



Functional genomics of psoriasis

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

Functional genomics of psoriasis

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. Several aspects of each project relied upon collaboration where part of the work was conducted by others.

Submitted Abstracts

Title	Year
Authors	

Other Publications

Title

Journal

Authors

Contents

Abstract	i
Acknowledgements	ii
Declarations	iii
Submitted Abstracts	iv
Contents	vi
List of Figures	xiv
List of Tables	xvii
Abbreviations	xviii
1 Introduction	1
1.1 Psoriasis and psoriatic arthritis	1
1.1.1 Epidemiology and global impact	2
1.1.2 Psoriasis and inflammatory dermatoses	3
1.1.3 PsA and spondyloarthropathies	3
1.2 Pathophysiology of psoriasis and psoriatic arthritis	5
1.2.1 Clinical presentation and diagnosis	5
1.2.2 Aetiology of psoriasis and PsA	8
1.2.3 Cell types involved in psoriasis and PsA pathogenesis	11
1.2.4 Therapeutic intervention	16
1.3 Genetics of psoriasis and psoriatic arthritis	18
1.3.1 Heritability	18
1.3.2 Non-GWAS and linkage studies	19
1.3.3 Genome-wide association studies	20
1.3.4 Relevance of non-coding variants in disease susceptibility .	24
1.3.5 The role of GWAS studies in highlighting immune-relevant cell types and pathways	25
1.3.6 Limitations and future of GWAS studies	31
1.4 Functional interpretation of genome-wide association studies in complex diseases	33
1.4.1 Overcoming the limitations of GWAS: post-GWAS studies .	33
1.4.2 Understanding the epigenetic landscape in complex diseases	34
1.4.3 The chromatin landscape	35
1.4.4 Transcriptional profiles in disease	41
1.4.5 Transcriptional regulation in complex diseases	45
1.4.6 Integration and interpretation of genomic data	46

CONTENTS

1.4.7	The use of fine-mapping to prioritise functional causal variants	48
1.4.8	Approaches to establish disease mechanisms and causality of genetic variant	50
2	Material and Methods	52
2.1	Ethical approval and recruitment of study participants	52
2.1.1	Psoriasis patient recruitment	52
2.1.2	PsA patient recruitment	53
2.1.3	Healthy volunteer recruitment	53
2.2	Sample processing	54
2.2.1	PBMCs and synovial fluid cells isolation	54
2.2.2	Skin biopsies processing and adherent assays	54
2.2.3	Fixation, cryopreservation and cell culture	55
2.2.4	Primary cell isolation using magnetic-activated cell sorting	56
2.2.5	Primary cell isolation using fluorescence-activated cell sorting	56
2.3	Experimental protocols	58
2.3.1	ATAC - Chromatin Accessibility	58
2.3.2	Chromatin immunoprecipitation with sequencing library preparation by Tn5 transposase	60
2.3.3	RNA extraction and gene expression quantification	61
2.3.4	DNA extraction and rs4672405 genotyping	64
2.3.5	Mass cytometry using cytometry by time of flight (CyTOF)	64
2.4	Computational and statistical analysis	65
2.4.1	ATAC data analysis	65
2.4.2	ChIPm data analysis	68
2.4.3	Gene expression analysis	69
2.4.4	Genomic region annotation, enrichment analysis, gene-network analysis and pathway visualisation	72
2.4.5	Statistical fine-mapping	73
2.4.6	Mass cytometry data analysis	75
3	Establishment of laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples	77
3.1	Introduction	77
3.1.1	Principle of ATAC-seq and compatibility with clinical samples	77
3.1.2	ATAC-seq limitations and advances in optimisation	78
3.1.3	Challenges of ATAC data analysis	79
3.1.4	The challenge of working with clinical samples	83
3.1.5	Aims	84
3.2	Results	85
3.2.1	Establishment of an ATAC-seq data analysis pipeline based on current knowledge	85
3.2.2	Assessment of ATAC-seq transposition times and comparison with Fast-ATAC protocol in relevant cell types	100

CONTENTS

3.2.3	Limitations of ATAC-seq and FAST-ATAC to assess chromatin accessibility in KC	103
3.2.4	Effect of cryopreservation and fixation in the chromatin landscape of immune primary cells	108
3.3	Discussion	120
3.3.1	Studying the chromatin landscape from psoriasis biopsies	124
3.3.2	Characterisation of the effect of preservative techniques in the chromatin landscape	125
3.3.3	Conclusions	126
4	Defining the genome-wide chromatin accessibility landscape and gene expression profile in psoriasis	128
4.1	Introduction	128
4.1.1	The systemic and skin-specific components in psoriasis	128
4.1.2	The personalised epigenome in disease	129
4.1.3	Transcriptional profiles in psoriasis	130
4.1.4	Chromatin accessibility, gene expression and genetic variability	134
4.1.5	Fine-mapping using summary stats	135
4.2	Aims	136
4.3	Results	137
4.3.1	Psoriasis and healthy controls: cohort description and datasets	137
4.3.2	Assessment of changes in the enhancer mark H3K27ac in psoriasis immune cells	141
4.3.3	Identifying global changes in immune cells chromatin accessibility between psoriasis patients and healthy controls	147
4.3.4	Gene expression analysis in psoriasis circulating immune cells	155
4.3.5	RNA-seq in epidermis from psoriasis patients	169
4.3.6	Comparison of the systemic and tissue-specific gene expression signatures in psoriasis	180
4.3.7	Integration of chromatin accessibility and expression data in circulating immune cells	182
4.3.8	Fine-mapping of psoriasis GWAS loci and functional interpretation	184
4.3.9	Maximising genetic commonalities across chronic inflammatory diseases: locus 2p15	189
4.4	Discussion	195
4.4.1	Chromatin accessibility and H3K27ac landscape in psoriasis immune cells	195
4.4.2	Dysregulation of gene expression in psoriasis circulating immune cells	197
4.4.3	Correlation between changes in chromatin accessibility and gene expression	200
4.4.4	Tranciptomic profiles in lesional and uninvolved psoriatic epidermis	201

CONTENTS

4.4.5	LncRNAs in psoriasis	204
4.4.6	Differences in transcriptional dysregulation in blood and skin	207
4.4.7	Fine-mapping and the chr2p15 locus	208
4.4.8	Limitations in the approach and future work	210
4.4.9	Conclusions	211
5	Cross-tissue comparison of chromatin accessibility, gene expression signature and immunophenotypes in PsA	213
5.1	Introduction	213
5.1.1	The relevance of cell type and tissue specificity in the study of PsA	213
5.1.2	Bulk transcriptomic studies in PsA and their limitations . .	214
5.1.3	Transcriptomics and proteomics at the single cell resolution	216
5.1.4	The challenges of using a multi-omics approach in the study of complex diseases	217
5.1.5	Integration of fine-mapping GWAS SNPs and functional data in PsA	218
5.2	Aims	219
5.3	Results	221
5.3.1	PsA patients cohort description and datasets	221
5.3.2	Immune cellular composition of blood and synovial fluid in the PsA cohort	223
5.3.3	Differential chromatin accessibility analysis in immune cells reveals differences between SF and PB	224
5.3.4	Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility	234
5.3.5	Differential gene expression analysis in paired circulating and synovial immune cells	236
5.3.6	Characterisation of the CD14 ⁺ monocyte heterogeneity in PsA using scRNA-seq	250
5.3.7	Mass cytometry reveals differences in protein expression consistent with the chromatin accessibility and transcriptomic profile in PsA CD14 ⁺ monocytes.	259
5.3.8	Prioritisation and interpretation of PsA GWAS SNPs	266
5.4	Discussion	277
5.4.1	Characterising chromatin accessibility in PsA samples . . .	277
5.4.2	Bulk gene expression profiling and integration with chromatin accessibility data	279
5.4.3	The relevance of monocytes and investigation of other cell types	283
5.4.4	Characterisation of monocytes by scRNA-seq	284
5.4.5	Mass cytometry in CD14 ⁺ monocytes and the integration with chromatin accessibility and transcriptomic data	286
5.4.6	Challenges and future perspectives in multi-omic approaches	287
5.4.7	The use of PsA functional data to inform fine-mapping GWAS loci	289
5.4.8	Conclusions	291

CONTENTS

Appendices	292
A Tables	294
A.1 Chapter 3 Tables	294
A.2 Chapter 4 Tables	294
A.3 Chapter 5 Tables	299
B Additional figures	302
B.1 Chapter 3 Figures	302
B.2 Chapter 4 Figures	305
B.3 Chapter 5 Figures	308
Bibliography	345

List of Figures

1.1 Environmental triggers and genetic predisposition leading to dysregulated innate and adaptive immune response in psoriasis and PsA.	13
3.1 Measurements for QC assessment in ATAC-seq samples	87
3.2 FRIP and peak calling at different sequencing depths in ATAC-seq libraries.	91
3.3 Peak calling filtering using IDR analysis in ATAC-seq samples.	93
3.4 Work flow illustrating the strategy to account for ATAC background noise prior to differential analysis.	95
3.5 Normalisation and differential chromatin accessibility analysis for different cut-offs using quantile normalisation limma voom and DESeq2.	97
3.6 Comparison of the DARs identified by differential analysis using limma voom or DESeq for the master list filtered at the optimal cut-off 80%.	98
3.7 Enrichment analysis for the significant DARs identified by DESeq2 between CD14 ⁺ monocytes and tCD4 ⁺ cells.	99
3.8 Assessment of the effect of transposition times on the ATAC-seq measurements	101
3.9 Differences in MT DNA abundance and TSS enrichment between ATAC-seq and Fast-ATAC protocols.	103
3.10 QC assessment of different ATAC protocols in psoriasis KCs and NHEKs.	105
3.11 QC assessment of Fast-ATAC and Omni-ATAC in cultured NHEK .	107
3.12 Comparison of the ENCODE DHSs overlap and signal density at the chr17 keratin family gene locus across different ATAC protocols.	109
3.13 Experimetal design to assess the impact of cryopreservation and fixation in the chromatin accessibility of immune primary cells. . .	111
3.14 Total number of ATAC-seq reads for the fresh, frozen and fixed CD14 ⁺ monocytes and total CD4 ⁺ samples.	112
3.15 Fragment size density distribution for ATAC-seq fresh, fixed and frozen in CD14 ⁺ monocytes and tCD4 ⁺ cells.	113
3.16 ATAC-seq enrichment of nucleosome-free and di-nucleosome fragments at the TSS and surroundings in CD14 ⁺ monocytes and tCD4 ⁺ samples for the three conditions.	114
3.17 Genomic features annotation for the ATAC-seq peaks called in each of the fresh,frozen and fixed samples from CD14 ⁺ monocytes and total CD4 ⁺	115
3.18 PCA analysis based on the ATAC-seq chromatin accessibility landscape in fresh, fixed and frozen samples.	116

LIST OF FIGURES

3.19 Comparison of the \log_2 normalised ATAC-seq counts at the consensus master lists peaks in fresh, fixed and frozen conditions.	118
3.20 Differential chromatin accessibility at the <i>TNFSF14</i> gene between ATAC-seq fresh and ATAC-seq frozen in CD14 ⁺ monocytes	120
4.1 Quality control evaluation of the H3K27ac ChIPm libraries in immune cells isolated from psoriasis and control samples.	143
4.2 PCA analysis and chromatin annotation states of the H3K27ac enriched sites in four immune primary cell types from psoriasis and healthy control samples.	144
4.3 Differential H3K27ac modification at a putative intergenic enhancer region in circulating CD14 ⁺ monocytes between psoriasis patients and healthy controls.	146
4.4 Quality control assessment of the ATAC libraries generated from circulating immune cells in psoriasis and control samples.	149
4.5 Clustered heatmap and conserved TFBS enrichment analysis in the consensus ATAC-seq regions identified in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ cells from the patients and controls cohort.	151
4.6 Epigenetic landscape at two ATAC differential accessible regions between patients and controls in CD8 ⁺ cells.	154
4.7 Epigenetic landscape at the only overlapping region identified as differentially accessible and differentially H3K27ac modified between psoriasis patients and controls.	156
4.8 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.	157
4.9 PCA analysis and sample distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles.	158
4.10 Magnitude and significance of the gene expression changes between psoriasis patients and healthy controls in four immune cell types.	160
4.11 RNA-seq expression levels of the lncRNA <i>HOTAIRM1</i> and its experimentally validated target <i>UPF1</i> in psoriasis and healthy controls CD14 ⁺ monocytes.	164
4.12 Mapping of the DEGs identified in CD8 ⁺ cells between psoriasis patients and healthy controls onto the TNF- α and the chemokine signalling pathways.	168
4.13 Mapping quality control and PCA analysis for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. .	170
4.14 Magnitude and significance of the gene expression changes between matched lesional and uninvolved epidermal biopsies from three psoriasis patients.	171
4.15 Overlap of the significantly differentially expressed genes between lesional and uninvolved epidermal sheets, split epidermis and whole skin biopsies.	173

LIST OF FIGURES

4.16 Correlation in gene expression between two dysregulated miRs in lesional skin and their putative target genes.	176
4.17 Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the HIF-I signalling pathway.	177
4.18 Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the NOD-like signalling pathway.	179
4.19 Differential gene expression and chromatin accessibility landscape for <i>TIAM1</i> gene in tCD8 ⁺ cells.	183
4.20 Epigenetic landscape at the location of the SNPs in the 95% credible set for the chr2p15 GWAS psoriasis locus.	191
4.21 Effect of rs4672505 genotype in tCD8 ⁺ cells chromatin accessibility at chr2:62,559,749-62,561,442.	192
4.22 Potential role of rs4672505 in regulating <i>B3GNT2</i> gene expression.	194
5.1 Comparative percentages of PB and SF immune cellular composition from the PsA cohort.	224
5.2 Quality control assessment of ATAC data generated in four immune cell types isolated from PB and SF of PsA patients samples.	226
5.3 PCA analysis based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and SF.	227
5.4 Enrichment of eQTLs publicly available data in the combined cell type and tissue chromatin accessibility master list for the PsA cohort.	228
5.5 Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states.	231
5.6 Enrichment of PsA DARs for the FANTOM5 eRNA dataset.	232
5.7 Differentially accessible regions located within gene bodies in CD14 ⁺ monocytes and NK cells from PsA patients.	233
5.8 Distinct functional pathways enriched for DARs open in SF or open in PB in CD14 ⁺ monocytes, CD4m ⁺ ,CD8m ⁺ and NK.	236
5.9 Heatmap of gene expression FCs between SF and PB for those gene significantly modulated (<i>pval</i> <0.05) in at least one cell type.	238
5.10 Gene expression changes in immune-relevant genes between SF and PB in CD14 ⁺ monocytes, mCD4 ⁺ and mCD8 ⁺ cells.	240
5.11 Chromatin accessibility landscape at the qPCR differentially expressed <i>FN1</i> gene in CD14 ⁺ monocytes.	242
5.12 Protein network analysis based on the immune qPCR array expression data.	246
5.13 Comparison of immune-relevant gene expression modulation across PsA tissues (SF vs PB) and in PsA patients versus healthy controls.	248
5.14 Identification of two main CD14 ⁺ monocytes subpopulations in the SF and PB combined analysis	252
5.15 Sc-RNA-seq differential gene expression results between SF and PB in the CC-mixed and CC-IL7R CD14 ⁺ monocytes subpopulations.	254
5.16 Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14 ⁺ monocytes.	256

LIST OF FIGURES

5.17 Comparison of TNF- α expression in SF and PB CD14 $^{+}$ monocytes before and after protein transport blockade with BFA using mass cytometry.	261
5.18 Percentage of CD14 $^{+}$ monocytes producing TNF- α , osteopontin and MCP-1 in SF and PB in a validation cohort of ten PsA samples.	263
5.19 Chromatin landscape in CD14 $^{+}$ monocytes upstream the differentially expressed gene CCL2.	265
5.20 Experimental Factor Ontology terms enriched in GWAS Catalog SNPs overlapping ATAC regions in four cell types.	272
5.21 Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal.	275
B.1 Distribution of the background read counts from all the master list peaks absent in each sample.	302
B.2 Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples.	303
B.3 Fast-ATAC and Omni-ATAC NHEK tapestation profiles.	304
B.4 Percentage of MT reads in the ATAC-seq libraries generated in CD14 $^{+}$ monocytes, CD4 $^{+}$, CD8 $^{+}$ and CD19 $^{+}$ isolated from psoriasis patients and healthy controls.	305
B.5 Genomic annotation of the consensus master list of ATAC-seq enriched sites built for downstream differential chromatin accessibility analysis in CD14 $^{+}$ monocytes, CD4 $^{+}$, CD8 $^{+}$ and CD19 $^{+}$	306
B.6 PCA analysis illustrating batch effect in ATAC and RNA-seq samples.	307
B.7 Permutation analysis SF vs PB in CD14 $^{+}$ monocytes, CD4m $^{+}$, CD8m $^{+}$ and NK.	308
B.8 Heatmap for the top 20 marker genes of the CC-mixed and CC-IL7R CD14 $^{+}$ monocytes subpopulations.	309
B.9 Identification of the CD14 $^{+}$ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes.	310
B.10 Quantification of cytokine levels in SF from ten PsA patients.	311

List of Tables

1.1	Variables and scoring used in the Psoriasis Area and Severity Index (PASI)	7
1.2	Variables and scoring used in the Psoriatic Arthritis Response Criteria (PsARC)	7
1.3	Main GWAS studies in psoriasis and PsA	22
2.1	Antibody panel used for FACS separation of primary cell populations in Chapter 3 controls and Chapter 5 PsA samples. . .	57
2.2	Molecules targeted by the mass cytometry ICS staining panel in whole blood and SF.	65
3.1	Summary table of ATAC-seq analysis methodology for peak calling, filtering and differential analysis.	81
3.2	ATAC-seq percentage of MT reads and fraction of reads in called peaks (FRiP).	88
3.3	Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies. . .	106
3.4	Summary results from the differential chromatin accessibility analysis comparing ATAC-seq frozen or fixed chromatin landscape to the reference ATAC-seq fresh.	119
4.1	Summary table of the most comprehensive transcriptional studies in psoriasis skin and blood.	131
4.2	Description and metadata of the psoriasis patients cohort.	138
4.3	Description of the healthy control cohort.	140
4.4	Datasets generated for the psoriasis and control cohort samples. .	141
4.5	Summary results from the differential H3K27ac analysis between psoriasis patients and healthy controls in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ cells.	145
4.6	Summary results from the differential chromatin accessibility analysis between psoriasis patients and healthy controls in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ cells.	153
4.7	Summary results from the DGE analysis between psoriasis patients and healthy controls in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ cells.	159
4.8	Overlap between putative psoriasis GWAS genes and the reported significantly DEGs in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ cells.	162
4.9	Functional interactions and overlap with another study for the differentially expressed lncRNAs in each cell type.	163
4.10	Most relevant pathways enriched for DEGs between psoriasis patients and healthy controls in CD14 ⁺ monocytes and CD8 ⁺ cells.	165

LIST OF TABLES

4.11	Summary results of the DGE analysis between uninvolved and lesional psoriatic epidermal biopsies.	171
4.12	Most relevant pathways enriched for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	177
4.13	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).	181
4.14	Summary table of the psoriasis Immunochip GWAS loci presenting $\log_{10}ABF > 3$ for the fine-mapping lead SNP.	185
4.15	SNPs from the 90% credible set of the successfully fine-mapped psoriasis loci overlapping ATAC accessible chromatin in four cell types.	188
5.1	Description and metadata of the PsA patients cohort.	222
5.2	Datasets generated for each sample in the PsA cohort.	223
5.3	Summary results of the differential chromatin accessibility analysis between SF and PB in PsA samples.	229
5.4	Characterisation of the DARs located within genes in each of the four cell types from PsA samples.	232
5.5	Immune genes with significant modulated expression in SF and proximal to a DAR in Fast-ATAC.	241
5.6	Pathway enrichment analysis for the modulated genes between SF and PB in CD14 ⁺ and mCD4 ⁺	244
5.7	Most relevant enriched pathways for the DEGs between SF and PBCD14 ⁺ monocytes in CC-mixed and CC-IL7R.	257
5.8	scRNA-seq DEGs in SF versus PB CD14 ⁺ monocytes proximal to a DAR in Fast-ATAC.	259
5.9	Summary table of the PsA GWAS loci presenting $\log_{10}ABF > 3$ for the fine-mapping lead SNP.	268
5.10	PsA fine-mapped SNPs from the 90% credible sets overlapping accessible chromatin identified by ATAC in four cell types.	271
5.11	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.	273
A.1	Enrichment of ATAC-seq reads across the TSS for the CD14 ⁺ monocytes and CD4 ⁺ samples fresh, frozen and fixed.	294
A.2	Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChiPm assay.	295
A.3	Additional enriched pathways for DEGs between psoriasis and healthy controls in CD14 ⁺ monocytes and CD8 ⁺ cells.	296
A.4	Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	297
A.5	Loci from the psoriasis GWAS Immunochip presenting $\log_{10}ABF <$ 3 for the fine-mapping lead SNP in the association analysis.	298
A.6	Additional enriched pathways for the DEGs between SF and PB CD14 ⁺ monocytes from the CC-mixed and CC-IL7R subpopulations.	299

LIST OF TABLES

Abbreviations

Abbreviation	Definition
Ab	Antibody
ATAC-seq	
Atopic dermatitis	AD
ChIPm	
CLE	cutaneous lupus erythematosus
DMARDs	disease-modifying antirheumatic drugs
Fast-ATAC	
IDR	
GWAS	Genome-wide association studies
KC	Keratinocytes
NSAID	nonsteroidal antiinflammatory drug
Omni-ATAC	
PCA	
PI	Protein inhibitor
PsA	
QC	
qPCR	quantitative polymerase chain reaction
RA	Rheumatoid arthritis
ROS	Reactive oxygen species
SDS	Sodium dodecyl sulfate
SF	Synovial fluid

Chapter 1

Introduction

In this thesis I have used a functional genomics approach to characterise the regulatory landscape in psoriasis and psoriatic arthritis, with the aim of uncovering the mechanism of action of disease risk variants. In this chapter, I begin by reviewing the current knowledge of disease pathophysiology, the role of genetic variation, and then describe cutting edge functional genomics approaches in complex disease research that are relevant to the subsequent chapters.

