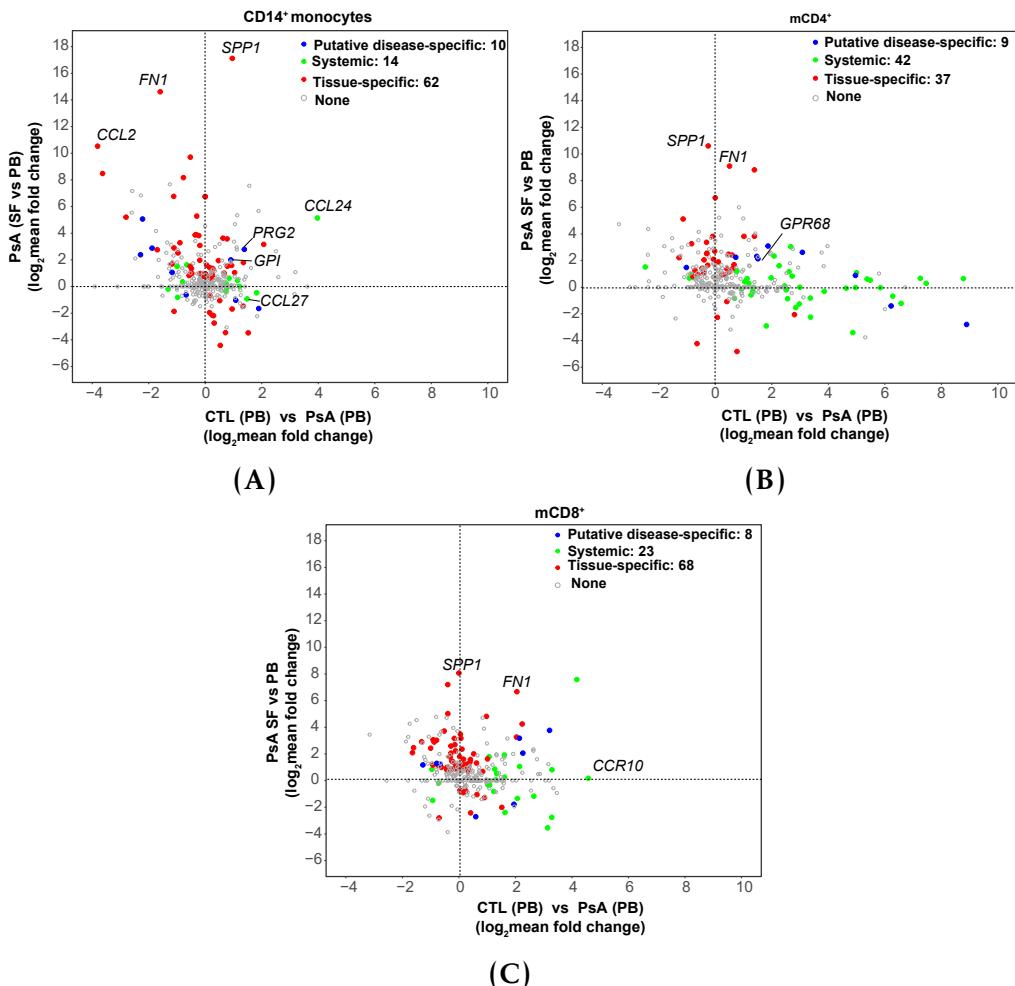
## **Cross-tissue comparison analysis in PsA**

---

significantly modulated genes (72 and 77, respectively) compared to mCD4<sup>+</sup> 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<sup>+</sup> monocytes showed mean fold change>1.5 versus the 14 genes with mean fold change<1.5 (Figure 5.9A).

Gene expression in peripheral blood of the three PsA patients was also compared to peripheral blood of the three healthy controls for CD14<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> cells. P-values and fold changes were calculated between the mean peripheral blood expression across 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).

## Cross-tissue comparison analysis in PsA



**Figure 5.10: Comparison of immune-relevant gene expression modulation across PsA tissues (synovial fluid vs peripheral blood) and in PsA patients verus 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<sup>+</sup> monocytes, (B) mCD4<sup>+</sup> and (C) mCD8<sup>+</sup> 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 ( $p\text{-value} < 0.05$ ) for either of the two contrasts.

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<sup>+</sup> monocytes. A second group of genes were designated as tissue-specific, since they were significantly

## **Cross-tissue comparison analysis in PsA**

---

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 *CCR10* in mCD8<sup>+</sup> cells plus *SPP1* and *FN1* in the three cell types. These two 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<sup>+</sup> 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 was only found to be significant in CD14<sup>+</sup> monocytes (Fisher exact test p-value=0.028) and not in mCD4<sup>+</sup> and mCD8<sup>+</sup> T cells (Fisher exact test p-value=0.466 and 0.173, respectively).

In CD14<sup>+</sup> monocytes, 13 out of the 56 significantly up-regulated genes in synovial fluid also showed to be proximal to a synovial fluid open DARs. 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 synovial biopsies compared to peripheral blood has been previously reported (Dolcino et al. 2015). In this cohort, *FN1* expression was up-

## Cross-tissue comparison analysis in PsA

---

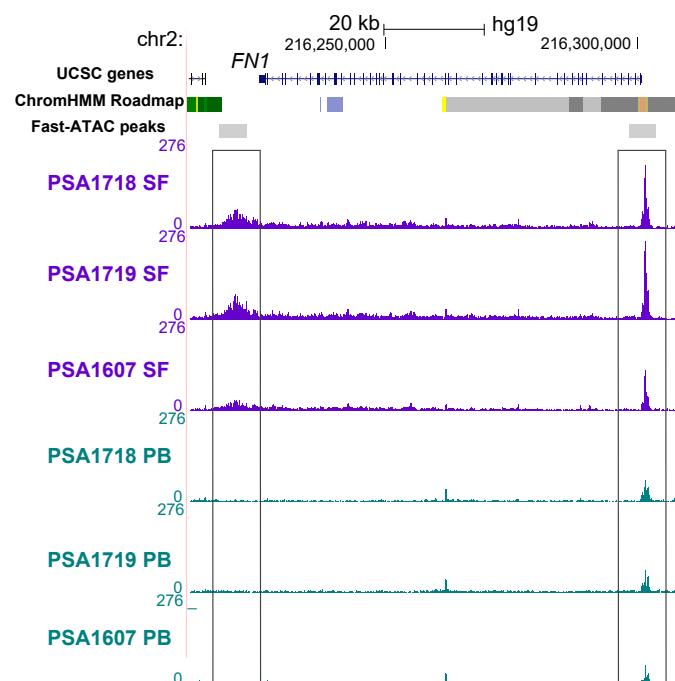
Cell type	Genes up-regulated and overlapping open chromatin in SF	Genes down-regulated and overlapping closed chromatin in SF
CD14 <sup>+</sup> monocytes	13 ( <i>BLNK, CCL2*, CCR1*, CD180, CXCL10, FN1, IL18, IL31RA*, IL7R*, NFKB1*, PRG2, SRGAP1, STAT3</i> )	2 ( <i>FOS, PROK2*</i> )
mCD4 <sup>+</sup>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

**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.

regulated in synovial fluid for all three cell types, with the highest fold change found in CD14<sup>+</sup> 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<sup>+</sup> monocytes in cytokine and chemokine production

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 abs 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<sup>+</sup> 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 degranulation" pathways. Of particular relevance in monocytes were the chemokine, NOD-like and TLR signalling pathways, involved in the



**Figure 5.11: Chromatin accessibility landscape at the qPCR differentially expressed *FN1* gene in CD14<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> monocytes is also shown. All DARs were significant based on FDR<0.01 and fold change>1.5. SF=synovial fluid; PB=peripheral blood.

## **Cross-tissue comparison analysis in PsA**

---

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 *NFKB*, contributed to the enrichment of these 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<sup>+</sup> monocytes (Figure 5.9A). Genes such as *TNF*, *IRF7* and *MYD88* particularly highlighted the cross-link between the NOD-like and the TLR signalling pathways.

The enrichment for the chemokine pathway in CD14<sup>+</sup> 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 (Figure 5.9) as well as genes only consistently modulated between synovial fluid and peripheral blood in CD14<sup>+</sup> monocytes (e.g *CCL28*, Figure 5.8 green box).

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

## Cross-tissue comparison analysis in PsA

---

**Table 5.6: Pathway enrichment analysis for the modulated genes between synovial fluid and peripheral blood in CD14<sup>+</sup> and mCD4<sup>+</sup>.**  
 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 <sup>+</sup>	Chemokine signalling (fold change=1.53)	CCL17, CCL18, CCL2, CCL28, CCL5, CCR1, CCR5, CXCL10 CXCL12, CXCL16, CXCL5, CXCL9, NFKB1, PPBP, PF4V1, STAT3, XCR1
	NOD-like receptor signalling (fold change=1.85)	CCL2, CCL5, IFNAR1, IL18, IRF7, MEFV, MYD88, NFKB1, NAMPT, TNF
mCD4 <sup>+</sup>	TLR signalling (fold change=1.64)	CCL5, CXCL10, CXCL9, IFNAR1, IRF7, MYD88, NFKB1, SPP1, FOS, TLR1, TLR2, TLR8, TNF
	IL-10 signaling (fold change=2.13)	CCL3, CCL4, CCL5, CCR1, CCR2, CSF1, CSF3, IL10RA, IL1R1, IL1R2

## **Cross-tissue comparison analysis in PsA**

---

**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<sup>+</sup> monocytes and mCD4<sup>+</sup> cells. This reinforced the tissue-specificity of the pathways enriched for differentially expressed between synovial fluid and peripheral blood, and suggests a more pronounced inflammatory phenotype of the pathological CD14<sup>+</sup> monocytes in synovial fluid compared to peripheral blood.**

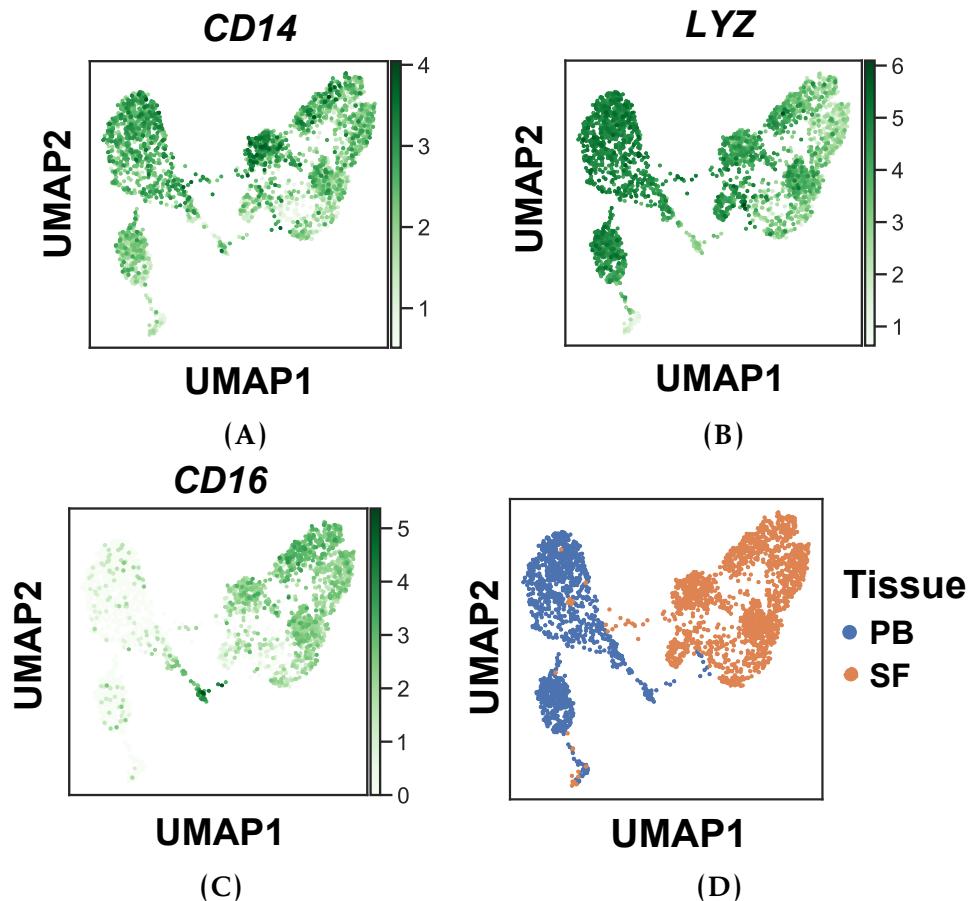
Overall, the integration of the chromatin accessibility and immune transcriptional data reinforced a relevant role of synovial CD14<sup>+</sup> monocytes in the production of cytokines and chemokines, likely leading to activation of the innate immune response and the recruitment of T cells to this site of inflammation.

### **5.3.6 Characterisation of CD14<sup>+</sup> monocyte heterogeneity in PsA using scRNA-seq**

According to the analysis of chromatin accessibility and immune-related gene expression in this pilot cohort, CD14<sup>+</sup> 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 highlight 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

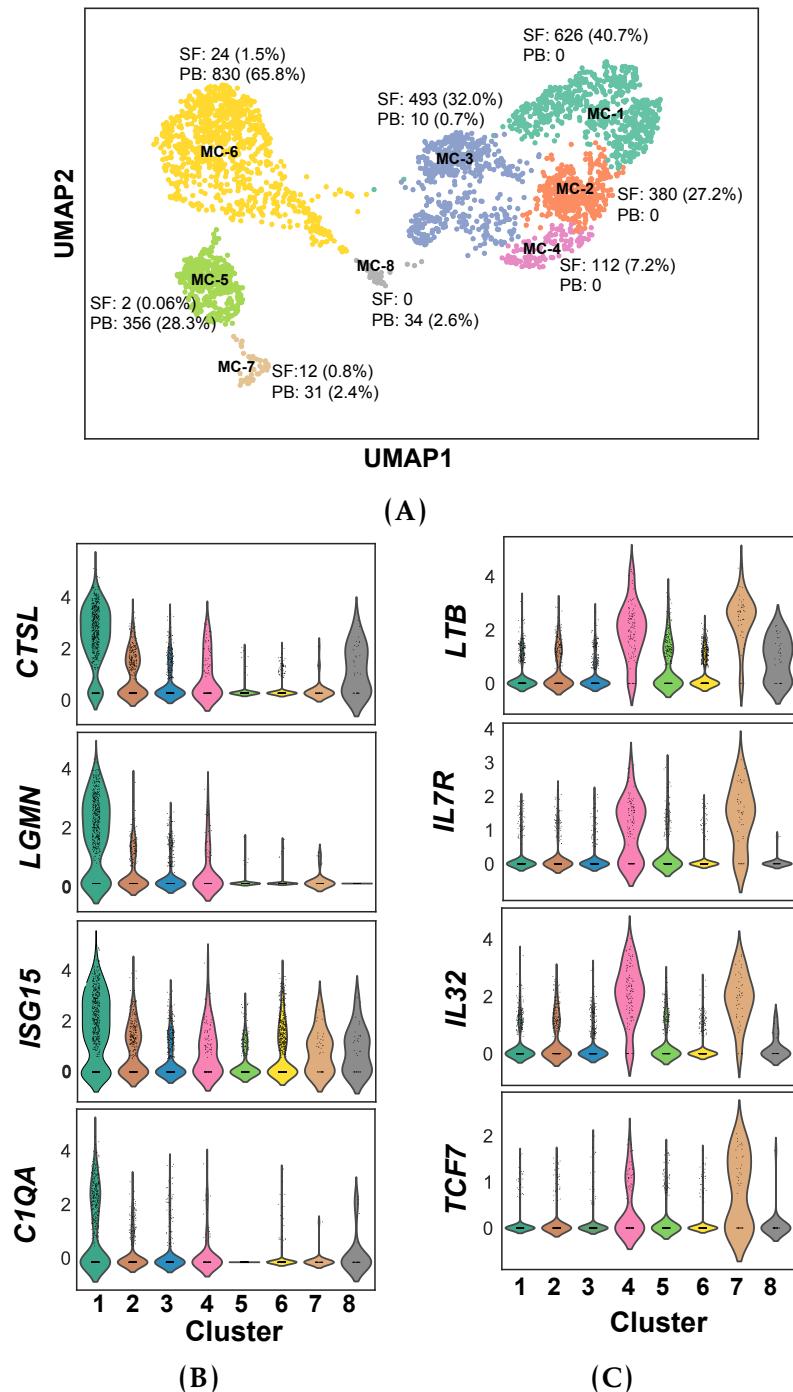
First, scRNA-seq analysis was performed in paired PBMCs and SFMCs isolated from three PsA patients (Table 5.2). scRNA-seq data from each of the PBMCs and SFMCs samples were downsampled to 3,000 cells and filtered to remove genes that were expressed only in few cells and those cells expressing low number of genes (as described in Chapter 2). The monocyte population sorted by FACS was characterised for expressing the surface marker CD14<sup>+</sup>. For consistency, the scRNA-seq monocyte population for downstream analysis was identified based on expression of *CD14* and *LYZ*, two canonical markers of CD14<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> monocytes previously reported using cell surface markers by mass cytometry (Figure 5.1). Interestingly, expression of the low-affinity immunoglobulin G Fc $\gamma$  receptor (Fc $\gamma$ RIIa) CD16<sup>+</sup> by the synovial fluid CD14<sup>+</sup> monocytes suggested a more intermediate phenotype (CD14<sup>+</sup> CD16<sup>+</sup>) in this tissue compared to peripheral blood monocytes showing a more classical phenotype (CD14<sup>+</sup> CD16<sup>-</sup>) (Figure 5.12C). Subsequent visualisation using uniform manifold approximation and projection (UMAP) clearly separated the synovial fluid and peripheral blood CD14<sup>+</sup> monocyte populations (Figure 5.12D).



**Figure 5.12: Visualisation using UMAP dimensional reduction of the synovial fluid and peripheral blood CD14<sup>+</sup> monocytes.** Visualisation using UMAP dimensional reduction of (A) *CD14*, (B) *LYZ* and *CD16* expression intensities in the combined synovial fluid and peripheral blood CD14<sup>+</sup> monocytes cells. (C) Overlaid colour-coded tissue identity for each of the CD14<sup>+</sup> 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<sup>+</sup> monocytes using standard default resolution of the Leiden clustering algorithm (detailed in Chapter ??). Using these parameters, 8 CD14<sup>+</sup> monocytes 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).

## Cross-tissue comparison analysis in PsA



**Figure 5.13: Identification and characterisation of CD14<sup>+</sup> monocytes subpopulations in synovial fluid and peripheral blood.** (A) Visualisation using UMAP dimensional reduction of the CD14<sup>+</sup> monocytes subpopulations identified in the combined synovial fluid and peripheral blood CD14<sup>+</sup> 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 characteristic of the (B) MC-1 and (C) MC-4 subpopulations. SF=synovial fluid; PB=peripheral blood.

## **Cross-tissue comparison analysis in PsA**

---

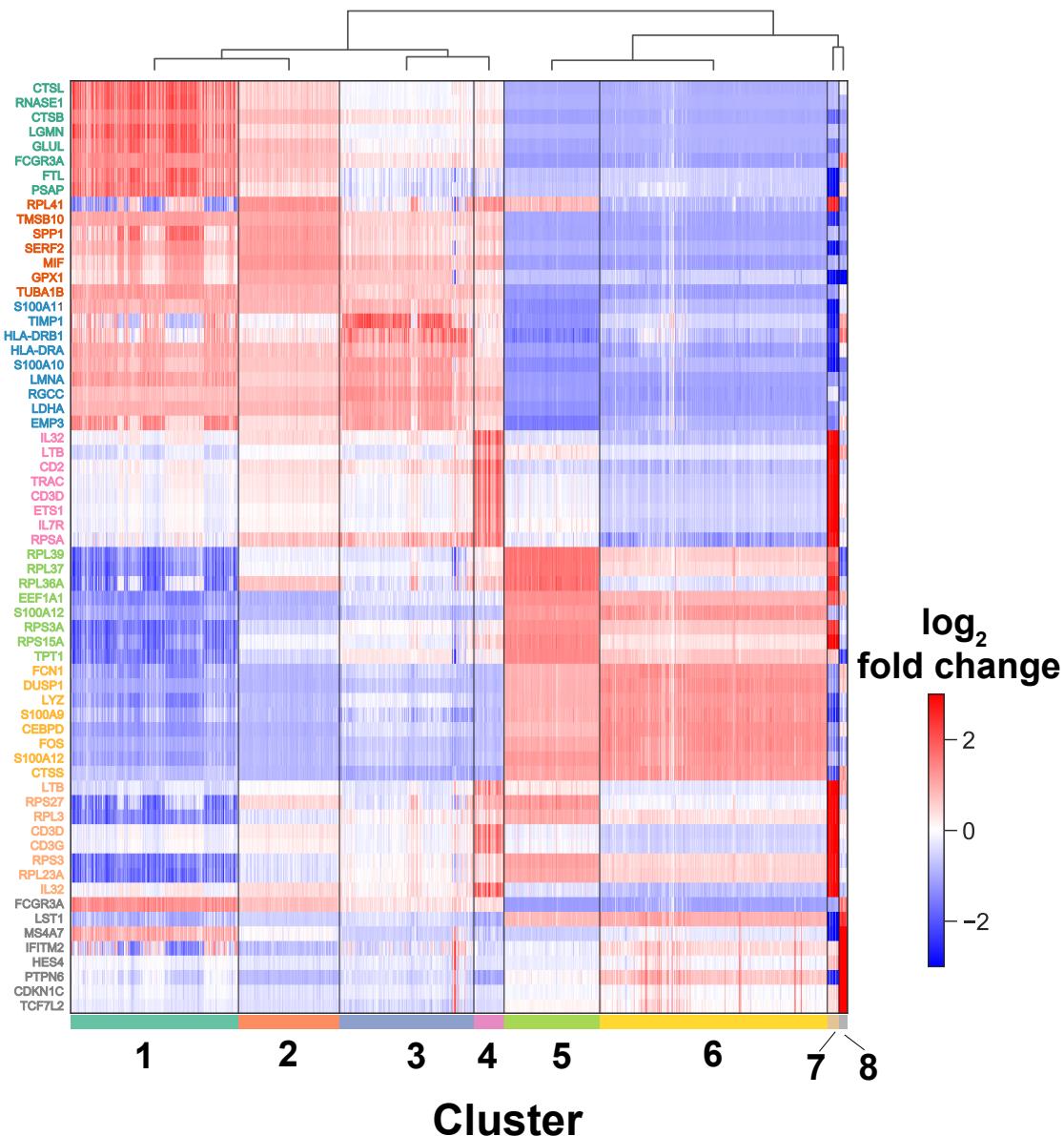
MC-7 was the cluster with the most even contribution of CD14<sup>+</sup> monocytes from the two tissues (0.8 and 2.4% from synovial fluid and peripheral blood, respectively). MC-7 may be a real cluster or result from remaining unfiltered layouts or cell miss classification by the clustering algorithm and additional sample size will be required to confirm this cluster identity. Each of the 8 CD14<sup>+</sup> monocytes clusters was characterised by a number of marker genes, a set of 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 highlighted only moderate differences across clusters mostly formed by monocytes of the same tissue, with very distinct pattern of expression of the markers genes across clusters from different tissues (Figure 5.14).

### **Functional exploration of synovial fluid and peripheral blood CD14<sup>+</sup> 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- $\gamma$  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 (Figure 5.15A).

MC-4 represented 7.2% of synovial fluid CD14<sup>+</sup> monocytes and was characterised by high expression of *IL7R*, the IL-7 regulated cytokines *LTB*, the TF *TCF7*, the pro-inflammatory cytokine induced of TNF- $\alpha$  *IL-32* (**Kim2005**)

## Cross-tissue comparison analysis in PsA



**Figure 5.14: Heatmap and hierarchical clustering for the top eight marker genes of the CD14<sup>+</sup> 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. Dendrogram indicating hierarchical relationship between clusters is indicated at the top and cluster identity for each of the cells is specified at the bottom.

## Cross-tissue comparison analysis in PsA

---

and the chemokine *CCL5* (Figures 5.13C and 5.14). These genes were also up-regulated by the CD14<sup>+</sup> monocyte subpopulation expressing IL7R<sup>+</sup> upon *in vitro* LPS stimulation conducted by Al-Mossawi and colleagues (in revision, Al-mossawi *et al.*, 2018). MC-4 cluster showed differential expression of 340 genes (228 and 112 up- and down-regulated, respectively) when compared to the rest of monocytes from the other clusters. These included the immunoregulatory non-coding RNA *MALAT1* and the stabiliser of the TCR *TRAT1*, 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<sup>+</sup> CD14<sup>+</sup> monocytes upon LPS stimulation in the aforementioned study (Al-Mossawi et al. 2018) (overlap of 116 genes, p-value=1.9x10<sup>-13</sup>, 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<sup>+</sup> monocytes responsive genes 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<sup>+</sup> monocytes, suggesting that the subpopulation of CD14<sup>+</sup> monocytes expressing *IL7R* (amongst the other markers) may be predominant and distinctive of the PsA synovial CD14<sup>+</sup> monocytes. This is consistent with the synovial fluid open DARs found at the promoter and 3'UTR of the transcriptional up-regulation (Figures 5.6A and 5.9A).

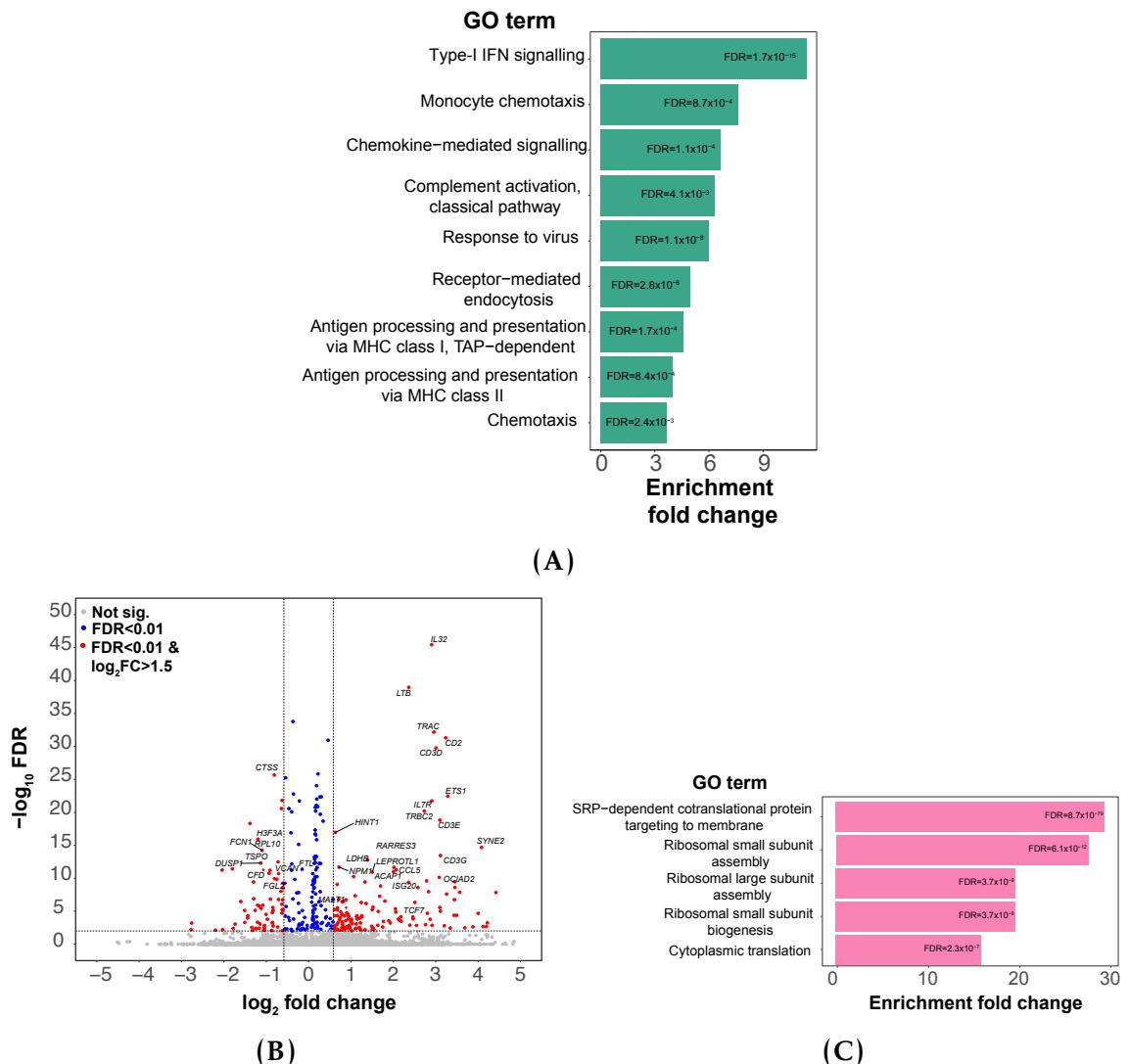
Synovial fluid CD14<sup>+</sup> monocytes, particularly the MC-1 and MC-3 subpopulations expressed significant higher levels of a number of MHC-II genes, including *HLA-DRB1*, *HLA-DPB1* or *HLA-DRB5*, involved in the monocytes antigen presentation and T cell activation, in line with the intermediate

## **Cross-tissue comparison analysis in PsA**

---

monocytes phenotype (Figures B.17A and 5.14). Moreover, peripheral blood CD14<sup>+</sup> 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 closer classic phenotype compared to the synovial fluid counterparts (Figure B.17B), as well as genes from the S100 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.