1.1 Psoriasis and psoriatic arthritis

Psoriasis and psoriatic arthritis (PsA) have been described as two distinct common complex disease entities that share nonetheless certain clinical features and genetic architecture. Psoriasis is a chronic inflammatory dermatose disease with episodes of relapse and remittance (Nestle et al. 2009). On the other hand, PsA is a seronegative chronic inflammatory disease within the spondyloarthritis family that usually develops after psoriasis skin manifestations(Villanova2016; Moll et al. 1973; Coates et al. 2016). The study of similarities and differences between the conditions at the pathological and genetic levels will benefit the understanding of genetic variability in the risk of developing psoriasis and PsA as well as the identification of new therapeutic targets.

Introduction

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). Onset of disease seems to have a bimodal distribution with psoriasis patients being classified as early-onset/ type I (around 16-22 and 30-39 years) or late-onset / type II (between 50-60 years) (Henseler and Christophers 1985). This classification based on the age of onset also has correlates with distinctive clinical features including severity, relapse frequency and family history. As reflected in the minor correlation with geographic latitude, the risk of developing psoriasis has a strong ethnicity component (Jacobson et al. 2011). In fact, the prevalence of psoriasis in adults is lower among African, African American and Asian (between 0.4 and 0.7%) compared to American and Canadian populations (4.6 and 4.7%, respectively).

Psoriasis prevalence ranges between 2 and 3%, affecting approximately 1.8 million people in the UK(Perera et al. 2012). On the other hand, cases of PsA in the general population varies between 0.04 and 1.2% (Perera et al. 2012) but dramatically increase up to 10 to 30% within psoriasis patients, evidencing the strong association between the two diseases (**Reich2008**; Gelfand et al. 2005). In the UK, 14% of psoriasis patients develop chronic inflammatory arthritis in the form of PsA during the course of the disease (Ibrahim et al. 2009). Overall, data suggest an steady increase in both psoriasis and PsA prevalence over the last thirty years (**Springate2007**; Organization 2016).

Several comorbidities have also been associated with psoriasis and PsA, with comparatively greater prevalence in PsA. For example, intraocular inflammation known as uveitis affects 8% of PsA patients compared to only 2% of psoriasis patients (Husted et al. 2011; Oliveira et al. 2015). Other comorbidities include IBD, cardiovascular disease (CVD), type II diabetes (T2D) and metabolic syndrome (**Cohrn20017**; Gelfand et al. 2006; Shapiro et al. 2007). Psoriasis

Introduction

and PsA have also been associated with an increased prevalence of depression and suicidal ideation (Sampogna et al. 2012). Accordingly, psoriasis and PsA represent a burden for the economy due to treatment costs and associated morbidity. Treatment and management-associated costs for psoriasis in 2015 in the UK accounted for £4,000 to £14,000, before and after requirements of biological therapy, respectively (Burgos-Pol and Dermo 2016; Poole et al. 2010) and the costs are further enhanced in the case of PsA.

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, due to exogenous or endogenous factors, can lead to development of inflammatory dermatose conditions, such as psoriasis or atopic dermatitis (AD) (Johnson-Huang et al. 2009; Proksch et al. 2008). The group of inflammatory dermatoses affects up to 70% of the UK population and represents the 4th leading cause of nonfatal burden (ICD-10; Roderick2014).

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 psoriasis vulgaris, guttate, pustular, erythroderma and nail pitting have been defined (Marrakchi et al. 2011).

1.1.3 PsA and spondyloarthropathies

PsA belongs to the family known as spondylarthropathies (SpA), which includes phenotypes such as ankylosing spondylitis (AS), reactive arthritis (ReA), idiopathic inflammatory bowel disease (IBD) and undifferentiated SpA (Baeten et al. 2013). All these SpA subtypes are characterised by structural damage

Introduction

(bone formation and erosion) as well as inflammation of joints and extra-articular sites such as eyes, gut and skin. Broadly, SpA has been classified into axial and peripheral based on the affected joints (spine/sacroiliac or peripheral) and the presence of extra-articular features (**Runwaleit2001; Runwaleit2001**).

The major histocompatibility class I (MHC-I) constitutes a family of genes involved in the presentation of intracellular peptides (self or from infectious agents) to T cells, which includes the human leukocyte antigen (HLA) A, B and C genes. HLA-B27 subtype has been identified as 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 forms, such as psoriasis and inflammatory bowel disease (IBD), within a single family or individual (**Said-Nahal2000**; Hammer et al. 1990). These observations support the hypothesis that SpA subtypes may in fact be a single multifaceted condition with shared genetic, immunopathological and structural features and dynamic phenotypes (Baeten et al. 2013). Conversely, some studies suggest that the immunopathological differences between axial and peripheral arthritis could be partially explained by genetic factors (Porcher et al. 2005; Appel et al. 2011; Noordenbos et al. 2012).

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 et al. 2009). Overall, this phenotypic heterogeneity of PsA increases the difficulty in the design and achievement of meaningful outcomes from clinical studies.

1.2 Pathophysiology of psoriasis and psoriatic arthritis

1.2.1 Clinical presentation and diagnosis

Of the aforementioned phenotypes, 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 Lancet 2007). The second most common clinical presentation is psoriasis guttate (10% of all cases) characterised by acute onset of small droplike papules usually in the trunk and proximal extremities (Vence et al. 2015). Despite psoriasis vulgaris and guttate representing an important burden for patient wellbeing, they are not life-threatening forms of disease. Conversely, the least prevalent phenotype, pustular psoriasis has a 50% increased risk of mortality compared to the milder psoriasis phenotypes (Gelfand et al. 2007; Moura et al. 2015).

As mentioned above, in addition to the difference in time of onset, type I and type II psoriasis also differ in clinical presentation. Type I psoriasis patients, with greater prevalence of HLA-C*06:02 (85.4% of the cases) and stronger family history, commonly present guttate lesions followed very often by bacterial infection, particularly *Streptococcus* throat infection (Telfer et al. 1992). In contrast, type II psoriasis, where individuals are positive for HLA-C*06:02 only in 14.6% of the cases, involves spontaneous chronic plaques such as 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

Introduction

PsA (around 30%), which exclusively affects single or few 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 and intergluteal skin lesions constitute a predictive biomarker for development of joint inflammation (**Moll** 1976; **McGonagle**; Griffiths and Lancet 2007; medicine 2011). This observation reinforces the need for appropriate coordination between dermatologists and rheumatologists for an early diagnosis and treatment that could prevent functional joint disability.

The diagnosis of psoriasis and PsA is primarily based on clinical assessment of the patient's symptoms due to the lack of appropriate 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 quantitative rating score of skin lesion severity in research and clinical trials (Fredriksson and Dermatology 1978; Finlay 2005). PASI quantifies the lesional burden weighted by body part based on area of affected surface and the degree of erythemas severity, 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 Dermatology 2005; Langewouters et al. 2008).

To diagnose PsA, modified Moll and Wright criteria known as Classification Criteria for Psoriatic Arthritis (CASPAR) are most widely used in a clinical setting (Taylor et al. 2006). A positive diagnosis based on CASPAR requires display of inflammatory arthritis, enthesitis, and/or spondylitis and three points from a list of associated elements. In terms of disease activity and treatment efficacy, the PsA Response Criteria (PsARC) is the preferred measure (Mease 2011; Clegg

Introduction

PASI	Description
Body location	Head and neck, upper limbs, trunk and lower limbs
Feature	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: For each of the four body locations the test quantifies the percentage of affected area and the severity of three intensity features: redness, thickness and scaling. The score ranges from 0 (no disease) to 72 (maximal disease).

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
TJC	Number of tender joints over 68
SJC	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 health 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.

Introduction

1.2.2 Aetiology of psoriasis and PsA

Psoriasis and PsA are complex chronic inflammatory diseases characterised by a dysregulated immune response initiated as the result of genetic predisposition and exposure to particular environmental cues (Figure 1.1). The origin of both pathologies, as well as the connection between skin and joint inflammation, still remain controversial. In the specific case of psoriasis, it is also unclear whether disruption of the skin triggers activation of the immune response or vice versa.

Environmental factors and disease

Several environmental factors are known to be associated with increased risk of development and worsening of psoriasis and PsA. A wide range of drugs including antidepressant, antihypertensive and anticytokine therapies have been clinically associated with initiation, exacerbation and worsening of psoriasis (Kim et al. 2010). Bacterial agents such as streptococcal throat infection have also been associated with development of type I psoriasis (Gudjonsson and others 2003; Valdimarsson and others in 2009; Diluvio et al. 2006). Consistent with other chronic inflammatory diseases such as IBD and AS, recent studies have also observed perturbation in the composition of the gut and skin microbiota of psoriasis and PsA patients (**add reference**). Furthermore, physical trauma and mechanical stress can trigger the appearance of skin lesions and digit joint inflammation (Weiss et al. 2002; Nestle et al. 2009). Lastly, behavioral factors including smoking, alcohol and stress have been linked to psoriasis and PsA but no clear association with disease development has been established yet (Meglio et al. 2014).

Introduction

Histopathological alterations in skin and joints

The epidermis is the most external compartment of the skin, formed by approximately 90% keratinocytes (KCs) and organised in a layer-like structure that self-renews in an spatial and time-dependent manner (Wikramanayake et al. 2014). KC 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, KCs undergo upregulation in the proliferation rate (hyperplasia) that causes aberrant cell differentiation (parakeratosis), thickening of the epidermis and subsequent scale formation (Ruchusatsawat2011). Concomitantly, inflammation causes immune cell infiltration and hypervascularisation of the lesion driven by upregulation in the expression of angiogenic factors and activation of the endothelium (Perera et al. 2012).

In PsA, the joint affection, arising after skin lesions in the majority of the cases, involves a wide range of histological changes (Haddad and Chandran 2013). One of the most common structural changes is arthritis caused by the swelling and inflammation of the joints (Schett et al. 2011). As a result of this inflammation, alterations in bone remodeling lead to osteolysis with subsequent bone resorption and erosion at the affected joints (**Mensah2017**). 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, alike in RA, causing structural debilitation of the joints parencteBenjamin2009,Finzel2014.

Introduction

Dysregulation of the innate and adaptive immune response

The dysregulated immune response in psoriasis and PsA is the result of the interaction between innate and adaptive immune cells through feedback loops and a complex cytokine milieu (Figure 1.1). Interferon (IFN)- α and γ are amongst the most relevant innate immune cytokines involved in disease initiation (**Leanne2009**). Both cytokines are mainly produced by circulating plasmacytoid dendritic cells (pDCs) and myeloid DC (mDCs), respectively, upon activation by KCs pro-inflammatory cytokines (Perera et al. 2012). Nevertheless production by T cells in the lesional skin has also been demonstrated (Hijnen et al. 2013). Increased mRNA levels for both IFNs have been detected in lesional skin and demonstrated to contribute to lymphocyte recruitment and maintenance of DCs activation (Schmid et al. 1994). TNF- α is another pivotal cytokine involved in the psoriasis and PsA dysregulated innate immune. TNF- α is produced by activated KCs, mast cells, NK and also adaptive immune cell types, including T helper (Th)-1 and Th-17 lymphocytes infiltrated in the skin lesions and inflamed joints (Perera et al. 2012; Lizzul et al. 2005). TNF- α causes activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B), a master transcriptional regulator the innate and adaptive immune system, which induces expression of pro-inflammatory cytokines, antiapoptotic genes and genes involved in maintenance of chronic inflammation (**Johansen2010**; Lizzul et al. 2005). Moreover, TNF- α has a prominent role in bone turnover and bone remodeling, key features of the histopathological PsA joint alterations(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 homing the inflamed skin and in a lesser amount by psoriatic KCs (Lee et al. 2004; Li et al. 2018). IL-23 exerts

Introduction

its function through binding to the IL-23 receptor (IL-23R), 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 KCs and T cells, where activation of the IL-23 pathway importantly leads Th-17 cell differentiation and increased IL-17 cytokine levels as a result of NF- κ B activation (McGeachy 2009). IL-17 maintains the perpetuation of the Th-17 immune mediated response through recruitment and activation of neutrophils, induction of pro-inflammatory cytokines, including interleukine-1 β (IL-1 β) and interleukine-6 (IL-6), and sustains KCs activation (Doyle and Arthritis 2012).

More recently, interleukin 22 (IL-22) has gained relevance as mediator of the dysregulated crosstalk between the innate and adaptive immune response. IL-22 levels are increased in the skin lesions and plasma of psoriatic patients and is mainly produced by a subset of CD4⁺ cells known as Th-22 (Wolk et al. 2006). IL-22 contributes to some of the histological changes in skin as well as to AMP production by KCs (Eyerich et al. 2009).

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.

Several studies have shown the role of KCs as immune sentinels through MHC-II antigen presentation and production of antimicrobial peptides (AMP), cytokines and chemokines (Black and of 2007). Indeed, complex formation between the cationic AMP LL-37 and self-DNA/RNA released by KCs has been observed upon damage triggered by environmental factors (Lande et al. 2007). This complex acts as an antigen for activation of the skin-resident DCs that initiate and perpetuate the skin inflammatory response through secretion of pro-inflammatory cytokines, including IL-1, IL-6 and TNF- α (Feldmeyer

Introduction

et al. 2007; Arend et al. 2008; Nestle et al. 2009; Nestle et al. 2005). Furthermore, *in vivo* studies have described the development of psoriatic lesions in immunodeficient mice upon human xenotransplant of psoriatic skin(Boymen et al. 2004). Altogether, these findings support the role of epidermis dysfunction in the initiation of the psoriatic chronic inflammatory response (**Proskch2008**). The relevance of KCs at early stages of psoriasis pathogenesis is reinforced by the genetic association between KC-specific genes from the late cornified envelope (LCE) family and increased psoriasis risk (Tsoi et al. 2012)

mDCs and pDCs are also considered important innate immune cells in disease initiation through antigen presentation, T-cell activation and the subsequent adaptive immune response(**Mahil20016**). pDCs are circulating professional antigen presentation cells (APCs) that only upon activation by the KCs self-DNA-LL-37 complex infiltrate into the lesional and uninvolved dermis of psoriasis patients (Nestle et al. 2005; Lande et al. 2007). In contrast, quiescent mDCs are epidermal resident cells that undergo maturation in presence of the IFN- α secreted by pDCs, expanding up to 30-fold only in the lesional skin (**Zaba2007**). The activated mDCs mediate the Th-1 and Th-17 response as well as perpetuation of KC activation through IL-23 production (Lee et al. 2004). Studies in immunodeficient psoriasis mouse models have shown that blockage of downstream IFN- α signaling or IFN- α production by pDCs failed to induce T-cell activation and psoriasis onset (Nestle et al. 2005).

Neutrophils are also thought to be closely involved in disease initiation through their ability to form neutrophil extracellular traps (NET) that contain host DNA and LL-37 (Hu et al. 2016). There is evidence of increased NET formation in peripheral blood and lesional skin of psoriasis patients and they seem to be contributing to pDC and CD4 $^{+}$ T cell activation (Hu et al. 2016). Neutrophils have also been identified in recent studies as one of the main sources

Introduction

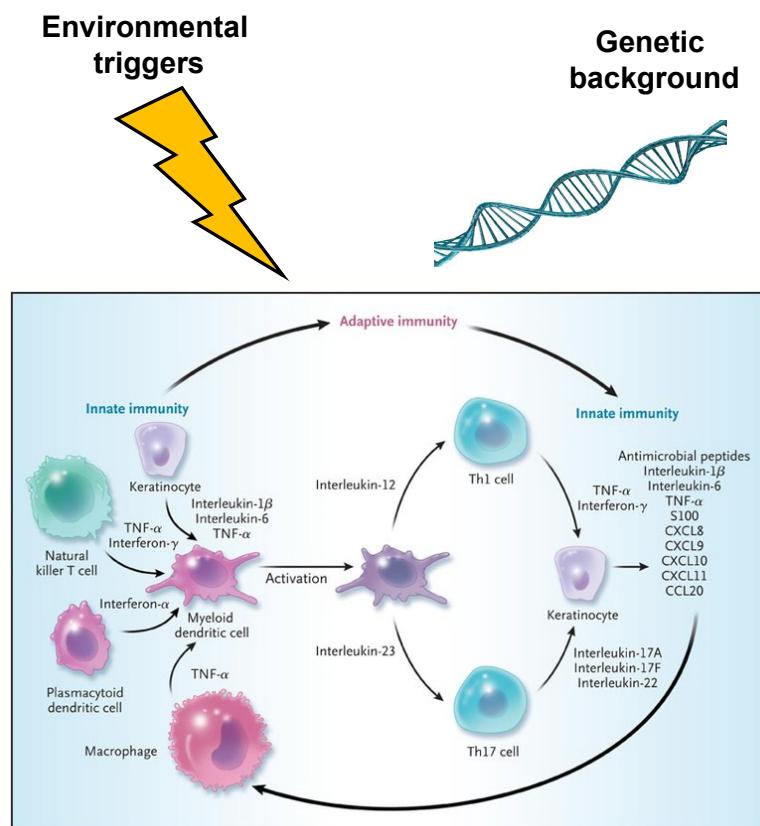


Figure 1.1: Environmental triggers and genetic predisposition leading to dysregulated innate and adaptive immune response in psoriasis and PsA. Figure adapted from (Nestle et al. 2009). As a result of interaction between environmental factors and genetic predisposition dysregulated cross-talk between the innate and adaptive immune response can resolve in chronic inflammation and development of psoriasis and PsA. Cells from the innate immunity. The main cell types, ctkines and chemokines involved in the psoriasis and PsA inflammatory response are depicted here and some of them further detailed in the main text.

Introduction

of IL-17 production in the skin lesions (Lin et al. 2011) and they also release a wide range of proteases which some induce KC proliferation (**Mahil2006**).

The involvement of monocytes and macrophages in psoriasis and PsA innate immunity has not been extensively explored. Resident macrophages in the healthy dermis undergo a 3-fold increase upon skin lesion and contribute to disease development through TNF α production (Perera et al. 2012; Mahil et al. 2016). Similarly, mouse models for chronic psoriasiform skin inflammation have shown the role of macrophage migration into the affected skin and production of TNF- α in maintenance of the skin lesions (Stratis et al. 2006; Wang et al. 2006). Initial studies showed greater phagocytic and bactericidal activity of peripheral blood mononuclear cells (PBMCs) isolated monocytes from psoriasis patients compared to those from healthy individuals (**Bar-Eli1979**). 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 (**Golden2015**). 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 rheumatoid arthritis (RA) joints (Hitchon et al. 2002).

Regarding the adaptive immunity, T lymphocytes have been considered the most relevant cell types in the initiation and maintenance of psoriasis and PsA. Report cases in humans have demonstrated that bone marrow transplantation can initiate or terminate psoriasis (**Gardembas1990**; Eedy et al. 1990). Reduced numbers of circulating T cells but increased percentages of the memory populations CD4 $^+$ CD45RO $^+$ and CD8 $^+$ CD45RO $^+$ have been observed in moderate-to-severe and severe psoriasis patients when compared to milder phenotypes and healthy controls (**Lecewicz-Toru2001**; Langewouters et al. 2008). Different studies have reported controversial results regarding the total abundance and ratios of CD4 $^+$ and CD8 $^+$ in PBMC, likely due to the phenotype

Introduction

heterogeneity of the psoriasis cohorts between studies (Lecewicz-Toru 2001; Cameron and of 2003; Langewouters et al. 2008). In PsA, no differences in abundance of circulating T cells have been identified when compared to healthy individuals (Costello et al. 1999).

In homeostasis, CD8⁺ and CD4⁺ lymphocytes are found in the epidermis and dermis, respectively (Clark et al. 2006). An increase in activated memory CD4⁺CD45RO⁺ and CD8⁺CD45RO⁺ cells can be detected by the third day from the lesion appearance (Clark et al. 2006; Perera et al. 2012). *In vivo* studies showed that development of psoriasis following engrafted human pre-lesional skin was only dependent on local T cell proliferation, highlighting the importance of circulating T cells recruitment during the priming event rather than at later stages of the disease (Wrone-Smith and clinical 1996; Nickoloff and Wrone-Smith 1999; Perera et al. 2012). The relative importance of CD4⁺ versus CD8⁺ cells in psoriasis initiation has been explored in pre-lesional skin mouse xenografts where CD4⁺ but not CD8⁺ T cells were required in the transition from uninvolved to lesional skin (Nickoloff and Wrone-Smith 1999). Interestingly, the injection of activated CD4⁺ cells in mice was followed by an acute increase in activated resident CD8⁺ T cells. Overall, these results supported the hypothesis of skin CD4⁺ cells being drivers of resident T-cell activation and the population of resident activated CD8⁺ the main effector of the immune response. In synovial tissues of PsA patients, CD4⁺ are significantly more abundant than CD8⁺ (Diani et al. 2015). However, amongst the CD8⁺ populations, the memory cells are prevalent in the patients synovial fluid (SF) with a significant enrichment compared their counterparts in PsA PB and RA SF(Costello et al. 1999). The contribution of regulatory T (Treg) remains controversial in both, psoriasis and PsA (Perera et al. 2012).

Based on the cytokine profile, psoriasis and PsA have been classified as a type 1 Th/Tc disease, where activation of naive CD4⁺ and CD8⁺ cells is driven

Introduction

by IL-12 and IFN- γ (Austin et al. 1999; Perera et al. 2012). In addition, T-cell subsets including Th-17/Tc-17 and Th-22/Tc-22, producing high levels of IL-17 and IL-22, respectively, have been identified to be relevant for the perpetuation of the inflammatory response (Mahil et al. 2016). The importance of Th-17 cells and IL-17 production has been evaluated in skin, joints and blood, with elevated mRNA and protein levels of IL-17 and also IL-23 reported in psoriasis and PsA patients compared to controls (Cai et al. 2012; Dolcino et al. 2015). The relevance of IL-17 has been further highlighted by the presence of CD8 $^{+}$ populations in patients SF that are predominantly IL-17 producers and whose abundance correlates with markers of inflammation and structural changes in the joint (Menon et al. 2014). This finding is in line with observations in skin and suggests a prominent role for CD8 $^{+}$ IL-17-producing cells in the different stages of both pathologies. Studies directed to understand the importance of IL-17 have led to the discovery of other immune cells producing this pivotal cytokine, including innate immune lymphoid (ILC) cells and $\gamma\delta$ T cells, opening new research avenues in the context of psoriasis and PsA pathophysiology and treatment (Meglio et al. 2014; Leijten et al. 2015). IL-17-producing cells have also been hypothesised to be at the link between skin and joint lesions. Although the precise mechanisms for transition between psoriasis and PsA is still poorly understood, the study of psoriasis and RA in mouse models revealed that skin lesions facilitate arthritis and joint inflammation (Yamamoto et al. 2015).

1.2.4 Therapeutic intervention

Psoriasis and PsA are currently incurable diseases, with the different treatments available solely focused on alleviating the symptoms. For instance, topical therapies are advocated in cases of mild-to-moderate psoriasis, represented by extended emollients and short-term corticosteroids, due to associated side-effects (Menter et al. 2009). Other treatments may be used in

Introduction

combination with corticosteroids, such as ultraviolet (UV) light therapy and vitamin D analogues, directed to inhibit T-cell and KC proliferation and stimulate KC differentiation (Rizova and Coroller 2001). In the case of PsA patients presenting with swelling of two or fewer joints, intra-articular injection of glucocorticosteroids together with joint aspiration is used as a short-term solution to reduce pain and inflammation (Coates et al. 2016). However, treatment of most forms of PsA and moderate-to-severe psoriasis require the use of systemic therapies. Patients presenting with mild cases of PsA commonly receive nonsteroidal anti-inflammatory drugs (NSAID) to control the inflammatory symptoms (Coates et al. 2016). More severe forms of PsA require the use of disease-modifying antirheumatic drugs (DMARDs) including the antagonist of folic acid methotrexate (MTX) and the phosphodiesterase 4 inhibitor Apremilast, which act as immunosuppressors of activated T cells and cytokine production, respectively (**Keating2017**; Schmitt and of 2014; Gossec et al. 2016; Polachek et al. 2017).

Remarkably, biologic systemic agents represent the most specific treatment option for severe psoriasis and PsA. This category encompasses an array of cell-based molecular species that modulate the immune response in a physiological manner (Perera et al. 2012). Specifically, the relevance of TNF-alpha in psoriasis and PsA has led to the extensive therapeutic use of TNF-alpha inhibitors (TNFi) during the past five decades, making them the drugs of choice amongst all the biologic agents targeting cytokine pathways. Three TNFi have been approved for the treatment of psoriasis: etanercept, infliximab and adalimumab (**Ahil2016**). In addition to those, certolizumab pegol and golimumab are often used in the management of PsA and other rheumatoid diseases (**Coates2016b**). Despite TNF- α blockade being one of the most effective treatments, side effects such as increased risk of infection or reactivation of latent infections have been identified (Nickoloff and Nestle 2004). Moreover, between 20 to 50% of patients

Introduction

fail to respond to the first TNFi administrated, requiring switching to an alternative TNFi (Abramson and and 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 (**Coates2016b**; Mahil et al. 2016).

1.3 Genetics of psoriasis and psoriatic arthritis

As complex diseases, the risk of developing psoriasis and PsA is not only influenced by environmental conditions but also by the genetic background of each individual. Determining the magnitude of the contribution of genetic factors in the development of these diseases and identifying the exact genes or genomic regions involved in the predisposition to psoriasis and PsA remains challenging. Several studies have shown an increase in psoriasis and PsA prevalence over the last 30 years in different countries (Organization 2016). This importantly reflects changes in life style habits and it highlights the need to better understand the genetic factors that predispose to disease upon interaction with environmental stresses.

1.3.1 Heritability

The contribution of genetics in the development of psoriasis has been demonstrated in several twins studies. The concordance of psoriasis is greater in monozygotic (33-55%) compared to dizygotic twins (13-21%), giving a heritability estimate of 80% in this condition (**Faber1974; Penderson2008**; Duffy et al. 1993). Conversely, similar concordance across mono- and di- zygotic twins has been reported in the case of PsA, probably due to lack of statistical power and appropriate diagnosis (**Penderson2008**). In the general population, approximately 40% of patients with psoriasis or PsA have family history in first degree relatives (Gladman et al. 1986). Interestingly, the recurrence rate in first-

Introduction

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). This suggests differences in the heritability between the two phenotypes and a stronger genetic contribution in PsA.

1.3.2 Non-GWAS and linkage studies

The study of psoriasis and PsA genetic architecture started with linkage analyses in family pedigrees presenting an autosomal dominant condition. This approach yielded nine psoriasis susceptibility loci (PSORS1-9) with PSORS1 showing the strongest genetic association (**International2003**; Capon 2017). PSORS1 locus lies within the MHC I region in chromosome 6p21, previously associated with psoriasis susceptibility in serological studies (**Rusell1972**; Tiilikainen et al. 1980). Importantly, Mendelian forms of disease with rare highly penetrant mutations have also been identified in family studies for two genes within PSORS2 (17q25): zinc finger protein 750 (*ZNF750*) and caspase domain family member 14 (*CARD14*) (Tomfohrde et al. 1994; Jordan et al. 2012a). Rare gain of function and *de novo* mutations together with common variants in *CARD14* have been identified in psoriasis and PsA patients, suggesting an important role of genetic variation in this gene in Mendelian and multi-factorial forms of disease (Jordan et al. 2012a; Tsoi et al. 2012). Additionally, gene based studies in psoriasis and PsA disclosed the importance of genetic variability in the activating killer immunoglobulin receptors 2DS1 (*KIR2DS1*) gene, also reported for AS and RA, which interestingly is mainly triggered by interaction with HLA-Cw*06:02 (uszczek et al. 2004; Williams et al. 2005; Carter et al. 2007; Yen et al. 2001). Nevertheless, the inability of independent studies to reproduce these results for regions other than PSOR1, 2 and 4, highlights the limitations of linkage studies to understand the genetics of complex diseases (Capon 2017).

1.3.3 Genome-wide association studies

Recent dramatic advances in sequencing and genotyping technologies have allowed the implementation of association studies at a genome-wide scale. Genome-wide association studies (GWAS) have benefitted from the understanding of common single base-pair changes known as single nucleotide polymorphisms (SNPs) in different populations through whole genome sequencing (WGS) projects such as HapMap (**The international HapMaP Consortium**) and the 1000 Genomes Project (**The 1000 Genomes**). GWAS generally focus on identifying disease-associated common SNPs (with minor allele frequency (MAF) >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, and are designed to have greater power than the previous linkage studies to identify multiple loci with low penetrance and moderate to small effects (Schork et al. 2009; Cui et al. 2010).

The genotyped SNPs in GWAS are solely used a proxy for the disease causative variant, for instance non-genotyped SNPs or other type of genetic variability such as copy number variants (CNVs) (Hirschhorn 2005; Ku et al. 2010). Since 2007, when the first psoriasis and PsA GWAS were published, a total of sixty-three genetic associations have been identified at a genome-wide significance ($p\text{val}>5\times 10^{-8}$) which explain 28% of the psoriasis and PsA heritability (Table ??) (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). The early GWAS performed in discrete size cohorts with moderate power confirmed association with loci overlapping the PSOR1 and PSOR2 genomic regions identified in the linkage studies (Cargill et al. 2007; Strange et al. 2010). Specifically, HLA-C has been consistently identified as the most significant locus with the greatest effect

Introduction

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 information extracted from GWAS studies was significantly enhanced with the use of the Immunochip genotyping platform, which covers 186 immune relevant loci identified in previous GWAS studies across different inflammatory diseases at a greater genotyping density (Tsoi et al. 2012). The psoriasis Immunochip study uncovered fifteen new associations, including the PSOR4 *CARD14* and also included meta-analysis with the largest available psoriasis cohorts at the time(Tsoi et al. 2012). This meta-analysis has been further expanded yielding sixteen additional associations in the latest study and reinforcing the importance of NF κ B and cytotoxicity pathways in disease pathophysiology (Tsoi et al. 2015; Tsoi et al. 2017). Meta-analysis of GWAS across Caucasian and Chinese populations has demonstrated the value of this trans-ethnic approach to identify new associations and understand the differences in the genetic associations contributing to disease risk in different populations (Yin et al. 2015). The importance of conducting independent GWAS studies for psoriasis and PsA has been shown not only by the discrepancy in first degree relatives heritability but also by differences in HLA-C and HLA-B alleles frequencies within each phenotype population (Winchester et al. 2012; Okada et al. 2014). In fact, cohort stratification confirmed specific GWAS association with PsA for previously identified psoriasis loci such as *TRAF3IP*, *IFNLR1*, *IFIH1* and *NFKBIA* as well as PsA-specific independent signals for *IL23R* and *TNFAIP3* (Ellinghaus et al. 2010; Stuart et al. 2015). Interestingly, the association for *LCE3C/B*, identified in combined phenotypic studies, showed greater strength in those patients presenting psoriasis for over ten years without developing joint affection (Stuart et al. 2015). Lately, PsA GWAS using the Immunochip platform revealed a PsA-specific association in chromosome 5q31 and an independent secondary signal in the *IL23R* region (Bowes et al. 2015).

Introduction

Table 1.3: Main GWAS studies in psoriasis and PsA. Summary table describing the most relevant psoriasis and PsA GWAS studies. 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>IL12B</i>
(Nair et al. 2009)	Eur	1,409/1,436	Psoriasis, PsA	<i>IL23A</i> , <i>IL23R</i> , <i>IL12B</i> , <i>TNIP1</i> , <i>TNFIP3</i> , <i>IL4</i> and <i>IL13</i>
(Stuart et al. 2010)	WA, Eur	1,831/2,546	Psoriasis, PsA	<i>NOS2</i> , <i>FBXL19</i> , <i>PSMA6</i> - <i>NFKBIA</i>
(Ellinghaus et al. 2010)	German	472/1,146	Psoriasis	<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 <i>LCE</i> gene family and <i>IL12B</i>) I)

Introduction

(Sun et al. 2010)	Chinese	8,312/12,919	Psoriasis, PsA	<i>ERAP1</i> , <i>PTTG1</i> , <i>CSMD1</i> , <i>GJB2</i> , <i>SERPINB8</i> , <i>ZNF816A</i>
(Tsoi et al. 2012)★	WA, Eur	10,588/22,806	Psoriasis, PsA	<i>CARDI4</i> (<i>PSOR4</i>), <i>RUNX3</i> , <i>B3GNT2</i> , <i>ELMO1</i> , <i>STAT3</i>
(Tsoi et al. 2015)★	WA, Eur	15,000/27,000	Psoriasis, PsA	1q31.1, 5p13.1, <i>PLCL2</i> , <i>NFKBIZ</i> , <i>CAMK2G</i>
(Bowes et al. 2015)	British, Australians	Irish, 1,962/8,923	PsA	5q31 PsA-specific
(Stuart et al. 2015)	WA and Eur	1,430/1,417	Psoriasis, PsA	PsA-specific secondary signals (main text), 1p36.23 psoriasis-specific, stronger psoriasis LCE 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

Overall, GWAS studies have demonstrated shared genetic susceptibility between psoriasis and PsA, but have also highlighted intrinsic specificity that may support a difference in the genetic architecture of both diseases. It is important to take into account that these results are affected by imprecise phenotyping of cases, which is one of the many challenges in the systematic comparison between the two diseases.

1.3.4 Relevance of non-coding variants in disease susceptibility

Approximately 88% of all GWAS associations map within non-coding regions, with only the remaining 12% comprising non-synonymous coding mutations impacting the protein function (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).

The association of non-coding variants with disease can be explained by their ability to regulate gene expression in a cell and context specific manner (Fairfax et al. 2012). These variants can be located in different regulatory elements, including enhancer, silencers, promoters and the 5' and 3' untranslated region (UTR) of genes (Ward and biotechnology 2012). Non-coding GWAS variants can alter the expression of target genes through different mechanisms including changes in chromatin accessibility, histone modifications, protein

Introduction

binding such as transcription factors (TFs), DNA methylation and binding of non-coding RNA molecules(Knight 2014) (1.4.2).

Identification of the target gene regulated by non-coding variants represents 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 an such approach revealed a *cis*-eQTL involving the TF *KLF4* and a haplotype of non-coding GWAS SNPs located 14kb up-stream (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, are required to demonstrate causality (Edwards et al. 2013).

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

GWAS represent a biologically unbiased approach to shed some light on pathophysiological relevant cell types and molecular pathways associated with disease. In the field of common immune-mediated diseases, GWAS have underlined some of the most important cell types for which genetic variation is functionally relevant. Better understanding of immune-related diseases has likewise led to identification of shared susceptibility loci and the use of therapeutic interventions across diseases, such as anti- IL-23 and anti- IL-17 antibodies to treat psoriasis, PsA, AS and IBD (Visscher et al. 2017).

Systematic comparison of the genetic architecture across different conditions has revealed psoriasis and PsA risk loci to be shared, in the same or opposite directions, with AS, Crohn's disease (CD), multiple sclerosis (MS),

Introduction

RA and type 1 diabetes (T1D) (<https://www.immunobase.org>). Interestingly, cross-disease association studies performed for AS, UC, primary sclerosing cholangitis (PSC), CD and psoriasis has revealed significant overlap of the multi-trait associated loci in regulatory elements in bone marrow, NK and T cells as well as immune response pathways (Ellinghaus et al. 2016).

In the case of psoriasis and PsA, the majority of GWAS risk loci have been linked to genes that belong to a limited number of pathways and show enrichment for regulatory elements in several cell types (Capon 2017).

Antigen presentation

In psoriasis *HLA-Cw*0602* represents the strongest GWAS association, shared with other diseases such as Hepatitis C, PSO 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 as the mechanism explaining disease association (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 studies 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). Similarly, the same study identified a dependent association between *HLA-Cw*0602* and SNPs in the vicinity of the zeta chain of T cell receptor associated protein kinase 70 (*ZAP70*) gene (Picard et al. 2009). These epistatic phenomena, whereby association of one gene is dependent on the presence of another, have also been reported between *HLA-B*27* and *ERAP1* in AS (Cortes2015b; Evans et al. 2011). Interestingly, the AS *ERAP1* GWAS association increases *ERAP1* expression and also alters splicing,

Introduction

resulting in an ERAP1 protein isoform with increased activity (**Constatino2015**; Hanson et al. 2018).

Skin barrier

GWAS have highlighted KC specific genes such the previously mentioned *LCE* gene cluster and genes with a key role in skin biology such as *CARD-14*. Further studies in the *PSORS4* region have revealed that association with disease is driven by a deletion in two of the genes within this family, *LCE3B* and *LCE3C* (*LCE3C_LCE3B_del*)(Cid et al. 2009). Expression of *LCE3B* and *LCE3C* is induced upon barrier disruption, where these proteins participate in the formation of the cornified envelope at the most external layer of the epidermis and are likely involved in KC terminal differentiation (Bergboer et al. 2011). The 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). In fact, the use of UVB radiation has been shown to upregulate *LCE3E* expression 48 hours after treatment, contributing to amelioration of the skin lesions(Jackson et al. 2005). Similarly to the *LCE* gene cluster, *CARD14* is primarily expressed in epithelial tissues mediating the recruitment and activation of the NF- κ B pathway in this tissue (Blonska and research 2011). Common and rare pathogenic mutations of *CARD14* in KC cell lines lead to increased activation of NF- κ B as well as overexpression of psoriasis-associated genes including *IL6*, *TNFA* and *TNFAIP2*, among others (Jordan et al. 2012b).

NF- κ B and TNF pathways

The NF- κ B pathway is involved in the regulation of the innate and adaptive immune response. In fact, dysregulation of the feedback loop between TNF- α and NF- κ B contributes to the development of many chronic inflammatory

Introduction

diseases and TNF- α i are widely used for treatment of immune-mediated diseases, as previously detailed (Liu et al. 2017). In psoriasis, for example, elevated levels of NF- κ B are present in lesional 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- κ B and TNF signalling pathways including *TNIP1*, *TNFAIP3*, *NFKBIA*, *REL*, *TRAF3IP2*(H“”–u”ffmeier2010; Nair et al. 2008; Ellinghaus et al. 2010; Wang et al. 2008; Idel et al. 2003; Bowes et al. 2012). 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- κ B activation and the IL-17/IL-23 axis(H“”–u”ffmeier2010; Ellinghaus et al. 2010). SNP variants in the *TRAF3IP2* locus have also been identified in CD and UC, reinforcing the relationship between these two pathways and chronic inflammation (<https://www.immunobase.org>). In addition, exome-sequencing studies have identified variants with predicted influence on protein structure and function at *TNFSF15*, a TNF ligand family protein induced by TNF- α , with a primary role in regulating NF- κ B and MAP kinases activation in endothelial cells (Dand et al. 2017; Wang et al. 2014).

The psoriasis and PsA GWAS associations with other members of these pathways, such as the NF- κ B inhibitor *NFKBIA* and the NF- κ B subunit *REL*, are solely supported by proximity to nearby intergenic SNPs, with no direct experimental evidence for a role of these genetic variants in regulating the expression of either gene (GWAS studies). Moreover, the latest psoriasis and PsA meta-analysis study revealed three additional associations with genes belonging to the NF- κ B pathway (*CHUK*, *IKBKE* and *FASLG*), further implicating NF- κ B activation in psoriasis and PsA development (Tsoi et al. 2017).

Introduction

Type I IFN and innate host defense

Members of the type-I IFN signalling pathway have also been associated to psoriasis and PsA by GWAS studies, highlighting the role of innate immunity, particularly genes contributing to the host response to viruses and bacteria, in the disease pathophysiology. Mapping of several GWAS loci to genes from the type-I IFN signalling pathway together with clinical and experimental data has reinforced the role of pathogen response in psoriasis and PsA (Nestle et al. 2005). GWAS associations involved in this IFN response include *IL28RA*, *IFIH1*, *TYK2*, *RNF114*, *ELMO1* and *DDX58*, some of which have been previously reported as susceptibility loci for other immune-mediated diseases (1.3). For instance, GWAS-lead SNPs causing missense mutations in *TYK2* have been identified in CD, IBD, T1D, RA and MS, in addition to psoriasis and PsA (<https://www.immunobase.org>). Exome-sequencing and GWAS studies 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 solely for use in RA treatment, is currently undergoing clinical trials in other immune-related diseases alongside the development of more specific JAK inhibitors (Baker and diseases 2017). Moreover, drugs targeting upstream type I IFN pathway members, such as inhibitors of the pathogen-sensing receptors *TLR7* and (*TLR9*), are being conducted in SLE and may be extended to other immune-mediated diseases (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

Introduction

the treatment of psoriasis and PsA (Griffiths et al. 2015; Blauvelt et al. 2017). The cytokine IL-23, involved in a wide range of pro-inflammatory processes as previously explained, is formed of two subunits: IL-23A/p19 and IL12-B/p40, the latter also being a component of IL-12 protein. Transcriptional studies have shown increased levels of p40 and p19 in psoriasis lesional skin and a role for both subunits in abnormal KC differentiation (Zhu2011; Lee et al. 2004). Psoriasis and PsA GWAS associations with the *IL-23R* have been reported in several studies, including a protective two SNP haplotype shared with CD in a German and American Caucasian cohort study (Nair et al. 2008; Strange et al. 2010; Tsoi et al. 2012). GWAS associations have also been established by proximity of non-coding lead SNPs to *IL23A* and *IL12* genes, but direct functional evidence in regulation of their expression has not yet been established (Cargill et al. 2007; Strange et al. 2010; Tsoi et al. 2012). Interestingly, an *IL-23* signal secondary to that reported by Tsoi *et al.*, 2012 has been specifically associated with PsA, and independence from AS secondary signals for the same locus has also been demonstrated *Tsoi2012,Bowes2015*. Regarding the genetics of the Th-17 pathway, its relevance is partly explained through the cross-talk with the IL-23 response, which mediates Th-17 cell differentiation and activation. Additionally, GWAS associations with intronic variants at TFs regulating the Th-17 polarisation, such as *IRF4* and *STAT3*, have also been identified for psoriasis and PsA (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 disclosed relevant biological processes by conducting genome-wide pathway analysis. In psoriasis, genome-wide pathway analysis has revealed association of novel processes, such as retinol metabolism, transport of inorganic ions and aminoacids and

Introduction

post-translational protein modifications (PTMs) not previously related with the disease aetiology (**Aterido2015**).

As previously mentioned, the majority of the non-coding GWAS associations are located at intergenic regions and often lack functional characterisation. Therefore these variants tend to be associated to the nearest gene within few Kb away and which function is known and in line with current understanding of the pathophysiology. Nevertheless, some of these intergenic non-coding SNPs are located at regions depleted of genes for 300Kb or more. Some examples in psoriasis and PsA include chr1p36.23, chr2p15 and chr9q31.2. One of the most interesting regions is chr2p15, where the lead SNP and direction of association is shared with AS (<https://www.immunobase.org>). Within this locus, genes with a role in the immune response involve *B3GNT2* and *CMMID1*, which are located at approximately 130Kb and 200Kb away from the GWAS lead SNP, respectively (Maine et al. 2007; Tsoi et al. 2012). Interestingly, the *B3GNT2* gene was identified by the aforementioned genome-wide pathway analysis to contribute to the significant enrichment of the post-translational protein modifications pathway. Another relevant association is the 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 immortal KC cell line HaCaT has revealed interaction of SNPs in this locus with the promoter of the *ERRFI1* gene, inhibitor of the epidermal growth factor receptor signalling required for normal KC proliferation (Ray-Jones et al. 2017).

1.3.6 Limitations and future of GWAS studies

GWAS have made a great contribution into the understanding of the genetic component of complex diseases. However, this approach presents limitations that need to be considered in the final result interpretation.

Introduction

One of the major GWAS limitations is the structure of the genome in LD blocks where hundreds to thousands variants in high LD due to low recombination rates within the segment are inherited together as haplotype blocks. The disease-associated loci expand large genomic regions containing hundreds to thousands variants in high LD with the GWAS lead SNP that are inherited together and separated from other regions. Therefore, an association between a genetic locus and a trait does not reveal the causal variant, which could potentially be any of the highly correlated SNPs in the same LD block as the lead. Moreover, GWAS association in non-coding regions also fail to identify the target gene and the mechanism driving the association. As a results, integration of dense genotyping, statistical fine-mapping and epigenetic data are required to identify the true GWAS causal SNP.

Another concern is the heritability missed to be explained by the GWAS associations in relation to the estimated heritability in 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). For example, in human height, most of the missing heritability could be explained by GWAS associated variants with nominal significance that failed to pass the stringent threshold due to their small effect size (Yang et al. 2010).