## Cross-tissue comparison analysis in PsA



**Figure 5.15: Functional characterisation of the MC-1 and MC-4 CD14<sup>+</sup> 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<sup>+</sup> monocytes interrogated by scRNA-seq. The significance (-log<sub>10</sub>FDR) of the differential expression (y-axis) is plotted against the log<sub>2</sub>fold change. Genes are coloured based on FDR and log<sub>2</sub>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.

### 5.3.7 Identification of additional pathways dysregulated between synovial fluid and peripheral blood CD14<sup>+</sup> monocytes using scRNA-seq data

Differential gene expression between the CD14<sup>+</sup> 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 ??A). Out of the 72 DEGs identified by qPCR, 38 were also differentially expressed (FDR<0.01) by scRNA-seq analysis (32 when further restricting on fold 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 (Figure A.7). Up-regulation in synovial fluid CD14<sup>+</sup> 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*, being the last one 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

## **Cross-tissue comparison analysis in PsA**

---

inflammasome were down-regulated in PsA synovial fluid CD14<sup>+</sup> monocytes compared to peripheral blood.

Other relevant enriched pathway identified in this analysis included IFN and IFN- $\gamma$  signalling, MHC-II antigen presentation, complement cascade, apoptosis and osteoclast differentiation (Figure 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 was also driven by up-regulation of the TF *STAT1* in synovial fluid CD14<sup>+</sup> 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<sup>+</sup> monocytes. This may suggest a increased resistance to apoptosis by PsA synovial fluid monocytes. Moreover, CD14<sup>+</sup> monocytes from the synovium showed large up-regulation of the *TNFRSF11A* (also know as *RANK*) (fold change=26.5, Figure ??A) involved in monocytes differentiation to osteoclasts and NF- $\kappa$ B activation (Mensah et al. 2008), and down-regulation of *NCF1*, which impairs extracellular degradation of inflammatory mediators by neutrophil proteases (Schauer et al. 2014).

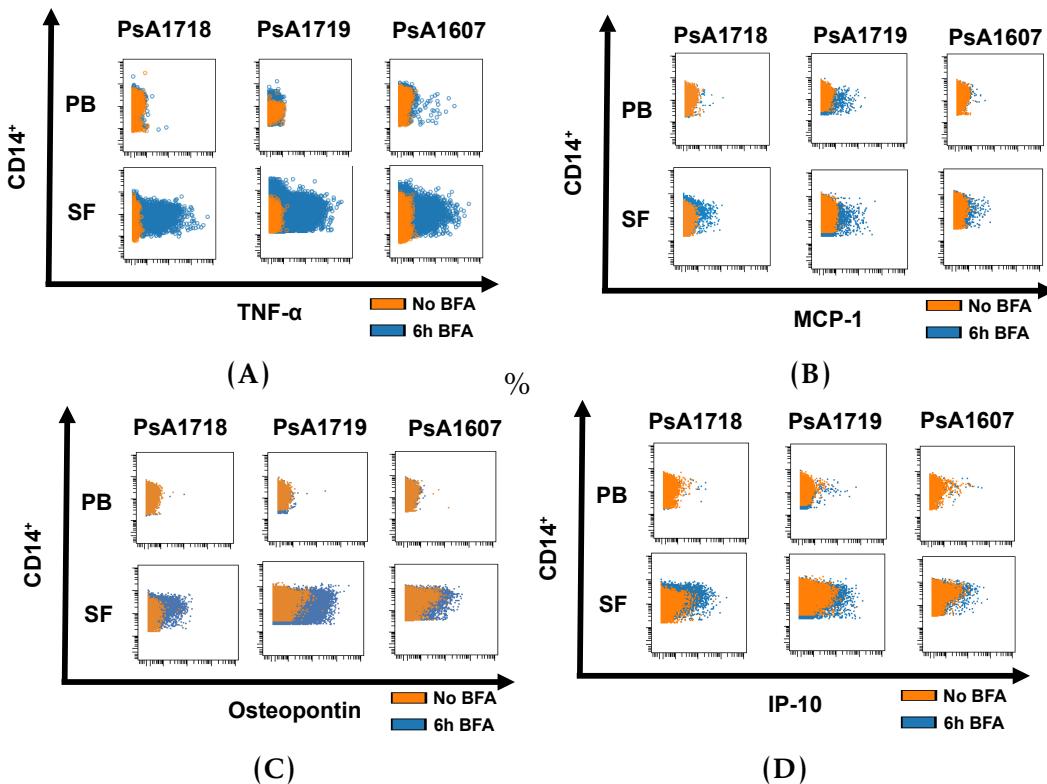
Lastly, genome-wide comparison between the log<sub>2</sub>fold changes for all the genes analysed by scRNA-seq expression and all the regions interrogated for differential chromatin accessibility between synovial fluid and peripheral blood only showed low correlation ( $R=0.162$ ,  $p\text{-value}=2\times 10^{-16}$ ) (Figure ??B), with 252 and 75 genes changing expression and chromatin accessibility in the same or opposite direction, respectively (significant enrichment,  $p\text{-value}=6.264\times 10^{-11}$  Fisher exact test). For example, *CCL2*, *FN1* and *FABP5* showed transcriptional up-regulation and were in the proximity of a synovial fluid open DAR (Figure ??C).

### **5.3.8 Integration of mass cytometry differences in protein expression with chromatin accessibility and gene expression data in PsA CD14<sup>+</sup> monocytes.**

To determine the effect of chromatin accessibility and genes expression differences in CD14<sup>+</sup> monocytes between synovial fluid and peripheral blood at the protein level, mass cytometry was conducted in collaboration with Dr Hussein Al-Mossawi and Dr Nicole Yager. This analysis included the six samples with either paired ATAC/qPCR data or scRNA-seq analysis presented in this chapter (Table 5.2) and additional 4 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- $\alpha$ , 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<sup>+</sup> monocytes compared to peripheral blood and have a key role in pathological inflammation. In the three PsA patients processed for ATAC and qPCR mass cytometry expression in CD14<sup>+</sup> of TNF- $\alpha$ , osteopontin, MCP-1 and IP-10 demonstrated a greater increase of cell staining for each cytokine after 6h of BFA treatment in CD14<sup>+</sup> monocytes from synovial fluid when compared to peripheral blood (Figure 5.16).

## Cross-tissue comparison analysis in PsA



**Figure 5.16: Biaxial representation of  $CD14^+$  monocytes expressing TNF- $\alpha$ , 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  $CD14^+$  versus (y-axis) versus (A) TNF- $\alpha$ , (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).

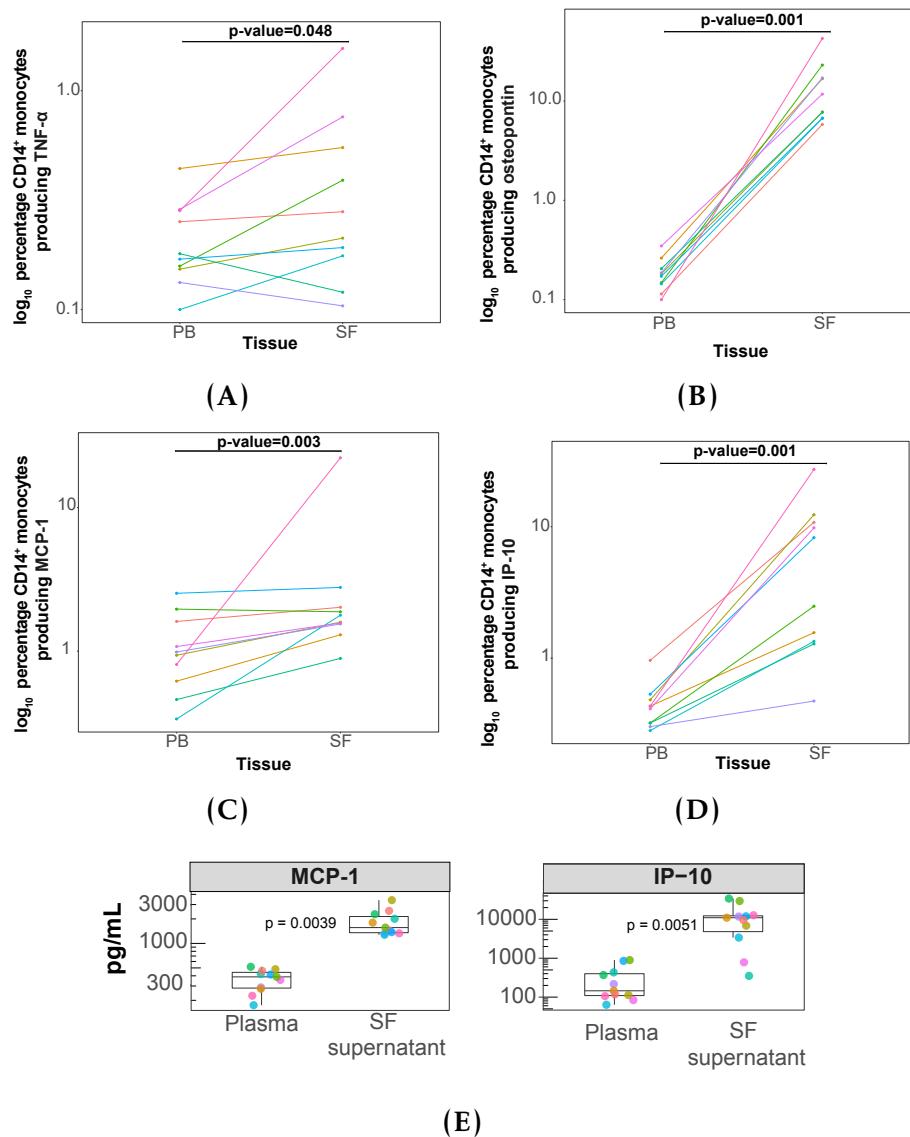
Quantification of the percentage of  $CD14^+$  monocytes producing these four inflammatory mediators was defined as the difference between percentage of cells with positive staining for each cytokine before and after BFA treatment in each of the ten PsA patients. The percentage of  $CD14^+$  monocytes producing TNF- $\alpha$  was significantly greater in synovial fluid compared to peripheral blood (Wilcoxon signed-rank test, p-value=0.048) (Figure 5.17A), consistent with increased chromatin accessibility nearby genes of the NF- $\kappa$ B signalling pathway and enrichment of DEGs for the TLR and NOD-like signalling pathways. Likewise, a larger 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) were also observed in synovial fluid compared to

## **Cross-tissue comparison analysis in PsA**

---

peripheral blood (Figure 5.17B, C and D). The increased percentage of cells producing MCP-1 and osteopontin 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, whereas up-regulation of CXCL10 in synovial fluid CD14<sup>+</sup> monocytes was only detected by qPCR. The differences between the percentage of cells producing cytokines in synovial fluid and peripheral blood was particularly moderate for TNF- $\alpha$  and MCP-1 for most of the samples (Figure 5.17A and C). PsA1807 presented particularly higher percentage of synovial fluid CD14<sup>+</sup> monocytes producing all 4 molecules when compared to the rest of the patients in the cohort (Figure 5.17A, B and C in pink). Nevertheless, the statistical significance remained when removing PsA1807 from the analysis.

## Cross-tissue comparison analysis in PsA

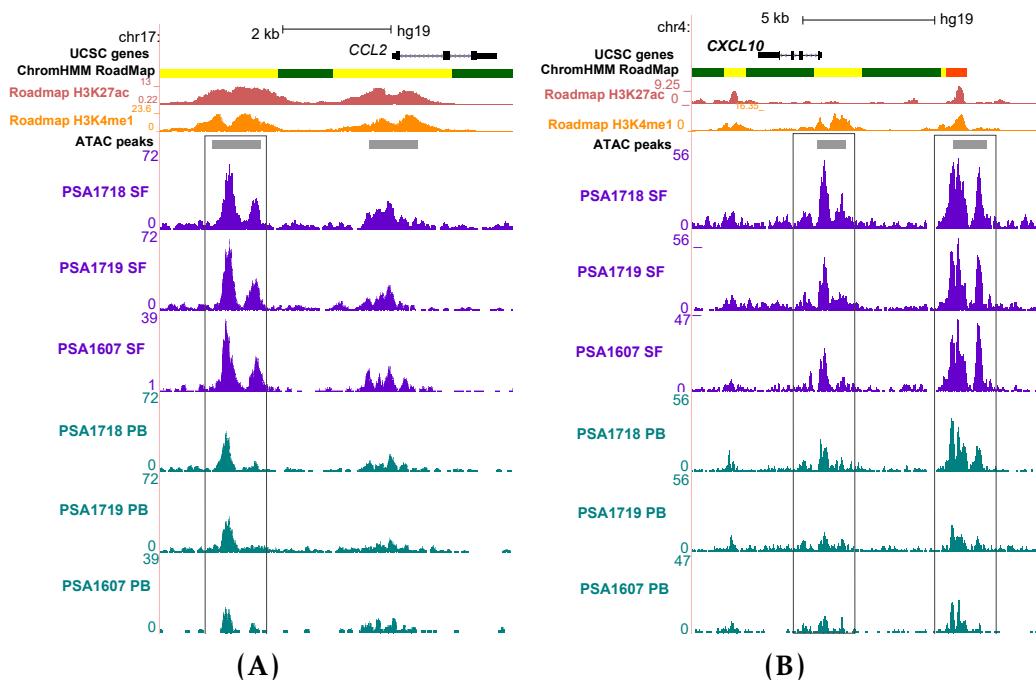


**Figure 5.17: Differences in the percentage of CD14<sup>+</sup> monocytes producing TNF- $\alpha$ , osteopontin, MCP-1 and IP-10 in synovial fluid and peripheral blood.** The percentage of CD14<sup>+</sup> monocytes producing (A) TNF- $\alpha$ , (B) osteopontin, (C) MCP-1 and (D) IP-10 in synovial fluid and peripheral blood are shown for each of the ten PsA samples assayed by mass cytometry. In each sample and tissue, the percentage of cells producing each of the cytokines is calculated as the increment in positive cells for the relevant cytokine before and after protein transport blockade with BFA (6h minus 0h). (E) Boxplots illustrating the concentration in pg/mL (x-axis log<sub>10</sub> 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<sup>+</sup> monocytes producing MCP-1 and IP-10 between synovial fluid and peripheral blood were of particular

## Cross-tissue comparison analysis in PsA

interest due to correlating changes in chromatin accessibility, gene expression and protein production. A statistically significant synovial fluid open DAR was found in CD14<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> monocytes, enzyme-linked immunosorbent assay

## **Cross-tissue comparison analysis in PsA**

---

(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.* 2015) 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. Bowes and colleagues also performed fine-mapping for a number of the 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.

Fine-mapping was performed in 36 loci reported in Bowes *et al.*, all showing at least nominal significance 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 ( $p\text{-value}<10^{-4}$ , e.g *RSPH3/TAGAP* and *ELMO1*) (Table A.8). Fine-mapping for GWAS signals with greater statistical significance ( $p\text{-value}<10^{-4}$ ), 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.

Cross-tissue comparison analysis in PsA

**Table 5.7: Summary table of the PsA GWAS loci presenting  $\log_{10}ABF \geq 3$  for the fine-mapping lead SNP.** For 12 PsA loci  $\log_{10}ABF$  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; ABF=approximate Bayes factor; PP=posterior probability.

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

## **Cross-tissue comparison analysis in PsA**

---

Amongst the loci that also failed to be fine-mapped in this thesis ( $\log_{10}\text{ABF}<3$ ), 4 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  $\log_{10}\text{ABF}\geq 3$  cut-off, 5 presented a fine-mapping lead SNPs in very 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, demonstrating 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 7 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 7 successfully fine-mapped loci, the size of credible sets ranged between 2 and 110 SNPs, with *TRAF3IP2* being the only locus refined to less than 10 SNPs and showing  $\log_{10}\text{ABF}>5$ , generally considered the genome-wide significant cut-off (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/ADRB1* 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 7 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

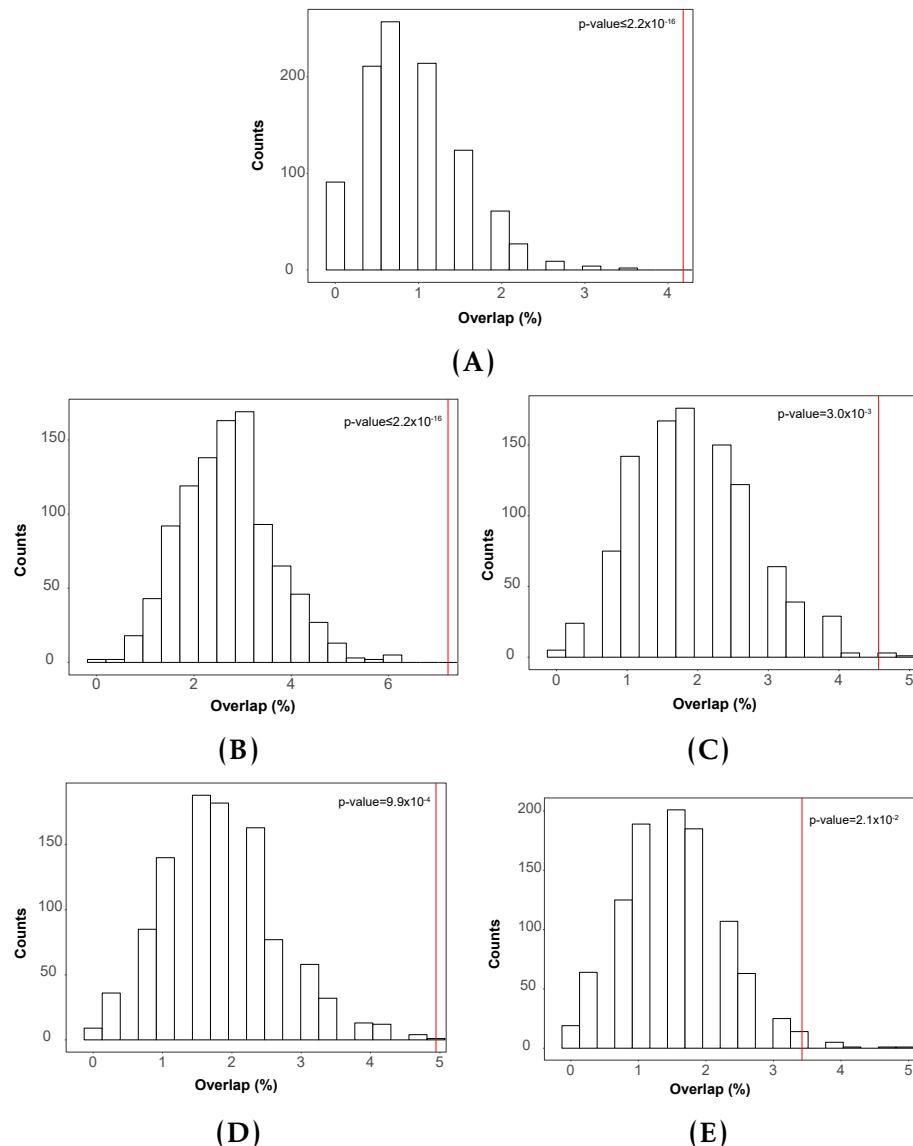
## **Cross-tissue comparison analysis in PsA**

---

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. 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 ( $p\text{-value}>0.05$ ).

## Cross-tissue comparison analysis in PsA

---



**Figure 5.19: Enrichment of PsA GWAS fine-mapped SNPs for DARs and ATAC peaks in CD14<sup>+</sup> monocytes, mCD4<sup>+</sup>, mCD8<sup>+</sup> 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. Histograms showing the percentage of the same fine-mapped SNPs overlapping ATAC peaks from the consensus list (red line) of (B) CD14<sup>+</sup> monocytes, (C) mCD4<sup>+</sup>, (D) mCD8<sup>+</sup> and (E) NK cells versus the distribution of overlaps identified in each of the 1,000 sets of background SNPs. Significance of the enrichment is indicated as p-value calculated using binomial test.

Additional overlap was performed between the 270 SNPs of the credible sets from the 7 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

## Cross-tissue comparison analysis in PsA

---

and SNPs from the 90% credible sets (17 SNPs) was found in CD14<sup>+</sup> monocytes, followed by mCD8<sup>+</sup>, mCD4<sup>+</sup> and NK cells (Table 5.8) with significant enrichment found in the 4 cell types when compared to matched backgrounds of SNPs (Figure 5.19B, C, D and E). The overlap of the fine-mapped SNPs with eQTL SNPs from GTEx and relevant immune cell types showed significant enrichment for a number of datasets when compared matched backgrounds of SNPs (Figure B.19). Amongst the immune-related eQTL datasets, significant enrichment was only found for monocytes stimulated with LPS (2h or 24h) or IFN- $\gamma$ . A number of GTEx eQTL datasets also showed significant enrichment, including whole blood and transformed fibroblasts, the last ones also being involved in the PsA structural changes (Espinoza et al. 1994).

ATAC cell type master list	90% credible set overlapping SNPs (number)	Cell type specific overlap of fine-mapped SNPs
CD14 <sup>+</sup> monocytes	32	<i>STAT2</i> (5), <i>TYK2</i> (2) <i>RUNX3</i> (1), <i>TRAF3IP2</i> (1)
mCD4 <sup>+</sup>	29	<i>CSF2</i> (1), <i>IL23R</i> (1)
mCD8 <sup>+</sup>	28	<i>RUNX3</i> (1)
NK	19	<i>TYK2</i> (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 7 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.

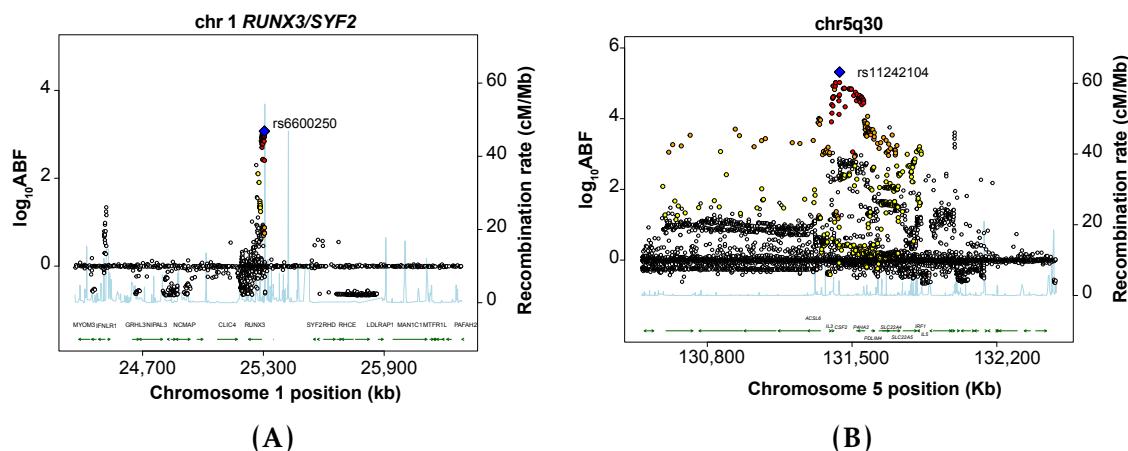
The 43 unique SNPs from the 90% credible sets of the successfully fine-mapped regions within ATAC peaks were located across the *CSF2* (8), *IL12B* (3), *IL23R* (4), *RUNX3* (6), *STAT2* (14), *TRAF3IP2* (1) and *TYK2* (7) loci. A number of these SNPs were found to only overlap accessible chromatin in one particular cell type. Interestingly, for the *TRAF3IP2* locus, which was fine-mapped with

## Cross-tissue comparison analysis in PsA

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 locus 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<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> cells a discrete consensus ATAC peaks annotated as an enhancer in the three cell types according to the corresponding chromatin segmentation maps. rs11249213 is located at an intergenic region, approximately 20Kb downstream the *RUNX3* gene and 1Kb away from Bowes GWAS lead SNP rs7523412.



**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  $\log_{10}$ 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).

In mCD8<sup>+</sup> the peak harbouring rs11249213 was identified as a DAR when using a threshold for significance of FDR<0.5, showing greater accessibility in synovial fluid when compared to peripheral blood (fold change=1.7) (Figure 5.21

## **Cross-tissue comparison analysis in PsA**

---

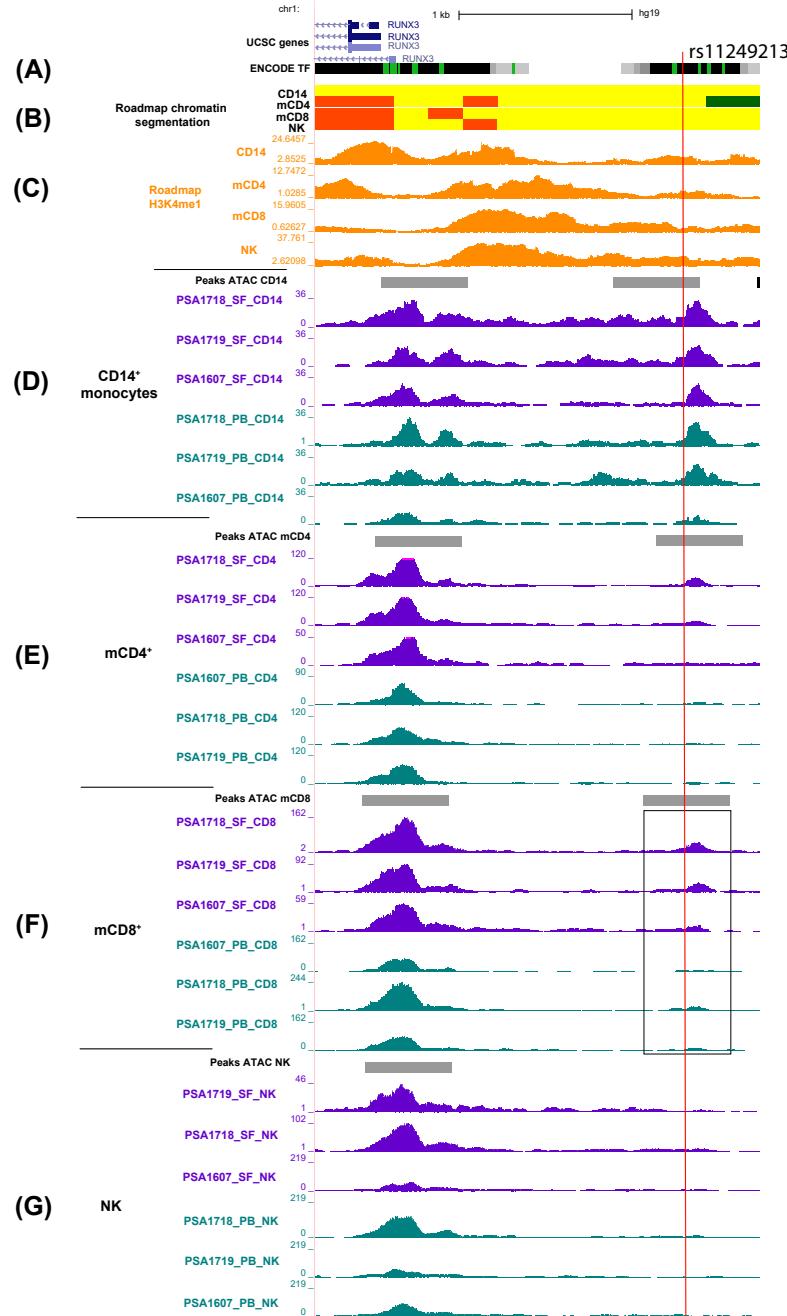
black box). In mCD4<sup>+</sup> the same ATAC peak was only found in cells from synovial fluid only in two of the patients (PsA1618 and PsA1619). In contrast, in CD14<sup>+</sup> monocytes all 3 PsA patients showed accessible chromatin at the location of the SNP in both tissues. No significant eQTL signal was found for rs11249213 in immune-relevant cell types, including unstimulated CD4<sup>+</sup> and CD8<sup>+</sup> cells or unstimulated and treated monocytes (LPS or IFN- $\gamma$ ). None of the additional 5 SNPs of the credible set that overlapped ATAC peaks in at least one of the 4 cell types showed significant eQTL effect for any gene in in 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<sup>+</sup> and mCD4<sup>+</sup> T cells.

## **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. 2015). The fine-mapping conducted here revealed rs11242104 as the lead SNP for this locus (Table 5.7, log<sub>10</sub>ABF=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 4 cell types analysed in this chapter (Table 5.9 and Figure 5.22 top panel). Either the lead SNP of this analysis or the one from Bowes and colleagues fine mapping overlapped ATAC peaks. Three of the variants highlighted by Bowes and colleagues as functionally relevant based on ENCODE data (rs10065787, rs27437 and rs7721882) were within an ATAC peak of at least one of the cell types isolated from PsA samples (Table 5.9).

## Cross-tissue comparison analysis in PsA



**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<sup>+</sup> 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<sup>+</sup> monocytes, (E) mCD4<sup>+</sup>, (F) mCD8<sup>+</sup> and (G) NK cells isolated from synovial fluid and peripheral blood.