Another source of unexplained heritability may be rare putative causal variants poorly tagged by common SNPs in the genotyping platforms due to differences in the allele frequencies (Wray 2005). Such limitations have partly been overcome by improved genotyping arrays like the Immunochip, which incorporates SNPs with MAF<1% (Cortes and Brown 2011). Moreover, exome studies have also demonstrated the contribution of coding and intronic rare variants (MAF<5%) in the genetic architecture of complex traits such as

Introduction

height or psoriasis (Marouli et al. 2017; Dand et al. 2017). In addition to rare variants, other sources of common variation such as CNV, small (<1Kb) insertions/deletions (indels) and inversions could contribute to the missing heritability. Incorporation of new genotyping platforms has allowed the genome-wide identification of CNV with autism and schizophrenia, among others (Glessner et al. 2009; Marshall et al. 2017). The accurate detection of translocations and inversions relies on the implementation of long reads WGS technologies (Visscher et al. 2017). Lastly, the missing heritability may also be the consequence of the overestimated heritability in complex traits as the result of assuming additive genetic effect instead of epistatic interaction between the different associated loci (Zuk et al. 2012).

1.4 Functional interpretation of genome-wide association studies in complex diseases

1.4.1 Overcoming the limitations of GWAS: post-GWAS studies

GWAS studies shed limited light on the link between genetic variants and disease mechanisms. As previously mentioned, GWAS report associations with disease for a particular locus but they fail to identify the true causal variant(s) within the haplotype block (Edwards et al. 2013). Statistical fine-mapping approaches have been designed to partially overcome those limitations and further refine the association of each GWAS locus towards the most likely causal variant driving disease association within each 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 its potential in regulating gene expression (Petronis 2010). Additionally, the incorporation of gene expression, eQTL analysis and chromatin

Introduction

interaction data can establish a relationship between non-coding variants and putative gene targets. Final validation of the functional relationship between the genetic variant and the disease phenotype involves conducting appropriate cellular assays and *in vivo* experiments using animal models.

1.4.2 Understanding the epigenetic landscape in complex diseases

Epigenetic modifications consist of heritable changes in the phenotype and/or gene expression that do not involve changes in the DNA sequence (Feil and Fraga 2012). These changes include a wide range of modification in the proteins which serve as scaffold for the chromatin, known as histones, and DNA methylation and other epigenetic processes. Environmental and intrinsic factors can trigger changes in the epigenome that result in dysregulation of expression and, consequently, in alteration of the gene function.

In addition, genetic background can increase the predisposition to epigenetic changes caused by extrinsic factors. In fact, 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 a number those 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 and particularly important in the immune system to ensure adaptation and response to different pathogen infections (Yosef and Regev 2016). The role of cell type specificity in the epigenetic landscape has been

Introduction

demonstrated in eQTL studies, where between 50 to 90% of the genetic variants regulating gene expression are cell type and stimulus dependent (Dimas et al. 2009; Nica et al. 2011; Fairfax et al. 2012; Fairfax et al. 2014; Raj et al. 2014; Naranbhai et al. 2015; Kasela et al. 2017). Recent methodological advances in the field have made the personalised study and understanding of the epigenome possible by the implementation of low-cell-input high-throughput techniques coupled to NGS (**Buenrostro2013**; **Schmidl2015**; Oudelaar et al. 2017). One step further, the understanding of cell-to-cell epigenomic heterogeneity is also being addressed with single-cell methods and may help to elucidate the impact of genetic variability in regulation of gene expression and disease mechanisms (Buenrostro et al. 2015; Cusanovich et al. 2015; Rotem et al. 2015; Nagano et al. 2013; Smallwood et al. 2014).

1.4.3 The chromatin landscape

In the cell nucleus, DNA is compacted into a highly organised structure known as chromatin. The nucleosome is the basic repeating unit of chromatin and is formed by a 147bp segment of DNA wrapped around an octamere core of histone proteins regularly spaced by 10bp of linker DNA (Luger et al. 1997). In general, highly compacted DNA will remain more inaccessible for the assembly of the transcriptional machinery, consequently preventing gene expression. Chromatin accessibility can be altered by PTM of the histone proteins that affects their affinity with the DNA within the nucleosome as well as the interaction between nucleosomes in the vicinity (Polach et al. 2000; Pepenella et al. 2014). Additionally, chromatin structure can also be influenced by adenosin triphosphate (ATP)-remodelling complexes that facilitate sliding of individual nucleosomes to neighboring DNA segments, increasing temporary chromatin accessibility at particular sites (Cosma et al. 1999). From the biochemical point of view, the signature of chromatin accessibility, histone

Introduction

modifications, transcription factor occupancy and DNA methylation has been used to identifying *cis*-regulatory elements such as promoters, enhancers, silencers, insulators and locus control regions, and define the cellular chromatin landscape (Boyle et al. 2012; Kundaje et al. 2015).

Methods to ascertaining chromatin accessibility

Accessible chromatin constitutes about 1% of the human genome and represents a very robust marker for histone modifications, early replication regions, transcription start sites (TSS) and TF binding sites (TFBSs) (ENCODE 2007). The informativeness of chromatin accessibility for understanding 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 on nucleosome-free regions known as 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 TFBS, generating footprints that identify TF binding in relation to chromatin structure (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 retain only the nucleosome-depleted regions that undergo NGS (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 tagged by DHSs (ENCODE 2007; Buck et al. 2014; Gaulton et al. 2010; Maurano et al. 2012). Indirect measurement of the chromatin

Introduction

accessibility has also been performed using micrococcal nuclease-sequencing (MNase-seq). In this approach chromatin-free DNA on cross-linked nuclei is degraded and only the nucleosome-bound material is retained for downstream sequencing , providing a qualitative and quantitative comprehensive map for nucleosome positioning and also TF occupancy (**Axel1975**; Ponts et al. 2010). The high number of cells (5 to 10 millions or more) required by these assays for good quality data limits their application to particular biological and clinical samples.

Recently, a new technique known as assay for transposase-accessible chromatin using sequencing (ATAC-seq) has represented a groundbreaking step in characterisation of the genomic regulatory landscape (**Buenrostro2013**). ATAC-seq is based on an engineered hyperactive transposase enzyme, known as Tn5, that preferentially accesses nucleosome-free and inter-nucleosomal DNA inserting sequencing adapters at both end of those fragments (Gradman et al. 2008; Adey et al. 2010). The main advantage of ATAC-seq over DNase-seq is the lower number of cells and the simplicity of the protocol. These two aspects make ATAC-seq a very versatile technique to interrogate the chromatin landscape in a clinical set-up, where sample availability and time-efficiency are key factors (Scharer et al. 2016; Qu et al. 2015; Qu et al. 2017).

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

Identifying the combination of histone modifications and binding of TF to the DNA is essential to characterise regulatory regions of the genome and fully understand the transcriptional regulation. Histone modifications take place in the NH₂terminal tail that protrudes from the nucleosome, being the most common modifications acetylation, phosphorylation and methylation. The co-localisation of different histone marks modulate the affinity for DNA-

Introduction

binding proteins and the interaction with neighboring nucleosomes in varied manners, contributing to the overall chromatin accessibility landscape of the cells (Jenuwein and Science 2001; Bannister and research 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 permanent silenced, respectively. Facultative heterochromatin is enriched for H3K27me3 and the polycomb repressor complexes (PRCs), whilst constitutive heterochromatin is marked by H3K9me3 (Hansen et al. 2008; Bannister et al. 2001).

Several types of chromatin corresponding to different regulatory elements have also been defined. Enhancers and promoters, regardless of their activation state, are tagged by high levels of H3K4me1 or H3K4me3, respectively, and both features co-localise with H3K4me2 modifications (Heintzman et al. 2007; Hon et al. 2009). H3K9ac is specifically enriched at active promoters whereas H3K27ac generally designates activation at both promoters and enhancers (Hon et al. 2009; Creyghton and the 2010). Conversely, H3K27me3 together with the heterochromatin mark H3K9me3 indicates gene repression at promoter elements (Hansen et al. 2008; Bannister et al. 2001; Pan et al. 2007). Interestingly, GWAS variants for different complex diseases have 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). Overall, functional understanding and interpretation of histone mark co-localisation still remains challenging and incorporation of additional epigenetic information is usually required. Together with histone modifications, TF 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

Introduction

al. 2009; Bell and Nature 2000). TF occupancy is indirectly tagged by chromatin accessibility assays, such as DHS, through reduced cutting sensitivity of DNase I due to protein binding and steric hindrance.

Chromatin immunoprecipitation sequencing (ChIP-seq) has been widely used in the last few years to precisely locate histone modifications and TF binding in the genome. This technique assays protein-DNA binding *in vivo* using Abs that specifically recognise histone modifications or TF after DNA-protein cross-linking and sonication. Following immunoprecipitation of the desired DNA-protein complexes with the appropriate Ab, 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 (**ENCODE2012**; Bernstein and Nature 2010; Adams et al. 2012). Similarly to the first generation of chromatin accessibility techniques, 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, a wide range of protocols have been developed, of which ChIPmentation (ChIPm) stands out as the simplest and most cost-effective method, only requiring 10,000 and 100,000 cells to assay histone modifications or TF binding, respectively (**Schmid2015**). ChIPm involves the use of the Tn5 transposase to simultaneously fragment and add adapters to the immunoprecipitated DNA, 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 are

Introduction

found along the entire genome and their methylation generally associates with repression of gene expression (Herman and Medicine 2003). Together with histone modifications, DNA methylation has a pivotal role in orchestrating the immune system, importantly in the differentiation of haematopoietic stem cells and the maturation and activation of immune cells (Sellars et al. 2015; Lai et al. 2013). The pathogenicity of changes in the methylome has been studied in a wide range of complex diseases including RA, SLE, psoriasis and PsA (Lei et al. 2009; Liu et al. 2013; Zhang et al. 2010). For example, regulation of TNF- α production upon inflammatory stimuli involves a complex network of DNMTs that alter the methylation signature at the locus (Sullivan et al. 2007). This ability to ascertain the epigenome profile in clinical samples has also enabled epigenome-wide association studies (EWAS) that identify CpG methylation changes between patients and controls in a cell type specific manner (Zhou et al. 2016).

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 expression of the closest gene, functional interpretation of GWAS variants requires genome-wide mapping of those 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 the corresponding 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 additional evidence for physical contact between enhancers and gene promoters coordinating assembly of the transcriptional machinery and consequently regulating expression. As an example, obesity risk

Introduction

non-coding variants located within the *FTO* gene appeared to regulate expression through chromatin looping of the *IRX3* gene, 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, showing differences in performance and suitability depending on the application (Davies et al. 2017). Of particular interest, Capture-C has simultaneously scaled up the number of interactions investigated at high resolution and minimised the number of input cells required (Davies 2016; 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 2016). Lately, HiChIP has improved the integration of ChIP and chromatin interaction methods to enhance the specificity of the assay while reducing sequencing depth and input material (Mumbach et al. 2016).

1.4.4 Transcriptional profiles in disease

The role of environmental and genetic factors in altering gene expression regulation in complex diseases has been investigated through extensive comparison of case-control transcriptional profiles. The informativeness of this approach is conditional on studying the relevant disease tissue, which sometimes remains challenging due to a lack of pathophysiological understanding of disease mechanisms or difficulties in accessing it. In immune-mediated diseases, PBMC differential gene expression (DGE) analysis between patients and controls has enabled identification on relevant pathways and biochemical functions in RA, UC, SLE, AS, psoriasis and PsA, amongst others (Miao et al. 2013; Junta and 2009; Baechler et al. 2003; Assassi et al. 2010; Batliwalla et al. 2005). Similarly, the growing evidence supporting cell type and context specificity in the regulation of gene expression has driven more disease-specific targeted studies. Such studies

Introduction

include synovial-isolated macrophages in RA, B cells and monocytes in SLE or skin biopsies in psoriasis (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 (Londsdale 2013; Fagny et al. 2017). Lately, eQTL studies have also expanded in the disease context. For instance, an eQTL study in five immune relevant cell types isolated from IBD and anti-neutrophil cytoplasmic antibody-associated vasculitis patients have revealed disease specific eQTLs, some of which disappear following treatment (Peters et al. 2016).

Long non-coding RNAs and enhancer RNAs

In the last few years, the understanding of transcription has experienced a profound revolution, with the revelation that most of the genome undergoes transcription (ENCODE 2007). In addition to protein coding mRNAs, a number of non-coding RNAs have been characterised and demonstrated to have a role in regulation of transcription and gene expression. One category of non-coding RNAs are the long non-coding RNAs (lncRNAs), transcripts between 200 and 100Kb long that undergo splicing, 5' capping and 3' poly-adenylation (Derrien et al. 2012). lncRNAs can positively and negatively regulate transcription through different mechanisms including guidance of chromatin modifiers such as 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

Introduction

(**Pandei2008**; Faghihi et al. 2008; Gong and Maquat 2011; Carrieri et al. 2012; Kino et al. 2010).

According to the latest GENCODE annotation release, 15,778 lncRNAs have been identified in humans (Derrien et al. 2012). 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 (**Rossetto2009**; Pang et al. 2009). Moreover, differential case-control gene expression analyses have underscored the contribution of lncRNAs in several chronic inflammatory conditions, including RA, SLE and psoriasis (**Muller2014**; **Li2014**; Shi et al. 2014; Ahn et al. 2016).

A particularly relevant type of lncRNAs are the enhancer RNAs (eRNAs), shorter molecules compared to the canonical lncRNAs (approximately 346 nucleotides) that do not undergo splicing or poly-adenylation (Fantom 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 Santa2010**; Kim et al. 2010). Importantly, the transcriptional activity of enhancers has been demonstrated to be an excellent proxy to identify functionally active regulatory regions, which have also been successfully validated by reporter assays (**Anderssen2014**; Fantom et al. 2014).

micro-RNAs

Another class of non-coding RNAs are micro-RNAs (miRNAs). miRNAs are generated as larger precursors through transcription of non-coding regions of the genome and undergo processing into RNA species 21 to 24 nucleotides long (Lee et al. 2002). Under particular conditions, expression of genes containing complementary sequences to miRNAs are commonly negatively regulated through assembly of the miRNA-induced silencing complex followed by mRNA

Introduction

degradation, mRNA destabilisation or translational repression (Ameres et al. 2010; Braun et al. 2011; Petersen et al. 2006). In fact, between 30 and 80% of human genes are predicted to be under transcriptional control of micro-RNAs (miRNAs) (Lewis et al. 2005; Friedman et al. 2008).

Methods to assay gene expression

The use of micro-array based methods to perform genome-wide expression studies has been increasingly replaced by RNA sequencing (RNA-seq), as result of NGS technologies becoming more cost-effective. 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). This method overcomes the bias introduced by the use of pre-designed complementary probes in hybridisation techniques as it sequences all the RNA species contained in the sample. 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 to the expression profile, including the identification of new exons, alternative splicing events and allele-specific expression (ASE). For example, regulation of gene activity through differential isoform usage is very common between different tissues and during particular biological processes. RNA-seq isoform quantification has highlighted that differentiation of CD4⁺ T cells into the pro-inflammatory Th-17 is particularly driven by one of the nuclear receptor RORC γ isoforms (Zhao et al. 2014). RNA-seq has also allowed quantification of ASE for transcripts in individuals heterozygous for exonic SNP haplotypes in a particular gene, avoiding performance of additional molecular assays (Yan et al. 2002). Importantly, ASE has provided direct evidence for local/*cis*-eQTLs driven by allele-specific mechanism, showing significant differences in haplotype

Introduction

transcript abundance for up to 88% of the genes with an associated *cis*-eQTL (Pickrell et al. 2010). Furthermore, the 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).

Variations of the RNA-seq methodology such as cap analysis of gene expression (CAGE) and other 5'end RNA-sequencing methods have enabled the precise identification of TSS and the associated promoters for each transcript (Yamashita et al. 2011; Fantom et al. 2014). The CAGE data generated by the functional annotation of the mammalian genome 5' (FANTOM5) Consortium includes thousands of eRNAs and has contributed to better definition of enhancers and their spatial and temporal specificity in hundreds of human primary cells and tissues (Andersson et al. 2014).

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 hands, non-coding variants at enhancers, silencers and promoters can dysregulate gene expression by altering affinity at TFBS, histone modifications and chromatin accessibility. For instance, in thyroid autoimmunity, the risk allele of an intronic SNP in the tyroid stimulating hormone receptor (*TSHR*) gene reduces *TSHR* protein expression in IFN- α stimulated thyroid cells (Stefan et al. 2014). The risk SNP increases the affinity of the repressor promyelocytic leukemia zinc finger protein (*PLZF*) that

Introduction

recruits histone acetylases (HDACs) to the locus, resulting in impaired tolerance to thyroid auto-antigens. Alterations in TF binding can also affect looping and long-range chromatin interactions between enhancers and promoters. For instance, in prostate cancer this phenomenon causes upregulated expression of the oncogene *SOX9* due to increased enhancer activity and enhancer-promoter interaction (Zhang et al. 2012).

Alternatively, non-coding SNPs can regulate gene expression by creating a new promoter-like element, as in the α - thalassemia disease, where this phenomenon leads to dysregulated downstream activation of all α -like globin genes in erythroid cells (Gobbi et al. 2006). Genetic variants at eRNAs can also affect regulation of gene expression as it has been demonstrated in the nuclear receptor for anti-diabetic drugs PPAR γ 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. This is the case of a CD associated variant at the 3'UTR of the gene immunity related GTPase M *IRGM* which reduces binding of the miR-196, increasing its mRNA stability and translation, ultimately resulting in disrupted autophagy (Brest et al. 2011). In psoriasis and PsA, 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 Integration and interpretation of genomic data

The evolution of different omics methods towards generation of paired datasets at a high-throughput scale presents a challenge in terms of interpretation and integration. This is particularly important in the field of complex diseases resulting from the interaction of many risk variants with small or moderate effect that involve several genes and signaling pathways

Introduction

through alteration of epigenetic features and subsequent dysregulation of gene expression.

Tools such as RegulomeDB allow the querying of a large number of human publicly available epigenetic and functional data, including DHSs, TFBS, histone modification and DNA-protein interactions, at the SNP level (Boyle et al. 2012). Other powerful tools is the University of California Santa Cruz (UCSC) genome browser, a resource to display in-house and publicly accessible annotation data (Kent et al. 2002). In addition to this, international consortia generating large-scale epigenetic and expression data such as ENCODE, Blueprint, The epigenome Roadmap, GTEx or FANTOM have created comprehensive website resources for browsing and downloading data (**Londsdale2013; Adams2012 ; ENCODE 2007; Fantom et al. 2014**). These collaborations have also led to the integration of epigenetic datasets and assembling of cell type specific chromatin states maps. This consists in the segmentation and labelling of the genome with a chromatin state based on concurrence of several epigenetic marks using Hidden Markov Model algorithms such as ChromHMM, amongst others (**Kundaje2015 ; Ernst and Kellis 2010; Ernst et al. 2011; Hoffman et al. 2013**).

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 differentially epigenetic modified regions. 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) within the entities of interest. For instance, pathway enrichment analysis uses functional units containing related genes defined by prior knowledge. Amongst the most comprehensive and informative pathways sources are The Kyoto Encyclopedia of Genes and Genomes (KEGG) and the REACTOME, which also considers biochemical reactions such as binding, activation or protein translocation (Kanehisa and Goto 2000; Fabregat

Introduction

et al. 2018). Such annotation sources may be used to interpret, for example, a set of differentially expressed genes or a list of genes obtained through annotation of non-coding regions using proximity, chromatin interaction data or eQTL studies. Similarly, type of analysis can be used to find enrichment of genomic regions of interest for a varied collection of epigenomic features tagging regulatory elements in relevant cell types.

From the number of tools designed to perform this type of analysis, eXploring Genomic Relations (XGR) is particularly powerful (Fang2016). XGR is an open source R package and web-app that allows handling of different types of input data (SNPs, genes and regions). XGR integrates a wide range of ontologies and up to date publicly available functional data to perform different types of annotation and enrichment analysis, facilitating background customisation for reliable and meaningful output results. Moreover, XGR also performs gene network analysis from the same inputs as the pathway analysis. This leverages experimentally validated interaction information to identify gene networks modulated by putative pathogenic variants, improving interpretation through consideration of network connectivity.

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

Fine-mapping strategies can partially overcome two of the main limitations of the GWAS studies: the association of hundreds of SNP per locus due to LD and the incomplete coverage of the human genetic variation. The aim of fine-mapping analysis is reducing the size of the 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 genetics 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

Introduction

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 in a cost-effective manner (Trynka et al. 2011). Similarly, imputation methods using WGS reference panels, such as the aforementioned HapMap and 1000 Genomes Project, have offered genome-wide coverage for SNPs and CNVs with MAF >1% across different ancestry groups (Abecasis et al. 2012). More recently, the UK10K project has improved the quality of imputation specifically for rare variants with MAF between 0.01% and 0.5% (Chou et al. 2016).

Bayesian statistical analysis has been chosen over the frequentist approach (based on p-value calculations) to increase the resolution of the GWAS associations and facilitate the identification of relevant genes and disease mechanism. 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 approximately 97.6% of loci, the Bayesian strategies report a credible set of SNPs that account for 95 or 99% of the overall PP in each loci(Bunt et al. 2015).

The inclusion of functional data from publicly available sources as priors of the approximate Bayesian model have demonstrated 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 Genetics 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

Introduction

within TF binding sites (TFBS) (Fahr2015). Fine-mapping can also benefit from the integration of epigenetic data generated in clinical samples and rare populations using the latest methodological improvements previously reviewed. The advances in the field of epigenetics have also led to the generation of more accurate chromatin states maps based on larger number of samples that have been integrated with fine-mapping strategies highlighted promising potential GWAS causal variants in T2D (Thurner et al. 2018)

1.4.8 Approaches to establish disease mechanisms and causality of genetic variant

Prioritisation of non-coding variants by integrating fine-mapping, epigenetics and expression data, as previously described, still does not unequivocally addresses the functional mechanisms conferring them pathogenic effect in a cell type and context specific manner. To overcome this, a wide range of experimental approaches can be performed to functionally validate and test the predicted effect of the variant in regulating gene expression.

In vitro assays to investigate the effect of genetic variants in regulating gene expression, involve for example transfection of constructs containing the promoter or enhancer element to test followed by luciferase expression (Niimi et al. 2002). Other molecular assays to interrogate allelic differences in affinity for TF binding include electrophoretic mobility shift assay (EMSA) and ChIP using Ab for the particular TF of interest (Vernes et al. 2007). The need to perform these assays at a genome-wide scale has yielded to development of high-throughput technologies, such as massively parallel reporter assays (MPRAs), which test putative enhancers and the effect of genetic variability in their functionality (Kheradpour et al. 2013). In addition to this, mass spectrometry (MS) techniques have been used to perform allele-specific quantitative proteomics and have revealed allele-dependent binding of TF and co-regulators at the T2D PPARG

Introduction

GWAS locus (Lee et al. 2017). *In vivo* validation has traditionally involved the use of mice models, including knock-outs for the potentially pathological genes or the regulatory elements containing GWAS prioritised variants. Nevertheless, the use of mice models to study human genotype-phenotype relationships has shown to have limitations that need to be taken into account when interpreting the results (Ermann and Glimcher 2012).

Both, *in vitro* and *in vivo*, models for functional studies have benefited from the development a the genome-editing technology known as clustered regularly interspaced short palindromic repeats (CRISPR/Cas) (Cong et al. 2013). CRISPR/cas enables monoallelic and biallelic modifications of primary cells and embrionic stem cells (ESCs) for the particular SNP or region of interest. The limitations of CRISPR to edit certain primary cells is being overcome by the use of human induced pluripotent stem cells (hiPSCs), which can undergo terminal differentiation into the cell type of interest after CRISPR modification (Ding et al. 2013).

Chapter 2

Material and Methods

2.1 Ethical approval and recruitment of study participants

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).

Psoriasis patients were eligible for recruitment when aged 18 years or older, previously or newly diagnose fulfilling the Psoriasis Area and Severity Index (PASI) classification and in a flare. Recruited patients were required to present moderate to severe disease (PASI>5), not having taken antibiotics in the two weeks before sampling and naïve for biological therapy. Availability of clinical information and written consent were also required. Detailed clinical information of the psoriasis cohort is included in Chapter ?? Table 4.2.

Material and Methods

2.1.2 PsA patient recruitment

Sample recruitment was performed as part of the Immune Function in Inflammatory Arthritis (IFI) 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).

Eligibility of the PsA patients required when aged 18 years or older, previously or newly diagnosed according to the PsARC, including 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. Written consent and clinical data were also collected. Further details about the cohort and clinical information can be found in Chapter 5 Table 5.1.

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 sample recruitment. Written consent was required.

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 (SF) 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 (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 adherent assays

KCs enrichment from skin biopsies was performed as described in Gutowska-Owsiaik and colleagues (GutowskaOwsiaik 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

Material and Methods

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 Ca²⁺ and cultured in a collagen IV coated 96-well plate at 37°C 5% CO₂ 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 40-50x10⁶ PBMCs was carried out using a modified version of the (Kent 2009) protocol, where cells were pre-conditioned in RPMI 1640 complete medium (Lonza) supplemented with 2 mM L-glutamine (Sigma), 100U penicillin and streptomycin 100µg/mL (Sigma) and 50% FCS for 30 minutes and, afterwards, diluted 1 in 2 in complete RPMI 1640 (supplemented as previously described) with 20% dimethyl sulfoxide (DMSO, Sigma). PBMCs followed slow cryopreservation at -80°C at -1°C per minute and then transferred and stored for a minimum of two weeks in liquid nitrogen.

PBMCs were thawed quickly in a 37°C water bath, resuspended in supplemented complete RPMI 1640 with 10% FCS at a density of 10⁶ cells/mL and rested for 30 min at 37°C, 5% CO₂ in 25mL non-adherent polypropylene cell culture flasks (Greiner) followed by filtering through a 40µm cell strainer to obtain an 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 5x10⁶ 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

Material and Methods

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% CO₂ 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 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⁺, CD8⁺, CD14⁺ monocytes and CD4⁺ cells (catalogue numbers 130-050-201, 130-045-101, 130-045-201 and 130-050-301, respectively) and AutoMACS Pro (Miltenyi) followed by a manual cell count with trypan blue. MACS separation was chosen over fluorescence-associated cell sorting (FACS) due to time and logistic constraints during sample processing.

2.2.5 Primary cell isolation using fluorescence-activated cell sorting

Isolation of cell subpopulations from PBMCs and SFMCs to study the effect of cryopreservation in the chromatin landscape (Chapter 3) (Chapter 5 was performed by FACS. PBMCs and SFMCs were resuspended in 1mM EDTA PBS (FACS buffer) at 10x10⁶ cells/mL, stained with the appropriate antibody cocktail (Table 2.1) 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, a modified FACS buffer supplemented with 3mM EDTA , 2% FCS and 25mM 4-

Material and Methods

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

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

Table 2.1: 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⁺ monocytes, CD4⁺ and CD8⁺ T cells, CD19⁺ B cells and CD56⁺ NK cells, as well as in epidermal KCs isolated from skin biopsies. Fast-ATAC and Omni-ATAC were two subsequent versions published following the standard ATAC-seq protocol from (**Buenrostro2013**), aiming to reduce the amount of mitochondrial DNA in the sequencing libraries and improve the signal-to-noise ratio of the original protocol. When using MACS separation, primary cells were manually counted, as specified above, and resuspended in PBS 1% FCS.

ATAC-seq

ATAC-seq was used to generate data from NHEKs, skin biopsies (Chapter ??) and healthy volunteers to test the effect of cryopreservation in the chromatin landscape (Chapter ??) as well as cohort 1A primary immune cells isolated from blood of psoriasis and control samples (Chapter ??). ATAC-seq was performed using an estimated number of 50,000 cells as described in Buenrostro *et al.*, 2013 with minor modifications. 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 fixative remains prior to 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

Material and Methods

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 KCs 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 Corces *et al.*, 2016, called Fast-ATAC. Optimised for hematopoietic cells it combined cell lysis and transposition into a single step. In this thesis, Fast-ATAC was performed in skin biopsies (Chapter ??), cohort 1A primary immune cells isolated from blood of psoriasis and control samples (Chapter ??) and primary immune cells isolated from blood and SF of PsA patients (Chapter ??). Fast-ATAC was conducted as described by Corces *et al.*, 2016 with minor modifications. Following the advice from Corces and colleagues, 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, as specified in (Corces *et al.* 2016). Following transposition DNA was prepared and purified as per ATAC-seq except with 13 cycles of PCR amplification after appropriate cell cycle determination in a pilot set of samples.

Material and Methods

Omni-ATAC

Omni-ATAC was performed in 50,000 viable NHEKs in suspension as described by (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 by the Oxford Genomics Centre, at the Wellcome Centre for Human Genetics (WCHG), aiming for 30 million paired-end reads.

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 (Schmidl2015). The H3K27ac histone mark (active enhancer and promoter marker) was assayed in four cells types (CD14⁺ monocytes, total CD4⁺, total CD8⁺ and CD19⁺). For ChIPm, 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, 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.

Material and Methods

Aliquots as needed were thawed on ice and then processed downstream for ChIPm as in Schmidl *et al.*, 2015. Immunoprecipitation was carried out with 1µg of the Diagenode Ab (C15410196). For each sample, an aliquot of chromatin was processed in parallel without incubation with the anti-H3K27ac Ab (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 *et al.* 2015). A pool of 64 libraries (including control input samples) were sequenced over a number of lanes in the HiSeq4000 Illumina platform by the Oxford Genomics Centre, WCHG, 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 between 2-3x10⁶ cells were resuspended in 350µL 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 ?? Table 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 ?? Table 4.2 and 4.3) and PsA samples (Chapter 5), cells were resuspended in 0.1% BM supplemented RLT

Material and Methods

buffer, which lysis cells and prevents RNA degradation. When starting from RNAProtect preserved material, cells were centrifuged at 300g for 10 min at room temperature, the supernatant were removed and the pellets were resuspended in 350µL of RLT 0.1% BM buffer. All cell lysates were homogenised using the QIAshredder (Qiagen) prior to RNA extraction.

Total RNA was extracted using the AllPrep DNA/mRNA/microRNA Universal kit (Qiagen) following the manufacturer's instructions. RNA extractions were performed in batches of 12 samples, including all cell types from each individual processed and a 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, library preparation and sequencing were carried out by Oxford Genomics Centre at the WCHG in two independent batches of samples, each including Cohort 1A or Cohort 1B, respectively. 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 was profiled by qPCR using the RT2 Profiler PCR Array (PAHS-3803Z, Qiagen) in collaboration with UCB. This platform included primers to test expression for 370 key genes involved in immune response during autoimmunity and inflammation, as well as appropriate house-keeping genes for normalisation. In brief, RNA was extracted, as detailed previously, from CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells. Reverse-transcription for cDNA synthesis and qPCR gene expression quantification was performed by UCB following the PCR array's manufacturers instructions.

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 WCHG. Briefly, PBMCs and SFMCs were made into a cell suspension. Approximately 3,000 cells from the PBMCs or SFMCs suspensions were partitioned into single-cell gel beads in emulsion (GEMs) using the 10X Chromium controller system. Reverse-transcription for cDNA synthesis was performed within the GEMs, which included a 16bp 10x barcode, a 10bp unique molecular identifier (UMI) and poly-dT primers. The cDNA was released from the GEMs, followed by PCR amplification, enzymatic fragmentation and size selection. 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.

Material and Methods

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

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

Mass cytometry assay was performed by Dr Nicole Yager in collaboration with UCB following their in-house protocol for the CyTOF instrument. Briefly, an aliquot of whole blood and SF 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 whole blood and SF 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 Abs against the cell surface markers of the intra-cellular staining (ICS) panel (Table 2.2). The samples were further

Material and Methods

permeabilised and stained with Abs of the ICS panel against the intracellular targets (Table 2.2).

Markers from the ICS mass cytometry panel

CD248, CD19, GP38, FAP, CD8a, IL8, CD16, CD25, CD123,
IL-17F, IL-17A, IL-10, CD11c, CD14, IL6, IFN- γ , CD-11b
CD45RO, CD56, HLA-DR, IL-13, CD117, CD4, IL4, IL-2, TNF α ,
IL-21, FceR, CD3, CD161, MCP-1^(*), GM-CSF, CD45,
MCP-1^(*), osteopontin^(*)

Table 2.2: Molecules targeted by the mass cytometry ICS staining panel in whole blood and SF. The molecules targeted by the Abs used in the ICS staining panel are listed. Note the panel also included Abs recognising surface markers to identify the cell populations of interest for further analysis of the intracellular cytokine production. (*) indicates two proteins which staining was only performed for additional eleven patients in the validation panel.

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, towards which I made an important contribution. 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 analysis

NGS data for each sample was trimmed for low quality base pairs and Nextera adapter sequences using cutadapt (Martin2011) before general QC assessment using fastqc (S. 2010). Trimmed reads were aligned to the reference genome (build hg19) using bowtie2 (Langmead2006) and the

Material and Methods

following parameters -k 4 -X 2000 -I 38 --mm -1, consistent with other publications (Buenrostro2013; 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 Bioinformatics 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 range of pvals (from 0 to 10^{-45}) and IDR analysis done for the resulting pairs

Material and Methods

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, SF or PB). Reads overlapping each of the peaks in the master list were retrieved for each sample using the HTSeq-count algorithm (Anders et al. 2015). Principal component analysis (PCA) was performed on all counts normalised with the vsd function from the DESeq2 v1.20 R package (Love et al. 2014).

Material and Methods

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 ??). 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 analysis

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 appropriate --shift parameter. For ChIPm PCA, filtering and downstream analysis peak homogenisation was performed similarly to Section 2.4.1 to build

Material and Methods

a combined master list for all samples and cell types from cohort 1B (Chapter ?? Table 4.2 and 4.3).

Sample quality was determined by a combination of measurements. For library complexity the non-redundant fraction (NRF) and PCR bottleneck coefficients (PBC1 and PBC2) were calculated following ENCODE guidelines (<https://www.encodeproject.org/chip-seq/histone/>) from unfiltered bam files. 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 SF and PB 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 PB from PsA patients versus healthy controls, the FC for each gene and each individual

Material and Methods

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. 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.

Material and Methods

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 fully processed downstream using the R package Seurat 2.3.4 (Butler et al. 2018). Each of the PBMCs and SFMCs individual count matrices were downsampled to 3,500 cells after removing all the genes expressed in fewer than 30 cells (drop-out events). Additional filtering was conducted to remove cells presenting more than 7.5% of mitochondrial reads, a number of genes larger than $500 \pm 1SD$ (approximately 1,800 genes in all the samples) and a maximum of 10,000 UMI. After filtering, individual PBMCs and SFMCs count matrices were processed for data scaling and normalisation (regressing number of UMI and percentage of mitochondrial counts) and PCA analysis. The first eight PCs (capturing most of the variation in the data according to jackstraw analysis) were then used to identify clusters (groups of cells with similar expression profiles) built by a shared nearest neighbor (SNN) modularity optimisation algorithm (default resolution 0.6) followed by visualisation using t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensional reduction. CD14⁺ monocyte clusters were selected based on co-expression of appropriate cell specific markers (*CD14* and *LYZ*). For each of the CD14⁺ monocyte populations, the top variable genes were quantified by

Material and Methods

dispersion of expression (variance/mean) ratio, using a cut-off for maximum mean expression of 4 and minimum dispersion of 0.25. The union of the 1,000 most variable genes across the six CD14⁺ monocyte samples (three from PBMCs and three from SFMCs) were used to perform canonical correlation analysis (CCA). The first nine canonical correlation vectors (CCs) were used to align all the CD14⁺ monocyte populations. This alignment using CCA was applied in order to merge all the CD14⁺ monocyte datasets, removing batch effects and allowing further downstream analysis. Cluster identification and visualisation using t-SNE in the combined CD14⁺ monocytes from the six samples was performed (further detailed in Chapter 5) followed by DGE analysis between the SFMCs and PBMCs CD14⁺ monocyte populations.

2.4.4 Genomic region annotation, enrichment analysis, gene-network 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).

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. 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

Material and Methods

analysis master list or all the gencode hg19 detected genes, depending on the analysis.

Gene network analysis was carried out with xSubneterGenes XGR functionality using as the input list all the qPCR array genes and as significance-level the best pval across the three cell types where the expression was assayed. This list of genes was superposed onto the STRING interaction network (including known and predicted proteinprotein interactions) to obtain a maximum-scoring gene subnetwork (30 genes) containing as many highly significant (highly scored) genes as possible and a lesser number of non-significant genes as linkers.

Visualisation of the signalling-pathway enrichment from the RNA-seq results was performed using the R package Atlas and Analysis of systems-biology-led pathways (A2), developed by Dr. Hai Fang and towards which I have contributed (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

PsA fine-mapping using Immunochip genotyping data

Fine-mapping was performed for a number of non-MHC PsA Immunochip GWAS susceptibility loci using 1,103 patients and 8,900 controls from the Bowes et al. PsA Immunochip UK cohort. 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 SNPTEST 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 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 *et al.*, 2015.

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 *et al.*, 2012, for which only the summary statistics of the GACP Immunochip cohort (2,997 cases and 9,183 controls)

Material and Methods

were available. 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 (σ^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 (β) and standard error (SE). SE is calculated as $\sqrt{\sigma^2}$ and β 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.

2.4.6 Mass cytometry data analysis

Mass cytometry analysis was performed by Dr Nicole Yager. Cell populations in SF and PB at 0h and 6h were identified using manual annotation. For the CD14⁺ monocyte population identified within each tissue at 0h and 6h, the percentage of TNF- α , MCP-1 and osteopontin positive staining cells were calculated. Significance of the differences in the percentage of positive staining

Material and Methods

cells for each cytokine between SF and PB CD14⁺ monocytes was determined using Wilcoxon signed rank test.

Chapter 3

Establishment of 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, as previously 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 (Buenrostro2013; 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 yield information regarding 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 the advantages in terms of reduced input material and processing time, early data from ATAC-seq revealed two major limitations involving high percentage of mitochondrial (MT) 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 protocol particularly for hematopoietic cells, named Fast-ATAC, replaced the NP-40 detergent used in ATAC-seq with digitonin. This prevents solubilisation of the MT membrane, and performs lysis and transposition in a single step. This modification efficiently reduced the percentage of MT reads down to approximately 10% and increased ATAC signal at annotated TSS, using only 5,000 cells as input material (Corces et al. 2016). Optimisation of the ATAC-seq protocol for KCs 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, 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, a 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. Moreover, greater signal-to-noise ratios were also achieved with Omni-ATAC by modifying the transposition buffer. Notably, this versatile protocol presented particular improvement in KCs with data demonstrating the inability of ATAC-seq and Fast-ATAC to yield good quality data in this cell type.

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 10×10^6 cells), ATAC-seq and Fast-ATAC showed lower signal-to-noise ratios and higher variability across samples, as previously mentioned. This required appropriate implementation of quality control (QC) 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 (Table 3.1). Determining 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. Using false discovery rate (FDR) has been the most widely applied

Establishment of methods to assess genome-wide chromatin accessibility

criterion except for ENCODE data, where technical replicates are generated and IDR analysis has been used to identify robust peaks (Table 3.1).

Table 3.1: Summary table of ATAC-seq methodology analysis for peak calling, filtering and differential analysis. NA indicates the study did not perform or detail that aspect of the analysis.

Publication	Peak calling and filtering	Master list	Differential analysis
Corces <i>et al.</i> , 2016	MACS2 (-nomodel), summit extension +/-250bp, overlapping peaks. rank summits by pval	Maximally significant	non-Quantile normalisation and unsupervised hierarchical clustering.
ENCODE	MACS2 -nomodel, pairwise IDR analysis, filtering IDR<10%	Choosing longest IDR filtered list or only peaks present in the two samples	pairwise NA
Turner <i>et al.</i> , 2018	MACS2 (-nomodel -q 0.01)	pseudoreplicates.	<i>De novo</i> :DiffReps with fragment from the different cell types. size 50bp.
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 limma voom (FDR<0.01).

Qu <i>et al.</i> , 2017	ZINBA PP>0.99.	Merging of filtered peaks from each individual sample.	Quantile 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 exact text (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).

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 logistic problem regarding the handling of clinical samples. Often sample recruitment takes place at distant geographical locations or out of normal working hours. This requires the application of preservation methods that the alteration of the *in vivo* cellular characteristics and avoid confounders in the biological significance of the results generated. 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 for ATAC processing, 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 WCHG 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, alike formaldehyde fixation. DSP preservation does not require sample freezing after fixation and samples can undergo successful immuno-staining as well as FACS cell separation (**Espina2013**).

3.1.5 Aims

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

The specific aims of this chapter are:

1. To establish a 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 KCs.
5. To determine the effect of cryopreservation and DSP fixation on ATAC-seq data quality and on the overall chromatin accessibility landscape of CD14⁺ monocytes and tCD4⁺ cells.

3.2 Results

3.2.1 Establishment of an ATAC-seq data analysis pipeline based on current knowledge

At the time of the first ATAC-seq publication (**Buenrostro2013**), well established protocols for complete processing and data analysis 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). In the process of analysing ATAC-seq data, several limiting aspects are encountered, including QC assessment, peak calling and filtering, and identification of differential chromatin accessibility between groups of samples. Using the current knowledge in the field as well as custom analysis, the most appropriate criteria and parameters to implement in the in-house pipeline were established. For this purpose different types of analysis were performed using ATAC-seq data generated with the (**Buenrostro2013**) protocol in paired CD14⁺ monocytes and total CD4⁺ T cells (tCD4⁺) from three healthy individuals (ATAC-seq fresh samples generated for).

Sample QC

The variability in performance of ATAC-seq and Fast-ATAC necessitated the identification of appropriate parameters to determine the quality of the samples before proceeding with downstream differential analysis. This has been a dynamic process during the project that has benefited from the increase in the number of publications including ATAC data analysis as well as understanding the technical limitations from ATAC-seq and Fast-ATAC protocols. After continuous review of the different read-outs implemented across different publications, as well as the recently ENCODE updates, a comprehensive analysis was performed in order to identify the most informative QC measures as well as equivalence and correlation between them.

First, ATAC-seq fragment size distribution was analysed in each of the six libraries following down-sampling to 30 million reads after filtering to facilitate the comparison. The fragment size distribution showed presence of nucleosome periodicity protecting the DNA during the transposition event (Figure 3.1 a), one of the indicators of chromatin integrity and thus good quality ATAC-seq libraries. All the samples presented appropriate nucleosome periodicity (every ~200bp) up to 600bp, clearly distinguishing chromatin organisation into mono-, di- and tri-nucleosomes. The relative intensity of nucleosome-free fragments (NFF, ≤ 147 pb, approximately) compared to nucleosome-bound DNA was greater for some of the samples (e.g CTL1 CD4 $^{+}$ and CD14 $^{+}$) and equal or lower for others (e.g CTL3 CD4 $^{+}$ and CD14 $^{+}$). NFF were clearly distinguished in all of the samples, which is considered a compulsory feature for ATAC-seq libraries to pass QC, according to the latest ENCODE recommendations ([ENCODE](#)).

Another QC measurement that was investigated and implemented was the enrichment of ATAC-seq signal over a random background of reads across all the TSS identified for Ensembl genes (Figure 3.1b). It is well established that

nucleosome repositioning and an increase in chromatin accessibility occur at the TSS to allow TF binding and initiation of gene transcription.