## Cross-tissue comparison analysis in PsA

---

SNP	Cell type ATAC overlap	Top eGene	Cell type (adj p-value)
rs10065787	CD14 <sup>+</sup> , mCD4 <sup>+</sup>	<i>P4HA2</i>	CD14 <sup>+</sup> LPS2 ( $2.4 \times 10^{-15}$ ), CD14 <sup>+</sup> LPS24 ( $5.9 \times 10^{-23}$ ) CD14 <sup>+</sup> IFN- $\gamma$ ( $4.1 \times 10^{-20}$ )
		<i>SLC22A5</i>	CD14 <sup>+</sup> UT ( $9.3 \times 10^{-33}$ ) CD14 <sup>+</sup> IFN- $\gamma$ ( $6.6 \times 10^{-26}$ ) CD4 <sup>+</sup> ( $1.6 \times 10^{-3}$ ) CD8 <sup>+</sup> ( $8.0 \times 10^{-3}$ )
		<i>SLC22A4</i>	CD14 <sup>+</sup> UT ( $2.3 \times 10^{-11}$ )
rs11242104	All	NA	
rs11242105	All	NA	
rs2069803	All	<i>SLC22A5</i>	CD4 <sup>+</sup> ( $< 2.2 \times 10^{-308}$ ) CD8 <sup>+</sup> ( $4.5 \times 10^{-4}$ )
rs27437	CD14 <sup>+</sup> , mCD4 <sup>+</sup>	<i>SLC22A5</i>	CD4 <sup>+</sup> ( $1.9 \times 10^{-3}$ ) CD8 <sup>+</sup> ( $2.6 \times 10^{-2}$ )
rs4705908	All	<i>SLC22A5</i>	CD4 <sup>+</sup> ( $< 2.2 \times 10^{-308}$ ) CD8 <sup>+</sup> ( $8.7 \times 10^{-4}$ )
rs2089855	All	<i>P4HA2</i>	CD14 <sup>+</sup> LPS2 ( $3.5 \times 10^{-25}$ ) CD14 <sup>+</sup> LPS24 ( $1.1 \times 10^{-34}$ ) CD14 <sup>+</sup> IFN- $\gamma$ ( $3.1 \times 10^{-34}$ )
		<i>SLC22A5</i>	CD14 <sup>+</sup> UT ( $8.8 \times 10^{-51}$ ) CD14 <sup>+</sup> IFN- $\gamma$ ( $4.1 \times 10^{-36}$ ) CD4 <sup>+</sup> ( $< 2.2 \times 10^{-308}$ ) CD8 <sup>+</sup> ( $< 2.2 \times 10^{-308}$ )
rs7721882	mCD4 <sup>+</sup>	<i>SLC22A5</i>	CD4 <sup>+</sup> ( $< 2.2 \times 10^{-308}$ ) CD8 <sup>+</sup> ( $1.5 \times 10^{-4}$ )

**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<sup>+</sup> monocytes (UT, LPS 2h, LPS 24h, IFN- $\gamma$ ) (Fairfax et al. 2014), B cells (Fairfax et al. 2012), total CD4<sup>+</sup> and total CD8<sup>+</sup> (Kasela et al. 2017). adj p-value refers to the corrected p-value for multiple testing reported by each of the studies.

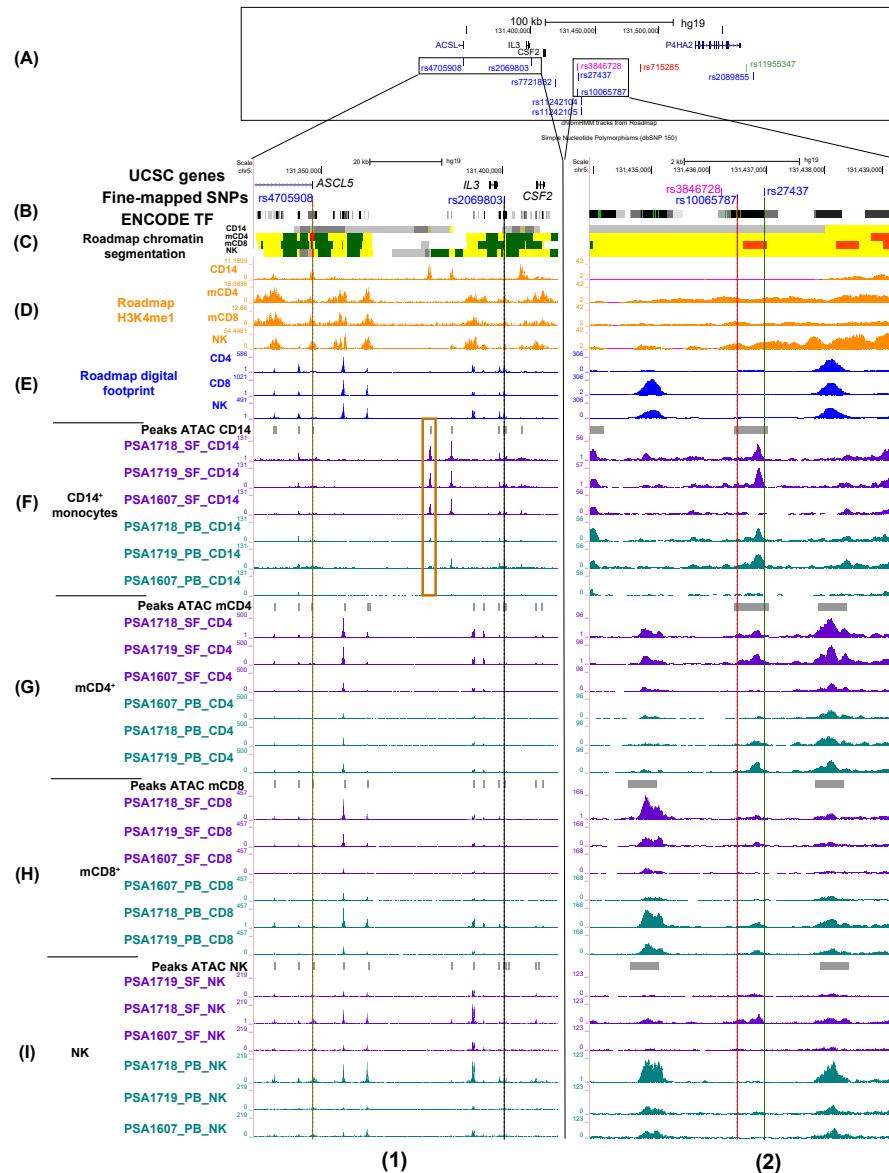
## Cross-tissue comparison analysis in PsA

---

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). Moreover, 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<sup>+</sup>, mCD8<sup>+</sup> and NK cells, according to chromatin segmentation maps (Figure 5.22 panel 1, C). Although both SNPs overlapped accessible chromatin in all 4 cell types, the most significant *cis*-eQTL signals were found for regulation of *SLC22A5* expression in CD4<sup>+</sup> or CD8<sup>+</sup> 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 CD4<sup>+</sup> CD8<sup>+</sup> cells 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 cell types showing extremely significant eQTL effect in regulating *SLC22A5* expression in CD4<sup>+</sup> and CD8<sup>+</sup> 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<sup>+</sup> and CD8<sup>+</sup> 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- $\gamma$ ), respectively (Table 5.9). Similarly, rs10065787 mapped to an ATAC peak in CD14<sup>+</sup> monocytes and mCD4<sup>+</sup> cells with moderate enrichment for the enhancer histone mark H3K4me1 in mCD4<sup>+</sup> (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- $\gamma$ ), respectively (Table 5.9). Chromatin conformation data using promoter capture-HiC (Javierre *et al.* 2016) in unstimulated monocytes does not clearly reveal interaction for

## Cross-tissue comparison analysis in PsA



**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 4 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<sup>+</sup> monocytes, (G) mCD4<sup>+</sup>, (H) mCD8<sup>+</sup> and (I) NK cells isolated from synovial fluid and peripheral blood. The yellow box highlights a DAR found in CD14<sup>+</sup> monocytes near rs2069803 and rs4705908.

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.

## 5.4 Discussion

### 5.4.1 Characterising chromatin accessibility in PsA samples

The epigenetic landscape of immune cells isolated from peripheral blood and synovial fluid in PsA patients has been characterised and compared in this chapter. The cell types interrogated ( $CD14^+$  monocytes, m $CD4^+$ , m $CD8^+$  and NK) are key players in the innate and adaptive immune response dysregulated in PsA pathophysiology. In particular, expansion of m $CD4^+$  and m $CD8^+$  cells in the synovium of PsA patients has been reported previously, and GWAS have highlighted significant association with MHC-I and other pivotal genes involved in T cell immune response (Taams 2018).

$CD14^+$  monocytes were found to show the largest number of DARs between synovial fluid and peripheral blood, with 23.3% of the ATAC peaks included in the consensus list of this cell type showing differences in chromatin accessibility. For all the cell types, the majority of the DARs were located at intergenic and intronic regions annotated as enhancers and were enriched for eRNAs identified in the FANTOM project (Forrest et al. 2014), consistent with differential regions having long-range effects on regulation of gene expression (Qu et al. 2015). Pathway enrichment analysis of the genes proximal to the DARs revealed differences between synovial fluid and peripheral blood in the 4 cell types. In

## **Cross-tissue comparison analysis in PsA**

---

synovial fluid, enrichment for genes in the NF- $\kappa$ B pathway was found in CD14 $^{+}$  monocytes, consistent with the TNF- $\alpha$  signature and the efficacy of anti-TNF therapies already reported for PsA (Mahil et al. 2016). Interestingly, in mCD8 $^{+}$  T cells enrichment of DARs open in peripheral blood was found proximal to genes in the Wnt signaling pathway, such as *SMAD3*. Wnt signalling is involved in many biological processes and its dysregulation is associated with a number of autoimmune diseases, such as 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 cells compared to the tissue-resident (Boudousqui et al. 2014). In NK cells increased chromatin accessibility in the proximity of genes involved in NK-mediated cytotoxicity was found only in peripheral blood. 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 relevant immune cells isolated from PsA patients. This is in line with other studies that have revealed changes in chromatin accessibility between patients and healthy controls and also across different tissues involved in disease (Scharer et al. 2016; Wang et al. 2018; Corces et al. 2016). Although these findings suggest meaningful functional differences in chromatin accessibility between circulating and affected tissue cell populations, a limitation of pathway analysis in ATAC is linking putative regulatory regions to genes. Annotation of ATAC peaks with genes in proximity has been widely 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 interactions, assuming accessible chromatin regulates neighbouring regions.

### **5.4.2 Bulk qPCR expression profiling and integration with paired chromatin accessibility data**

In contrast to chromatin accessibility profiling, recent research has investigated differences in the transcriptional profile between PBMCs, bulk T cells, SFMCs and synovial tissue from PsA patients (Dolcino et al. 2015; Fiocco et al. 2015). However, comparative analysis in matched discrete cell subpopulations from synovial fluid and peripheral blood from the same patient have yet to be reported. The pilot study presented in this chapter has characterised the expression of relevant immune genes in CD14<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> cells isolated from synovial fluid and peripheral blood.

In this study, CD14<sup>+</sup> monocytes and mCD8<sup>+</sup> showed the largest number of statistically significant DEGs between synovial fluid and peripheral blood. Integration of transcriptome profiling with paired-ATAC data in CD14<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> revealed that a small number of the DEGs between synovial fluid and peripheral blood were nearby DARs showing changes in the same direction. This overlap only presented significant enrichment for CD14<sup>+</sup> monocytes.

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 of the three cell types. 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. Other genes were also only found to be differentially expressed in CD14<sup>+</sup> monocytes across tissues, for example *CXCL10*, *CCL7*, and *CCL17*, which plays a critical role in the migration of monocytes from bone marrow to tissues

## Cross-tissue comparison analysis in PsA

---

(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<sup>+</sup> and/or mCD8<sup>+</sup> when compared to peripheral blood.

Comparison of DEGs in peripheral blood between PsA patients and controls with the DEGs between synovial fluid and peripheral blood in patients identified a number of systemic, tissue-specific and putative-disease specific DEGs. In CD14<sup>+</sup> monocytes, a much larger number of immune-relevant genes were found to be differentially expressed between the two tissues of patients when compared to those changing expression between PsA patients and healthy controls in peripheral blood. This is likely due to the plasticity of monocytes and their transition into macrophage differentiation at the inflamed synovium (Yoon et al. 2014; Park et al. 2016). A particularly interesting systemic DEG in mCD8<sup>+</sup> cells was *CCR10*, a chemokine receptor co-expressed by a subset of memory cells that preferentially migrate to skin, and was identified as an up-regulated gene in CD8<sup>+</sup> cells in the psoriasis cohort (Chapter 4) and in patients with atopic dermatitis (Hijnen et al. 2005). In mCD4<sup>+</sup>, *GPR68*, a G protein-coupled receptor (GPCR) was found to be a putative disease-specific gene 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 tissues under inflammation (Biniecka et al. 2016; Saxena et al. 2012). Another member of the acid-sensing GPCR family, *GPR65*, has been associated with a number of immune mediated diseases (including AS, CD and MS) and found to be a marker of pathogenic Th17 cells (Cortes et al. 2013; Lassen et al. 2016; Wirasinha et al. 2018; Gaublomme et al. 2015; Al-Mossawi et al. 2017).

In the three cell types, *SPP1* and *FN1* were the most significantly DEGs between synovial fluid and peripheral blood of PsA patients, in line with the findings from Dolcino and colleagues (Dolcino et al. 2015). Significant up-

## **Cross-tissue comparison analysis in PsA**

---

regulation of both genes was tissue-specific in the three cell types, without significant changes in peripheral blood between PsA patients and controls. This was consistent with Dolcino and colleagues results for *FN1* but not for *SPP1*, where moderate up-regulation in peripheral blood in PsA patients was also detected. In CD14<sup>+</sup> monocytes ATAC data has also revealed two synovial fluid open DARs at the promoter and downstream *FN1* only. *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 oxyde and IL-1 $\beta$  (Gramoun et al. 2010).

Pathway analysis using significant DEGs between synovial fluid and peripheral blood in this data revealed enrichment for chemokine, TLR and NOD-like signalling pathways in CD14<sup>+</sup> monocytes. This was consistent with the well-known role of TLR and NOD-like receptors for rheumatic diseases (McCormack et al. 2009). NOD-like signaling has also been highlighted as one of the dysregulated pathways in the genome-wide transcriptomic analysis in lesional and uninvolved psoriatic skin presented in Chapter 4. Down-regulation of the TLR inhibitor *SIGIRR* (K et al. 2014) and up-regulation of *TLR1*, *TLR2*, *MYD88* in the qPCR array may lead to increased activation of NF $\kappa$ B TF in synovial fluid CD14<sup>+</sup> monocytes. This is consistent with the significant reduction of NF- $\kappa$ B 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- $\kappa$ B pathway may be the priming the transcriptional up-regulation of downstream genes such as *TNFA*, *CCL2* and *CCL5* and subsequent enrichment of chemokine signalling in synovial fluid CD14<sup>+</sup> monocytes. In mCD4<sup>+</sup>, enrichment of the IL-10 signalling pathway was interesting in the context of the expansion of Tregs in synovial fluid, and was

## **Cross-tissue comparison analysis in PsA**

---

revealed by flow cytometry analysis in the same samples. Tregs are characterised by the expression of anti-inflammatory cytokines, including IL-10 (O'Garra et al. 2004) and qPCR transcriptomic data showed in a significant increase in expression of the IL-10 receptor subunit  $\alpha$  *IL10RA* and a similar trend for *IL10* ( $p$ -value=0.004 and 0.14, respectively) in synovial fluid mCD4 $^{+}$  and mCD8 $^{+}$  compared to peripheral blood. This may suggest that inflammation in PsA is refractory to the immunomodulatory effects for IL-10 signalling 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).

### **5.4.3 The relevance of monocytes in the PsA synovial fluid**

In this exploratory study, CD14 $^{+}$  monocytes showed the largest number of differences in chromatin accessibility and confidently modulated genes 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 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 involved in this project 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 Charcaterisation of synovial and peripheral blood CD14<sup>+</sup> monocytes by scRNA-seq: subpopulations and additional relevant pathways**

Monocytes are very plastic cell population known for being precursors of DCs, macrophages and osteoclast cells and importantly undergo phenotypic changes at the site of inflammation (Qu2014; Rivollier2012). Therefore interrogation at the single-cell level can reveal distinct functionally relevant subpopulations. In addition to this, scRNA-seq was also used to perform differential gene expression of synovial fluid vs peripheral blood complementary to the biased qPCR array and unveil additional relevant pathways dysregulated between the two tissues.

Combined analysis of synovial fluid and peripheral blood PsA CD14<sup>+</sup> monocytes transcriptome by scRNA-seq showed the largest differences driven by tissue of origin. Synovial fluid monocytes appeared as more intermediate monocytes expressing *CD14* and *CD16* in contrast to the peripheral blood ones showing a more classical phenotype lacking expression of *CD16<sup>+</sup>*. This transcriptional finding was consistent with mass cytometry data showing expansion of CD14<sup>+</sup> CD16<sup>+</sup> monocytes in the synovium of the same ten patients presented in 5.3.8 (data not shown) and is consistent with the more pro-inflammatory phenotype characteristic of intermediate monocytes and the 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 circulation in PsA and other inflammatory diseases such as sepsis and SLE (Chiu et al. 2010; Yoon et al. 2014; Mukherjee et al. 2015).

## **Cross-tissue comparison analysis in PsA**

---

The 8 subpopulations (MC-1 to 8) of monocytes identified by cluster analysis were mostly formed by cells from one of the tissues, being MC-7 the only one with a mixed contribution, reinforcing the large variability across monocytes of the two tissues. Of particular interest was the MC-4 subpopulation showing distinctive expression of *IL7R*, *IL32* and *CCL5*, amongst other markers. Expression of *IL7R* by this synovial monocytes subpopulation is consistent with the increased chromatin accessibility and gene expression of this receptor in synovial fluid monocytes of PsA patients previously shown. Although expression of a number of MC-4 markers, including *IL7R*, was also found in the mixed MC-7 cluster, the peripheral blood monocytes only represented 2.4% vs 7.2% of the synovial monocytes in MC-4 population, suggesting that  $\text{IL7R}^+$  monocyte subpopulation is expanded in the PsA synovium. This observation has been reinforced by flow cytometry analysis conducted by Al-Mossawi and colleagues showing up to 35% of the total synovial fluid CD14 $^+$  monocyte to be  $\text{IL7R}^+$  in PsA and AS patients versus approximately 1% in peripheral blood (in revision Al-Mossawiet *et al.*, 2018). The identification this  $\text{IL7R}^+$  monocyte subpopulation in synovial PsA monocytes 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). Although the role of IL-7 and IL7R in mediating T cell homeostasis and survival naïve and memory T cell subsets, expression of *IL7R* and production of soluble IL7R (sIL7R) is genetically regulated by the MS risk allele (rs6897932) under LPS stimulation and TNF- $\alpha$  only in monocytes but not T cells (Kondrack *et al.* 2003; Fairfax *et al.* 2014; Al-Mossawi *et al.* 2018). Interestingly, the open synovial fluid DARs found at the 3' UTR of the *IL7R* gene is nearby this risk variant and is monocyte specific. *In vitro* work conducted by Al-Mossawi and colleagues has also demonstrated that  $\text{IL7R}^+$  monocytes are transcriptionally responsive to IL-7 and the main source of sIL7R, which may suggest a role of

## **Cross-tissue comparison analysis in PsA**

---

monocytes in prolonging IL-7 response in T cells (**Lundstrom2013**). The gene signature of the scRNA-seq IL7R<sup>+</sup> monocytes subpopulation showed enrichment for IL-7 responsive genes in the *in vitro* LPS stimulated, reinforcing the *in vivo* implications of this subpopulations, which may be related with the role of IL-7 in osteoclastogenesis (Weitzmann et al. 2000). Nevertheless, additional work will be required to understand the role of this monocyte subpopulation in PsA pathophysiology.

Amongst the other subpopulations, MC-1 appeared as the most functionally interesting and it was characterised by expression of genes involved in IFN signalling, viral response, antigen presentation and chemokine signalling. A recent study in RA have also identified a subset of monocyte expressing *IFI6*, amongst other markers, and showing enrichment for IFN signalling and the IFN- $\gamma$  mediated pathway (Zhang et al. 2008). Moreover, synovial fluid monocytes, and particularly those from the MC-1 and MC-2 subpopulations, showed increased expression for MHC-II *HLA-DR*, *HLA-DP* and *HLA-DQ* genes compared to peripheral blood monocytes that contributed to enrichment of the MHC-II antigen presentation pathway for DEGs between the two tissues. Increased expression of these MHC-II genes is characteristic of the pro-inflammatory intermediate and non-classical (CD14<sup>-</sup> CD16<sup>++</sup>) monocytes with a role in T cell recruitment and activation through antigen presentation (**Grage-Griebenow2001**). The enrichment of synovial fluid vs peripheral blood DEGs for the chemokine signalling pathway not only recapitulated the qPCR results but also reinforced the increased chemotaxis and T cell activation driven by up-regulation of *CCL3*, *CCL2* and *CXCL10* in synovial fluid monocytes (Baba and oncology 2014). Altogether, this suggests that the inflammatory cytokine milieu in the joints leads to an enhanced IFN activated phenotype of the synovial monocytes, importantly of the MC-1 subpopulation, with functional relevance in antigen presentation and T cell activation. This has been demonstrated in RA,

## Cross-tissue comparison analysis in PsA

---

where intermediate synovial fluid monocytes showed ability to promote Th-1 and Th-17 pathogenic response in autologous mCD4<sup>+</sup> 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 the inflammasome activation was down-regulated in synovial fluid monocytes but no transcriptional up-regulation of *IL-1B* has been identified in peripheral blood CD14<sup>+</sup> monocytes by scRNA-seq. Interestingly, a recent study has shown that type I Interferon can inhibit NLRP1-mediated inflammasome activation through repression 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 were also unveiled by this analysis. For example, up-regulation of inhibitors and down-regulation of positive regulators of the apoptosis pathway suggests resistance to apoptosis by synovial fluid CD14<sup>+</sup> monocytes. This is consistent with a functional study on RA and PsA synovial fluid monocytes showing resistance to the extrinsic and intrinsic apoptosis pathway and may suggest that monocytes in the synovium are more prone to macrophage differentiation (Fall207; Rajasekhar et al. 2017). The osteoclast differentiation pathway also presented very significant up-regulation in synovial fluid cells of for the RANKL receptor *TNFRSF11A*, the key axis 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), consistent with the more intermediate phenotype of the synovial monocytes in this data.

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

## **Cross-tissue comparison analysis in PsA**

---

monocytes by qPCR and scRNA-seq was only moderate. As an example, up-regulation of *TLR2* observed in synovial fluid CD14<sup>+</sup> monocytes by qPCR array was not reproduced by the scRNA-seq differential analysis. This could be caused by differences in the way of selecting monocytes (FACS based on CD14<sup>+</sup> for qPCR vs *CD14* expression in scRNA-seq), the sensitivity of each of the two techniques and patient heterogeneity in each of the two assays (Table 5.1), requiring validation in a larger cohort.

### **5.4.5 Mass cytometry data supports changes in chromatin accessibility and gene expression of pro-inflammatory mediators in PsA CD14<sup>+</sup> monocytes**

Single-cell mass cytometry data allowed to investigate the expression of a number of cytokines produced by CD14<sup>+</sup> monocytes in synovial fluid and peripheral blood in a total of ten PsA patients (including those with paired qPCR and/or ATAC data demonstrating increased percentage of monocytes producing TNF- $\alpha$ , MCP-1, osteopontin and IP-10. This was consistent with the transcriptional up-regulation of these four genes detected by qPCR and/or scRNA-seq in PsA CD14<sup>+</sup> monocytes from synovial fluid compared to peripheral blood. A previous study using ELISA reported an increased production of TNF- $\alpha$  in PsA synovial fluid compared to osteoarthritis, consistent with the therapeutic effect of TNFi (Partsch et al. 1997).

Interestingly, osteopontin, MCP-1 and IP-10 are induced in monocytes upon IFN stimulation, in line with the enrichment of this pathways by DEGs between synovial and peripheral blood cells as well as the functional biological processes enriched for 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 it has been demonstrated to have a role in a number of chronic inflammatory and autoimmune diseases

## **Cross-tissue comparison analysis in PsA**

---

(Morimoto et al. 2010; Rittling and research 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<sup>+</sup> monocytes. *CCL2/MCP-1* and *CXCL10/IP-10* are two examples of pro-inflammatory molecules showing correlating differences between synovial fluid and peripheral blood CD14<sup>+</sup> monocytes at the chromatin accessibility, transcriptomic and protein expression levels in PsA patients. Up-regulation of these two genes in synovial fluid monocytes could be driven by the simultaneous 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). Dolcinos study did not identify up-regulation of *CCL2* or *CXCL10* expression in PsA synovial membranes when compared to osteoarthritis, likely due to the mixed composition of cells in this tissue. The open synovial fluid DARs upstream *CCL2* and *CXCL10* in CD14<sup>+</sup> 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<sup>+</sup> 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 involved in monocytes migration from peripheral blood into 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 levels of infiltrated T cells in the

## **Cross-tissue comparison analysis in PsA**

---

synovium (Ross et al. 2000). Nevertheless, CD14<sup>+</sup> monocytes-specific gene and protein expression of *CCL2/MCP-1* in PsA had not been investigated. The transcriptional up-regulation of the receptor of MCP-1, *CCR2*, in PsA synovial fluid mCD8<sup>+</sup> vs peripheral blood suggests a biological role of *CCL2-CCR2* in mediating T cell infiltration in the synovium and the expansion of mCD8<sup>+</sup> 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 inflammation (Antonelli et al. 2014). Significant up-regulation of *CXCL10* by qPCR was found in CD14<sup>+</sup> monocytes between synovial fluid and peripheral blood and 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<sup>+</sup> and mCD8<sup>+</sup> 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<sup>+</sup> monocyte specificity of *CXCL10* dysregulation between the two tissues in PsA.

### **5.4.6 The use of PsA functional data to inform fine-mapping GWAS loci**

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). Successful fine-mapping, in terms of  $\log_{10}\text{ABF}$  being greater than the cut-off of 3, was not achieved for all the loci successfully fine-mapped by Bowes and colleagues due to the reduced sample size (Bunt et al. 2015). Moreover, only the *TNFAIP3* locus, which presented the largest OR (1.71) was refined to less than 10

## Cross-tissue comparison analysis in PsA

---

SNPs in the 90% credible set, in line with Bunt and colleagues observation where the ability to refine the credible sets is dependent on the OR, with greater success for loci showing OR>1.2 (Bunt et al. 2015).

The significant enrichment of fine-mapped SNPs for accessible chromatin in PsA immune cell types from peripheral blood and synovial fluid reinforces the role of these cell types in the genetics of PsA. Interestingly, enrichment with eQTL SNPs from immune-relevant cell types was only significant for monocytes under pro-inflammatory stimulation, highlighting the importance of context in the putative functional role of risk SNPs in complex diseases (Walsh et al. 2016; S et al. 2017).

DARs in 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 mCD8<sup>+</sup>-specific DAR only significant when using FDR<0.05, showing the same trend in mCD4<sup>+</sup>. Similar results have been found by colleagues in the group comparing chromatin accessibility at this location in unstimulated and activated total CD4<sup>+</sup>. This, together with the absence of eQTL signal for this SNP in the unstimulated T cell dataset from Kasela and colleagues may suggest a putative role for this SNP in regulating gene expression, potentially *RUNX3*, only in the context of the inflamed synovium in T cells. *RUNX3* is one of the key TFs in CD8<sup>+</sup> T cells development and activation, and its role in chromatin remodelling and CD8<sup>+</sup> 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 Bowes and colleagues lead variant (rs7523412) and also with rs11249213 (Reveille et al. 2010; Dubois et al. 2010). Interestingly, Apel and colleagues identified a PsA genome-wide significant

## **Cross-tissue comparison analysis in PsA**

---

association for rs4649038, at an intron of *RUNX3*, independent from AS and CD GWAS lead SNP ( $r^2 \leq 0.5$ ) and not significant in psoriasis vulgaris (Apel et al. 2013). This may suggest two different signals with independent contribution to PsA risk nearby *RUNX3*.

Integration of in-house ATAC data at the 5q31 PsA-specific signal reduced the 58 SNPs of the 90% credible set to 8 variants, a number of which were at accessible chromatin for PsA patients mCD4<sup>+</sup> and mCD8<sup>+</sup> cells and showed significant eQTL effects in T cells *SLC22A5* expression (Kasela et al. 2017). The role of 5q31 PsA association in regulating *SLC22A5* expression in total CD4<sup>+</sup> and CD8<sup>+</sup> cells had already been shown by Bowes and colleagues in a pilot eQTL study. *SLC22A5* is a cell membrane transporter of carnitine involved in fatty acids metabolism and has been prioritised by an in-house pipeline as the third most promising druggable candidate for the treatment of psoriasis ([http://galahad.well.ox.ac.uk:3010/pidb/discovery/PS0/SLC22A5#bookmark\\_details\\_genomic](http://galahad.well.ox.ac.uk:3010/pidb/discovery/PS0/SLC22A5#bookmark_details_genomic)) and a potential as a druggable therapeutic target in PsA (Leung et al. 2006). Data integration also suggested a role of fine-mapped SNPs overlapping PsA CD14<sup>+</sup> 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 to reduce inflammation and the production of MCP-1, which role in PsA synovial CD14<sup>+</sup> monocytes has also been highlighted in these thesis (Natarajan et al. 2007).