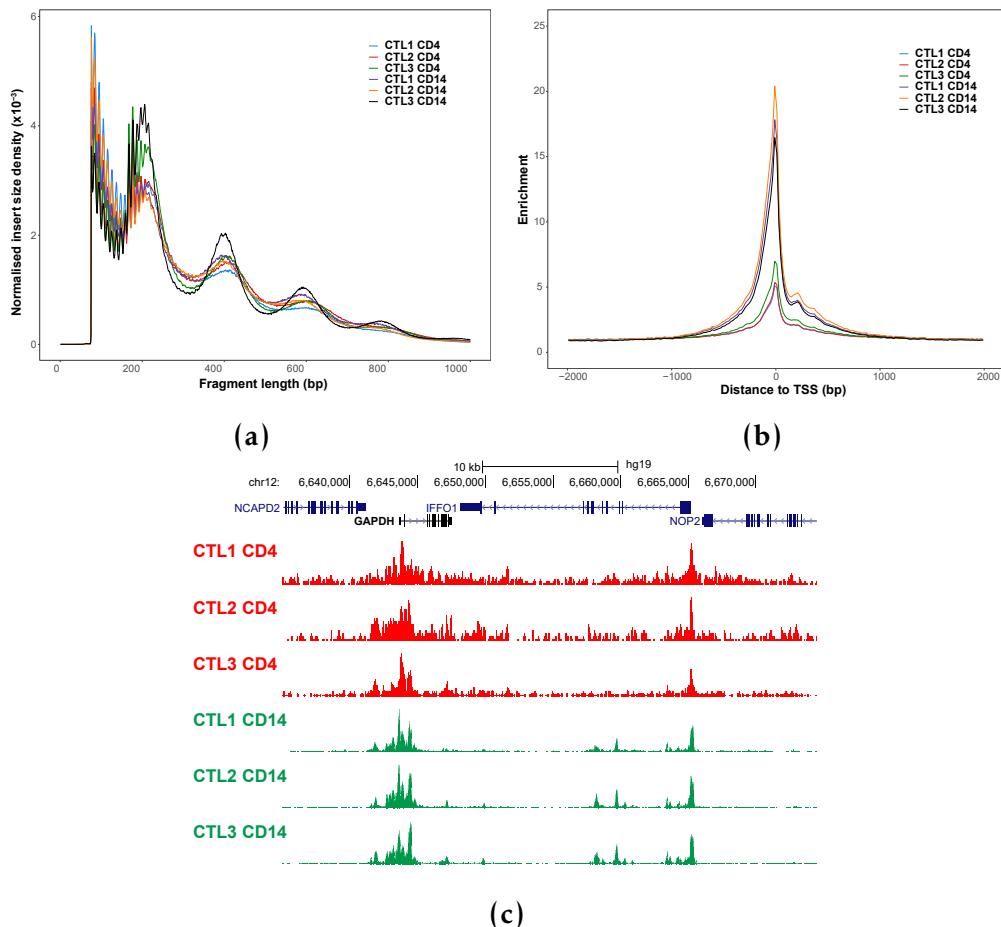


Figure 3.1: Measurements for QC assessment in ATAC-seq samples. For each of the CD14⁺ monocytes and tCD4⁺ samples used to establish ATAC analysis QC measures include a) density distribution of ATAC-seq fragment sizes, b) enrichment of ATAC-seq fragments across the TSS of all Ensembl genes and c) UCSC Genome Browser view illustrating the ATAC-seq normalised read density (y-axis) at the promoters of *GAPDH* and *NOP2* genes. In c) CD14⁺ monocytes and tCD4⁺ tracks are colour-coded in green and red, respectively.

Fold-enrichment signals over the TSS ranged from 5 to 7 for the tCD4⁺ samples, and were much higher (17 to 20) in the CD14⁺ samples. The lower sample quality of the tCD4⁺ compared to CD14⁺ samples indicated by the TSS enrichment values were recapitulated by the ATAC-seq read pile up at the promoters of the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and the

NOP2 Nucleolar Protein (*NOP2*) gene, showing more background reads and lower signal for the tCD4⁺ samples in red (Figure 3.1c).

As part of the QC assessment, the percentage of mitochondrial reads and the fraction of reads in peaks (FRiP) were also investigated (Table 3.2).

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

Table 3.2: ATAC-seq percentage of MT reads and fraction of reads in called peaks (FRiP). The percentage of MT 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).

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 (data not shown), suggesting both are appropriate inter-dependent QC measures to evaluate sample noise. The MT content ranged between 14.9-43.3%, was higher in CD14⁺ than in tCD4⁺ and was not directly related with any of the other QC measurements. Therefore, MT reads in this range did not appear to reflect the samples quality and the main inconvenience related to the need for deeper sequencing to achieve the desired number of non-MT reads for downstream analysis.

In summary, both TSS and FRiP are appropriate signal-to-noise measures, with recommended threshold values from ENCODE and Alsooo *et al.*, 2018 being of FRiP between 10-20% and TSS between 6-10. Importantly ENCODE has prioritised the use of TSS over FRiP as a more stable measure to determine the noise in the sample and will also be the chosen measure in this study.

Establishment of methods to assess genome-wide chromatin accessibility

According to this analysis, all six samples showed appropriate ATAC-seq patterns of fragment size distribution, FRIP and TSS, with exception of tCD4⁺ CTL1 and CTL2, being borderline for the 6 fold-enrichment TSS threshold. Importantly, this differences in the enrichment around the TSS successfully recapitulated the differences observed in the ATAC-seq density signal of the UCSC Genome Browser tracks between the CD14⁺ and tCD4⁺ samples. These differences in ATAC-seq quality observed in these samples reflected the variability in performance of ATAC-seq and were useful in determining the influence of borderline sample quality in downstream analysis in order to choose the most robust strategy maximising the use of precious clinical samples.

Peak calling and filtering

As part of the ATAC-seq pipeline implementation, criteria for peak calling and filtering were another two aspects to define. Although different peak callers have been used to analyse ATAC-seq data, MACS2 has been methodology preferred by ENCODE and most publications (Table 3.1). MACS2 was initially developed for ChIP data, but it 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, it was set to 200bp which was 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, such as DHS or ATAC-seq. Later publications based on clinical samples have also used similar parameter for ATAC peak calling (Wang et al. 2018).

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 every 5 million, ranging from 5 to 30 total million reads, followed by peak calling with arbitrary filtering for false discovery rate (FDR)<0.01. The number of called peaks passing filtering showed a steady increase with read depth (Figure 3.2 a), beginning to *plateau* at approximately 25 million reads (Figure 3.2 b). Moreover, lower number of peaks were detected in tCD4⁺ samples compared to CD14⁺ when using standard FDR<0.01 filtering, highlighting the influence of sample quality on the total number of significant called peaks. Interestingly, sample quality as measured by FRiP (which relies on peak calling) showed very discrete changes with read depth and was stable from 15 million reads onwards for all six samples (Figure 3.2 c), similarly to TSS (data not shown). Overall, this confirmed that measurement of sample quality using FRiP or TSS was not affected by the sequencing depth.

For peak filtering, most of the ATAC-seq publications using MACS2 have consistently used an FDR<0.01 (Table 3.1). However, this arbitrary way of filtering may not remove low quality peaks equally successfully in samples at the lower side of the acceptable quality spectrum. Moreover, filtering based on MACS2 FDR<0.01 does not take into account the reproducibility of the called peaks. In collaboration with Dr. Gabriele Migliorini and following the ENCODE pipeline, 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 pvals was calculated (Figure 3.3 a 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 pval giving the largest percentage of IDR shared peaks between the two pseudoreplicates varied more erratically

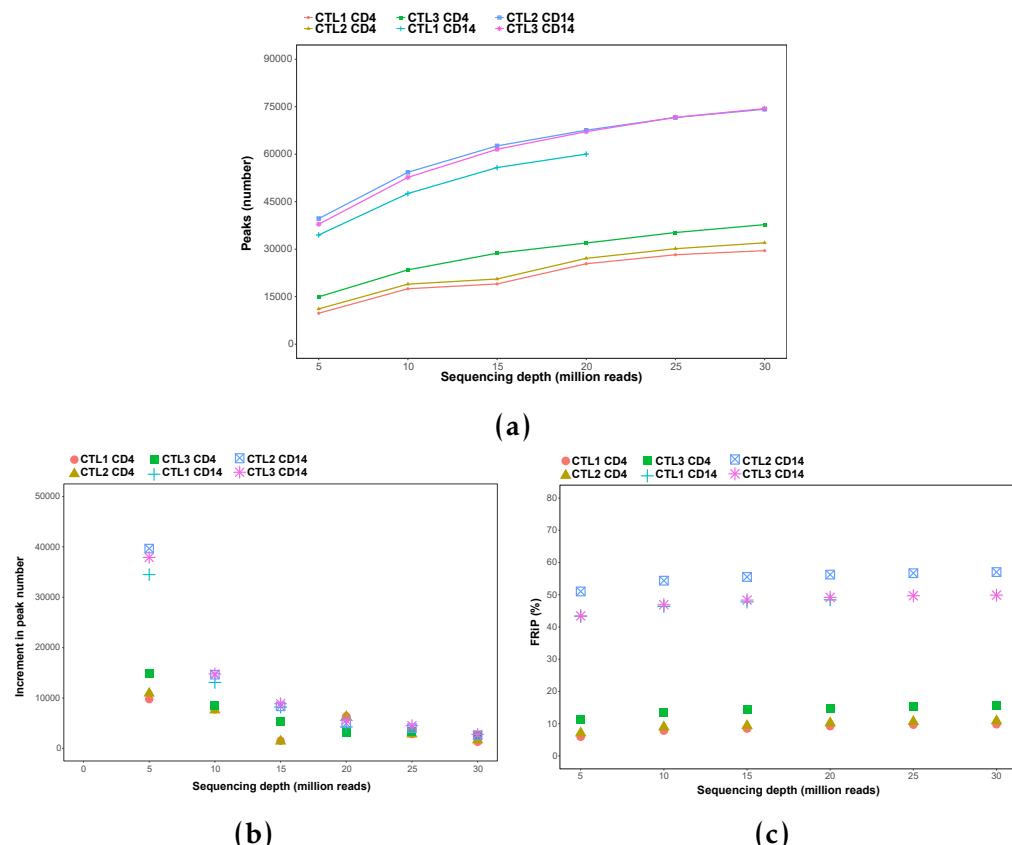


Figure 3.2: FRIP and peak calling at different sequencing depths in ATAC-seq libraries. For a serie of sequencing depths (from 5 to 30 million reads after filtering) representation of a) number of called peaks (standard filtering using FDR<0.01, b) the increment on the number of called peaks and c) FRIP as a function of the sequencing coverage in the six samples included for the analysis in this section.

Establishment of methods to assess genome-wide chromatin accessibility

when the sequencing depth was lower than 10 million reads (Figure 3.3 a and b), suggesting this analysis was not appropriate for lower read depths. This variation was more pronounced and extended in the tCD4⁺, which presented the lower TSS values, when compared to the CD14⁺ monocytes ATAC-seq libraries (representative examples of CTL2 CD4⁺ and CD14⁺ monocytes in Figure 3.3 a and b). The shape of the curves were also influenced by the sample quality. For appropriate sequencing depth (15-20 million reads), the CD14⁺ monocytes (TSS enrichment ≥ 10) samples presented a profile reaching a single maximum of shared IDR peaks for a particular filtering pval (Figure 3.3 b), which was -log₁₀ pval 8 in all three samples (data not shown). In contrast, the same analysis in tCD4⁺ samples revealed two pvals for which the percentage of IDR shared peaks reached two local maxima (Figure 3.3 a).

Filtering the CD4⁺ peaks at the pval of the first local maximum (10^{-11}) reduced the percentage of peaks overlapping regions suggestive of noise (e.g heterochromatin, repetitive sequences and repressed regions) 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 pval 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 master list of regions across all the samples and perform differential chromatin accessibility analysis.

Differential chromatin accessibility analysis

In this project, a peak-based approach using the number of read counts overlapping the peaks included in the consensus master list (ML_all) was implemented to perform differential analysis. One of the main limitations of the ATAC-seq and Fast-ATAC protocols (discussed in the next section) is the

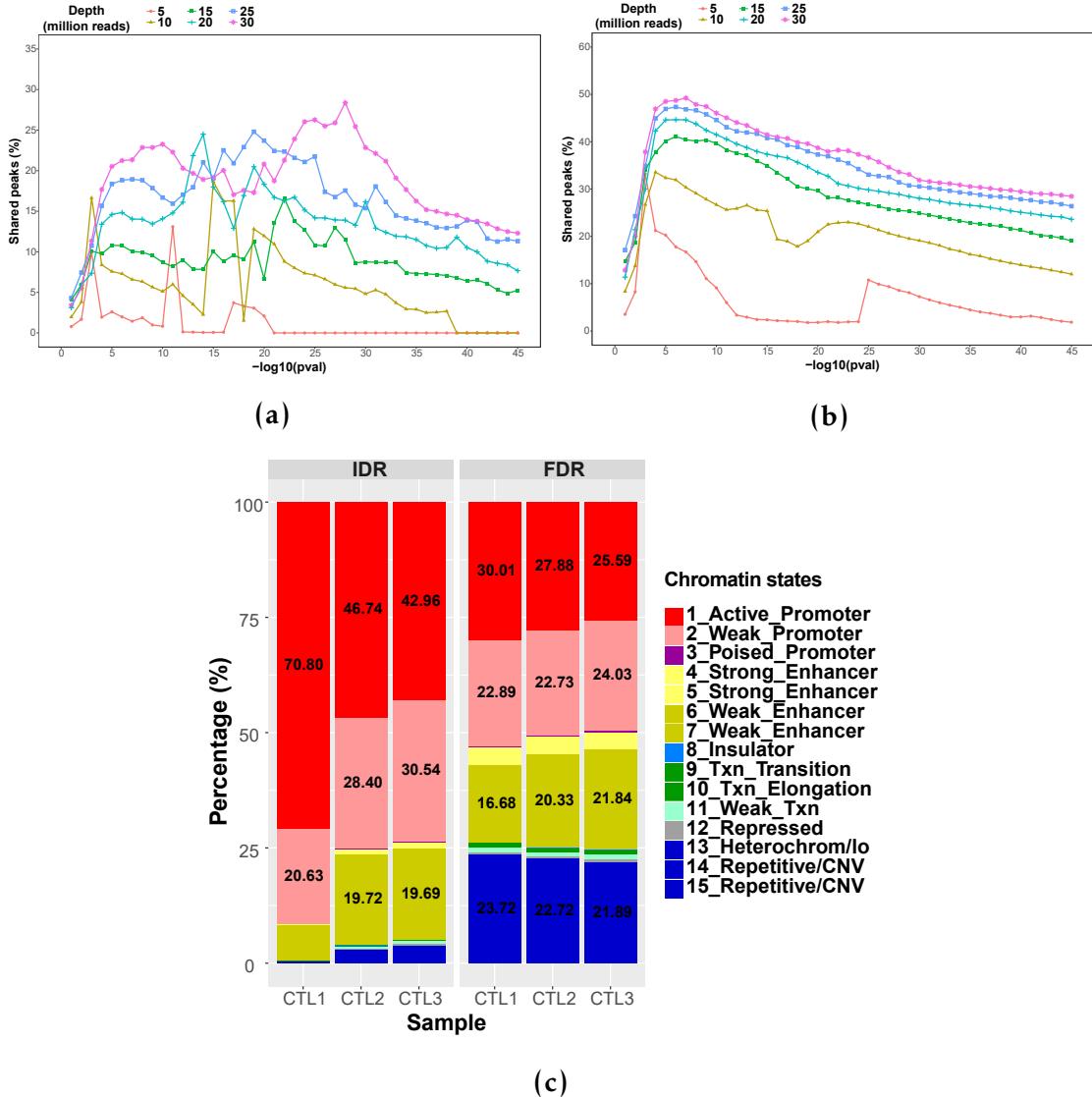


Figure 3.3: Peak calling filtering using IDR analysis in ATAC-seq samples. For each of the sequencing depths tested (from 5 to 30 million reads after filtering), illustration of the percentage of peaks sharing IDR rank between the two pseuroreplicates when using different pval filtering thresholds in CTL2 a) CD14⁺ monocytes and b) tCD4⁺ (used as representative samples differing in quality for this analysis). c) Annotation (as percentage of total) of the CTL1, CTL2 and CTL3 tCD4⁺ ATAC-seq peaks filtered for FDR<0.01 or optimal pval from the IDR analysis (pval=10⁻¹¹) with the corresponding cell type specific Epigenome Roadmap chromatin segmentation map.

Establishment of methods to assess genome-wide chromatin accessibility

background signal. Therefore, in collaboration with Dr Hai Fang, 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 (Table 3.1), two strategies were tested.

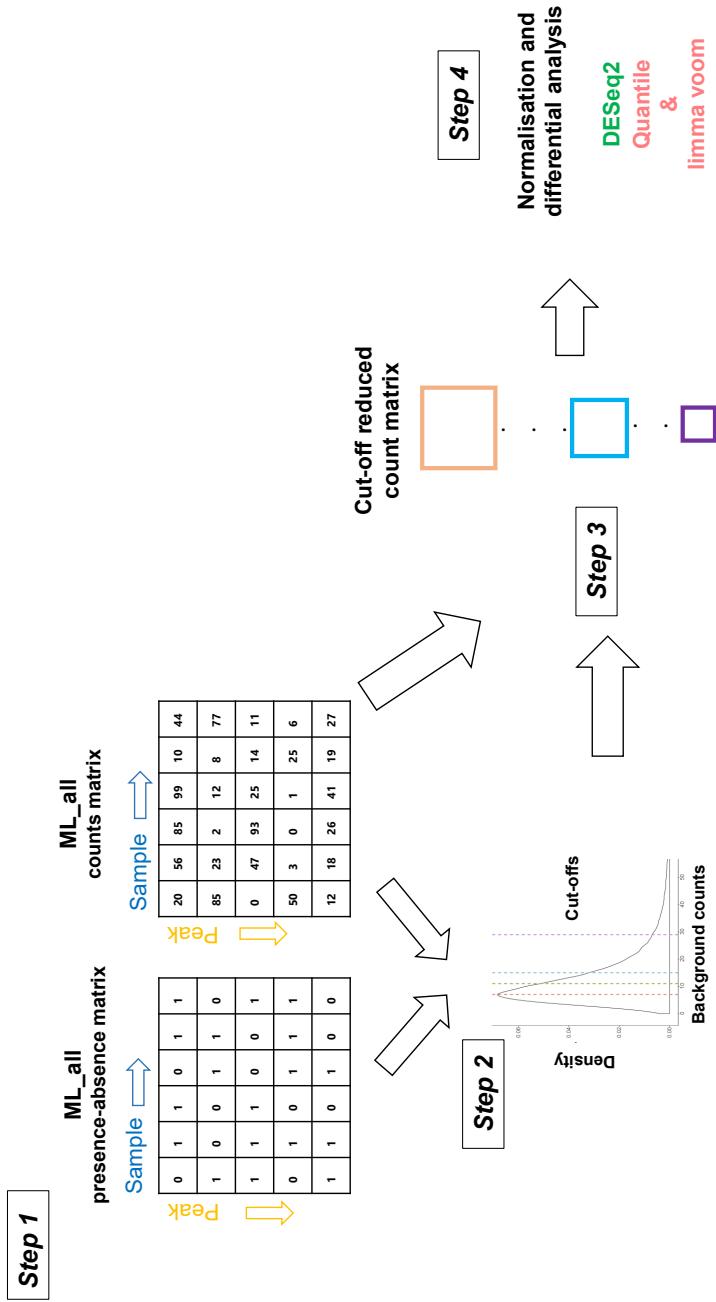


Figure 3.4: Work flow illustrating the strategy to account for ATAC background noise prior to differential analysis. A combined consensus peak master list (ML_all) was built as explained in Chapter 2, including all the significant peaks from each sample that were shared by at least 30% of the total number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

From the count matrix for the six samples defined above, the read counts from those peaks that were absent in each sample (since the ML_all includes peaks present in at least 30% of the total samples)(Figure 3.4 Step 1) were used to generate a density distribution plot (Figure 3.4 Step 2). From this plot, a sequence of twenty cut-offs were defined, with each representing the number of counts showed by a particular percentage of the total absent peaks (Figure B.1). Each cut-off was used to filter out peaks from the ML_all raw count matrix whose values in more than three samples were lower than the background counts (Figure B.1 Step 3). A filter of three samples were chosen, as it corresponds to the smallest group of replicates in this particular experimental design and ensures peaks absent in one condition were retained. As a result, each cut-off generates a reduced matrix of low-noise peaks that was normalised using quantile or DESeq2 (library size and variance stabilisation (Love et al. 2014)) and used to conduct differential analysis with limma voom or DESeq2, respectively. Both normalisation methods performed appropriately for all the reduced master lists across the six samples, with slightly greater consistency for the quantile normalisation across the two groups (Figure ?? a and b). Differential chromatin accessibility analysis using quantile normalisation counts& limma voom showed a greater number of significant ($FDR < 0.01$ and absolute FC ($\text{abs}(FC) > 1.5$) differentially accessible regions (DARs) compared to DESeq2, across all filtering cut-offs (Figure ?? c). The two approaches presented a progressive decrease in the number of DARs from the 75% cut-off onward, suggesting a reduction in the number of false positive hits reported for peaks with read counts close to the background noise cut-off. Further increases in the cut-off value however are expected to also remove true positives, so an intermediate value is chosen here. Depending on the noise inherent to an experiment this threshold may vary.

From this analysis, 80% was chosen as a conservative filtering cut-off, with the vast majority of the 19,855 DARs called as significant using the

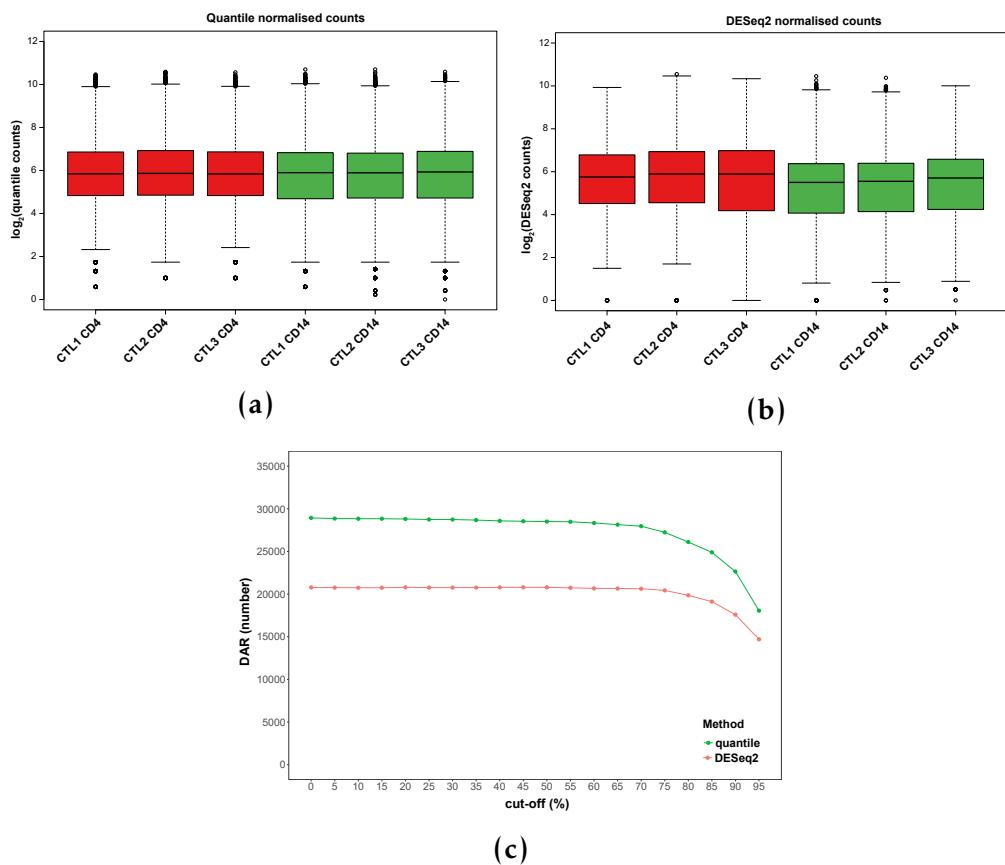


Figure 3.5: Normalisation and differential chromatin accessibility analysis for different cut-offs using quantile normalisation&limma voom and DESeq2. Boxplots representing the \log_2 read counts for each of the peaks from the unfiltered master list normalised by a) quantile or b) DESeq2 in the three $CD14^+$ monocytes and total $CD4^+$ healthy control paired samples. c) Representation of the number of significant DARs ($FDR < 0.01$ and $\text{abs}(FC) > 1.5$) 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 peak master list.

more conservative method (DESeq2) recapitulated by limma voom at the same significance threshold ($\text{FDR} < 0.01$ and $\text{abs(FC)} > 1.5$) (Figure 3.6 a). FDR rank revealed that out of the first 19,855 limma voom DARs 18,768 were the same as those retrieved by DESeq2. Moreover, very significant positive correlation was found between FCs for all the regions included in the 80% cut-off reduced matrix reported by the two methods ($R=0.999$, $p\text{-val}=2.2 \times 10^{-16}$) (Figure ?? b). 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.

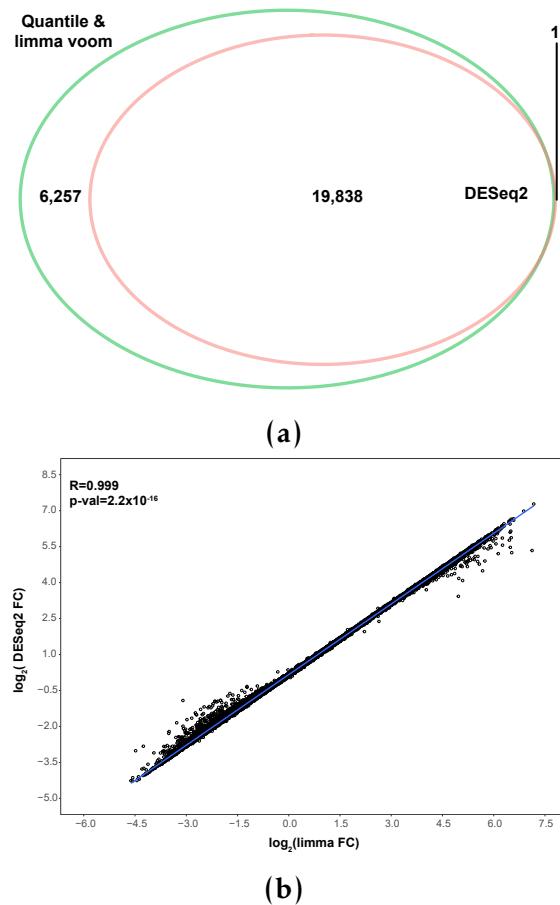


Figure 3.6: Comparison of the DARs identified by differential analysis using limma voom or DESeq for the master list filtered at the optimal cut-off 80%. a) Venn diagram illustrating the common and distinct significant ($\text{FDR} < 0.01$ and $\text{abs(FC)} > 1.5$) DARs identified by differential analysis in the filtered master list for the 80% optimal cut-off using limma voom or DESeq2. b) Representation of the correlation between limma voom and DESeq2 $\log_2\text{FCs}$ (no FDR filtered) in each the peaks from the filtered master list at the 80% optimal cut-off. Pearson correlation coefficient (R) and significance ($p\text{-val}$) are indicated. Limma voom was applied to quantile normalised count data.

Establishment of methods to assess genome-wide chromatin accessibility

Lastly, the significant DARs identified by DESeq2 and filtered for FDR<0.01 and abs(FC)>1.5 were divided in those more accessible in CD14⁺ monocytes (open in monocytes) or tCD4⁺ (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 DARs open in CD14⁺ monocytes included as the top hits for each of the categories T cell eRNAs, CD4⁺ H3Kme1 and H3K27ac and CD14⁺ DHSs (Figure 3.7). Conversely, the top enriched features for DARs open in monocytes included eRNAs, H3K27ac and DHSs in monocytes. Overall, this enrichment analysis confirmed the ability of this differential analysis method to identify significant and robust DARs that highlight cell-type specific regulatory regions for CD14⁺ monocytes and tCD4⁺ cells.

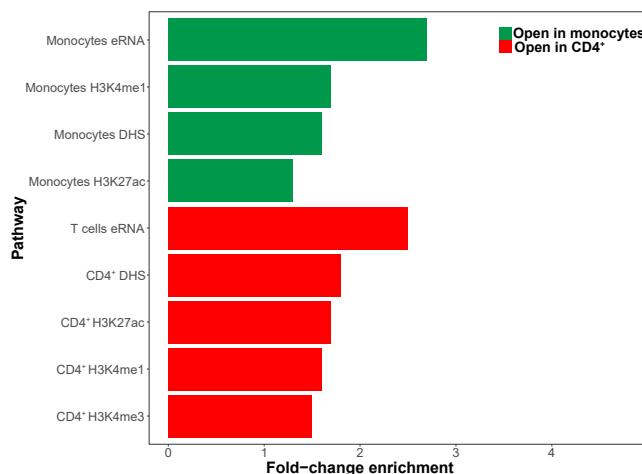


Figure 3.7: Enrichment analysis for the significant DARs identified by DESeq2 between CD14⁺ monocytes and tCD4⁺ cells. Barplot representing the FC for the top significantly enriched (FDR<0.01) FANTOM5 eRNAs and, histone marks and DHSs from Blueprint. Enrichment analysis was performed separately for the significant (FDR<0.01 and abs(FC)>1.5) DARs more accessible in CD14⁺ monocytes (open in monocytes) or tCD4⁺ (open in CD4⁺).

3.2.2 Assessment of ATAC-seq transposition times and comparison with Fast-ATAC protocol in relevant cell types

In addition to establishing an appropriate pipeline to analyse ATAC-seq data (NGS data processing, QC and differential analysis), the effect of the duration of the transposition reaction was evaluated in the main immune cell types of interest for this thesis. ATAC-seq was performed for three transposition times (20, 30 and 40 min) in CD14⁺ monocytes, tCD4⁺, tCD8⁺ and CD19⁺ cells and the impact on various ATAC-related measurements explored. All three transposition times produced appropriate fragment size distributions which recapitulated the nucleosome periodicity pattern (Figure 3.8 a). 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. In order to explore this effect in the different cell types, the ratio between NFR and nucleosome bound fragments ($\text{NBF} > 150\text{bp}$) was calculated (Figure 3.8 b). The relationship of this ratio with transposition time was heterogenous between cell types. For example, tCD8⁺ presented the greatest proportion of NFR for 20 min of transposition whereas the NFR/NBR reached a maximum at 40 min for the CD14⁺ monocytes. However, the change in this ratio across transposition times was very moderate in all cell types but tCD8⁺ cells.

Transposition times did not significantly affect the signal-to-noise ratio measured as fold-enrichment at the TSS, with the largest enrichment values corresponding to different times across the four cell types (Figure 3.8 c). For all cell types, 30 and 40 min yielded the ATAC-seq libraries with the largest enrichment, with the differences between the two being very moderate (not more than 4 units difference) across all cell types (data not shown). Before performing this formal comparison of transposition times, some sample recruitment had

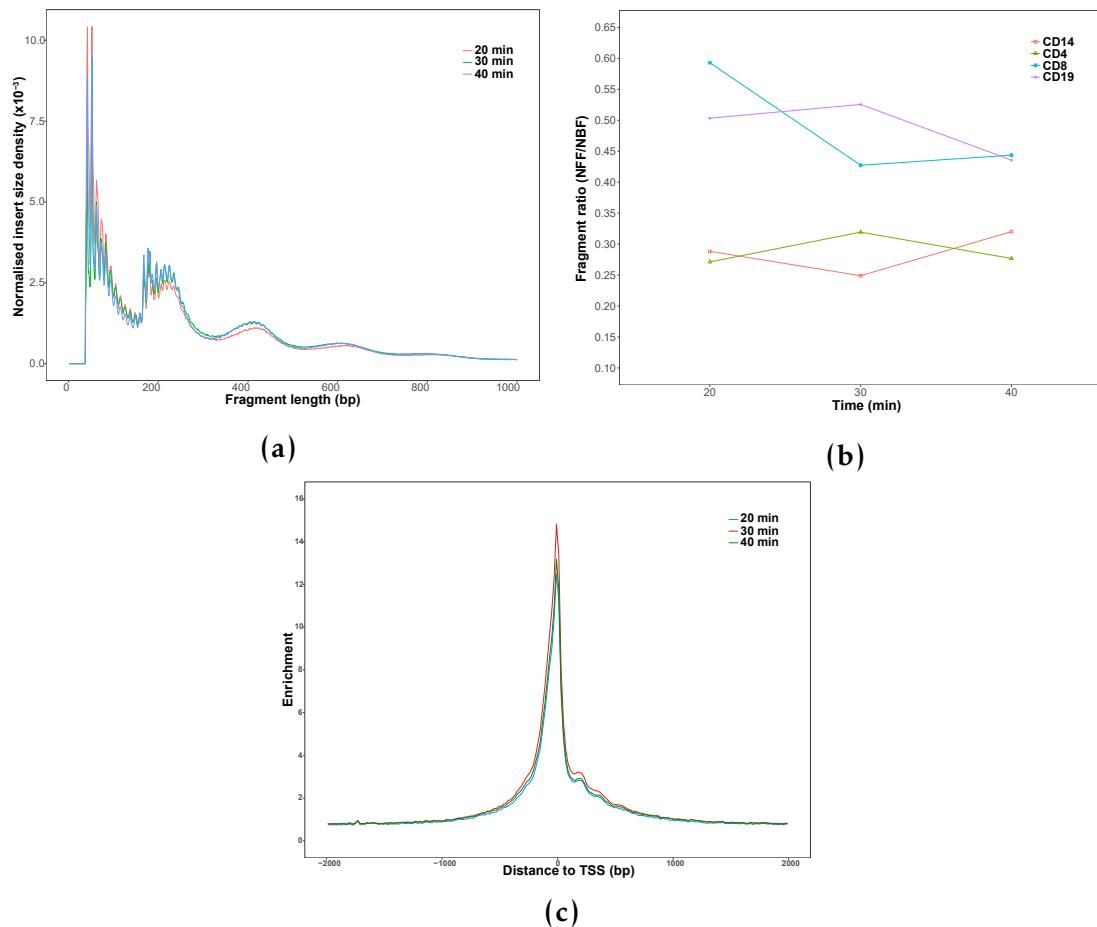


Figure 3.8: Assessment of the effect of transposition times on the ATAC-seq measurements

a) Representative plot of the ATAC-seq fragment sizes density distribution following 20, 30 and 40 min of transposition in healthy total CD8⁺ cells.

b) Changes in the ratio between nucleosome-free fragments (NFF) (fragments ≤ 150 bp) and long (> 151 bp) ATAC-seq nucleosome-bound fragments (NBF) across different transposition times in CD14⁺ monocytes, tCD4⁺, tCD8⁺ and CD19⁺ cells.

Establishment of methods to assess genome-wide chromatin accessibility

already been conducted using standard ATAC-seq with transposition for 40 min, as it was found to be the most appropriate condition based on the tapestation profiles from library QC (here not shown). Although in this analysis 30 min was the best condition for most of the cell types, the differences between 30 min and 40 min were minor and 40 min was used for the rest of patients samples processing using ATAC-seq.

In addition to testing different transposition times, comparison between the standard ATAC-seq and the improved Fast-ATAC protocol (Corces et al. 2016) was also conducted. The aim was to reproduce the two main improvements (reduction of MT reads and signal enhancement) from the Fast-ATAC publication before implementing the replacement of the ATAC-seq protocol, used at the time for patients' samples processing. Similarly to the data of Corces and colleagues, the percentage of MT reads in the Fast-ATAC libraries was lower for all the cell types compared to ATAC-seq (Figure 3.9 a). Importantly, the MT percentage in tCD4⁺ and tCD8⁺ cells showed a reduction from 50% to less than 20%. With regard to the background signal, opposing trends were observed across cell types (Figure 3.9 b). ATAC-seq TSS enrichment was improved by the Fast-ATAC protocol in CD14⁺ monocytes (6.9 to 23.2) and tCD4⁺ cells (14.8 to 17.9). The improvement was particularly relevant in CD14⁺ monocytes for this particular samples, where the TSS fold enrichment in ATAC-seq was just above the threshold cut-off of 6. In contrast, Fast-ATAC tCD8⁺ and CD19⁺ libraries presented a reduction when compared to the ATAC-seq libraries. When comparing across the different ATAC-seq samples generated up to that time, wide variation on TSS enrichment was observed and therefore no final conclusion on the signal improvement by the Fast-ATAC protocol could be drawn. Conversely, the large decrease in MT reads together with the reduced duration of the protocol supported the replacement of ATAC-seq by Fast-ATAC for future patient recruitments (see Chapter ??).

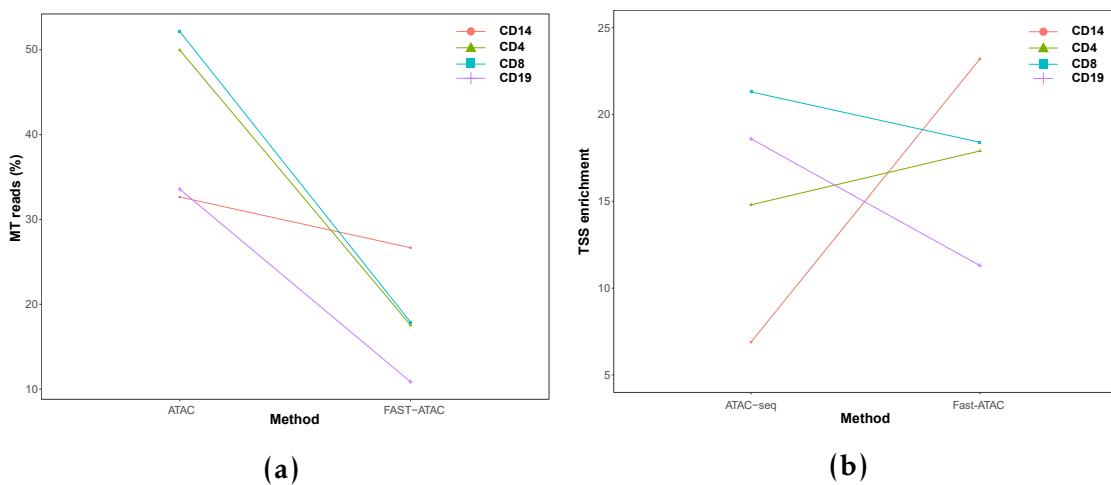


Figure 3.9: Differences in MT DNA abundance and TSS enrichment between ATAC-seq and Fast-ATAC protocols. Representation of changes in a) percentage of MT reads and b) TSS fold-enrichment between ATAC-seq and Fast-ATAC libraries for CD14⁺ monocytes, tCD4⁺, tCD8⁺ and CD19⁺ cells.

3.2.3 Limitations of ATAC-seq and FAST-ATAC to assess chromatin accessibility in KC

This project also aimed to characterise the regulatory landscape in KCs, one of the most relevant cell types in psoriasis pathophysiology. In order to determine the performance of the standard ATAC-seq protocol from Buenrostro *et al.*, 2013 (referred to ATAC 1 in this subsection), a cell suspension from a psoriatic lesional skin biopsy was generated and ATAC 1 was performed in 50,000 cells at two different transposition times (30 and 40 min). Since biopsy handling and lesional epidermal KCs are particularly challenging, this was considered the best system to test the performance of the standard protocol in the clinical setting of interest for the study. Library QC based on tapestation profiles for the two samples revealed expected DNA fragment sizes that recapitulated the characteristic nucleosome pattern every ~200bp generated by transposition of nucleosome-free and nucleosome-bound DNA (Figure 3.11 a). 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 3.11 b).

Establishment of methods to assess genome-wide chromatin accessibility

However the relative abundance of the mono-nucleosome fragments appeared to have approximately equal to the NFF, which is not observed in higher quality libraries. Regarding the signal-to-noise ratio, 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 3.11 c). Cell suspensions obtained from skin biopsies using trypsinisation of the epidermal layer are enriched in KC (~90%). However, they also contain significant amounts of dead cells and free-DNA released by apoptotic cells which could contribute to the increased background noise observed above.

Following Buenrostro and colleagues' ATAC-seq protocol, a modified version for KCs by Bao and colleagues (named ATAC-seq 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's paper, 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 adherent assay allows the isolation of viable undifferentiated KCs. 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 presence of the NFR and a poorly defined nucleosome pattern, particularly for the ATAC 2 protocol in both NHEKs and adherent KCs from skin biopsies (Figure ?? a). Similarly to the results in cell suspensions from

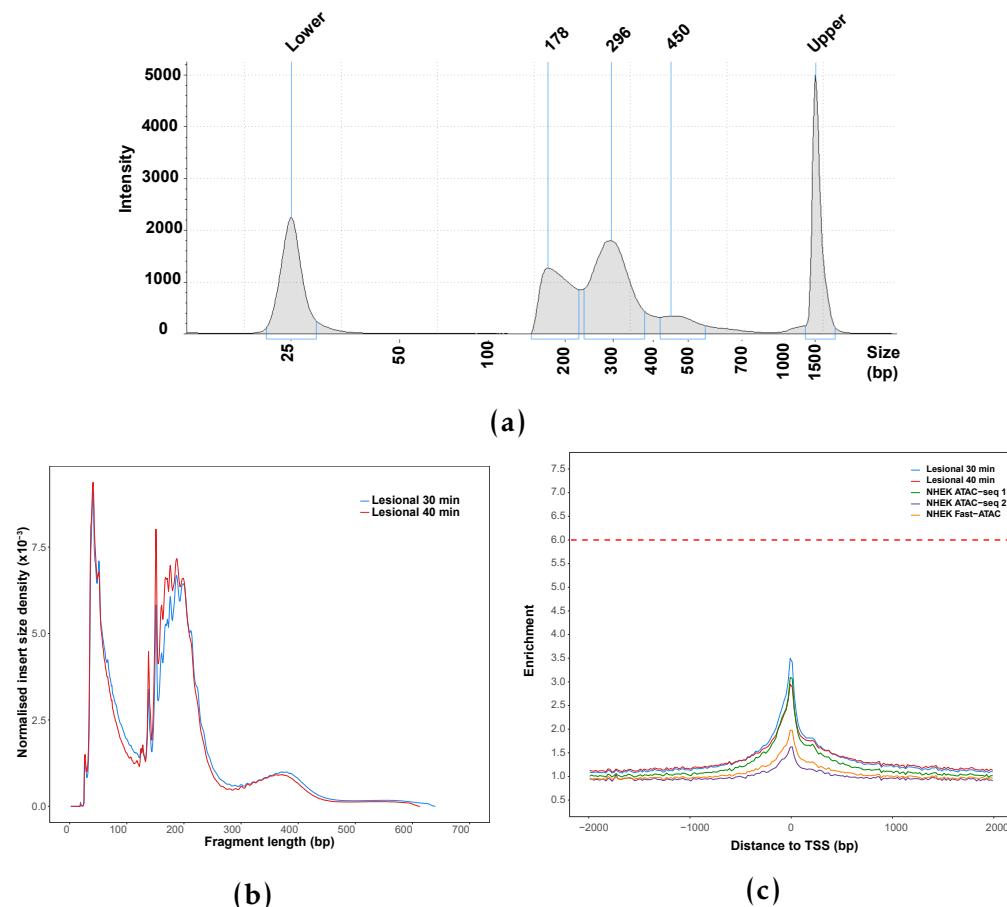


Figure 3.10: QC assessment of different ATAC protocols in psoriasis KCs and NHEKs. a) Pre-sequencing quantification of DNA fragment sizes in the ATAC libraries (tapestation profile) and b) the density distribution of sequenced fragments for ATAC 1 libraries generated in 50,000 KCs in suspension isolated from one psoriasis patient lesional skin. Two transposition times (30 and 40 min) were tested in the same sample KCs suspension and only the tapestation profile for the 30 min transposition has been included as the representative one. c) Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from the ATAC 1 psoriasis lesional KCs 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 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 time for all the different protocols was 30 min. (*) corresponds to the original Fast-ATAC conditions from Corces *et al.*, 2016.

psoriatic lesional epidermis, TSS enrichments were very low, particularly for the ATAC 2 protocol, and in all the instances under the acceptable cut-off in the two cell systems (Figure 3.11 c and Figure B.2).

Further optimisation of Fast-ATAC was performed modifying the original concentration of the NP-40 detergent and the Tn5 enzyme (Table 3.3). Library QC using tapestation profiles to assess the DNA fragment size distributions failed to show the nucleosome pattern profile expected in ATAC, thus the samples did not proceed to NGS (Figure B.3).