### **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, a very powerful strategy to dissect disease pathophysiology in a cell type specific manner. Nevertheless, a number of

## **Cross-tissue comparison analysis in PsA**

---

limitations and challenges were encountered and need to be taken into account to contextualise these results. The small sample size ( $n=3$ ) and the lack of paired data across all the techniques represent one of the limitations together with lack of peripheral blood samples from healthy individuals or synovial fluid from another autoimmune or non-inflammatory joint disease, as incorporated by other studies (Mizoguchi et al. 2018; Dolcino et al. 2015; Zhang et al. 2018). This would allow pinpointing changes in chromatin accessibility, transcriptomic and protein expression specific for the PsA inflamed synovium and will be conducted in the next stages of the project. The use of RNA-seq instead of the biased qPCR array would have allowed a genome-wide overview of transcriptomic differences between tissues in the same samples interrogated with 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 in other publications (Zhang et al. 2018).

Another challenge in this study relates to the analysis and integration of scRNA-seq and mass cytometry data. Both techniques still represent emerging fields where there is no "gold" standard on the strategy 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 out has been conducted to compensate for the reduced sample size. However, this imputation methods may introduce bias in the data and may be avoided when larger cohort size is available (Andrews 2018). Cluster identification is another aspect that lacks of consensus and the results can be affected depending on the resolution parameters and the algorithm of choice as well as the sample size (Kiselev et al. 2019). This work represents and attempt to identify potential relevant subpopulations of monocytes in synovial fluid from PsA. However, the identification of robust and stable subpopulations through cluster

analysis will benefit from the implementation of algorithms designed for cluster validation such as Silhouettes, which has recently been used successfully in the field of single-cell (Rousseeuw 1987; Zhang et al. 2018). Similarly, different methodologies for cluster identification, annotation and cytokines quantification by in mass cytometry are available and additional data analysis is in progress to identify distinct subpopulations of monocytes based on cytokine and chemokine expression. Moreover, staining for some of the markers stated in Chapter 2 Table 2.3 showed high levels of background noise and their results are not compelling. Therefore additional mass cytometry in a validation cohort together with expansion of the panel of markers will be required to further confirm some of the hypothesis formulated based on the transcriptional data.

Regarding fine-mapping, the analysis presented in this chapter is limited by the sample size and its effect in the Bayesian association analysis (Bunt et al. 2015). This explains the inability to fine-map a number of loci that were successfully analysed by Bowes and colleagues when using the entire PsA Immunochip cohort, as well as discrepancies in the identity of the lead SNPs and the size of some of the credible sets.

## **5.5 Conclusions**

The strategy and analysis presented in this chapter illustrates 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 functional pathways. In this study, both data types highlighted CD14<sup>+</sup> monocytes as the cell type presenting the largest number of significant changes in chromatin accessibility and consistent modulation of gene expression between synovial fluid and peripheral blood, with

## **Cross-tissue comparison analysis in PsA**

---

enrichment for pathways that may lead to NF- $\kappa$ B 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 enriched for DEGs that were not unveiled by qPCR analysis. Moreover subpopulations of monocytes had been identified, of particular relevance the IL7R $^{+}$  and the IFN-activated subsets, highlighting the heterogeneity of the monocytes population in the synovium. Lastly, relatively basic 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 some correlation between differences in chromatin accessibility, gene expression and protein production between synovial fluid and peripheral blood PsA CD14 $^{+}$  monocytes. 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 distinguish circulating cells from those infiltrated homing 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 in psoriasis and PsA. This required the establishment of methodological and analytical basis of ATAC and assessment of the effect that relevant preservation 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 methods and the incorporation of GWAS fine-mapping data provides an integrated understanding of previous 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 for 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 potential clinical application in the future.

### **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;

## **General discussion**

---

Fairfax et al. 2014). In psoriasis and PsA, the enrichment of prioritised GWAS SNPs in putative enhancers and regulatory regions has highlighted a variety of immune and non-immune cell types and tissues, including Th-1, Th17, memory CD4<sup>+</sup> and CD8<sup>+</sup> cells, monocytes, keratinocytes, B cells and adipose tissue (Farh et al. 2015; Tsoi et al. 2017; Bowes et al. 2015; Ellinghaus et al. 2016; Lin et al. 2018).

In this thesis, due to logistic and budget constraints, only a number of the pathophysiological relevant cell types have been investigated in psoriasis (CD14<sup>+</sup> monocytes, total CD4<sup>+</sup> and CD8<sup>+</sup> T cells and CD19<sup>+</sup> B cells) and PsA (CD14<sup>+</sup> monocytes, mCD4<sup>+</sup> and mCD8<sup>+</sup> 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 fully understand disease pathophysiology and aetiology. Of particular interest would be investigating plasmacytoid DCs, key players in the skin inflammatory process for which psoriasis GWAS SNPs have shown context specific effects in their regulation of gene expression (Lee et al. 2014). Moreover, the epigenetic and transcriptional changes between synovial fluid and peripheral blood of PsA patients in four disease relevant cell types has demonstrated the importance of the inflammatory environment at the site of inflammation to better understand disease pathophysiology. Therefore, future investigation of psoriatic skin immune-infiltrated cells using FACS or single-cell technologies (Ahn et al. 2017) and comparing them with the circulating counterparts from psoriasis patients will be of great relevance.

## **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, with total CD8<sup>+</sup> cells demonstrating the largest number of differences. In contrast, the comparison of chromatin accessibility between synovial fluid and peripheral blood in PsA revealed a larger number of differentially accessible regions with fold-changes greater than 1.5 in all the cell types, with CD14<sup>+</sup> monocytes being the cell type showing greater 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 in the psoriasis and PsA study suggest that the physiological inflammatory environment at the PsA inflamed joints has an effect in driving chromatin accessibility changes of greater effect size than those observed in circulating cells between patients and control, and reinforce the need of studying 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 when comparing 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<sup>+</sup> and CD8<sup>+</sup> populations. The success in assessing chromatin accessibility by Omni-ATAC in NHEKs (Chapter 3) and the development of scATAC-seq opens the door to

## **General discussion**

---

interrogate lesional skin keratinocytes and mapping the chromatin landscape of small number 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<sup>+</sup> cells showed paired changes in chromatin accessibility and H3K27ac between psoriasis patients and healthy controls, likely as a result of H3K27ac being only one of the diverse 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, respectively, 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 PBMCs or SFMCs or in a particular cell type undergoing exogenous inflammatory stimuli. This has provided a very broad overview of the dysregulated gene expression within mixed cell populations. Similarly, the study of lesional skin from psoriasis patients has been 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 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. CD14<sup>+</sup> and CD8<sup>+</sup> were demonstrated as the main cell types undergoing dysregulation of gene expression, highlighting immune-relevant

## General discussion

---

pathways with both up- and down-regulation of pro-inflammatory genes, and a comprehensive characterisation of cell type specific differentially expressed lncRNAs. Particularly unexpected was the up-regulation in at least one of the four cell types of two anti-inflammatory genes, *TNFAIP3* and *NFKBIA*, which are also psoriasis GWAS risk loci. Interestingly, the study of 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 physiological inflammatory environment at the main disease affected tissue in driving functional changes. Moreover, the results of the transcriptomic study in this thesis has also demonstrated 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 the duration of this project (Tervaniemi et al. 2016).

The qPCR transcriptional study in three relevant cell types isolated synovial fluid and peripheral blood allowed, for the first time, to compare cell type specific dysregulation of inflammatory genes with changes in chromatin accessibility in PsA. According to this data generated in this thesis CD14<sup>+</sup> monocytes revealed the largest number of DEGs between the two tissues. Similarly to the transcriptional results in skin, DEGs between synovial fluid and peripheral blood monocytes highlighted dysregulation of the NOD-like receptor signalling pathway in the qPCR and scRNA-seq data and reinforces the prominent role of this pathway at the site of inflammation, as previously reported (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

## **General discussion**

---

is particularly important for the phenotypic characterisation of very plastic cells as the monocytes, which can differentiate into macrophages, DCs and osteoclast upon different environmental stimuli (Qu2014; Rivollier2012). In this thesis, a pilot exploratory study of subpopulations of monocytes in the synovium and peripheral blood of PsA has been conducted for the first time and suggested the existence of functionally relevant subpopulations within the joints. Simultaneously, colleagues involved in this collaborative project are also exploring clonal expansion in T cells, and their results suggest relevant mCD8<sup>+</sup> clonal expansion in the joints compared to peripheral blood (manuscript in preparation). This work will require increased sample size and further optimisation of the analytical procedures to fully dissect. scRNA-seq has also been used to heterogeneity in pathphysiologically relevant cell populations RA, MS and Chron's disease (Mizoguchi2018; Schafflick2018; Zhang et al. 2018). 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 (Itoh2018; Joost2016).

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<sup>+</sup> cells from psoriasis when compared to controls and also in mCD8<sup>+</sup> cells from synovial fluid in PsA. *CXCL10* presented higher expression in PsA peripheral blood CD14<sup>+</sup> when compared to controls, showing a monocyte-specific up-regulated protein expression in synovial fluid when compared to matched peripheral blood. No changes were found in psoriasis vs healthy controls, in line with a recent conference abstract (Muntyanu et al. 2017), and supporting the role of *CXCL10* as potential PsA biomarker involved in the initiation and progression of joint inflammation previously hypothesised by another study (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 ability of linking accessible chromatin to the target gene to which it may be regulating. It has been demonstrated that 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 has been performed based on genes in physical proximity, in line with most of the 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 have 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, changes in chromatin accessibility can precede changes in transcript levels that also should be taken into account when correlating both (Alasoo et al. 2018; Calderon et al. 2018).

The analysis of differential chromatin accessibility and histone modifications in psoriasis and PsA will benefit from incorporating genotyping data to the disease status, which could eventually be expanded to perform chromatin accessibility QTL in disease (Mo et al. 2018; Alasoo et al. 2018). This would allow the identification of other accessible regions similar to the ATAC peak

## **General discussion**

---

harbouring rs4672505 at chr2p15 GWAS locus in psoriasis, 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 was limited by the sample size, technical aspects and time scale, but provides a platform for future validation studies. To better understand the correspondence between transcriptional profiles and gene expression 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, but their work is still under peer review (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 to further characterise and correlate 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**

In order to reduce the number of candidate causal SNPs at a particular GWAS locus, fine-mapping is a widely accepted strategy for refining putative causal SNPs for functional follow-up. Fine-mapping using Bayesian methods have been conducted 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, which explained the variation of results when compared to the fine-mapping conducted by Bowes and colleagues in PsA (Bowes et al. 2015). Further refinement of the number of putative causal SNPs

## **General discussion**

---

was achieved by integration of the in-house ATAC data generated in psoriasis and PsA samples and highlighted two putative causal variants at the *B3GNT2* and *RUNX3* loci in psoriasis and PsA, respectively. The location of rs11249213 at a mCD8<sup>+</sup> synovial fluid open DAR under the significance threshold of FDR<0.01 but FDR<0.05 suggest the potential for finding additional overlaps with DARs of smaller effect size that may become significant in larger studies.

## **6.6 Clinical relevance and translation**

Technological advances in sequencing and the latest low-input and time-efficient epigenetic techniques, such as ATAC, enable the implementation of molecular tools based on gene expression and chromatin accessibility in a clinical set up. In this context, 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, patients stratification, treatment decisions and drug design. Nevertheless, clinical translation of these finding is very challenging due to variability in the results across studies which are influenced by sample size, age, sex, patient heterogeneity and interindividual variation (**Shen-Orr2010**). Interindividual heterogeneity has also 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 (**Alizadeh2000; Rendeiro2015**). In inflammatory diseases such as eosinophilic esophagitis, in which symptoms are common to a number of other gastrointestinal affections, the expression of 94 genes in esophageal biopsies has been successfully used to diagnose disease and assess treatment response (**Dellon2017**). Currently, no molecular diagnosis based on

## **General discussion**

---

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 dermatitis and irritant contact dermatitis (Kamsteeg2010). A pilot metabolomic study showed that PsA patients had decreased level of alpha ketoglutaric acid and an increased level of lignoceric acid 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 will contribute to progress into 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 sample 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 will be conducted. Expansion of the psoriasis recruitment should include peripheral blood, lesional and uninvolved skin from patients not undergoing biological treatment and also

## **General discussion**

---

prospective matched samples from those individuals undergoing biological therapy such as TNFi. Regarding skin samples, separation of the different infiltrated immune cells at the site of inflammation should be conducted to perform scATAC-seq and scRNA-seq and compare to those from peripheral blood immune cells. The continued PsA naïve-treated patients recruitment will also aim to include skin biopsies and prospective matched samples from the same patients when undergoing biologic therapy. Moreover, peripheral blood and synovial fluid from healthy controls and individuals suffering from arthritic pathologies, such as RA and osteoarthritis, will also be included to enable comparison 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 will ensure appropriate classification of patients as psoriasis non-converters into PsA.

### **Further exploration of chromatin accessibility and histones modification analysis**

Successful implementation of the optimised Omni-ATAC protocol in NHEKs (Chapter 5) will allow the generation accessibility data in keratinocytes isolated from lesional and uninvolved skin biopsies from psoriasis and PsA samples. Moreover, ChIPm will be used to interrogate additional histone marks, including H3K4me1 (enhancer) and H3K4me3 (promoter), RNA-pol II and TFs such as CTCF. Performing ATAC and ChIPm in the same relevant cell types isolated from patients and controls, followed by stimulation will allow to focus the analysis in a set of promoters and regulatory regions undergoing modifications under inflammation (Peeters et al. 2015).

## **General discussion**

---

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 between patients and controls and also across tissues (Shen et al. 2013). Moreover, genotyping of all the patients and controls will 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. In addition to A2 functionality for RNA-seq pathway interpretation, this tool will also 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 will be performed in the same cell types currently interrogated by qPCR, and also in additional cell populations of interest. Available RNA-seq data and new data generated in future recruitment will be used to conduct allele-specific expression (ASE) analysis to assess the *cis* effects of genetic variants in gene expression and removing confounders such as *trans*-acting environmental

## **General discussion**

---

effects. Moreover, miRNAs will also be interrogated in future to further investigate the role of non-coding RNAs in disease.

## **Data integration**

Integration of histone marks and chromatin accessibility changes in a larger cohort may reveal co-regulated changes for a number of epigenetic features in a disease or genotype specific manner, revealing additional information about the mechanism of action of risk variants. Prospective ASE data could be further characterised by integration with epigenetic data from the same individuals in each of the cell types and the SNPs from fine-mapping credible sets. This would help to better formulate hypotheses for the mechanisms of action of putative causal variants. Interpretation of this data will also benefit from the incorporation of psoriatic skin methylation data, and also from publicly available chromatin accessibility QTLs and methyl-QTLs generated in different tissues (Chandra2018; Kumasaka2016; 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, implementation of an algorithm that allows to correlate the subpopulations of cell identified at the transcriptomic and proteomic level could involve, amongst others, CCA (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 prioritised putative variants at GWAS loci that merits further investigation. One such as example is SNP rs4672505, which is hypothesised to have a role in the distal regulation of *B3GNT3* expression in CD8<sup>+</sup> T cells. Genomic-editing using CRISPR-

## **General discussion**

---

cas9 (Ding et al. 2013) could be used to generate CD8<sup>+</sup> cells from induced-pluripotent stem (iPS) cells lacking this variant to determine the effect in *B3GNT2* expression in homeostasis and also under anti-CD3 and anti-CD28 stimulation, in combination with ionomycin and PMA in the cellular context. Chromatin interaction experiments using capture-C could also be performed in untreated and stimulated cells to assess the effect of rs4672505 in the interaction with *B3GNT2* promoter described by Javierre and colleagues (Javierre et al. 2016). In addition to this, molecular assays to interrogate allelic differences in affinity for TF binding include electrophoretic mobility shift assay (EMSA) and ChIP using antibodies for TF of interest (Vernes et al. 2007). In terms of genome-wide screening, self-transcribing active regulatory region sequencing (STARR-seq) (Arnold 2013) can be used to create a genome-wide quantitative enhancer map that can be co-localised with epigenetic data and non-coding fine-mapped SNPs.

## **6.8 Concluding remarks**

The overall aim of this thesis was to establish a methodological and analytical pipeline to interrogate chromatin accessibility and to start investigating psoriasis and PsA disease mechanisms at the epigenetic and transcriptional level in a cell type and tissue specific manner. 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

## **General discussion**

---

further refined, the understanding of disease mechanisms governing psoriasis and PsA in important cell types and states will no doubt continue.

# Appendices

<b>A Tables</b>	<b>296</b>
A.1 Chapter 3 Tables . . . . .	296
A.2 Chapter 4 Tables . . . . .	296
A.3 Chapter 5 Tables . . . . .	300
<b>B Additional figures</b>	<b>303</b>
B.1 Chapter 3 Figures . . . . .	303
B.2 Chapter 4 Figures . . . . .	305
B.3 Chapter 5 Figures . . . . .	309
<b>Bibliography</b>	<b>353</b>

## List of Tables

A.1 Enrichment of ATAC-seq reads across the TSS for the CD14 <sup>+</sup> monocytes and CD4 <sup>+</sup> samples fresh, frozen and fixed. . . . .	296
A.2 Size of the master lists generated by DiffBind for the H3K27ac differential analysis between psoriasis patients and healthy controls in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> cells. . . . .	297
A.3 Evaluation of ChIPm library complexity for the psoriasis and control chort 1B ChIPm assay. . . . .	297
A.4 Functional interactions and overlap with another study for the differentially expressed lncRNAs in each cell type. . . . .	298
A.5 Additional enriched pathways for DEGs between psoriasis and healthy controls in CD14 <sup>+</sup> monocytes and CD8 <sup>+</sup> cells. . . . .	298
A.6 Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	299
A.7 Enriched pathways for the scRNA-seq DEGs between synovial fluid PsA CD14 <sup>+</sup> monocytes. . . . .	300
A.8 PsA GWAS Immunochip loci presenting log <sub>10</sub> ABF<3 for the fine-mapping lead SNP in the association analysis. . . . .	301

## List of Figures

B.1	Distribution of the background read counts from all the master list peaks absent in each sample. . . . .	303
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. . . . .	303
B.3	Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples. . . . .	304
B.4	Fast-ATAC and Omni-ATAC NHEK pre-sequencing profiles of relative abundance of DNA library fragment sizes. . . . .	304
B.5	Differences in TSS fold-enrichment and percentage of mirochondrial between the ATAC-seq and Fast-ATAC libraries from all the samples generated in the thesis. . . . .	305
B.6	Percentage of MT reads and called peaks after IDR p-value filtering in the ATAClibraries generated in CD14 <sup>+</sup> monocytes, CD4 <sup>+</sup> , CD8 <sup>+</sup> and CD19 <sup>+</sup> isolated from psoriasis patients and healthy controls. .	305
B.7	PCA analysis illustrating batch effect in ATAC and RNA-seq samples.	306
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. . . . .	306
B.9	Mapping quality control for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. . . . .	307
B.10	Distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles for all 72 samples. . . . .	307
B.11	Correlation in gene expression between the miRs <i>MIR31HG</i> and the target gene <i>HIF-1A</i> in lesional and uninvolved skin. . . . .	308
B.12	Assessment of the percentage of mitochondrial 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. . . . .	309
B.13	Enrichment of eQTL SNPs for the consensus list of ATAC peaks used to perform differential chromatin accessibility analysis in each cell type. . . . .	309
B.14	Permutation analysis SF vs PB in CD14 <sup>+</sup> monocytes,CD4m <sup>+</sup> ,CD8m <sup>+</sup> and NK. . . . .	310
B.15	Enrichment of DARs in the 15 chromatin states of the cell type specific Roadmap Epigenomic Project segmentation maps. . . . .	311
B.16	Identification of the CD14 <sup>+</sup> monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. . . . .	311
B.17	Expression levels of antigen presentation and phagocytosis characteristic genes in synovial fluid and peripheral blood CD14 <sup>+</sup> monocytes. . . . .	312

## **General discussion**

---

B.18 Differential gene expression between synovial fluid and peripheral blood PsA CD14 <sup>+</sup> monocytes and correlation with qPCR and chromatin accessibility . . . . .	313
B.19 Enrichment of PsA GWAS fine-mapped SNPs for publicly available eQTL datasets. . . . .	314

# 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<sup>+</sup> monocytes and CD4<sup>+</sup> samples fresh, frozen and fixed.

### A.2 Chapter 4 Tables

## Tables

---

Cell type	Master list size genome-wide	Master list size enhancers
CD14 <sup>+</sup>	99,862	60,962
CD4 <sup>+</sup>	110,353	56,282
CD8 <sup>+</sup>	137,194	51,607
CD19 <sup>+</sup>	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<sup>+</sup> monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> and CD19<sup>+</sup> 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 cohort 1B ChIPm assay.** NRF, PBC1 and PBC2 are the three measures used according to the ENCODE standards as referred in Chapter 2.  $0.5 \leq \text{NRF} < 0.8$  acceptable;  $0.8 \leq \text{NRF} \leq 0.9$  compliant;  $\text{NRF} > 0.9$  ideal;  $0.5 \leq \text{PBC1} < 0.8$  and  $1 \leq \text{PBC2} < 3$  moderate bottlenecking;  $0.8 \leq \text{PBC1} < 0.9$  and  $3 \leq \text{PBC2} < 10$  mild bottlenecking. NRF= non-redundant fraction; PBC= PCR bottle-neck coefficient.

## Tables

---

<b>Cell type</b>	<b>LncRNAs with functional interactions</b>	<b>LncRNAs overlapping Dolcino <i>et al.</i>,2018</b>
CD14 <sup>+</sup>	24	4 ( <i>HOTAIRM1*</i> , <i>ILF3-AS1*</i> , <i>MMP24-AS1</i> , <i>RP11-325F22.2</i> )
CD4 <sup>+</sup>	10	1 ( <i>MMP24-AS1</i> )
CD8 <sup>+</sup>	21	1( <i>CTB-25B13.12</i> )
CD19 <sup>+</sup>	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.*.

<b>CD14<sup>+</sup> monocytes Pathway</b>	<b>FDR</b>	<b>Fold change</b>
Generic transcription	$1.1 \times 10^{-6}$	2.66
RNA transport	$4.5 \times 10^{-5}$	2.98
GnRH signalling	$2.4 \times 10^{-4}$	3.61
Ribosome biogenesis in eukaryotes	$3.0 \times 10^{-4}$	3.42
Neurotrophin signaling	$4.0 \times 10^{-4}$	2.96
Spliceosome	$3.2 \times 10^{-3}$	2.51
Autophagy	$8.9 \times 10^{-3}$	2.18
Protein processing endoplasmic reticulum	$8.9 \times 10^{-3}$	2.05

<b>CD8<sup>+</sup> Pathway</b>	<b>FDR</b>	<b>fold change</b>
Epstein-Barr virus infection	$1.5 \times 10^{-2}$	2.33
RNA-pol I and III and	$3.1 \times 10^{-2}$	2.79
Apoptosis	$4.5 \times 10^{-3}$	2.41

**Table A.5:** Additional enriched pathways DEGs between psoriasis and healthy controls in CD14<sup>+</sup> monocytes and CD8<sup>+</sup> cells. Significant pathways for FDR<0.01. All the enriched pathways contained a minimum of ten DEGs FDR<0.05 from the analysis.

## Tables

---

<b>Lesional vs uninvolved Pathway</b>	<b>FDR</b>	<b>Fold change</b>
Genes encoding extracellular matrix proteins	$5.1 \times 10^{-9}$	3.3
Serine/threonine-protein kinase	$8.1 \times 10^{-7}$	4.6
Genes encoding secreted soluble factors	$8.1 \times 10^{-7}$	3.42
FOXM1 TF network	$2.3 \times 10^{-6}$	4.76
Phase 1 functionalization compounds	$8.1 \times 10^{-6}$	4.29
Biological oxidations	$2.9 \times 10^{-5}$	3.02
G2/M Checkpoints	$5.4 \times 10^{-5}$	3.86
Aurora B signaling	$8.1 \times 10^{-5}$	3.92
Chemical carcinogenesis	$2.0 \times 10^{-4}$	3.86
Serotonergic synapse	$2.0 \times 10^{-4}$	3.86
Drug metabolism-cytochrome P450	$2.6 \times 10^{-4}$	3.45
Mitotic M-M/G1 phases	$3.0 \times 10^{-4}$	2.16
MicroRNAs in cancer	$6.4 \times 10^{-3}$	2.18
Glycosaminoglycan metabolism	$1.9 \times 10^{-2}$	2.36
E2F TF network	$2.0 \times 10^{-3}$	2.49
p73 TF network	$2.3 \times 10^{-3}$	2.45
Fc- $\epsilon$ R I signalling in mast cells	$4.2 \times 10^{-3}$	2.54
Tight junction	$4.2 \times 10^{-3}$	2.13
Orc1 removal from chromatin	$4.7 \times 10^{-3}$	2.38

**Table A.6: Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.** Significant pathways for FDR<0.005. All the enriched pathways contained a minimum of ten DEGs FDR>0.05 from the analysis.

## Tables

---

### A.3 Chapter 5 Tables

Pathway	FDR	Fold change
Antigen processing and presentation	$2.5 \times 10^{-6}$	3.4
Chemokine signalling	$1.8 \times 10^{-5}$	2.3
GPCR ligand binding	$1.9 \times 10^{-5}$	2.2
Platelet activation and aggregation	$1.7 \times 10^{-7}$	2.1
NOD-like receptor signalling	$4.0 \times 10^{-3}$	1.8
IFN/IFN $\gamma$ signalling	$1.2 \times 10^{-3}$	2.4
Cytokine and interleukin sig.	$5.5 \times 10^{-4}$	1.7
Complement cascade	$3.8 \times 10^{-10}$	6.7
MHC-II antigen presentation	$2.8 \times 10^{-5}$	2.6
Osteoclast differentiation	$1.2 \times 10^{-3}$	2.0
Apoptosis	$9.9 \times 10^{-3}$	1.8
Phagosome	$1.2 \times 10^{-12}$	3.5
Cell adhesion molecules (CAMs)	$2.4 \times 10^{-5}$	2.7
PPAR signalling	$3.3 \times 10^{-3}$	2.7
VEGF signalling	$2.0 \times 10^{-4}$	2.4
Lymphoid and non-lymphoid cells interaction	$5.5 \times 10^{-4}$	2.3
IL-4 and 13 signalling	$1.7 \times 10^{-3}$	2.2
PI3K-Akt signalling pathway	$9.9 \times 10^{-6}$	2.1
Leukocyte transendothelial migration	$4.6 \times 10^{-3}$	2.1
FoxO signalling	$1.7 \times 10^{-3}$	2.1
Cytokine-cytokine receptor interaction	$5.9 \times 10^{-3}$	1.8
Signalling by receptor tyrosin kinases	$5.5 \times 10^{-4}$	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.

## Tables

**Table A.8: PsA GWAS Immunochip loci presenting  $-\log_{10} \text{ABF} < 3$  for the fine-mapping lead SNP in the association analysis.** For each of the signals chromosome (chr), genes nearby,  $\log_{10} \text{ABF} < 3$  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.*. OR=odd ratio

chr	Closer gene	FM lead SNP	$\log_{10} \text{ABF}$	GWAS lead SNP	SNP (p-value)	OR	lead SNP	Bowes FM	Bowes 99% credible set
2	<i>B3GNT2/TMEM17</i>	2:62501912(INS)	1.8	rs6713082 ( $4.59 \times 10^{-5}$ )		1.2	rs6713082	22	
17	<i>CARD14</i>	rs11150848	0.8	rs11652075 (0.014)		1.1	NA	NA	
9	<i>DDX58</i>	rs138398872	0.5	rs1133071 ( $3.36 \times 10^{-5}$ )		1.2	NA	NA	
7	<i>ELMO1</i>	rs10279209	1.1	rs73112675 (0.0041)		1.1	NA	NA	
6	<i>ERAP1/ERAP2</i>	rs58711860	2.7	rs62376445 (0.00017)		1.4	NA	NA	
1	<i>SLC45A1/TNFRSF9</i>	rs11367773	1.7	rs11121129 (0.00093)		1.1	NA	NA	
11	<i>ETS1/FLI1</i>	rs7935286	0.6	rs4936059 (0.0014)		1.1	NA	NA	
1	<i>LCE3B/LCE3A</i>	rs11205042	2.8	rs6693105 (0.0028)		1.1	NA	NA	
22	<i>LOC150223</i>	rs371643642	1.2	rs2298428 ( $4.38 \times 10^{-5}$ )		1.2	NA	NA	
11	<i>ZC3H12C</i>	rs1648153	0.2	rs4561177 (0.0037)		1.1	NA	NA	
9	<i>LOC392382</i>	rs36015268	0.8	rs12236285 (0.038)		1.2	NA	NA	
17	<i>NOS2A</i>	rs4795067	1.9	rs4795067 ( $1.94 \times 10^{-7}$ )		1.2	rs4795067	2	

## Tables

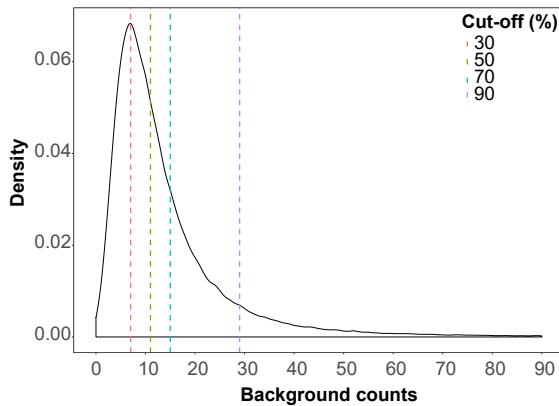
---

2	<i>PAPOLG/REL</i>	rs60685986	2.0	rs1306395 ( $2.99 \times 10^{-5}$ )	1.2	rs1306395	32
18	<i>POLI</i>	18:51926806	0.3	rs602422 (0.0047)	1.1	NA	NA
14	<i>NFKBIA</i>	rs35309046	0.9	rs8016947 ( $9.65 \times 10^{-5}$ )	1.2	NA	NA
11	<i>RPS6KA4</i>	rs146881600	1.3	rs645078 (0.00086)	1.1	NA	NA
6	<i>RSPH3/TAGAP</i>	rs11754601	1.3	rs1973919 (0.018)	1.1	NA	NA
6	<i>TNFAIP3</i>	rs1890370	2.0	rs610604 (0.00032)	1.1	NA	NA
5	<i>TNIP1/ANXA6</i>	rs75851973	2.8	rs76956521 ( $4.98 \times 10^{-9}$ )	1.5	rs76956521	24
10	<i>ZMIZ1</i>	rs2395526	0.9	rs1972346 (0.0082)	1.1	NA	NA
20	<i>ZNF313</i>	rs73129222	1.6	rs6063454 ( $2.90 \times 10^{-5}$ )	1.2	NA	NA
16	<i>ZNF668</i>	rs9939243	0.9	rs7197717 (0.0035)	1.1	NA	NA

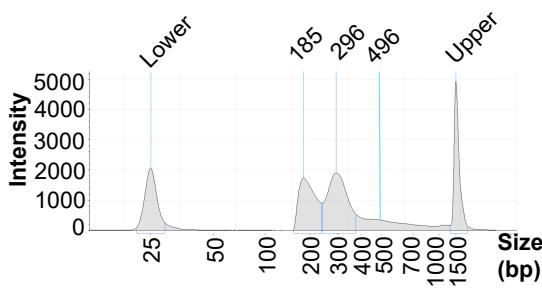
# Appendix B

## Additional figures

### B.1 Chapter 3 Figures



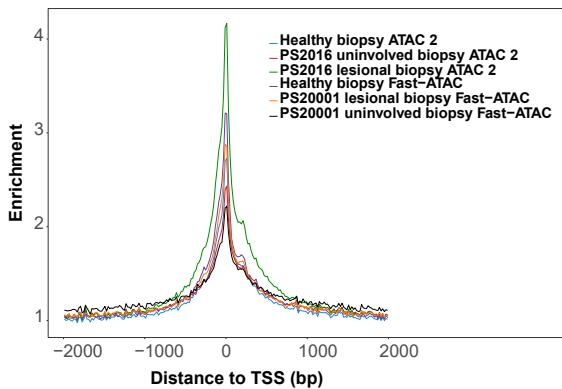
**Figure B.1: Distribution of the background read counts from all the master list peaks absent peaks in each sample.** Each cut-off corresponds to the number of background counts showed by a particular percentage of the total number of absent peaks.



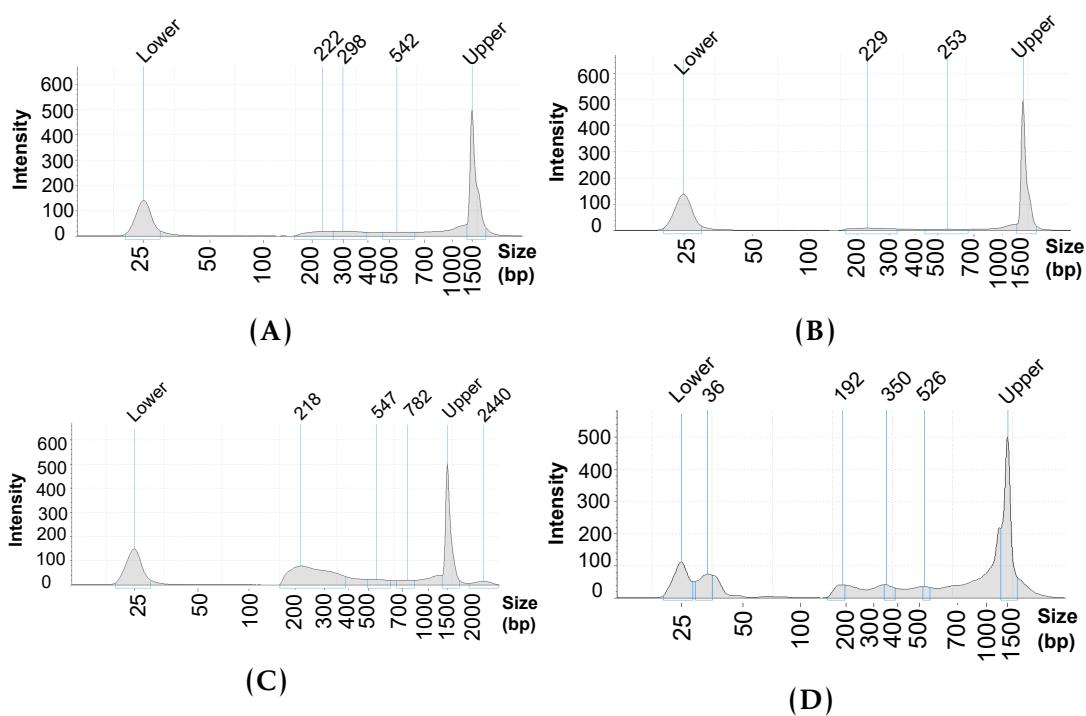
**Figure B.2: Pre-sequencing profiles of relative abundance of DNA library fragment sizes for a psoriatic lesional keratinocytes ATAC 1 library generated using 40 min of transposition.** Pre-sequencing quantification of DNA fragment sizes from the ATAC library generated using 50,000 keratinocytes isolated from a psoriatic lesional skin biopsy and the ATAC 1 protocol (detailed in Table 3.3) and transposition for 40 min.

## Additional figures

---

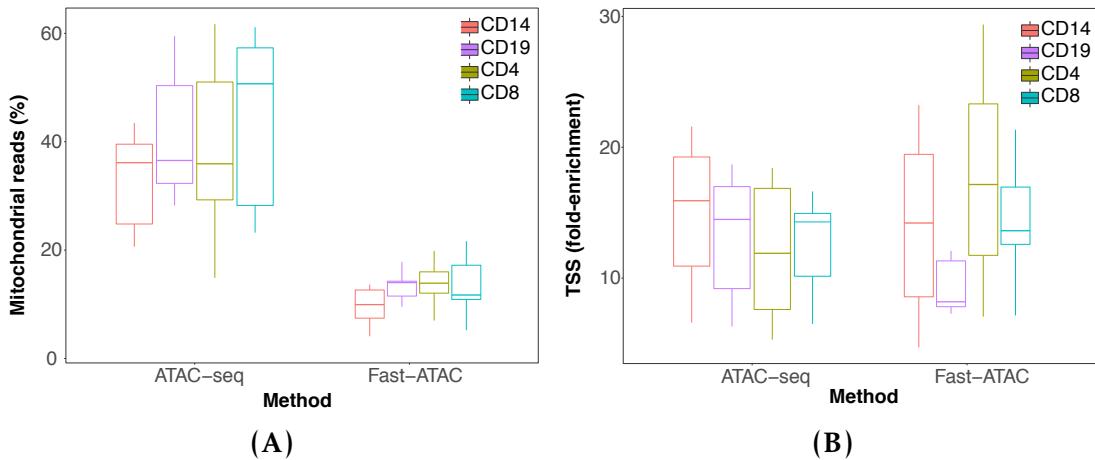


**Figure B.3: Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples.** Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from the different ATAC libraries.



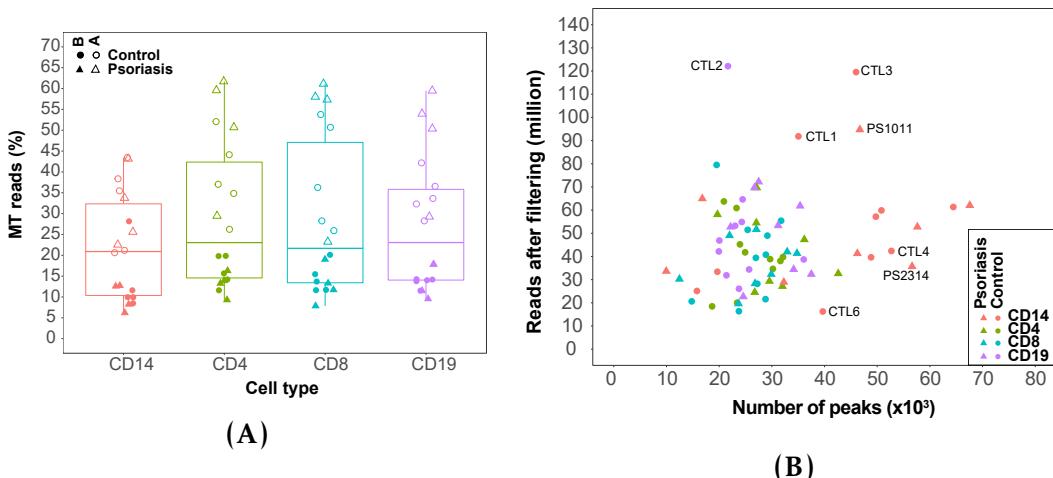
**Figure B.4: Fast-ATAC and Omni-ATAC NHEK pre-sequencing profiles of relative abundance of DNA library fragment sizes.** Pre-sequencing quantification of DNA fragment sizes from the libraries generated using the (A) C2, (B) C3, and (C) C4 versions of the Fast-ATAC protocol based on modifications in the detergent and Tn5 concentration and (D) Omni-ATAC. C2, C3 and C4 detergent and Tn5 concentrations are detailed in Table 3.3.

## Additional figures



**Figure B.5: Differences in TSS fold-enrichment and percentage of mirochondrial between the ATAC-seq and Fast-ATAC libraries from all the samples generated in the thesis.** Boxplot illustrating differences between ATAC-seq and Fast-ATAC by cell type for (A) TSS fold enrichment and (B) mitochondrial reads (percentage).

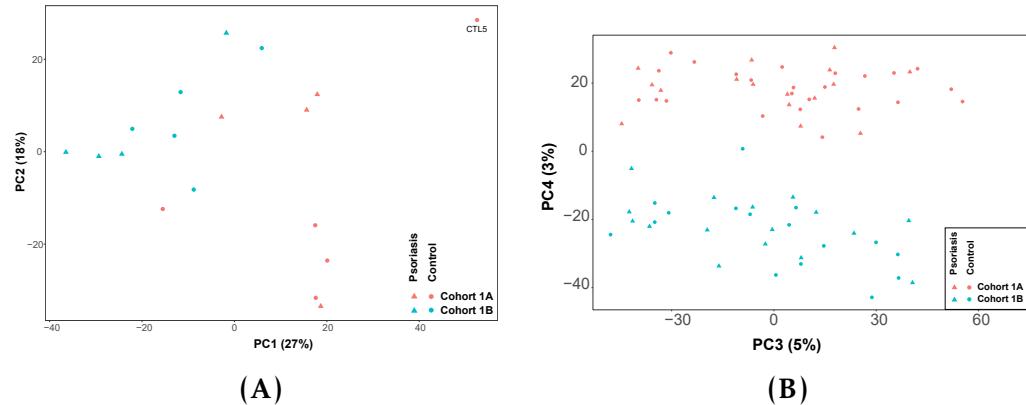
## B.2 Chapter 4 Figures



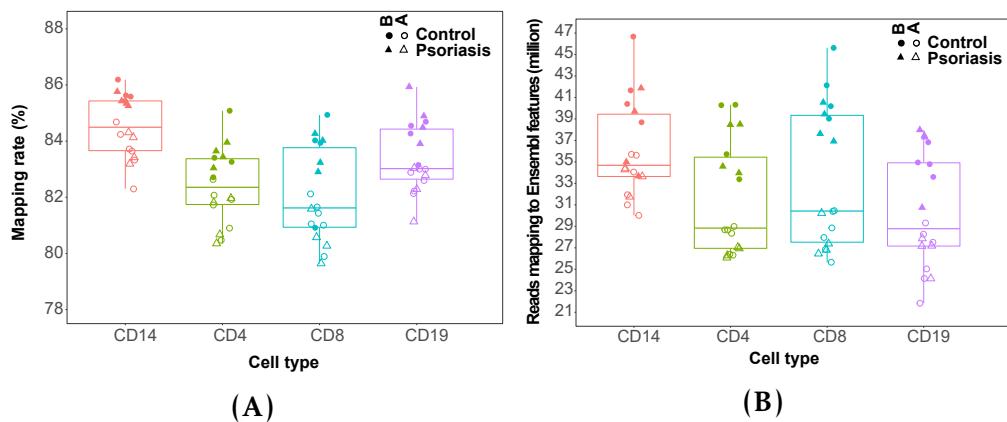
**Figure B.6: (A) Percentage of MT reads in the ATAC-seq samples generated in  $CD14^+$  monocytes,  $CD4^+$ ,  $CD8^+$  and  $CD19^+$  isolated from psoriasis patients and healthy controls.** Samples from cohort 1A (open circles and triangles) were generated with the standard ATAC-seq protocol from 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.

## Additional figures

---



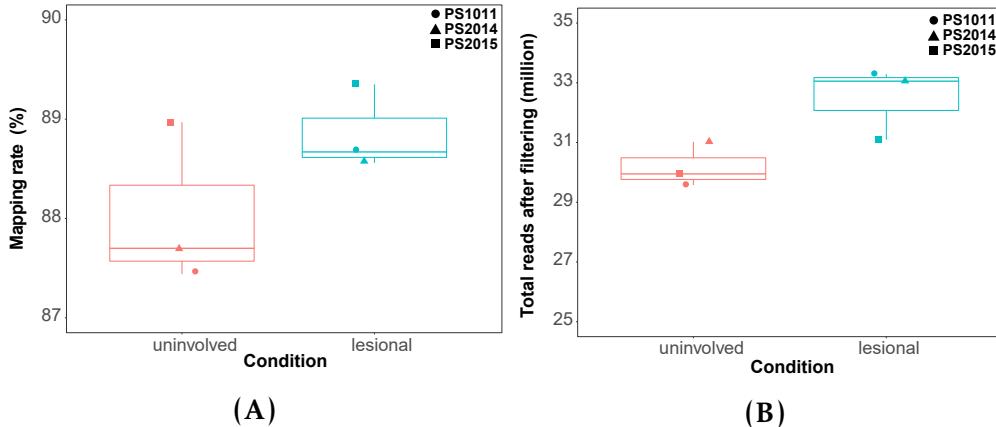
**Figure B.7: PCA analysis illustrating batch effect in ATAC and RNA-seq samples.** (A) First and second component of the PCA analysis performed using the normalised ATAC counts in a list of consensus peaks across all the combined CD8<sup>+</sup> samples from psoriasis patients and healthy controls. (B) Third and fourth component of the PCA analysis performed on the normalised number of reads mapping to the Ensembl list of mRNAs and lncRNAs detected in CD8<sup>+</sup> cells from psoriasis patients and healthy controls.



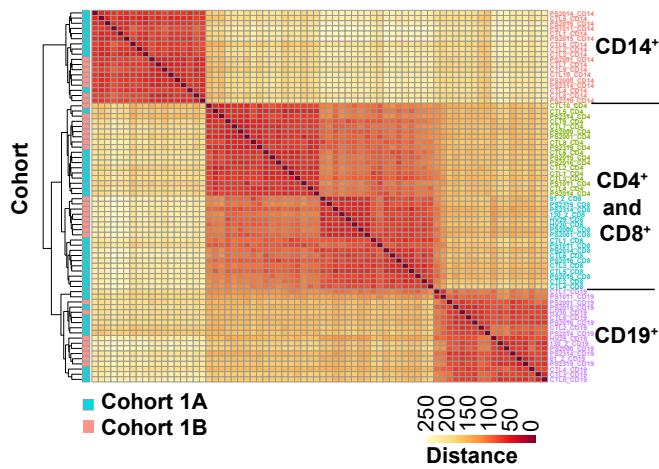
**Figure B.8: Mapping rate and total reads after filtering (million) mapping to Ensembl genes in all the RNA-seq samples from psoriasis patients and controls in four cell types.** (A) The mapping rate refers to the percentage of total sequenced reads from each sample that uniquely mapped to a particular 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.

## Additional figures

---



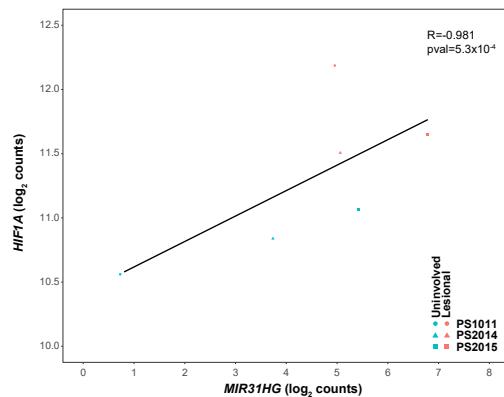
**Figure B.9: Mapping quality control for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients.** (A) Mapping rate calculated as the proportion of sequencing reads mapping uniquely to a particular region of the genome. (B) The total number of reads mapping to an Ensembl feature (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.

## Additional figures

---

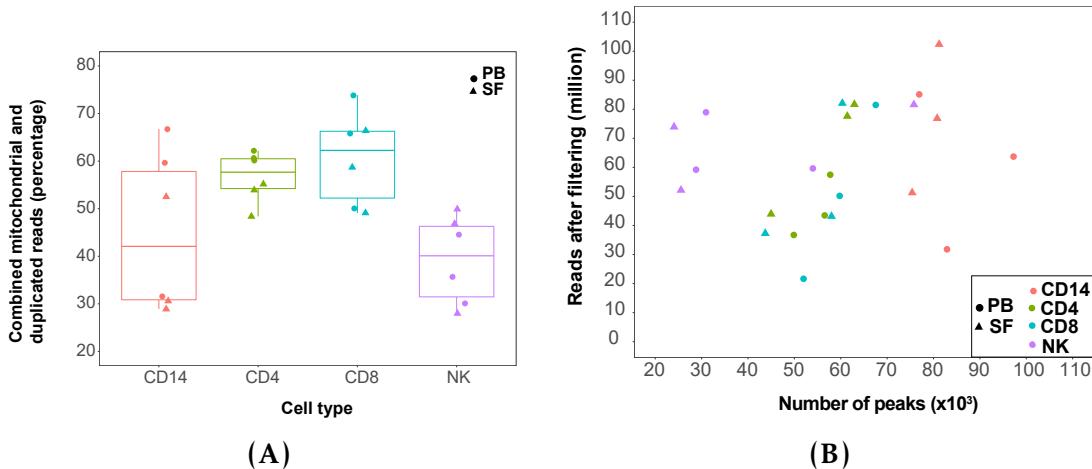


**Figure B.11: Correlation in gene expression between the miRs *MIR31HG* and the target gene *HIF-1A* in lesional and uninvolved skin.** Plot showing the correlation in log<sub>2</sub> normalised read counts for *MIR31HG* and *HIF-1A*. Pearson correlation values (R) and significance (p-value) are included. Each of the dots represents one samples, where colour represents condition (lesional or uninvolved) and shape corresponds to the patient ID.

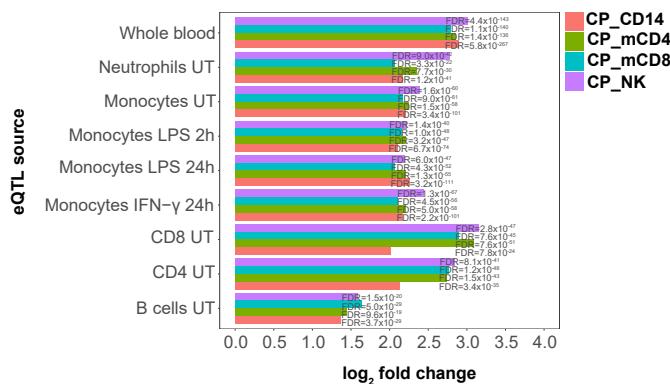
## Additional figures

---

### B.3 Chapter 5 Figures



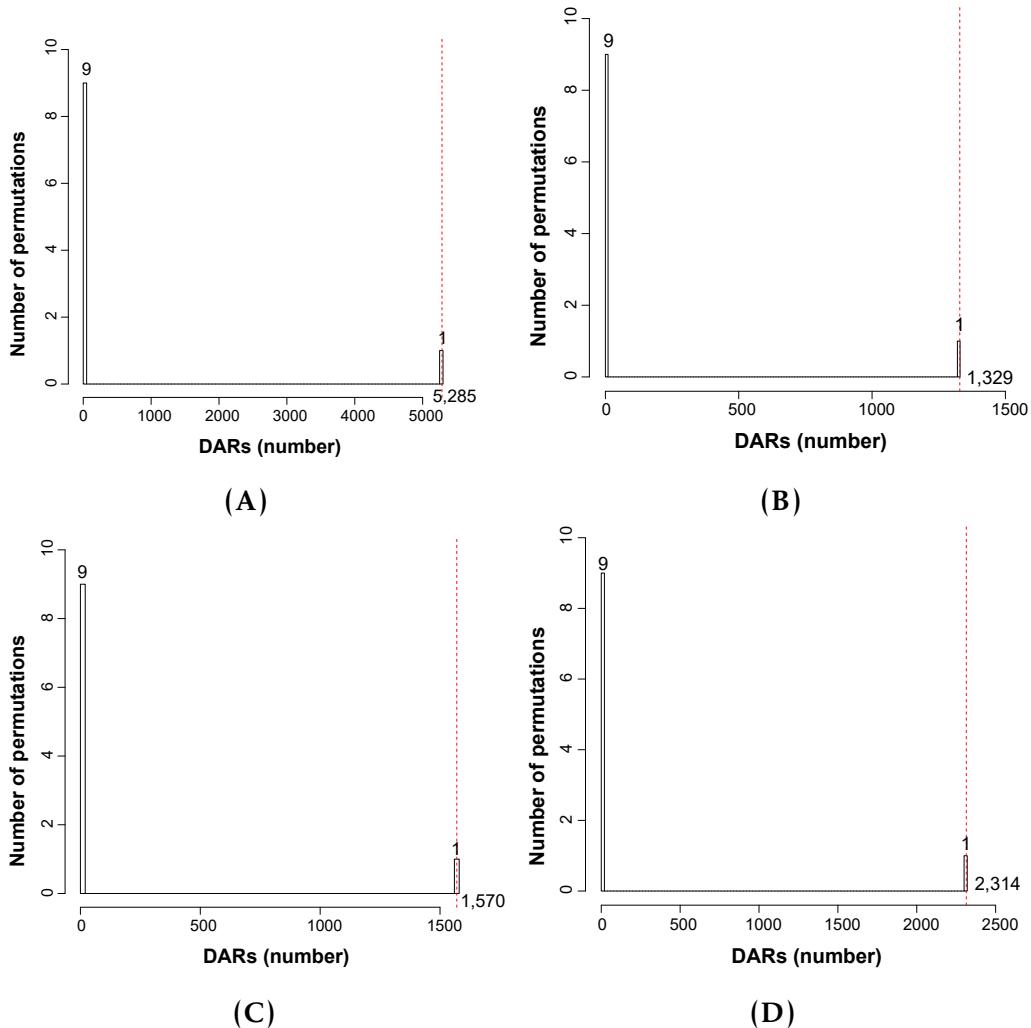
**Figure B.12: Assessment of the percentage of mitochondrial and duplicated reads and the number of called peaks after filtering in the ATAC data from peripheral blood and synovial fluid of PsA patients samples.** For each of the cell types and samples (A) boxplot representing percentage of duplicated 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.13: Enrichment of eQTL SNPs for the consensus list of ATAC peaks used to perform differential chromatin accessibility analysis in each cell type.** Barplot illustrating for the consensus list of ATAC peaks in of the 4 assayed cell types the log<sub>2</sub> fold change and significance (FDR) of the top enriched GTEx eQTL dataset (whole blood) and the publicly available eQTL datasets in immune cell types. Although enrichment of eQTL SNPs in the majority of GTEx datasets for the consensus list of peaks in the 4 cell types was found, whole blood presented the largest and most significant enrichment amongst all of them

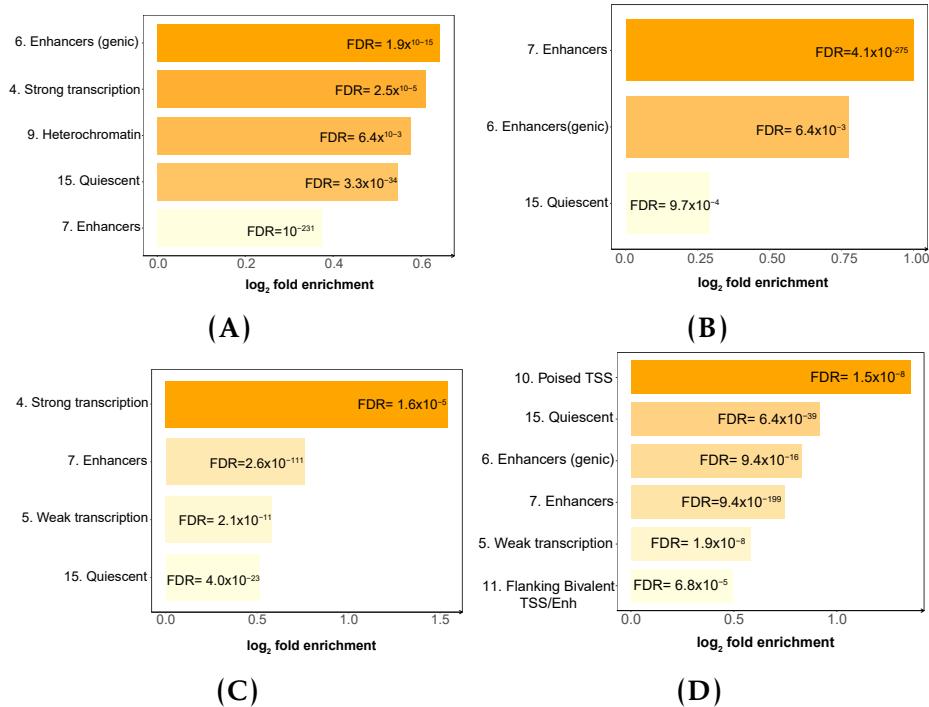
## Additional figures

---

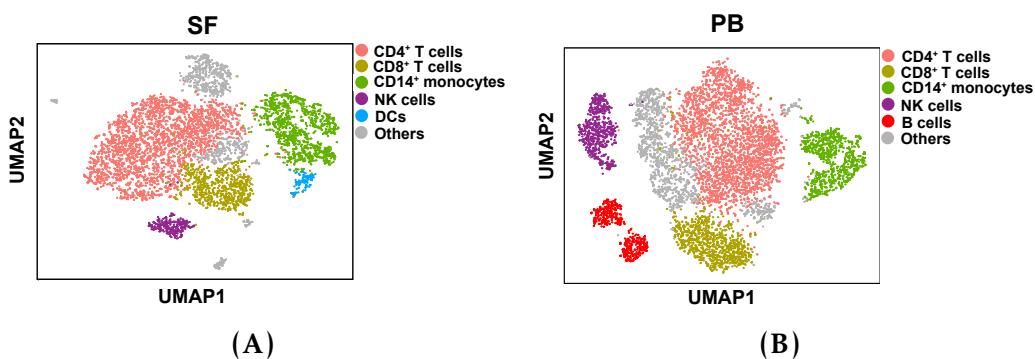


**Figure B.14: Permutation analysis SF vs PB in CD14<sup>+</sup> monocytes, CD4m<sup>+</sup>, CD8m<sup>+</sup> and NK.** Sample labels were permuted within each cell type to achieve the ten unique possible combinations and differential analysis was performed. The number of significant DARs (FDR<0.01 and no abs(FC)>1.5) across all permutations is plotted for (A) CD14<sup>+</sup> monocytes, (B) mCD4<sup>+</sup>, (C) mCD8<sup>+</sup> and (D) NK, demonstrating that the true observation (dashed red line) is significantly more than expected by chance ( $p\text{-value}<0.1$ , the lowest  $p\text{-value}$  for the maximum number of permutations that can be conducted with this sample size) in all four cell types.

## Additional figures



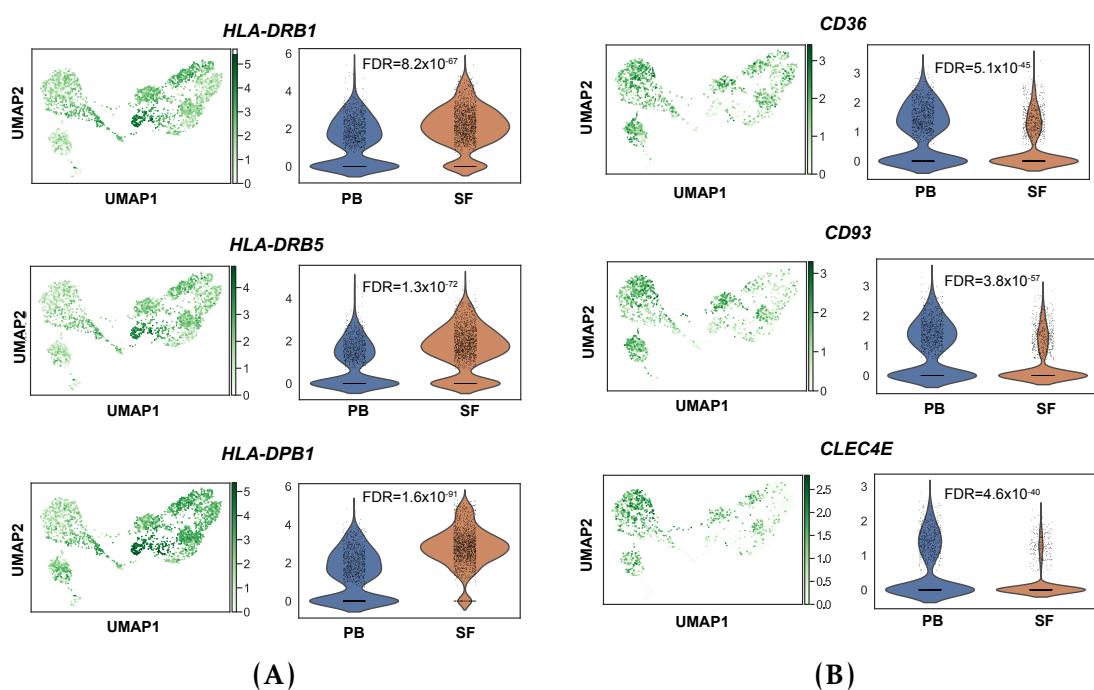
**Figure B.15: Enrichment of DARs in the 15 chromatin states of the cell type specific Roadmap Epigenomic Project segmentation maps.** Barplots illustrating the chromatin states significantly enriched ( $FDR < 0.01$ ) in DARs from the differential chromatin accessibility analysis between synovial fluid and peripheral blood in (A) CD14<sup>+</sup> monocytes, (B) mCD4<sup>+</sup>, (C) mCD8<sup>+</sup> 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<sup>+</sup> 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<sup>+</sup> (pink), CD8<sup>+</sup> (khaki), CD14<sup>+</sup> monocytes (green), NK (purple), DCs (blue), B cells (red) and others (grey).

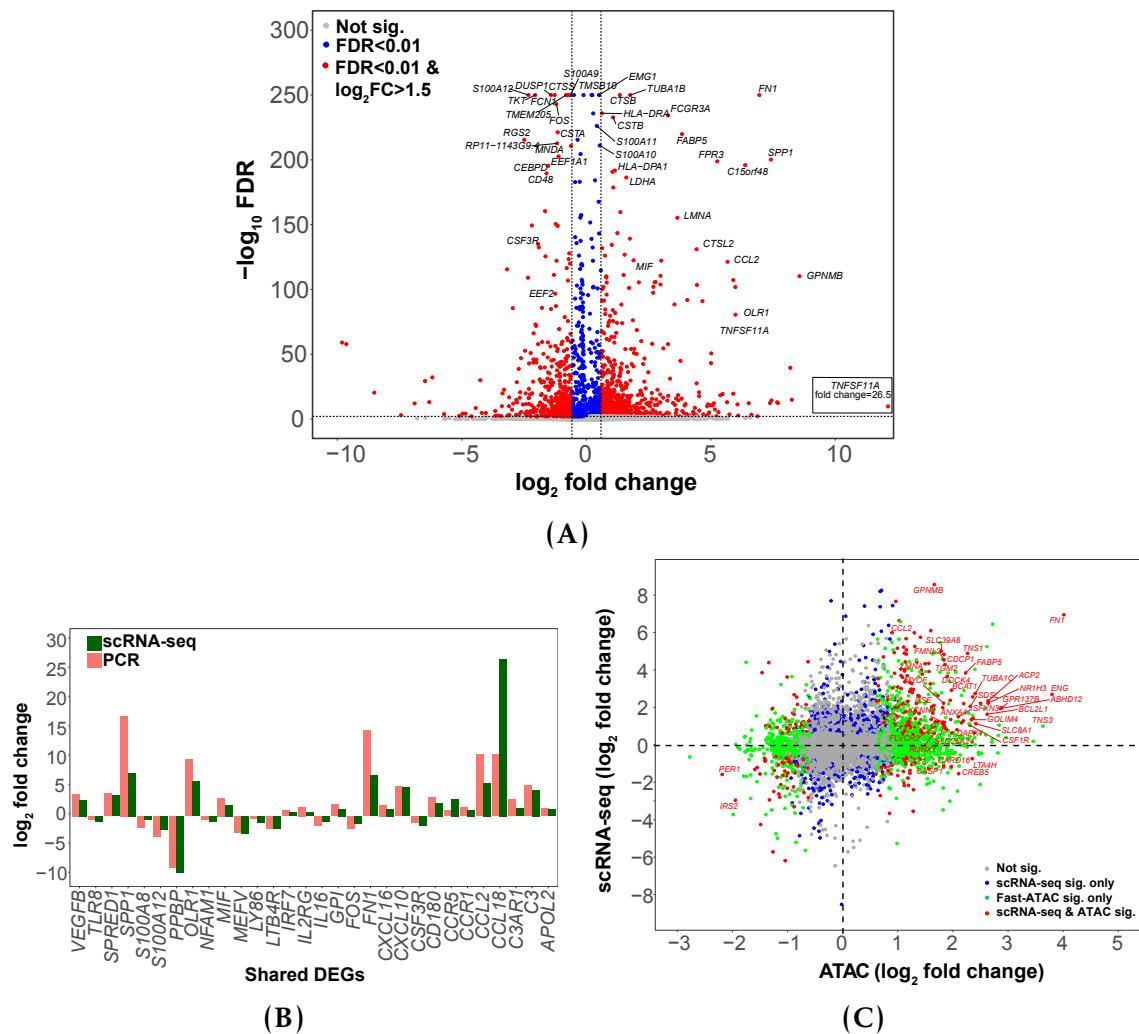
## Additional figures

---



**Figure B.17: Expression levels of antigen presentation and phagocytosis characteristic genes in synovial fluid and peripheral blood CD14<sup>+</sup> 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<sup>+</sup> monocytes analysed by scRNA-seq. On the right, violin plots illustrate for synovial fluid and peripheral blood CD14<sup>+</sup> monocytes (x-axis) the probability density of expression (y-axis) for each of the genes. SF=synovial fluid; PB=peripheral blood.

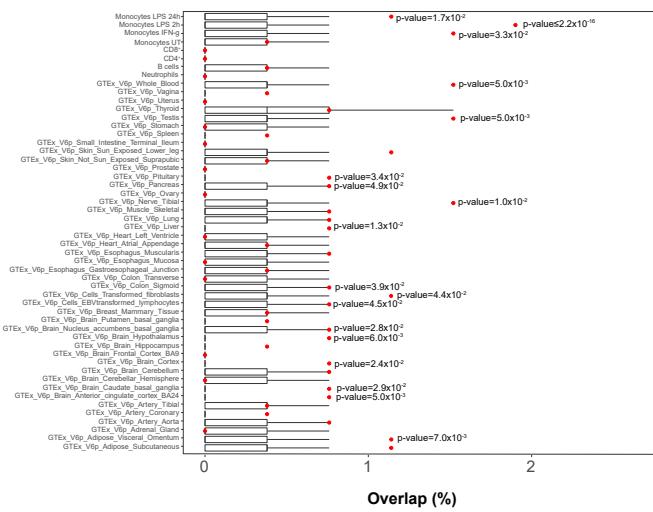
## Additional figures



**Figure B.18: Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14<sup>+</sup> monocytes.** (A) Volcano plots showing differences in gene expression between synovial fluid and peripheral blood CD14<sup>+</sup> monocyte using scRNA-seq data. The significance ( $\log_{10}\text{FDR}$ ) of the differential expression (y-axis) is plotted against the  $\log_2\text{fold change}$ , where a positive fold change indicates higher expression in CD14<sup>+</sup> monocytes from synovial fluid compared to peripheral blood and viceversa. Genes are coloured based on FDR and  $\log_2\text{fold changes}$  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<sup>+</sup> monocytes. The barplot illustrates the  $\log_2\text{fold change}$  for the 30 DEGs identified by scRNA-seq (based on  $\text{FDR}<0.01$  and  $\text{fold change}>1.5$ ) and qPCR array (based on  $p\text{-value}<0.05$ ) analysis. (B) Correlation plot comparing synovial fluid and peripheral blood differences in scRNA-seq expression and ATAC chromatin accessibility in CD14<sup>+</sup> monocytes. The  $\log_2\text{fold changes}$  for scRNA-seq differential expression of all transcripts in CD14<sup>+</sup> monocytes are plotted against the  $\log_2\text{fold change}$  for total CD14<sup>+</sup> monocytes ATAC differential chromatin accessibility analysis in regions proximal ( $\leq 5\text{Kb}$ ) 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<sup>+</sup> monocytes. Significance is based on FDR and fold change thresholds ( $\text{FDR}<0.01$  and  $\text{fold change}>1.5$ ) in both assays.

## Additional figures

---



**Figure B.19: Enrichment of PsA GWAS fine-mapped SNPs for publicly available eQTL datasets.** The percentage of SNPs from the combined 90% credible setx SNPs of the seven successfully fine-mapped PsA loci overlapping SNPs from publicly available eQTL datasets (immune cell types and GTEx) are shown (red dot). This overlap is compared to the distribution of overlaps with the same eQTL datasets for each of the 1,000 sets of background SNPs (boxplots). Significance of the enrichment is indicated as p-value calculated using binomial test.

# Bibliography

- Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489: pp. 57–74.
- Abji, F, Pollock, RA, Liang, K, and &, Chandran V (2016). "Brief report: CXCL10 is a possible biomarker for the development of psoriatic arthritis among patients with psoriasis". *Arthritis &*
- Abraham, Gad and Inouye, Michael (2014). "Fast principal component analysis of large-scale genome-wide data". *PloS one* 9.
- Abramson, SR. and Khattri, S (2016). "A Review of Current Evidence for TNF Inhibitor Switching in Psoriatic Arthritis". *Journal of Psoriasis and Psoriatic arthtritis.*
- Ackermann, A. M., Wang, Z., Schug, J., and Naji, A. (2016). "Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes". *Molecular*.
- Adams, D., Altucci, L., Antonarakis, S. E., and Nature, Ballesteros-J. (2012). "BLUEPRINT to decode the epigenetic signature written in blood". *Nature*.
- Adey, Andrew, Morrison, Hilary G., Asan, Xun, Xu, Kitzman, Jacob O., et al. (2010). "Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition". *Genome Biology* 11.
- Ahn, R, 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*.
- Ahn, RS, Taravati, Keyon KL, Lee, KM, and Nititham, J (2017). "Transcriptional landscape of epithelial and immune cell populations revealed through FACS-seq of healthy human skin". *Scientific*.
- Ai, R., Hammaker, D., Boyle, D. L., and Nature, Morgan-R. (2016). "Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes". *Nature*.
- Al-Mossawi, H., Yager, N., Taylor, C., Lau, E., and bioRxiv, Danielli-S. (2018). "Context-specific regulation of monocyte surface IL7R expression and soluble receptor secretion by a common autoimmune risk allele". *bioRxiv*.
- Al-Mossawi, M. H., Chen, L., Fang, H., Ridley, A., and Nature, Wit-J. (2017). "Unique transcriptome signatures and GM-CSF expression in lymphocytes from patients with spondyloarthritis". *Nature*.
- Alasoo, K., Rodrigues, J., Mukhopadhyay, S., and Nature, Knights-A. J. (2018). "Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response". *Nature*.

## BIBLIOGRAPHY

---

- Alrefai, H, Muhammad, K, Rudolf, R, and Nature, DAT (2016). "NFATc1 supports imiquimod-induced skin inflammation by suppressing IL-10 synthesis in B cells". *Nature*.
- Amatullah, H, Shan, Y, and, Beauchamp BL journal of (2017). "DJ-1/PARK7 impairs bacterial clearance in sepsis". *American journal of*.
- Ameres, S. L., Horwich, M. D., Hung, J. H., and, Xu-J. (2010). "Target RNAdirected trimming and tailing of small silencing RNAs" ..
- Anders, S., Pyl, P. T., and Bioinformatics, Huber-W. (2015). "HTSeqa Python framework to work with high-throughput sequencing data". *Bioinformatics*.
- Andersen, William M. Gordon, Michael, D. Zeller, Rachel, H. Klein, William, R. Swindell, Hsiang, Ho, et al. (2014). "A GRHL3-regulated repair pathway suppresses immune-mediated epidermal hyperplasia".
- Andersson, Robin, Gebhard, Claudia, Miguel-Escalada, Irene, Hoof, Ilka, Bornholdt, Jette, et al. (2014). "An atlas of active enhancers across human cell types and tissues". *Nature* 507: p. 455.
- Andrews, TS (2018). "False signals induced by single-cell imputation". *F1000Research*.
- Antonelli, Alessandro, Ferrari, Silvia, Giuggioli, Dilia, Ferrannini, Ele, Ferri, Clodoveo, et al. (2014). "Chemokine (CXCL motif) ligand (CXCL) 10 in autoimmune diseases". 13: pp. 272–280.
- Antonioli, L., Pacher, P., Vizi, E. S., and medicine, Hask-G. in molecular (2013). "CD39 and CD73 in immunity and inflammation". *Trends in molecular medicine*.
- Apel, M., Uebe, S., Bowes, J., and, Giardina-E. (2013). "Variants in RUNX3 contribute to susceptibility to psoriatic arthritis, exhibiting further common ground with ankylosing spondylitis". *Arthritis &*.
- Appel, H., Maier, R., and Arthritis, 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*.
- Arend, W. P., Palmer, G., and reviews, Gabay-C. (2008). "IL1, IL18, and IL33 families of cytokines". *Immunological reviews*.
- Armstrong, AW, Harskamp, CT, and, Dhillon JS journal of (2014). "Psoriasis and smoking: a systematic review and metaanalysis". *British journal of*.
- Assassi, S., Reveille, J. D., and of, Arnett-F. C. (2010). "Whole-blood gene expression profiling in ankylosing spondylitis shows upregulation of toll-like receptor 4 and 5". *The Journal of*.
- Aterido, Adrià, Julià, Antonio, Ferrández, Carlos, Puig, Lluís, Fonseca, Eduardo, 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., Bryer, C., and reports, Cubitt-L. (2018). "A practical solution for preserving single cells for RNA sequencing". *Scientific reports*.

## BIBLIOGRAPHY

---

- B, Uszczynska-Ratajczak, Lagarde, J, and resource, Frankish A (2018). "Towards a complete map of the human long non-coding RNA transcriptome". *resource*.
- Baba, T and oncology, Mukaida N cellular (2014). "Role of macrophage inflammatory protein (MIP)-1/CCL3 in leukemogenesis". *Molecular & cellular oncology*.
- Baechler, E. C., Batliwalla, F. M., and the, Karypis-G. of (2003). "Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus". *Proceedings of the*.
- Baeten, D., Breban, M., Lories, R., and, Schett-G. (2013). "Are spondylarthritides related but distinct conditions or a single disease with a heterogeneous phenotype?" *Arthritis &*.
- Bai, F, Zheng, W, Dong, Y, Wang, J, Garstka, MA, et al. (2018). "Serum levels of adipokines and cytokines in psoriasis patients: a systematic review and meta-analysis". *Oncotarget*.
- Baker, K. F. and diseases, Isaacs-J. D. of the rheumatic (2017). "Novel therapies for immune-mediated inflammatory diseases: What can we learn from their use in rheumatoid arthritis, spondyloarthritis, systemic lupus erythematosus". *Annals of the rheumatic diseases*.
- 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, and Nature, Miska EA (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain". *Nature*.
- Bao, Xiaomin, Rubin, Adam J, Qu, Kun, Zhang, Jiajing, Giresi, Paul G, 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. and functional, Chandra-T. in (2017). "Experimental design for single-cell RNA sequencing". *Briefings in functional*.
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., and Cell, Schones-D. E. (2007). "High-resolution profiling of histone methylations in the human genome". *Cell*.
- Batliwalla, F. M., Li, W., Ritchlin, C. T., and Molecular, Xiao-X. (2005). "Microarray analyses of peripheral blood cells identifies unique gene expression signature in psoriatic arthritis". *Molecular*.
- Belasco, J, Louie, JS, Gulati, N, and &, Wei N (2015). "Comparative genomic profiling of synovium versus skin lesions in psoriatic arthritis". *Arthritis &*.

## BIBLIOGRAPHY

---

- Bell, A. C. and Nature, Felsenfeld-G. (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene". *Nature*.
- Bensch, B., Ohtani, T., Herati, R. S., and of, Bovenschen-N. (2018). "Deep immune profiling by mass cytometry links human T and NK cell differentiation and cytotoxic molecule expression patterns". *Journal of*.
- Benham, H., Norris, P., and Arthritis, Goodall-J. (2013). "Th17 and Th22 cells in psoriatic arthritis and psoriasis". *Arthritis*.
- 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.
- Bergboer, J. G. M., Tjabringa, G. S., and of, Kamsteeg-M. journal (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*.
- Bernstein, B. E., Stamatoyannopoulos J. A. and Costello, JF., Ren, B., Milosavljevic, A., Meissner, A., et al. (2010). "The NIH roadmap epigenomics mapping consortium". *Nature Biotechnology*.
- 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., and the, Gao-W. of (2016). "Dysregulated bioenergetics: a key regulator of joint inflammation". *Annals of the*.
- Black, A. P. B. and ArdernJones, M. R. (2007). "Human keratinocyte induction of rapid effector function in antigenspecific memory CD4+ and CD8+ T cells". *European journal of*.
- Blais, M. E., Dong, T., and Immunology, RowlandJones-S. (2011). "HLAC as a mediator of natural killer and Tcell activation: spectator or key player?" *Immunology*.
- Blauvelt, A., Papp, K. A., Griffiths, C. E. M., and American, Randazzo-B. of the (2017). "anti-interleukin-23 monoclonal antibody, compared with adalimumab for the continuous treatment of patients with moderate to severe psoriasis: results from the phase". *Journal of the American*.
- Bonifati, V, Rizzu, P, Baren, Van MJ, and, Schaap O (2003). "Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism" ..
- Bouaziz, JD, Yanaba, K, and, Venturi GM of the (2007). "Therapeutic B cell depletion impairs adaptive and autoreactive CD4+ T cell activation in mice". *Proceedings of the*.
- Boudousqui, Caroline, Danilo, Maxime, Pousse, Laurne, Jeevan-Raj, Beena, Angelov, Georgi S., 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.

## BIBLIOGRAPHY

---

- Bowes, J., Ho, P., Flynn, E., and the, Ali-F. of (2012). "Comprehensive assessment of rheumatoid arthritis susceptibility loci in a large psoriatic arthritis cohort". *Annals of the*.
- Bowes, John, Budu-Aggrey, Ashley, Huffmeier, Ulrike, Uebe, Steffen, Steel, Kathryn, et al. (2015). "Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis". *Nature communications* 6: p. 6046.
- Boyle, A. P., Song, L., Lee, B. K., London, D., and Genome, Keefe-D. (2010). "High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells". *Genome*.
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., 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., Nickoloff, B. J., Suter, M., 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.
- Branisteanu, DE, Voicu, CM, and Medical-Surgical, Cretu A (2015). "Adverse reactions of biological therapy for psoriasis". *The Medical-Surgical*.
- Braun, J. E., Huntzinger, E., Fausser, M., and cell, Izaurrealde-E. (2011). "GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets". *Molecular cell*.
- Brest, P., Lapaquette, P., Souidi, M., and Nature, Lebrigand-K. (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*.
- Broome, AM, Ryan, D, and Histochemistry, Eckert RL of (2003). "S100 protein subcellular localization during epidermal differentiation and psoriasis". *Journal of Histochemistry*.
- Buck, M. J., Raaijmakers, L. M., Ramakrishnan, S., and Oncogene, Wang-D. (2014). "Alterations in chromatin accessibility and DNA methylation in clear cell renal cell carcinoma". *Oncogene*.
- Buenrostro, Jason D., Paul, G. Giresi, Lisa, C. Zaba, Howard, Y. Chang, and William, J. Greenleaf (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., and Greenleaf, William J. (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., and Morris, A. P. (2015). "Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci". *PLoS Genet* 11: pp. 1–14.

## BIBLIOGRAPHY

---

- Burgos-Pol, R. and Martnez-Sesmero, J. M. (2016). "The cost of psoriasis and psoriatic arthritis in 5 European countries: a systematic review". *Actas Dermo.*
- Burton P., WTCCC2 (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls". *Nature.*
- Calderon, D, Nguyen, MLT, Mezger, A, and A, bioRxiv, Kathiria (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.*
- Cañete, JD, Martínez, SE, Farrés, J, and, Sanmartí R of the (2000). "Differential Th1/Th2 cytokine patterns in chronic arthritis: interferon is highly expressed in synovium of rheumatoid arthritis compared with seronegative ". *Annals of the.*
- Cañete, JD, Celis, R, Hernández, V, and, Pablos JL of the (2010). "Synovial immunopathological changes associated with successful abatacept therapy in a case of severe refractory psoriatic arthritis". *Annals of the.*
- Cao, Qin, Anyansi, Christine, Hu, Xihao, Xu, Liangliang, Xiong, Lei, et al. (2017). "Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines." *Nat. Genet.*
- Capon, Francesca (2017). "The Genetic Basis of Psoriasis". *International Journal of Molecular Sciences* 18: p. 2526.
- Capon, Francesca, Bijlmakers, Marie-Jos J., Wolf, Natalie, Quaranta, Maria, Huffmeier, Ulrike, 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, Garcia, Veronica E., Brandon, Rhonda, 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.
- Carrieri, Claudia, Cimatti, Laura, Biagioli, Marta, Beugnet, Anne, Zucchelli, Silvia, 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., Miceli-Richard, C., Bourgain, C., et al. (2007). "Combined analysis of three whole genome linkage scans for ankylosing spondylitis". *Rheumatology* 46: pp. 763–771.
- Chandran, Vinod, Schentag, Catherine T., Brockbank, John E., Pellett, Fawnda J., Shanmugarajah, S., et al. (2009). "Familial aggregation of psoriatic arthritis". *Annals of the rheumatic diseases* 68: pp. 664–667.
- Chang, C. C., Chow, C. C., and, Lcam (2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets" ..

## BIBLIOGRAPHY

---

- 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". *Proceedings of the*.
- Chang, X, Yamada, R, Suzuki, A, and, Kochi Y (2005). "Citrullination of fibronectin in rheumatoid arthritis synovial tissue". *Rheumatology*.
- Chen, Xingqi, Shen, Ying, Draper, Will, Buenrostro, Jason D., Litzenburger, Ulrike, et al. (2016). "ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing". *Nature methods* 13: p. 1013.
- Chiu, YG, Shao, T, and Arthritis, Feng C (2010). "CD16 (FcRIII) as a potential marker of osteoclast precursors in psoriatic arthritis". *Arthritis*.
- Chou, W. C., Zheng, H. F., Cheng, C. H., Yan, H., and reports, Wang-L. (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, and &, Tocker J (2007). "Interferon regulates susceptibility to collageninduced arthritis through suppression of interleukin17". *Arthritis &*.
- Cid, De R, E, Riveira-Munoz, and Nature, PLJM (2009). "Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis". *Nature*.
- Clegg, Daniel O., Reda, Domenic J., Mejias, Edwin, Cannon, Grant W., Weisman, Michael H., 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., and and, Paul-C. in arthritis (2016a). "Psoriasis, psoriatic arthritis, and rheumatoid arthritis: Is all inflammation the same?" *Seminars in arthritis and*.
- Coates, Laura C., Kavanaugh, Arthur, Mease, Philip J., Soriano, Enrique R., Felquer, Maria, 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., and Genomics, Sinha-A. A. (2012). "Global transcriptional analysis of psoriatic skin and blood confirms known disease-associated pathways and highlights novel genomic hot spots for". *Genomics*.
- Cohen, A. D., Sherf, M., Vidavsky, L., Vardy, D. A., and Dermatology, Shapiro-J. (2008). "Association between psoriasis and the metabolic syndrome". *Dermatology*.
- Constant, S, Schweitzer, N, and, West J of (1995). "B lymphocytes can be competent antigen-presenting cells for priming CD4+ T cells to protein antigens in vivo." *The Journal of*.

## BIBLIOGRAPHY

---

- Corces, MR, Trevino, AE, Hamilton, EG, and Nature, Greenside PG (2017). “An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues”. *Nature*.
- Corces, Ryan M., Buenrostro, Jason D., Wu, Bejing, Greenside, Peyton G., Chan, Steven M., 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., and and, Sun-L. (2010). “Mitogen-activated protein kinase phosphatase 2 regulates the inflammatory response in sepsis”. *Infection and*.
- Cortes, A., Pulit, S. L., Leo, P. J., and Nature, Pointon-J. J. (2015). “Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1”. *Nature*.
- 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., Robinson, Philip C., Karaderi, Tugce, 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 Cell, Nasmyth-K. (1999). “Ordered recruitment of transcription and chromatin remodeling factors to a cell cycleand developmentally regulated promoter”. *Cell*.
- Creyghton, M. P. and the, Cheng-A. W. of (2010). “Histone H3K27ac separates active from poised enhancers and predicts developmental state”. *Proceedings of the*.
- Cui, Y., Li, G., Li, S., and Biology, Wu-R. in (2010). “Designs for linkage analysis and association studies of complex diseases”. *Statistical Methods in Molecular Biology*.
- Cusanovich, Darren A, Daza, Riza, Adey, Andrew, Pliner, Hannah A, Christiansen, Lena, 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”. *& Rheumatism: Official Journal of the*.
- Dand, N., Mucha, S., Tsoi, L. C., and molecular, Mahil-S. K. (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*.
- Dang, EV, Barbi, J, Yang, HY, Jinasena, D, and Cell, Yu H (2011). “Control of TH17/Treg Balance by Hypoxia-Inducible Factor 1”. *Cell*.
- Das, Sayantan, Stuart, Philip E., Ding, Jun, Tejasvi, Trilokraj, Li, Yanming, et al. (2014). “Fine mapping of eight psoriasis susceptibility loci”. *European Journal of Human Genetics*.

## BIBLIOGRAPHY

---

- Davies, James O., Oudelaar, A. M., Higgs, Douglas R., and Hughes, Jim R. (2017). "How best to identify chromosomal interactions: a comparison of approaches". *Nature methods* 14: pp. 125–134.
- De Santa, Francesca, Barozzi, Iros, Mietton, Flore, Ghisletti, Serena, Polletti, Sara, et al. (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers". *PLoS biology* 8.
- Delaneau, O., Marchini, J., and methods, Zagury-J. F. (2012). "A linear complexity phasing method for thousands of genomes". *Nature methods*.
- Derrien, Thomas, Johnson, Rory, Bussotti, Giovanni, Tanzer, Andrea, Djebali, Sarah, 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, Yim, K, and Kathail, P (2018). "Recovering Gene Interactions from Single-Cell Data Using Data Diffusion".
- Diluvio, L., Vollmer, S., Besgen, P., and of, Ellwart-J. W. (2006). "Identical TCR -chain rearrangements in streptococcal angina and skin lesions of patients with psoriasis vulgaris". *The Journal of*.
- Dimas, Antigone S., Deutsch, Samuel, Stranger, Barbara E., Montgomery, Stephen B., Borel, Christelle, 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, Oostrom, Leone A. A., Cowan, Chad A., 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, Alexander, Davis, Carrie A., Schlesinger, Felix, Drenkow, Jorg, Zaleski, Chris, et al. (2013). "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics* 29: pp. 15–21.
- Dolcino, M, Pelosi, A, Fiore, PF, and, Patuzzo G in (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*.
- Dolcino, Marzia, Ottavia, Andrea, Barbieri, Alessandro, Patuzzo, Giuseppe, Tinazzi, Elisa, 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., and Pennington, SR. (2012). "New insight into the functions of the interleukin-17 receptor adaptor protein Act1 in psoriatic arthritis". *Arthritis Research and Therapy*.
- Dozmorov, M. G., Dominguez, N., and insights, Bean-K. *biology* (2015). "B-cell and monocyte contribution to systemic lupus erythematosus identified by cell-type-specific differential expression analysis in RNA-seq data". *and biology insights*.

## BIBLIOGRAPHY

---

- Dubois, PCA, Trynka, G, Franke, L, and Nature, Hunt KA (2010). "Multiple common variants for celiac disease influencing immune gene expression". *Nature*.
- Duffy, D. L., Spelman, L. S., and of, Martin-N. G. of the (1993). "Psoriasis in Australian twins". *Journal of the American Academy of*.
- Duren, Zhana, Chen, Xi, Zamanighomi, Mahdi, Zeng, Wanwen, Satpathy, Ansuman T, 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, and Investigative, Robinson N of (2004). "S100 proteins in the epidermis". *Journal of Investigative*.
- Eder, L, Law, T, and &, Chandran V care (2011). "Association between environmental factors and onset of psoriatic arthritis in patients with psoriasis". *Arthritis care &*.
- Edwards, Alexis C, Bigdeli, Tim B, Docherty, Anna R, Bacanu, Silviu, Lee, Donghyung, 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., and Dunning, A. M. (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., and Jones, F. (1990). "Clearance of severe psoriasis after allogenic bone marrow transplantation". *BMJ* 300.
- Ellinghaus, David, Jostins, Luke, Spain, Sarah L., Cortes, Adrian, Bethune, Jrn, 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., and Nair, R. P. (2010). "Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2". *Nature genetics* 42: pp. 991–996.
- ENCODE (2007). "Identification and analysis of functional elements in the human genome by the ENCODE pilot project". *Nature* 447: p. 799.
- Eppinga, H. and Konstantinov, SR. (2014). "The microbiome and psoriatic arthritis". *Current rheumatology*.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., and Shores, N. (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.

## BIBLIOGRAPHY

---

- Espinoza, LR, Aguilar, JL, and, Espinoza CG of (1994). "Fibroblast function in psoriatic arthritis. I. Alteration of cell kinetics and growth factor responses." *The Journal of*.
- Evans, D. M., Spencer, C. C. A., Pointon, J. J., Su, Z., and Nature, Harvey-D. (2011). "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility". *Nature*.
- Fabregat, Antonio, Jupe, Steven, Matthews, Lisa, Sidiropoulos, Konstantinos, Gillespie, Marc, et al. (2018). "The reactome pathway knowledgebase". *Nucleic acids research* 46.
- Faghhihi, Mohammad A., Modarresi, Farzaneh, Khalil, Ahmad M., Wood, Douglas E., Sahagan, Barbara G., 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., and the, Kuijjer-M. L. of (2017). "Exploring regulation in tissues with eQTL networks". *Proceedings of the*.
- Fairfax, Benjamin P., Makino, Seiko, Radhakrishnan, Jayachandran, Plant, Katharine, Leslie, Stephen, 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, Naranbhai, Vivek, Wong, Daniel, 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., and Knight, Julian C. (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 dermatology, Watson-W. of (1974). "Natural history of psoriasis in 61 twin pairs". *Archives of dermatology*.
- Farh, Kyle, Carrasco-Alfonso, J., Dita, Mayer, Luckey, C., Nikolaos, A., et al. (2015). "Genetic and epigenetic fine mapping of causal autoimmune disease variants". *Nature* 518: 337343.
- 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, Hohl, D, and Biology, Werner S (2007). "The inflammasome mediates UVB-induced activation and secretion of interleukin-1 by keratinocytes". *Current Biology*.
- Ferlazzo, G, Pack, M, and, Thomas D of the (2004). "Distinct roles of IL-12 and IL-15 in human natural killer cell activation by dendritic cells from secondary lymphoid organs". *Proceedings of the*.

## BIBLIOGRAPHY

---

- Fert, I., Cagnard, N., Glatigny, S., and, Letourneur-F. (2014). "Reverse interferon signature is characteristic of antigenpresenting cells in human and rat spondyloarthritis". *Arthritis &*
- Finlay, Andrew (2005). "Current severe psoriasis and the rule of tens". *British Journal of Dermatology* 152: pp. 861–867.
- Finzel, S., Sahinbegovic, E., and, Kocjan-R. (2014). "Inflammatory bone spur formation in psoriatic arthritis is different from bone spur formation in hand osteoarthritis". *Arthritis &*
- Fiocco, Ugo, Martini, Veronica, Accordi, Benedetta, Caso, Francesco, Costa, Luisa, 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, Baillie, Kenneth J., Hoon, Michiel J. L. de, et al. (2014). "A promoter-level mammalian expression atlas". *Nature* 507.
- Fredriksson, T. and Pettersson, U. (1978). "Severe psoriasis oral therapy with a new retinoid". *Dermatology*.
- Friedman, R. C., Farh, K. K. H., and research, Burge-C. B. (2008). "Most mammalian mRNAs are conserved targets of microRNAs". *Genome research*.
- Gandhapudi, S. K., Murapa, P., and Md (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". *(Baltimore)*.
- Gao, J., Chen, F., Hua, M., Guo, J., and Biological, Nong-Y. (2018). "Knockdown of lncRNA MIR31HG inhibits cell proliferation in human HaCaT keratinocytes". *Biological*.
- Gardembas-Pain, M., Ifrah, R., Foussard, C., Boasson, M., Saint Andre, JP., et al. (1990). "Psoriasis after allogeneic bone marrow transplantation". *Archives of dermatology*.
- Gaublomme, JT, Yosef, N, Lee, Y, Gertner, RS, and Cell, Yang LV (2015). "Single-cell genomics unveils critical regulators of Th17 cell pathogenicity". *Cell*.
- Gaulton, K. J., Nammo, T., Pasquali, L., and Nature, Simon-J. M. (2010). "A map of open chromatin in human pancreatic islets". *Nature*.
- Gelfand, J. M., Gladman, D. D., Mease, P. J., and Academy, Smith-N. (2005). "Epidemiology of psoriatic arthritis in the population of the United States". *the American Academy*.
- Gelfand, J. M., Neumann, A. L., Shin, D. B., and Jama, Wang-X. (2006). "Risk of myocardial infarction in patients with psoriasis". *Jama*.
- Giresi, P. G., Kim, J., McDaniell, R. M., and Genome, Iyer-V. R. (2006). "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin". *Genome*.
- Gladman, D. D., Anhorn, K. A., Schachter, R. K., and Mervart, H. (1986). "HLA antigens in psoriatic arthritis". *The Journal of Rheumatology* 13: pp. 586–592.

## BIBLIOGRAPHY

---

- Gladman, D. D., Antoni, C., Mease, P., and rheumatic, Clegg-D. O. of the (2005). "Psoriatic arthritis: epidemiology, clinical features, course, and outcome". *Annals of the rheumatic diseases*.
- Glessner, J. T., Wang, K., Cai, G., Korvatska, O., and Nature, Kim-C. E. (2009). "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes". *Nature*.
- Gobbi, De M., Viprakasit, V., Hughes, J. R., and, Fisher-C. (2006). "A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter" ..
- Golden, JB, Groft, SG, and, Squeri MV of (2015). "Chronic psoriatic skin inflammation leads to increased monocyte adhesion and aggregation". *The Journal of*.
- 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, and Vinuesa, Carola G. (2005). "Cellular and genetic mechanisms of self tolerance and autoimmunity". *Nature* 435: p. 590.
- Gossec, L., Smolen, J. S., Ramiro, S., Wit, M. de, Cutolo, M., 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.
- Gradman, R. J., Ptacin, J. L., and Molecular, Bhasin-A. (2008). "A bifunctional DNA binding region in Tn5 transposase". *Molecular*.
- Gramoun, Azza, Azizi, Natoosha, Sodek, Jaro, Heersche, Johan N, Nakchbandi, Inaam, 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.
- Gregory, Adam P., Dendrou, Calliope A., Attfield, Katherine E., Haghikia, Aiden, Xifara, Dionysia K., 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, Oksenberg, Jorge R, Hart, John, 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*.
- Griffiths, C. E. M. and Barker, JN. (2007). "Pathogenesis and clinical features of psoriasis". *The Lancet*.

## BIBLIOGRAPHY

---

- Griffiths, C. E. M., Reich, K., Lebwohl, M., and Lancet, Kerkhof-P. van de (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*.
- Gu, J., MrkerHermann, E., Baeten, D., and, Tsai-W. C. (2002). “A 588gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients”..
- Guan, Tianxia, Dominguez, Claudia X., Amezquita, Robert A., Laidlaw, Brian J., Cheng, Jijun, 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, Tardivel, A, and Immunity, Mattmann C (2011). “Type I interferon inhibits interleukin-1 production and inflammasome activation”. *Immunity*.
- Gudjonsson, J. E. and Karason, A. (2003). “Psoriasis patients who are homozygous for the HLACw\* 0602 allele have a 2 5fold increased risk of developing psoriasis compared with Cw6 heterozygotes”. *British journal of*.
- Gudjonsson, J. E., Ding, J., Johnston, A., Tejasvi, T., Guzman, A. M., 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*.
- Gusev, Alexander, Lee, Hong S., Trynka, Gosia, Finucane, Hilary, Vilhjlmsson, Bjarni J., 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.
- GutowskaOwsiak, D. and Schaupp, A. L. (2012). “IL17 downregulates filaggrin and affects keratinocyte expression of genes associated with cellular adhesion”. *Experimental dermatology* 21: pp. 104–110.
- Grard, A., Kammen, R. A. van der, and Blood, Janssen-H. (2009). “The Rac activator Tiam1 controls efficient T-cell trafficking and route of transendothelial migration”. *Blood*.
- Hackshaw, Allan (2008). “Small studies: strengths and limitations”.
- Haddad, Amir and Chandran, Vinod (2013). “Arthritis Mutilans”. *Current Rheumatology Reports* 15.
- Haftek, M, Viac, J, Cordier, G, and investigative, Thivolet J of (1986). “Flow cytometry for separation of keratinocyte subpopulations from the viable epidermis”. *Journal of investigative*.
- Hammer, R. E., Maika, S. D., Richardson, J. A., and Cell, Tang-J. P. (1990). “Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human 2m: an animal model of HLA-B27-associated human disorders”. *Cell*.

## BIBLIOGRAPHY

---

- Hansen, K. H., Bracken, A. P., Pasini, D., and cell, Dietrich-N. (2008). "A model for transmission of the H3K27me3 epigenetic mark". *Nature cell.*
- Hanson, A. L., Cuddihy, T., Haynes, K., and, Loo-D. (2018). "Genetic Variants in ERAP1 and ERAP2 Associated With ImmuneMediated Diseases Influence Protein Expression and the Isoform Profile". *Arthritis &.*
- Hao, Yajing, Wu, Wei, Li, Hui, Yuan, Jiao, Luo, Jianjun, et al. (2016). "NPInter v3. 0: an upgraded database of noncoding RNA-associated interactions". 2016.
- Harris, DP, Haynes, L, Sayles, PC, and Nature, Duso DK (2000). "Reciprocal regulation of polarized cytokine production by effector B and T cells". *Nature.*
- Harris, T. J., Grosso, J. F., Yen, H. R., and of, Xin-H. (2007). "Cutting edge: An in vivo requirement for STAT3 signaling in TH17 development and TH17-dependent autoimmunity". *The Journal of.*
- Hecht, I, Toporik, A, Podojil, JR, and, Vaknin I of (2018). "ILDR2 is a novel B7-like protein that negatively regulates T cell responses". *The Journal of.*
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., and genetics, Ching-C. W. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". *genetics.*
- 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.*
- Hesselberth, J. R., Chen, X., Zhang, Z., and Nature, Sabo-P. J. (2009). "Global mapping of protein-DNA interactions in vivo by digital genomic footprinting". *Nature.*
- Hijnen, DJ, Nijhuis, E, and, Bruin-Weller - M de of investigative (2005). "Differential expression of genes involved in skin homing, proliferation, and apoptosis in CD4+ T cells of patients with atopic dermatitis". *Journal of investigative.*
- Hijnen, DJ, Knol, EF, Gent, YY, and Investigative, Giovannone B of (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.*
- 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., El-Gabalawy, Hani S., and Boumpas, Dimitrios T. (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., Kundaje, A., Harris, R. S., et al. (2013). "Integrative annotation of chromatin elements from ENCODE data". *Nucleic Acids Res* 41: pp. 827–41.

## BIBLIOGRAPHY

---

- Homey, B., Alenius, H., Mller, A., Soto, H., and medicine, Bowman-E. P. (2002). “CCL27CCR10 interactions regulate T cellmediated skin inflammation”. *Nature medicine*.
- 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., Gasparetto, M., and Gastroenterology, Ross-A. (2018). “DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes”. *Gastroenterology*.
- Howie, B. N., Donnelly, P., and genetics, 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., Lai, M. J., death, et al. (2014). “Preclinical anti-arthritis study and pharmacokinetic properties of a potent histone deacetylase inhibitor MPT0G009”. *Cell death &*.
- Hu, Stephen, Yu, Hsin-Su, Yen, Feng-Lin, Lin, Chi-Ling, Chen, Gwo-Shing, 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.
- Huber, M., Brstle, A., and the, Reinhard-K. of (2008). “IRF4 is essential for IL-21-mediated induction, amplification, and stabilization of the Th17 phenotype”. *Proceedings of the*.
- Hudak, S., Hagen, M., Liu, Y., and of, Catron-D. (2002). “Immune surveillance and effector functions of CCR10+ skin homing T cells”. *The Journal of*.
- Hundhausen, C., Bertoni, A., Mak, R. K., and Investigative, Botti-E. of (2012). “Allele-Specific Cytokine Responses at the HLA-C Locus: Implications for Psoriasis”. *Journal of Investigative*.
- Hunt, KA, Mistry, V, Bockett, NA, Ahmad, T, and Nature, Ban M (2013). “Negligible impact of rare autoimmune-locus coding-region variants on missing heritability”. *Nature*.
- Husted, J. A., care, and, Thavaneswaran-A. (2011). “Cardiovascular and other comorbidities in patients with psoriatic arthritis: a comparison with patients with psoriasis”. *Arthritis care &*.
- Hffmeier, U., Uebe, S., Ekici, A. B., Bowes, J., and Giardina, E. (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.
- Idel, S., Dansky, H. M., and the, Breslow-J. L. of (2003). “A20, a regulator of NFB, maps to an atherosclerosis locus and differs between parental sensitive C57BL/6J and resistant FVB/N strains”. *Proceedings of the*.

## BIBLIOGRAPHY

---

- Jabbari, Ali, Surez-Farias, Mayte, Dewell, Scott, and Krueger, James G. (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”..
- Jansen, R., Hottenga, J. J., and molecular, Nivard-M. G. (2017). “Conditional eQTL analysis reveals allelic heterogeneity of gene expression”. *Human molecular*.
- Javierre, BM, Burren, OS, Wilder, SP, and Cell, Kreuzhuber R (2016). “Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters”. *Cell*.
- Jenuwein, T. and Allis, CD. (2001). “Translating the histone code”. *Science*.
- Ji, Jong, Cheon, Hyeyoung, Jun, Jae, Choi, Seong, Kim, Ye, et al. (2001). “Effects of peroxisome proliferator-activated receptor- (PPAR- ) on the expression of inflammatory cytokines and apoptosis induction in rheumatoid synovial fibroblasts and monocytes”. *Journal of autoimmunity* 17: pp. 215–221.
- Jiang, X, Tian, H, Fan, Y, Chen, J, 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 Vaccine*.
- Johansen, C., Kragballe, K., and of, Rasmussen-M. (2004). “Activator protein 1 DNA binding activity is decreased in lesional psoriatic skin compared with nonlesional psoriatic skin”. *British Journal of*.
- John, S., Sabo, P. J., Canfield, T. K., and in, Lee-K. protocols (2013). “Genomescale mapping of DNase I hypersensitivity”. *Current protocols in*.
- Johnson, D. S., Mortazavi, A., Myers, R. M., and Science, Wold-B. (2007). “Genome-wide mapping of in vivo protein-DNA interactions”. *Science*.
- Johnson-Huang, LM., McNutt, NS., and Krueger, JG. (2009). “Cytokine-producing dendritic cells in the pathogenesis of inflammatory skin diseases”. *Journal of clinical immunology*.
- Jongbloed, Sarah L., Lebre, M. C., Fraser, Alasdair R., Gracie, J. A., Sturrock, Roger D., et al. (2006). “Enumeration and phenotypical analysis of distinct dendritic cell subsets in psoriatic arthritis and rheumatoid arthritis”. *Arthritis research & therapy* 8.
- Jonker, M. J., Leeuw, W. C. de, and acids, Marinkovi-M. (2014). “Absence/presence calling in microarray-based CGH experiments with non-model organisms”. *Nucleic acids*.

## BIBLIOGRAPHY

---

- Jordan, C. T., Cao, L., Roberson, E. D. O., and of, Duan-S. (2012). "Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis". *The American Journal of*.
- Jostins, Luke, Ripke, Stephan, Weersma, Rinse K, Duerr, Richard H, P McGovern, Dermot, et al. (2012). "Hostmicrobe interactions have shaped the genetic architecture of inflammatory bowel disease". 491: p. 119.
- Jung, M., Sabat, R., and of, Krtzschmar-J. journal (2004). "Expression profiling of IL10regulated genes in human monocytes and peripheral blood mononuclear cells from psoriatic patients during IL10 therapy". *European journal of*.
- Junta, C. M., SandrinGarcia, P., Fachin-Salatoratto, AL., Mello, S.S., Oliveira, R.D., et al. (2009). "Differential gene expression of peripheral blood mononuclear cells from rheumatoid arthritis patients may discriminate immunogenetic, pathogenic and treatment features". *Immunology*.
- K, Ueno-Shuto, Kato, K, Tasaki, Y, Sato, M, and Biological, Sato K of (2014). "Lipopolysaccharide decreases single immunoglobulin interleukin-1 receptor-related molecule (SIGIRR) expression by suppressing specificity protein 1 (Sp1) via the ". *Journal of Biological*.
- Kagami, Shinji, Rizzo, Heather L, Lee, Jennifer J, Koguchi, Yoshinobu, and Blauvelt, Andrew (2010). "Circulating Th17, Th22, and Th1 cells are increased in psoriasis". 130: pp. 1373–1383.
- Kaminsky, Z. A., Tang, T., Wang, S. C., Ptak, C., and Nature, G. H. T. (2009). "DNA methylation profiles in monozygotic and dizygotic twins". *Nature*.
- Kanehisa, M. and Goto, S. (2000). "KEGG: kyoto encyclopedia of genes and genomes". *Nucleic acids research* 28: pp. 27–30.
- Kasela, Silva, Kisand, Kai, Tserel, Liina, Kaleviste, Epp, Remm, Anu, 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., and, Qin-S. (2001). "Differential expression of chemokine receptors on peripheral blood, synovial fluid, and synovial tissue monocytes/macrophages in rheumatoid arthritis". *Arthritis &*.
- 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.
- Keermann, M., Kks, S., and Bmc, Reimann-E. (2015). "Transcriptional landscape of psoriasis identifies the involvement of IL36 and IL36RN". *BMC*.
- Kent, James W., Sugnet, Charles W., Furey, Terrence S., Roskin, Krishna M., Pringle, Tom H., 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*.

## BIBLIOGRAPHY

---

- Kent, W. J., Zweig, A. S., Barber, G., and, Hinrichs-A. S. (2010). "BigWig and BigBed: enabling browsing of large distributed datasets" ..
- Kichaev, G. and Pasaniuc, B. (2015). "Leveraging functional-annotation data in trans-ethnic fine-mapping studies". *The American Journal of Human Genetics*.
- Kim, G. K., of, clinical, and aesthetic, Del J. Q. (2010). "Drug-provoked psoriasis: is it drug induced or drug aggravated?: understanding pathophysiology and clinical relevance". *The Journal of clinical and aesthetic*.
- Kimball, AB, Kawamura, T, Tejura, K, and, Boss C of (2002). "Clinical and immunologic assessment of patients with psoriasis in a randomized, double-blind, placebo-controlled trial using recombinant human interleukin ". *Archives of*.
- Kino, Tomoshige, Hurt, Darrell E., Ichijo, Takamasa, Nader, Nancy, and Chrousos, George P. (2010). "Noncoding RNA gas5 is a growth arrest-and starvation-associated repressor of the glucocorticoid receptor". *Science signaling* 3.
- Kiselev, VY, Andrews, TS, and Genetics, Hemberg M (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*.
- Kondrack, Robyn M, Harbertson, Judith, Tan, Joyce T, E, McBreen, Meghan, Surh, Charles D, et al. (2003). "Interleukin 7 regulates the survival and generation of memory CD4 cells". 198: pp. 1797–1806.
- Kopydlowski, KM, Salkowski, CA, and, Cody MJ of (1999). "Regulation of macrophage chemokine expression by lipopolysaccharide in vitro and in vivo". *The Journal of*.
- Ku, C. S., Loy, E. Y., Salim, A., and genetics, Pawitan-Y. of human (2010). "The discovery of human genetic variations and their use as disease markers: past, present and future". *Journal of human genetics*.
- Kumasaka, N, Knights, AJ, and genetics, Gaffney DJ (2018). "High-resolution genetic mapping of putative causal interactions between regions of open chromatin". *Nature genetics*.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., and Nature, Yen-A. (2015). "Integrative analysis of 111 reference human epigenomes". *Nature*.
- Kurdi, A. T., Bassil, R., Olah, M., Wu, C., and Nature, Xiao-S. (2016). "Tiam1/Rac1 complex controls Il17a transcription and autoimmunity". *Nature*.
- Lande, R., Gregorio, J., Facchinetto, V., Chatterjee, B., Wang, Y., et al. (2007). "Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide". *Nature* 449.
- Landt, Stephen G., Marinov, Georgi K., Kundaje, Anshul, Kheradpour, Pouya, Pauli, Florencia, et al. (2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia". *Genome research* 22: pp. 1813–1831.

## BIBLIOGRAPHY

---

- Langewouters, A. M. G., Erp, P. E. J., Jong, Emgj, and Kerkhof, P. C. M. (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, Murano, T, and Immunity, Begun J (2016). “Genetic coding variant in GPR65 alters lysosomal pH and links lysosomal dysfunction with colitis risk”. *Immunity*.
- Lee, Donghyung, Bigdeli, Bernard T., Riley, Brien P., Fanous, Ayman H., and Bacanu, Silviu-Alin (2013). “DIST: direct imputation of summary statistics for unmeasured SNPs”. *Bioinformatics* 29: pp. 2925–2927.
- Lee, Edmund, Trepicchio, William L., Oestreicher, Judith L., Pittman, Debra, Wang, Frank, 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-Chlo C., Raj, Towfique, Li, Weibo, 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., and of, Seo-S. H. (2009). “A global gene expression analysis of the peripheral blood mononuclear cells reveals the gene expression signature in psoriasis”. *Annals of*.
- Lee, Yoontae, Jeon, Kipyong, Lee, Jun, Kim, Sunyoung, and Kim, Narry V. (2002). “MicroRNA maturation: stepwise processing and subcellular localization”. *The EMBO journal* 21: pp. 4663–4670.
- Lei, W., Luo, Y., Lei, W., Luo, Y., Yan, K., et al. (2009). “Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis”. *Scandinavian journal*.
- Leipe, Jan, Grunke, Mathias, Dechant, Claudia, Reindl, Christiane, Kerzendorf, Ulrike, et al. (2010). “Role of Th17 cells in human autoimmune arthritis”. 62: pp. 2876–2885.
- Leung, Euphemia, Hong, Jiwon, Fraser, Alan G, Merriman, Tony R, Vishnu, Prakash, et al. (2006). “Polymorphisms in the organic cation transporter genes SLC22A4 and SLC22A5 and Crohn’s disease in a New Zealand Caucasian cohort”. 84: pp. 233–236.
- Lewis, B. P., Burge, C. B., and cell, Bartel-D. P. (2005). “Conserved seed pairing, often flanked by adenines, indicates that thousands of human genes are microRNA targets”. *cell*.
- Li, B., Tsoi, L., Swindell, W., Gudjonsson, J., Tejasvi, T., 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.

## BIBLIOGRAPHY

---

- Li, Chun-xiao, Li, Hua-guo, Huang, Lin-ting, Kong, Yu-wei, Chen, Fu-ying, et al. (2017). "H19 lncRNA regulates keratinocyte differentiation by targeting miR-130b-3p". *Cell death & disease* 8.
- Li, H, Yao, Q, Mariscal, AG, Wu, X, and Nature, Hülse J (2018). "Epigenetic control of IL-23 expression in keratinocytes is important for chronic skin inflammation". *Nature*.
- Li, Heng, Handsaker, Bob, Wysoker, Alec, Fennell, Tim, Ruan, Jue, et al. (2009). "The sequence alignment/map format and SAMtools". *Bioinformatics* 25: pp. 2078–2079.
- Li, X, AW, O'Regan, and research, Berman JS interferon cytokine (2003). "IFN-Induction of Osteopontin Expression in Human Monocytoid Cells". *of interferon & cytokine research*.
- 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*.
- Lin, Andrew M., Rubin, Cory J., Khandpur, Ritika, Wang, Jennifer Y., Riblett, MaryBeth, 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, Y., Liu, L., Sheng, Y., Shen, C., Zheng, X., 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 of (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*
- Liontos, Larissa M., Dissanayake, Dilan, Ohashi, Pamela S., Weiss, Arthur, Dragone, Leonard L., 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, Grebe KM of (2005). "Signal transduction activator of transcription 5 (STAT5) dysfunction in autoimmune monocytes and macrophages". *Journal of*
- Liu, Ting, Zhang, Lingyun, Joo, Donghyun, and Sun, Shao-Cong (2017). "NF-B signaling in inflammation". *Signal Transduction and Targeted Therapy* 2: p. 17023.
- Liu, Y., Aryee, M. J., Padyukov, L., and Nature, Fallin-M. D. (2013). "Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis". *Nature*.
- Liu, Y, Chang, JC, Hon, CC, Fukui, N, Tanaka, N, et al. (2018). "Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis". *bioRxiv*.

## BIBLIOGRAPHY

---

- Lizzul, P. F., Aphale, A., Malaviya, R., and Investigative, Sun-Y. of (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*.
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., and genetics, Lo-E. (2013). "The genotype-tissue expression (GTEx) project". *genetics*.
- Lories, RJU, Derese, I, and, Luyten FP and (2008). "Activation of nuclear factor kappa B and mitogen activated protein kinases in psoriatic arthritis before and after etanercept treatment". *Clinical and*.
- Lotem, J, Levanon, D, Negreanu, V, and one, Leshkowitz D (2013). "Runx3-mediated transcriptional program in cytotoxic lymphocytes". *PloS one*.
- Love, M. I., Huber, W., and biology, Anders-S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome biology*.
- Lowes, MA, Kikuchi, T, and Investigative, Fuentes-Duculan - J of (2008). "Psoriasis vulgaris lesions contain discrete populations of Th1 and Th17 T cells". *Journal of Investigative*.
- Lu, J, Ding, Y, Yi, X, and and, Zheng J of (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*.
- Luger, K., Mder, A. W., Richmond, R. K., Sargent, D. F., and Nature, Richmond-T. J. (1997). "Crystal structure of the nucleosome core particle at 2.8 resolution". *Nature*.
- M, Bar-Eli, Gallily, R, and Investigative, Cohen HA of (1979). "Monocyte function in psoriasis." *Journal of Investigative*.
- M., Keating-G. (2017). "Apremilast: a review in psoriasis and psoriatic arthritis". *Drugs*.
- M., Martin (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". *EMBnet. journal*.
- Mahil, Satveer K., Capon, Francesca, and Barker, Jonathan N. (2016). "Update on psoriasis immunopathogenesis and targeted immunotherapy". *Seminars in Immunopathology* 38: pp. 11–27.
- 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., Vukcevic, D., Palin, K., 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*.
- Mao, Liming, Kitani, Atsushi, Similuk, Morgan, Oler, Andrew J, Albenberg, Lindsey, et al. (2018). "Loss-of-function CARD8 mutation causes NLRP3 inflammasome activation and Crohn's disease". *Journal of Clinical Investigation*.

## BIBLIOGRAPHY

---

- Marlow, GJ, Gent, D van, and of, Ferguson LR journal (2013). "Why interleukin-10 supplementation does not work in Crohn's disease patients". *World journal of.*
- Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., and Nature, Wood-A. R. (2017). "Rare and low-frequency coding variants alter human adult height". *Nature.*
- Marrakchi, S., Guigue, P., and of, Renshaw-B. R. (2011). "Interleukin-36receptor antagonist deficiency and generalized pustular psoriasis". *England Journal of.*
- Marshall, C. R., Howrigan, D. P., and Nature, Merico-D. (2017). "Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects". *Nature.*
- Martin-Fontecha, Alfonso, Thomsen, Lindy L, Brett, Sara, Gerard, Craig, Lipp, Martin, 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., and, Thurman-R. E. (2012). "Systematic localization of common disease-associated variation in regulatory DNA" ..
- McCormack, WJ., Parker, AE., and O'Neill, LA. (2009). "Toll-like receptors and NOD-like receptors in rheumatic diseases". *Arthritis.*
- McGeachy, Mandy J., Chen, Yi, Tato, Cristina M., Laurence, Arian, Barbara, Joyce-Shaikh, 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, Dennis, Ash, Zoe, Dickie, Laura, McDermott, Michael, and Aydin, Sibel (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 Enthesitis 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 TherapyFatigue (FACITF), 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, and one, Chu CC (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.

## BIBLIOGRAPHY

---

- Meijerink, Marjolein, Ulluwishewa, Dulantha, Anderson, Rachel C, and Wells, Jerry M (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, Walter-G. J. (2014). "Interleukin17+ CD8+ T cells are enriched in the joints of patients with psoriatic arthritis and correlate with disease activity and joint damage progression". *Arthritis & Rheumatology*.
- Mensah, K. A., Schwarz, E. M., and reports, Ritchlin-C. T. rheumatology (2008). "Altered bone remodeling in psoriatic arthritis". *Current rheumatology reports*.
- Menter, Alan, Korman, Neil J., Elmets, Craig A., Feldman, Steven R., Gelfand, Joel M., 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*.
- Messeguer, X., Escudero, R., Farr, D., and, Nez-O. (2002). "PROMO: detection of known transcription regulatory elements using species-tailored searches" ..
- Miao, Y. L., Xiao, Y. L., Du, Y., and of, Duan-L. P. (2013). "Gene expression profiles in peripheral blood mononuclear cells of ulcerative colitis patients". *World Journal of*.
- Michel, T., Poli, A., Cuapio, A, and, Briquemont B of (2016). "Human CD56bright NK cells: an update". *The Journal of*.
- 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*.
- Mino, T., Murakawa, Y., Fukao, A., and Cell, Vandenbon-A. (2015). "Regnase-1 and roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms". *Cell*.
- Mizoguchi, Fumitaka, Slowikowski, Kamil, Wei, Kevin, Marshall, Jennifer L., Rao, Deepak A., et al. (2018). "Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis". *Nature communications* 9: p. 789.
- Mo, Angela, Marigorta, Urko M, Arafat, Dalia, Chan, Lai, Ponder, Lori, et al. (2018). "Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease". 10.
- Moll, J. M. H., in, *arthritis, and rheumatism*, Wright-V. (1973). "Psoriatic arthritis". *Seminars in arthritis and rheumatism*.
- Morimoto, Junko, Kon, Shigeyuki, Matsui, Yutaka, and Uede, Toshimitsu (2010). "Osteopontin; as a target molecule for the treatment of inflammatory diseases". 11: pp. 494–505.

## BIBLIOGRAPHY

---

- Mortazavi, Ali, Williams, Brian A., McCue, Kenneth, Schaeffer, Lorian, and Wold, Barbara (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq". *Nature methods* 5: pp. 621–628.
- Moser, B, Desai, DD, Downie, MP, and, Chen Y of (2007). "Receptor for advanced glycation end products expression on T cells contributes to antigen-specific cellular expansion in vivo". *The Journal of*.
- Mueller, C., Edmiston, H., Carpenter, C., and Gaffney, E. (2013). "One-Step Cell and Tissue Preservative for Morphologic and Molecular Analysis".
- Mukherjee, R, Barman, PK, Thatoi, PK, and reports, Tripathy R (2015). "Non-classical monocytes display inflammatory features: validation in sepsis and systemic lupus erythematosus". *Scientific reports*.
- Müller, Nike, Dring, Frank, Klapper, Maja, Neumann, Katrin, Schulte, Dominik M., 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, Chandran, Vinod, and D. Gladman, Dafna (2017). "Abstract 156. Gene Expression in Cellular Subsets in Psoriatic Disease".
- Nagano, T., Lubling, Y., Stevens, T. J., and Nature, Schoenfelder-S. (2013). "Single-cell Hi-C reveals cell-to-cell variability in chromosome structure". *Nature*.
- Nair, R. P., Ruether, A., Stuart, P. E., Jenisch, S., Tejasvi, T., 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., Ding, J., Stuart, P. E., 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, Humburg, Peter, Wong, Daniel, et al. (2015). "Genomic modulators of gene expression in human neutrophils". *Nat Commun* 6: p. 7545.
- Natarajan, R, Salloum, FN, and, Fisher BJ of (2007). "Activation of hypoxia-inducible factor-1 via prolyl-4 hydroxylase-2 gene silencing attenuates acute inflammatory responses in postischemic myocardium". *American Journal of*.
- Nestle, Frank O., Conrad, Curdin, Tun-Kyi, Adrian, Homey, Bernhard, Gombert, Michael, et al. (2005). "Plasmacytoid dendritic 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, Nisbet, Barrett, 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.

## BIBLIOGRAPHY

---

- 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.
- Nora, Elphge P. P., Goloborodko, Anton, Valton, Anne-Laure L., Gibcus, Johan H., Uebersohn, Alec, 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, and Goldfeld, Anne E. (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., and of, Stuart-P. E. (2014). "Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes". *The American Journal of*.
- Oliveira, Mfsp, Rocha, B. O., and de, Duarte-G. V. brasileiros (2015). "Psoriasis: classical and emerging comorbidities". *Anais brasileiros de*.
- Organization, World Health (2016). *Global report on psoriasis*. Report.
- Ortega, C, S, FernándezA, and, Carrillo JM of leukocyte (2009). "IL17producing CD8+ T lymphocytes from psoriasis skin plaques are cytotoxic effector cells that secrete Th17related cytokines". *Journal of leukocyte*.
- 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*.
- 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*.
- Ouyang, Jing, Zhu, Xiaomei, Chen, Yuhai, Wei, Haitao, Chen, Qinghuang, 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, Juli, Antonio, Ferrndiz, Carlos, Puig, Llus, Fonseca, Eduardo, 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., Yang, C., Ruotti, V., et al. (2007). "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells". *Cell stem cell*.
- Pandey, R. R., Mondal, T., Mohammad, F., and cell, Enroth-S. (2008). "Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation". *Molecular cell*.
- Pang, Ken C., Dinger, Marcel E., Mercer, Tim R., Malquori, Lorenzo, Grimmond, Sean M., et al. (2009). "Genome-wide identification of long noncoding RNAs in CD8+ T cells". *Journal of immunology (Baltimore, Md. : 1950)* 182: pp. 7738–7748.

## BIBLIOGRAPHY

---

- Park, SY, Lee, SW, Kim, HY, Lee, SY, and Molecular, Lee WS of (2016). "SIRT1 inhibits differentiation of monocytes to macrophages: amelioration of synovial inflammation in rheumatoid arthritis". *Journal of Molecular*.
- Partsch, G, Steiner, G, Leeb, BF, and, Dunky A of (1997). "Highly increased levels of tumor necrosis factor-alpha and other proinflammatory cytokines in psoriatic arthritis synovial fluid." *The Journal of*.
- Pasaniuc, B and Genetics, Price AL (2017). "Dissecting the genetics of complex traits using summary association statistics". *Nature Reviews Genetics*.
- Pedersen, OB, Svendsen, AJ, and oEjstrup, L (2008). "On the heritability of psoriatic arthritis. Disease concordance among monozygotic and dizygotic twins". *Annals of the*.
- Peeters, J. G. C., Vervoort, S. J., Tan, S. C., and reports, Mijnheer-G. (2015). "Inhibition of super-enhancer activity in autoinflammatory site-derived T cells reduces disease-associated gene expression". *Cell reports*.
- Pène, J, Chevalier, S, and, Preisser L of (2008). "Chronically inflamed human tissues are infiltrated by highly differentiated Th17 lymphocytes". *The Journal of*.
- Pepeñella, S., Murphy, K. J., and Chromosoma, Hayes-J. J. (2014). "Intra-and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure". *Chromosoma*.
- 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., and PloS, Richard-A. C. (2016). "Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease". *PLoS*.
- Petersen, C. P., Bordeleau, M. E., Pelletier, J., and cell, Sharp-P. A. (2006). "Short RNAs repress translation after initiation in mammalian cells". *Molecular cell*.
- 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, K, and protocols, Winberg-G. (2014). "Full-length RNA-seq from single cells using Smart-seq2". *Nature protocols*.
- Pickrell, J. K., Marioni, J. C., Pai, A. A., and Nature, Degner-J. F. (2010). "Understanding mechanisms underlying human gene expression variation with RNA sequencing". *Nature*.
- Pivarcsi, A., Sthle, M., and dermatology, Sonkoly-E. (2014). "Genetic polymorphisms altering microRNA activity in psoriasis a key to solve the puzzle of missing heritability?" *Experimental dermatology*.
- Pogribny, I. P., Tryndyak, V. P., and of, Bagryukova-T. V. (2009). "Hepatic epigenetic phenotype predetermines individual susceptibility to hepatic steatosis in mice fed a lipogenic methyl-deficient diet". *Journal of*.

## BIBLIOGRAPHY

---

- Polach, K. J., Lowary, P. T., and biology, Widom-J. of molecular (2000). "Effects of core histone tail domains on the equilibrium constants for dynamic dna site accessibility in nucleosomes1". *Journal of molecular biology*.
- Polacheck, A., Li, S., care, and, Chandran-V. (2017). "Clinical Enthesitis in a Prospective Longitudinal Psoriatic Arthritis Cohort: Incidence, Prevalence, Characteristics, and Outcome". *Arthritis care &*.
- Ponts, N., Harris, E. Y., Prudhomme, J., and Genome, Wick-I. (2010). "Nucleosome landscape and control of transcription in the human malaria parasite". *Genome*.
- Poole, C. D., Lebmeier, M., Ara, R., and Rheumatology, Rafia-R. (2010). "Estimation of health care costs as a function of disease severity in people with psoriatic arthritis in the UK". *Rheumatology*.
- Porcher, R., SaidNahal, R., and, D'Agostino-M. A. (2005). "Two major spondylarthropathy phenotypes are distinguished by pattern analysis in multiplex families". *Arthritis Care &*.
- 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., and Embo, Welch-E. J. (2009). "A nonredundant role for MKP5 in limiting ROS production and preventing LPSinduced vascular injury". *The EMBO*.
- Qu, K., Zaba, L. C., Satpathy, A. T., Giresi, P. G., Li, R., et al. (2017). "Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors". *Cancer Cell*.
- Qu, Kun, Zaba, Lisa C., Giresi, Paul G., Li, Rui, Longmire, Michelle, 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, Ye, Chun, Lee, Mark N., 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*.
- Rao, D. A., Gurish, M. F., Marshall, J. L., and Nature, Slowikowski-K. (2017). "Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis". *Nature*.
- Rauch, I., Müller, M, and T, Jak-Stat, Decker (2013). "The regulation of inflammation by interferons and their STATs". *Jak-Stat*.
- Ray-Jones, H. F., McGovern, A., Warren, P., Duffus, Martin K., Eyre, S., et al. (2017). "[Abstract] Capture Hi-C identifies chromatin interactions between psoriasis-associated genetic loci and disease candidate genes." *UNPUBLISHED*.

## BIBLIOGRAPHY

---

- 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., Krger, K., and Dermatology, Mssner-R. of (2009). "Epidemiology and clinical pattern of psoriatic arthritis in Germany: a prospective interdisciplinary epidemiological study of 1511 patients with plaque-type psoriasis". *Journal of Dermatology*.
- Reich, Kristian, Papp, Kim A., Matheson, Robert T., Tu, John H., Bissonnette, Robert, et al. (2015). "Evidence that a neutrophilkeratinocyte crosstalk is an early target of IL17A inhibition in psoriasis". *Experimental Dermatology* 24: pp. 529–535.
- Rendeiro, Andr F., Schmidl, Christian, Strefford, Jonathan C., Walewska, Renata, Davis, Zadie, 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., Evans, D. M., and Nature, Leo-P. (2010). "Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci". *Nature*.
- Ritchlin, C., Haas-Smith, S. A., and of, Hicks-D. (1998). "Patterns of cytokine production in psoriatic synovium". *The Journal of*.
- Rittling, SR and research, Singh R of dental (2015). "Osteopontin in immune-mediated diseases". *Journal of dental research*.
- Rizova, E. and Coroller, M. (2001). "Topical calcitriolstudies on local tolerance and systemic safety". *British Journal of Dermatology* 144: pp. 3–10.
- Romanowska, M, Yacoub, Al N, and Investigative, Seidel H of (2008). "PPAR enhances keratinocyte proliferation in psoriasis and induces heparin-binding EGF-like growth factor". *Journal of Investigative*.
- 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., Shores, N., and Nature, Sperling-R. A. (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 pathology, Protjaroen P (2011). "Parakeratosis in skin is associated with loss of inhibitor of differentiation 4 via promoter methylation". *Human pathology*.

## BIBLIOGRAPHY

---

- Rudwaleit, M., Heijde, Van D., and the, Landew-R. of (2009). "The development of Assessment of Spondyloarthritis International Society (ASAS) classification criteria for axial spondyloarthritis (part II): validation and final selection". *Annals of the*.
- S., Andrews (2010). "FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>".
- S, Kim-Hellmuth, Bechheim, M, and Nature, Pütz B (2017). "Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations". *Nature*.
- Said-Nahal, R., Miceli-Richard, C., D'Agostino, MA., Dernis-Labous, E., Berthelot, JM., et al. (2001). "Phenotypic diversity is not determined by independent genetic factors in familial spondylarthropathy". *Arthritis Rheumatology*.
- 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*.
- Scharer, Christopher D., Blalock, Emily L., Barwick, Benjamin G., Haines, Robert R., Wei, Chungwen, 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, Zhao, Y, and medicine, Kienhöfer D (2014). "Aggregated neutrophil extracellular traps limit inflammation by degrading cytokines and chemokines". *Nature medicine*.
- Schett, G., Coates, L. C., and Arthritis, Ash-Z. R. (2011). "Structural damage in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: traditional views, novel insights gained from TNF blockade, and". *Arthritis*.
- Schmid, P., Itin, P., and interferon, Cox-D. of (1994). "The type I interferon system is locally activated in psoriatic lesions". *Journal of interferon*.
- Schmidl, Christian, Rendeiro, Andr F., Sheffield, Nathan C., and Bock, Christoph (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: metaanalysis of randomized controlled trials". *British Journal of*.
- 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*.
- Schork, N. J., Murray, S. S., Frazer, K. A., and Topol, E. J. (2009). "Common vs. rare allele hypotheses for complex diseases". *Curr Opin Genet Dev* 19: pp. 212–9.

## BIBLIOGRAPHY

---

- Shadidi, KR, Aarvak, T, and Scandinavian, HenrikSEN JE (2003). "The chemokines CCL5, CCL2 and CXCL12 play significant roles in the migration of Th1 cells into rheumatoid synovial tissue". *Scandinavian*.
- Shapiro, J., Cohen, A. D., David, M., and American, Hodak-E. of the (2007). "The association between psoriasis, diabetes mellitus, and atherosclerosis in Israel: a case-control study". *Journal of the American*.
- Shen, L., Shao, N. Y., Liu, X., Maze, I., Feng, J., 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., Wang, Wei, Wei, Zhi, 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.
- Shu, Jin, Li, Ling, Zhou, Lan-Bo, Qian, Jun, Fan, Zhi-Dan, 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, Föller, M, and reports, Mak TW (2015). "Differential effect of DJ-1/PARK7 on development of natural and induced regulatory T cells". *Scientific reports*.
- Small, K. S., Hedman, K, Grundberg, E., and Nature, Nica-A. C. (2011). "Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes". *Nature*.
- Smallwood, SA, Lee, HJ, Angermueller, C, and Nature, Krueger F (2014). "Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity". *Nature*.
- Smemo, S., Tena, J. J., Kim, K. H., Gamazon, E. R., and Sakabe, N. J. (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, and Dekker, Job (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, DeLay, Monica L., Inman, Robert D., 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., Safabakhsh, Pegah, Marinis, Jill M., 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., Purath, U., and Muller-Ladner, U. (2010). "Natural killer cells trigger osteoclastogenesis and bone destruction in arthritis". *Proceedings of the National Academy of Sciences of the United States of America*.

## BIBLIOGRAPHY

---

- 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*.
- Sos, Brandon, Fung, Ho-Lim, Gao, Derek, Osothprarop, Trina, Kia, Amirali, 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., Bernardini, G., Riccieri, V., 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*.
- Springate, D. A., Parisi, R., and of, Kontopantelis-E. (2017). "Incidence, prevalence and mortality of patients with psoriasis: a UK populationbased cohort study". *British Journal of*.
- Stefan, M., Wei, C., Lombardi, A., Li, C. W., Concepcion, E. S., 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 and, Ortmann-W. A. (2006). "A distinct inflammatory gene expression profile in patients with psoriatic arthritis". *Genes and*.
- Strange, A., Capon, F., Spencer, C. C., Knight, J., Weale, M. E., 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., Markur, Doreen, Hartmann, Karin, 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.
- Straus, DS and immunology, Glass CK in (2007). "Anti-inflammatory actions of PPAR ligands: new insights on cellular and molecular mechanisms". *Trends in immunology*.
- Stuart, Philip E., Nair, Rajan P., Ellinghaus, Eva, Ding, Jun, Tejasvi, Trilokraj, 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., Tejasvi, Trilokraj, Das, Sayantan, 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, and, Garaczi E of (2005). "Dysfunctional blood and target tissue CD4+ CD25high regulatory T cells in psoriasis: mechanism underlying unrestrained pathogenic effector T cell proliferation". *The Journal of*.

## BIBLIOGRAPHY

---

- Sullivan, KE, Reddy, ABM, and biology, Dietzmann K cellular (2007). "Epigenetic regulation of tumor necrosis factor alpha". *and cellular biology*.
- Sun, Lian-Dang, Cheng, Hui, Xue-Qin, Yang, Jian-Zhong, Zhang, Ai, E., et al. (2010). "Association analyses identify six new psoriasis susceptibility loci in the Chinese population". *Nature Genetics* 42.
- Swindell, W. R. and Remmer, H. A. (2015). "Proteogenomic analysis of psoriasis reveals discordant and concordant changes in mRNA and protein abundance". *Genome*.
- Swindell, W. R., Sarkar, M. K., Liang, Y., Xing, X., and reports, Baliwag-J. (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". 193: pp. 1–2.
- Taganov, K. D., Boldin, M. P., and the, Chang-K. J. of (2006). "NF-B-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses". *Proceedings of the*.
- Takagi, Rie, Higashi, Takehiro, Hashimoto, Kumiko, Nakano, Kazuhisa, Mizuno, Yosuke, et al. (2008). "B cell chemoattractant CXCL13 is preferentially expressed by human Th17 cell clones". 181: pp. 186–189.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., and Nature, Lee-C. (2009). "mRNA-Seq whole-transcriptome analysis of a single cell". *Nature methods* 6: p. 377.
- Tang, Fuchou, Barbacioru, Catalin, Nordman, Ellen, Li, Bin, Xu, Nanlan, 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., Jiang, H., Tang, X., et al. (2014). "A large-scale screen for coding variants predisposing to psoriasis". *Nature*.
- Taylor, W, Gladman, D, and &, Helliwell P (2006). "Classification criteria for psoriatic arthritis: development of new criteria from a large international study". *Arthritis &*.
- Telfer, NR, Chalmers, RJ, Whale, K, and Colman, G (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, Siitonens, Annika H., Vuola, Jyrki, et al. (2016). "NOD-like receptor signaling and inflammasome-related pathways are highlighted in psoriatic epidermis". *Scientific reports* 6: p. 22745.
- Thrastardottir, T and international, Love TJ (2018). "Infections and the risk of psoriatic arthritis among psoriasis patients: a systematic review". *Rheumatology international*.
- Thurner, M., Bunt, M. van de, Torres, J. M., and Elife, Mahajan-A. (2018). "Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci". *Elife*.

## BIBLIOGRAPHY

---

- Toberer, Ferdinand, Sykora, Jaromir, Gttel, Daniel, Ruland, Vincent, Hartschuh, Wolfgang, 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, and immunity, O'Farrelly - C of innate (2011). "Natural killer cells in psoriasis". *Journal of innate immunity*.
- Togayachi, Akira, Kozono, Yuko, Kuno, Atsushi, Ohkura, Takashi, Sato, Takashi, et al. (2010). "3GnT2 (B3GNT2), a major polylactosamine synthase: analysis of B3GNT2-deficient mice". 479: pp. 185–204.
- Tonel, Giulia, Conrad, Curdin, Laggner, Ute, Di Meglio, Paola, Grysz, Katarzyna, 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., Romanos, J., and Nature, Mistry-V. (2011). "Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease". *Nature*.
- 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., Ellinghaus, E., Stuart, P. E., 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., and Nature, Gudjonsson-J. E. (2017). "Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants". *Nature*.
- Tsoi, Lam C., Iyer, Matthew K., Stuart, Philip E., Swindell, William R., Gudjonsson, Johann E., et al. (2015). "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, Slaymaker, Sarah, Aslanian, Ara M, et al. (2007). "Critical roles for CCR2 and MCP-3 in monocyte mobilization from bone marrow and recruitment to inflammatory sites". 117: pp. 902–909.
- Valdimarsson, H., Thorleifsdottir, RH., Sigurdardottir, SL., Gudjonsson, JE., and Johnston, A. (2009). "Psoriasis is an autoimmune disease caused by molecular mimicry". *Trends in Immunology*.
- Valk, E., Leung, R., Kang, H., Kaneko, K., and Immunity, Rudd-C. E. (2006). "T cell receptor-interacting molecule acts as a chaperone to modulate surface expression of the CTLA-4 coreceptor". *Immunity*.

## BIBLIOGRAPHY

---

- Vandooren, B, Kruithof, E, Yu, DTY, and, Rihl M of (2004). "Involvement of matrix metalloproteinases and their inhibitors in peripheral synovitis and downregulation by tumor necrosis factor blockade in spondylarthropathy". *: Official Journal of*.
- Vence, L., Schmitt, A., and Medical, Meadows-C. E. (2015). "Recognizing guttate psoriasis and initiating appropriate treatment". *West Virginia Medical*.
- 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, Jrme, Groszer, Matthias, Taylor, Jennifer M., 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, Sandstrom, Richard, and Stamatoyannopoulos, John A. (2014). "Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH". *Nature Methods*.
- Villani, AC, Satija, R, Reynolds, G, and, Sarkizova S (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, and, Sklar P of (2017). "10 years of GWAS discovery: biology, function, and translation". *The American Journal of*.
- Walsh, Alice M, Whitaker, John W, Huang, Chris C, Cherkas, Yauheniya, Lamberth, Sarah L, et al. (2016). "Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations". 17: p. 79.
- Wang, H., Kess, D., Lindqvist, A. K. B., and of, Peters-T. (2008). "A 9-centimorgan interval of chromosome 10 controls the T cell-dependent psoriasisform skin disease and arthritis in a murine psoriasis model". *The Journal of*.
- Wang, Honglin, Peters, Thorsten, Kess, Daniel, Sindrilaru, Anca, Oreshkova, Tsvetelina, et al. (2006). "Activated macrophages are essential in a murine model for T cellmediated chronic psoriasisform skin inflammation". *Journal of Clinical Investigation* 116: pp. 2105–2114.
- Wang, Jie, Zibetti, Cristina, Shang, Peng, Sripathi, Srinivasa R, Zhang, Pingwu, 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, Liu, Hanzhe, Pober, Jordan S., 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 Rev, Mangan PR (2007). "IL-17 family cytokines and the expanding diversity of effector T cell lineages". *Annu. Rev.*

## BIBLIOGRAPHY

---

- 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, Brown, C, and Blood, Pacifici R (2000). "Interleukin-7 stimulates osteoclast formation by up-regulating the T-cell production of soluble osteoclastogenic cytokines". *Blood*.
- Welter, D., MacArthur, J., Morales, J., and acids, Burdett-T. (2013). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations". *Nucleic acids*.
- Westra, HJ, Peters, MJ, Esko, T, and Nature, Yaghootkar H (2013). "Systematic identification of trans eQTLs as putative drivers of known disease associations". *Nature*.
- Whitney, Adeline R., Diehn, Maximilian, Popper, Stephen J., Alizadeh, Ash A., Boldrick, Jennifer C., et al. (2003). "Individuality and variation in gene expression patterns in human blood". *Proceedings of the National Academy of Sciences* 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, Cook, Daniel, Fernandez-Vina, Marcelo, et al. (2005). "Activating Killer Cell Immunoglobulin-Like Receptor Gene KIR2DS1 Is Associated With Psoriatic Arthritis". *Human Immunology* 66: pp. 836–841.
- Winchester, R., Minevich, G., and, Steshenko-V. (2012). "HLA associations reveal genetic heterogeneity in psoriatic arthritis and in the psoriasis phenotype". *Arthritis &*.
- Wirasinha, RC, Vijayan, D, and, Smith NJ cell (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, and, Döcke WD journal of (2006). "IL22 regulates the expression of genes responsible for antimicrobial defense, cellular differentiation, and mobility in keratinocytes: a potential role in psoriasis". *European journal of*.
- Wong, W. F., Kohu, K., Chiba, T., Sato, T., and Immunology, Satake-M. (2011). "Interplay of transcription factors in Tcell differentiation and function: the role of Runx". *Immunology*.
- Woolacott, N., Hawkins, N., Kainth, A., and Khadjesari, Z. (2006). "Etanercept and Efalizumab for the Treatment of Moderate to Severe Psoriasis". *Etanercept and Efalizumab for the Treatment of Moderate to Severe Psoriasis*.
- Woolf, E, Xiao, C, Fainaru, O, and, Lotem J of the (2003). "Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis". *Proceedings of the*.

## BIBLIOGRAPHY

---

- 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*.
- Wu, Q, Jin, H, Yang, Z, Luo, G, Lu, Y, et al. (2010). "MiR-150 promotes gastric cancer proliferation by negatively regulating the pro-apoptotic gene EGR2". *Biochemical and Molecular Biology International*.
- Xia, Yu-Ping, Li, Baosheng, Hylton, Donna, Detmar, Michael, Yancopoulos, George D, et al. (2003). "Transgenic delivery of VEGF to mouse skin leads to an inflammatory condition resembling human psoriasis". *Journal of Investigative Dermatology* 120: pp. 161–168.
- Xinmin, L., Kim, J., Zhou, J., Gu, W., et al. (2005). "Use of signal thresholds to determine significant changes in microarray data analyses". *Genetics and Molecular Biology*.
- Yamazaki, H., Hiramatsu, N., and of, 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 Cell Biology*.
- Yan, Di, Issa, Naiem, Afifi, Ladan, Jeon, Caleb, Chang, Hsin-Wen, 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., Vogelstein, Bert, and Kinzler, Kenneth W. (2002). "Allelic variation in human gene expression". *Science* 297: pp. 1143–1143.
- Yanaba, K, Kamata, M, and, Ishiura N of leukocyte (2013). "Regulatory B cells suppress imiquimod-induced, psoriasis-like skin inflammation". *Journal of Leukocyte Biology*.
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., et al. (2010). "Common SNPs explain a large proportion of the heritability for human height". *Nature Genetics* 42: pp. 565–9.
- Yao, P and medicine, Fox PL molecular (2013). "Aminoacyl tRNA synthetases in medicine and disease". *EMBO Molecular Medicine*.
- Yen, J. H., Moore, B. E., Nakajima, T., and of, Scholl-D. (2001). "Major histocompatibility complex class I recognizing receptors are disease risk genes in rheumatoid arthritis". *Journal of Immunology*.
- Yin, Xianyong, Low, Hui Q., Wang, Ling, Li, Yonghong, Ellinghaus, Eva, 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, and, Sasaki S of medicinal (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*.

## BIBLIOGRAPHY

---

- Yoon, Bo, Yoo, Su-Jin, Choi, Yeon, Chung, Yeon-Ho, Kim, Jinhyun, 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, Gilleaudeau P of (2007). “Amelioration of epidermal hyperplasia by TNF inhibition is associated with reduced Th17 responses”. *Journal of*.
- Zgraggen, S., Huggenberger, R., Kerl, K., and One, Detmar-M. (2014). “An important role of the SDF-1/CXCR4 axis in chronic skin inflammation”. *PLoS One*.
- Zhang, Chen, Wang, Chen, Jia, Zhenyu, Tong, Wenwen, Liu, Delin, et al. (2017). “Differentially expressed mRNAs, lncRNAs, and miRNAs with associated co-expression and ceRNA networks in ankylosing spondylitis”. 8: p. 113543.
- Zhang, F., Wu, L., Qian, J., Qu, B., Xia, S., et al. (2016). “Identification of the long noncoding RNA NEAT1 as a novel inflammatory regulator acting through MAPK pathway in human lupus”. *Journal of*.
- Zhang, F., Wei, K., Slowikowski, K., Fonseka, C. Y., and bioRxiv, Rao-D. A. (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., Zhao, M., and dermatological, Lu-Q. of (2010). “Abnormal DNA methylation in skin lesions and PBMCs of patients with psoriasis vulgaris”. *Journal of dermatological*.
- Zhang, X., Cowper-Sal, R., Bailey, S. D., and Genome, Moore-J. H. (2012). “Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus”. *Genome*.
- Zhang, X. J., Huang, W., Yang, S., Sun, L. D., and Zhang, F. Y. (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 Genome, Meyer-C. A. (2008). “Model-based analysis of ChIP-Seq (MACS)”. *Genome*.
- Zhao, C, Zhang, H, Wong, WC, and, Sem X of proteome (2009). “Identification of novel functional differences in monocyte subsets using proteomic and transcriptomic methods”. *Journal of proteome*.
- Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., and one, Liu-X. (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*.

## BIBLIOGRAPHY

---

- Zhu, K. J., Zhu, C. Y., Shi, G., and Research, Fan-Y. M. (2012). "Association of IL23R polymorphisms with psoriasis and psoriatic arthritis: a meta-analysis". *Inflammation Research*.
- Ziegenhain, C, Vieth, B, Parekh, S, and cell, Reinius B (2017). "Comparative analysis of single-cell RNA sequencing methods". *Molecular cell*.
- Zuk, O., Hechter, E., and the, Sunyaev-S. R. of (2012). "The mystery of missing heritability: Genetic interactions create phantom heritability". *Proceedings of the*.
- Zuo, Xianbo, Sun, Liangdan, Yin, Xianyong, Gao, Jinping, Sheng, Yujun, et al. (2015). "Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis". *Nature communications* 6: p. 6793.
- uszczek, W., Maczak, M., Ciso, M., and immunology, Nockowski-P. (2004). "Gene for the activating natural killer cell receptor, KIR2DS1, is associated with susceptibility to psoriasis vulgaris". *Human immunology*.