Towards the end of experimental work for this project, a new protocol known as 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 optimised for hematopoietic cells (Buenrostro 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 NFR followed by mono and di-nucleosome fragments (Figure 3.11 b). Moreover, high TSS enrichment values (approximately 20 fold) were observed for the two replicates (Figure 3.11 c). When performing overlap between the

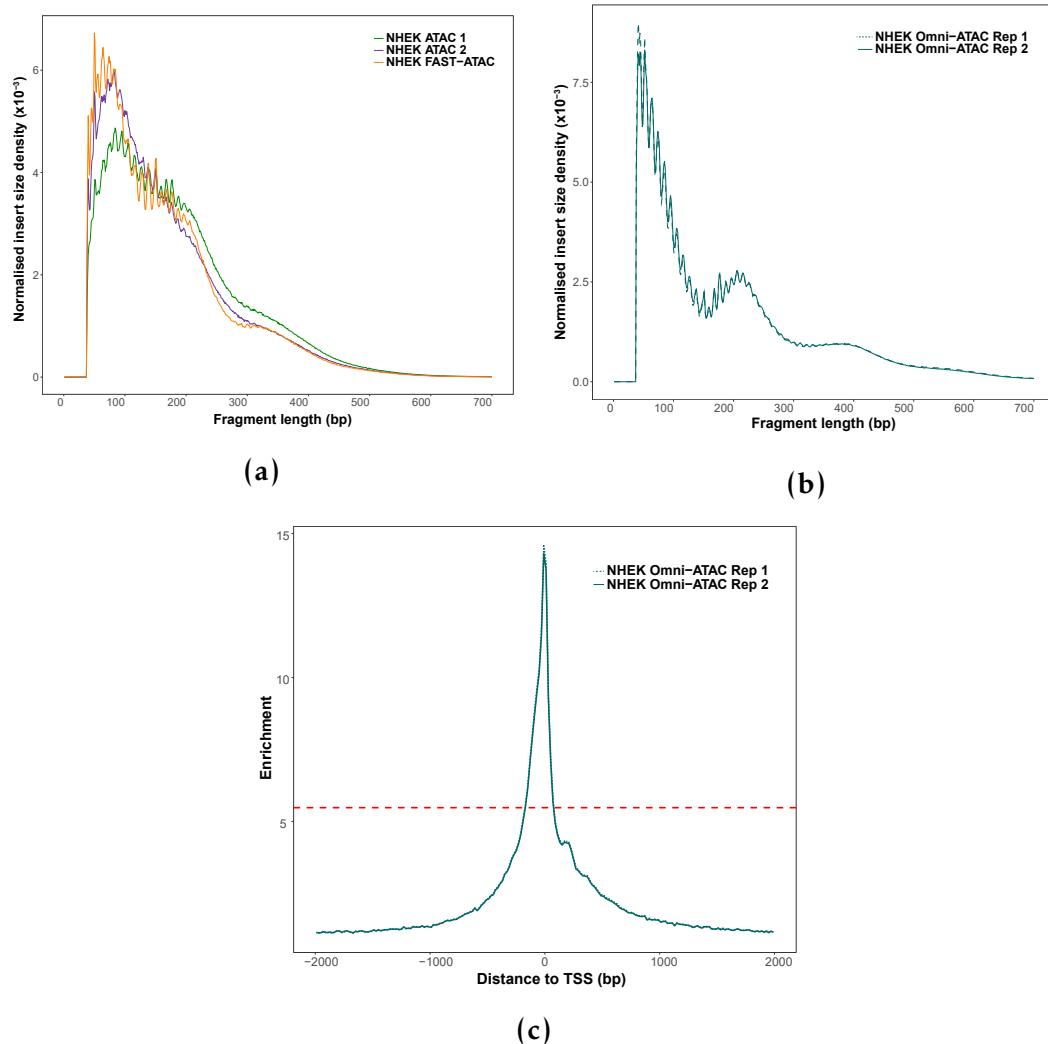


Figure 3.11: QC assessment of Fast-ATAC and Omni-ATAC in cultured NHEK. Representation of the fragment sizes density distribution in NHEKs libraries generated using a) ATAC1, ATAC2 and Fast-ATAC or b) Omni-ATAC protocols. c) Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from two Omni-ATAC technical replicates.

Omni-ATAC sample peaks filtered for a low stringency pval (pval=0.01) and the ENCODE DHSs from 125 cell types, the highest percentage (approximately 55%) was reported for NHEKs (Figure 3.12 a).

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 KCs, even at a very low pval filtering. The differences in the quality of ATAC signal between Omni-ATAC and the prior protocols was clearly observed at the chr17 locus harbouring a number of keratin genes, which are the main components of KCs' cytoskeleton (Figure 3.12 b). Omni-ATAC clearly presented the lowest background noise, the highest signal intensity and the greatest number of high quality peaks across the different keratin (KRT) genes when compared to the samples generated with the other ATAC protocols. Overall, this data was consistent with Corces *et al.*, 2017, where consistent successful results in NHEKs were shown, and it encourages future testing of Omni-ATAC in KCs from psoriasis patients biopsies processed through adherent assay to minimise the presence of dead cells.

3.2.4 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. In the context of this thesis two different approaches were of interest and a collaborative project to investigate these was established with High-Throughput Genomics at the WCHG. 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

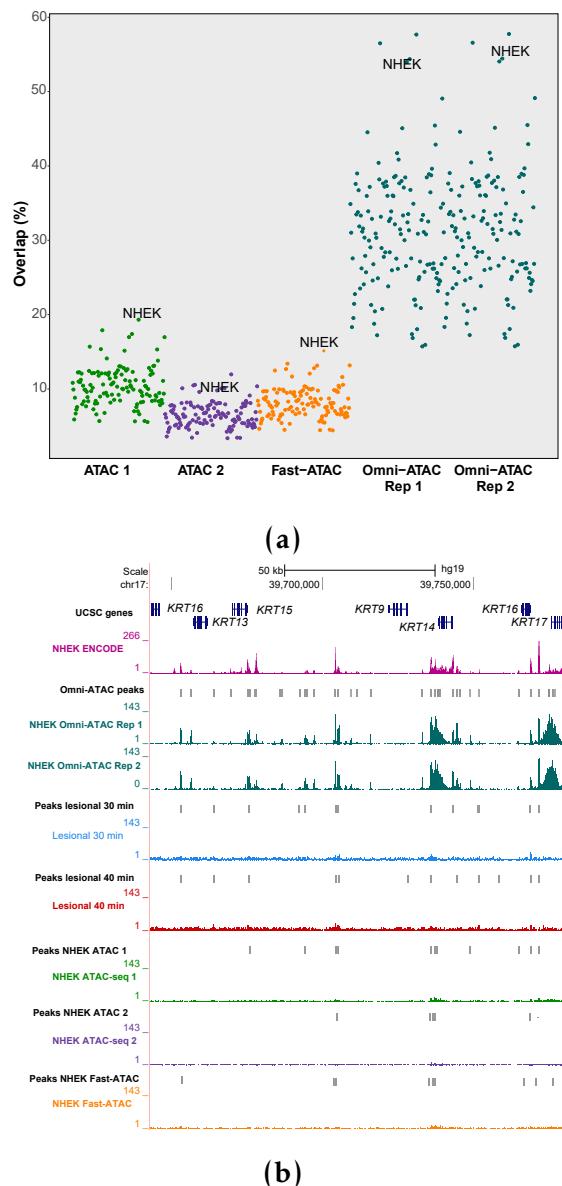


Figure 3.12: Comparison of the ENCODE DHSs overlap and signal density at the chr17 keratin family gene locus across different ATAC protocols. For the psoriasis lesional KCs and the NHEK libraries generated with the different ATAC protocols a) represents the percentage of overlap between significant peaks from each sample using low stringent pval and open DHS chromatin regions in 125 ENCODE cell types; and b) UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) at the chr17 locus (x-axis) containing several genes from the keratin (KRT) family.

Establishment of methods to assess genome-wide chromatin accessibility

investigated as a short term preservation method for FACS-isolated relevant cell types .

In order to investigate the performance of these two strategies, three healthy volunteers matched for sex and age were processed on different days to simulate the experimental design when using patient samples (Figure 3.13). PBMCs were isolated from blood and a fraction was stained with the appropriate panel of Abs to isolate CD14⁺ monocytes and tCD4⁺ T cells, as detailed in Chapter 2. ATAC-seq was performed on 50,000 CD14⁺ and tCD4⁺ freshly isolated cells (Figure 3.13 Day 0, ATAC-seq fresh). A fraction of the FACS-sorted CD14⁺ and tCD4⁺ cell were appropriately fixed with DPS, stored at 4°C for 24h and processed for ATAC-seq (Figure 3.13 Day 1, ATAC-seq fixed). Lastly, an aliquot of the 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, PBMCs were stained with the appropriate Ab panel and CD14⁺ and tCD4⁺ cells were isolated to perform ATAC-seq (Figure 3.13 Day 7, ATAC-seq frozen). For each control sample, three matched ATAC-seq libraries were generated in the two cell populations: ATAC-seq fresh, ATAC-seq fixed and ATAC-seq frozen.

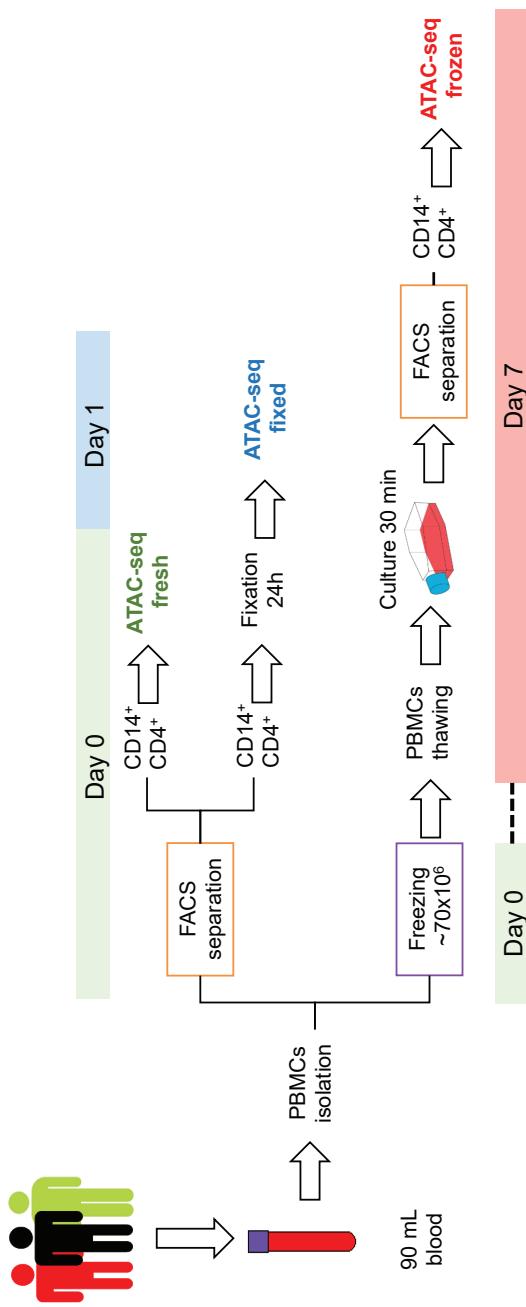


Figure 3.13: Experimental design to assess the impact of cryopreservation and fixation in the chromatin accessibility of immune primary cells. Three healthy control individuals were recruited on different days and PBMCs were isolated from 90mL of blood (Day 0), on Day 0, a fraction of PBMCs were used for FACS staining and isolation of 50,000 CD14⁺ and tCD4⁺, which were directly processed for ATAC-seq (ATAC-seq fresh). Also on Day 0, a 50,000 FACS-sorted CD14⁺ and tCD4⁺ cells were fixed with DPS, stored at 4°C for 24h and processed for ATAC-seq in Day 1 (ATAC-seq fixed). Lastly, on Day 0 a fraction of the PBMCs (70x10⁶ million cells) were cryopreserved in DMSO and slow-cooling. On Day 7 of storage in liquid nitrogen, PBMCs were thawed, recovered in culture for 30 min and stained with FACS Abs to isolate 50,000 CD14⁺ and tCD4⁺ cell to perform ATAC-seq (ATAC-seq frozen).

Chromatin structure characterisation in the different 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). For CD14⁺ monocytes the median reads across the fresh, frozen and fixed were more similar (58.6, 64.2 and 39.6 million reads, respectively) (Figure 3.14 a) than in the tCD4⁺ 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 3.14 b).

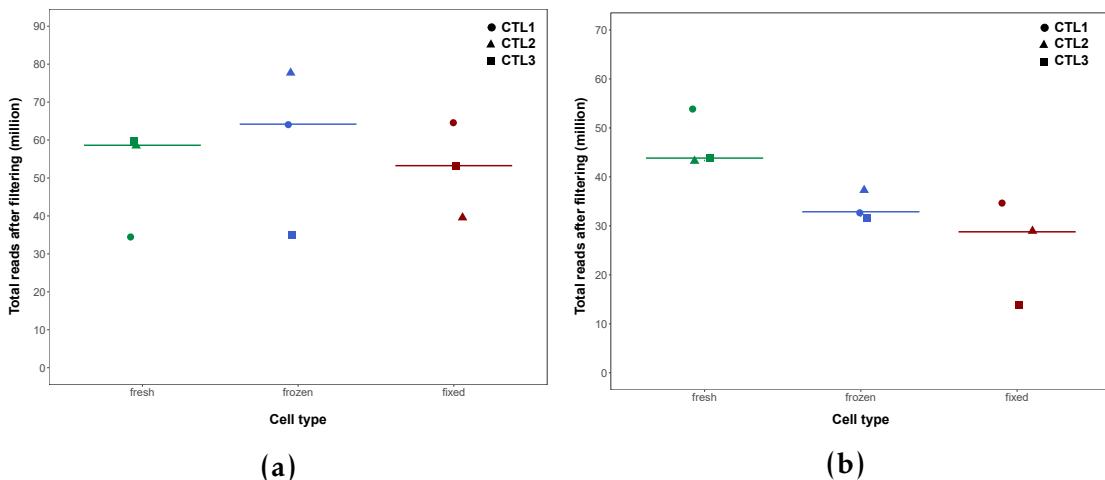


Figure 3.14: Total number of ATAC-seq reads for the fresh, frozen and fixed CD14⁺ monocytes and total CD4⁺ samples. Representation of million reads after filtering for the fresh, fixed and frozen ATAC-seq libraries in a) CD14⁺ monocytes and b)tCD4⁺ cells.

The ATAC-seq signal-to-noise ratios across the TSS presented a similar median for the fresh and fixed CD14⁺ monocytes libraries (17.4 and 16.5 fold-enrichment, respectively) and higher for the frozen samples(26.3 fold-enrichment)(Table A.1). The median TSS enrichments in the frozen and fixed tCD4⁺ 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. Interestingly, the CTL1 fixed samples of both cell types presented 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⁺ monocytes and tCD4⁺ from CTL1, were similar, showing NFR <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 appeared to show very similar distributions in both cell types as expected. In particular, fixed CD14⁺ monocytes and tCD4⁺ 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 ATAC-seq fixed previously highlighted.

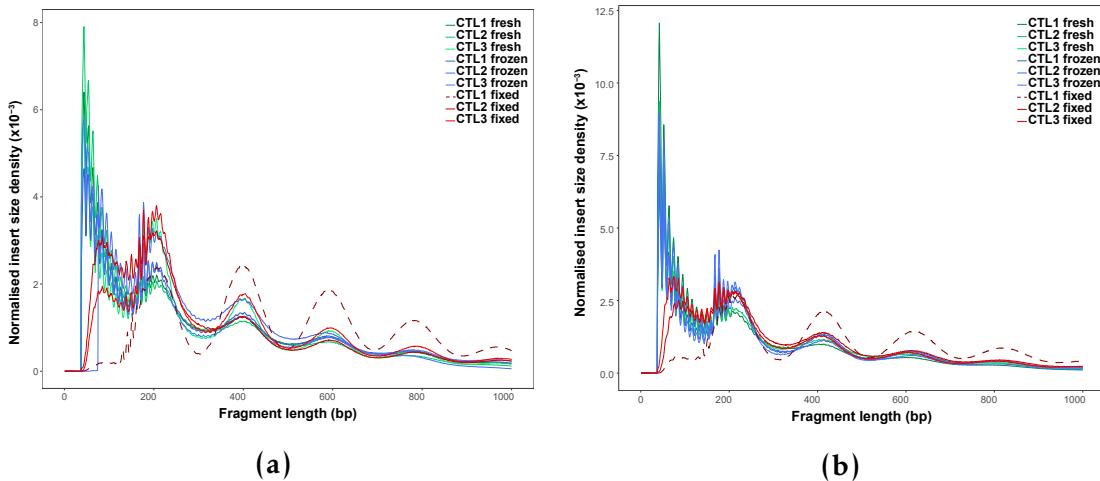


Figure 3.15: Fragment size density distribution for ATAC-seq fresh, fixed and frozen in CD14⁺ monocytes and tCD4⁺ cells. The distribution of ATAC-seq fragment lengths are illustrated for a) CD14⁺ monocytes and b) tCD4⁺ cells and colour-coded by condition (fresh=green, frozen=blue and fixed=red).

Study of maintenance of chromatin structure across and within the TSS was conducted as described in (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 tCD4⁺ fresh samples presenting the lowest enrichment (Figure 3.16 a 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 3.16 b and d).

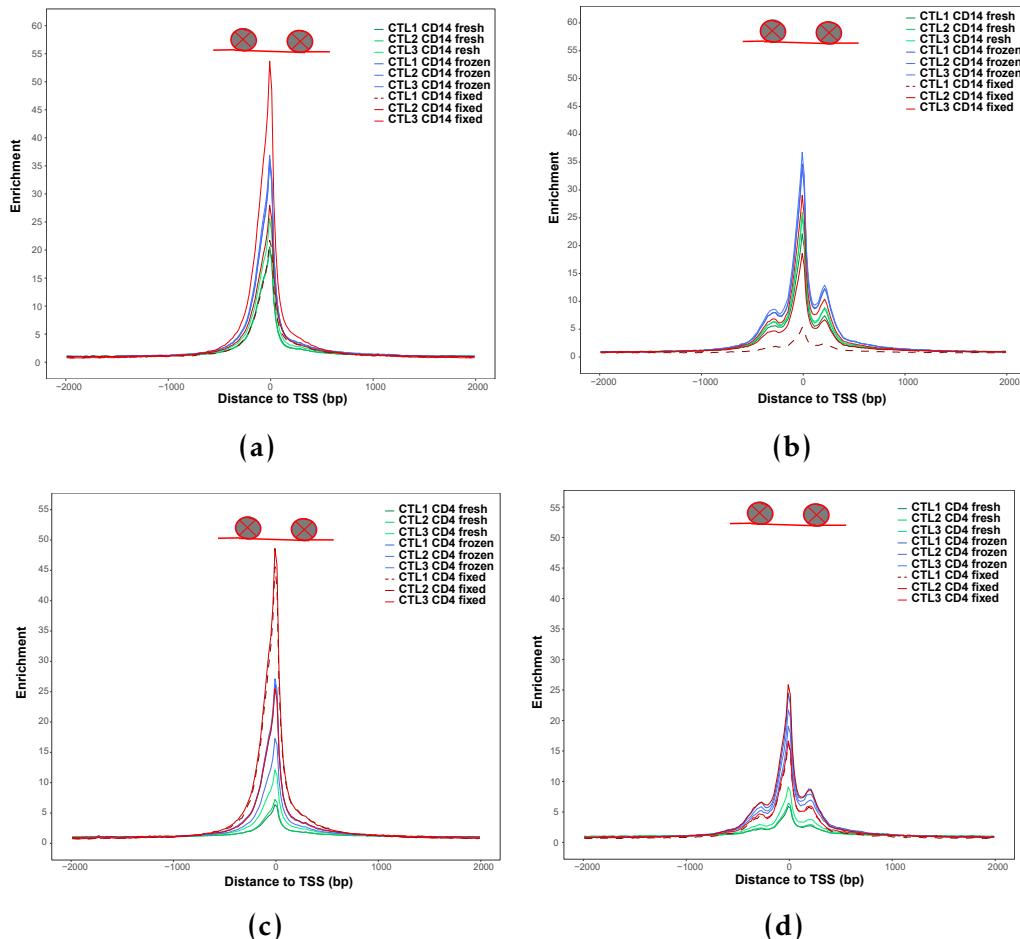


Figure 3.16: ATAC-seq enrichment of nucleosome-free and di-nucleosome fragments at the TSS and surroundings in CD14⁺ monocytes and tCD4⁺ samples for the three conditions. Nucleosome-free fragments (<150bp) and di-nucleosome (between 260 and 340bp) were selected *in silico* and enrichment analysis was carried out +/-1Kb across all the Ensembl annotated TSS.

Although fixation in CTL1 CD14⁺ 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⁺ 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⁺ 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. Altogether, 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 for the optimal IDR pval as explained in 3.2.1) demonstrated the greatest percentage of ATAC-seq peaks at promoters, introns and intergenic regions (Figure 3.17 a and b), consistent with previous studies (Buenrostro 2013; Scharer et al. 2016). The higher percentage of peaks annotated as promoters for some ATAC-seq libraries revealed the preferential location of strong good quality peaks at this feature when compared to other genomic features. Overall, the genomic annotation of the peaks across the three conditions revealed accessible ATAC-seq regions at meaningful genomic features.

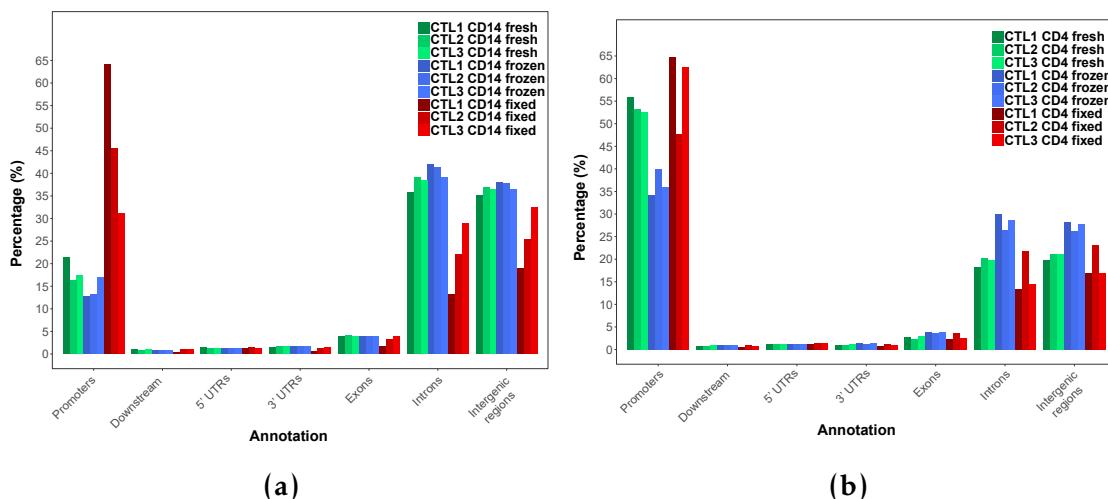


Figure 3.17: Genomic features annotation for the ATAC-seq peaks called in each of the fresh, frozen and fixed samples from CD14⁺ monocytes and tCD4⁺. Overlap was performed between the genomic features and the list of a) CD14⁺ monocytes and b) tCD4⁺ peaks filtered for FDR<0.01 in each sample from each 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 peaks from a consensus master list including the three conditions for each cell type (ML_CD14_all_cond and ML_CD4_all_cond). For this analysis CTL1 fixed samples from CD14⁺ and tCD4⁺ cells were removed from the analysis given the low quality and alterations in the chromatin structure previously described.

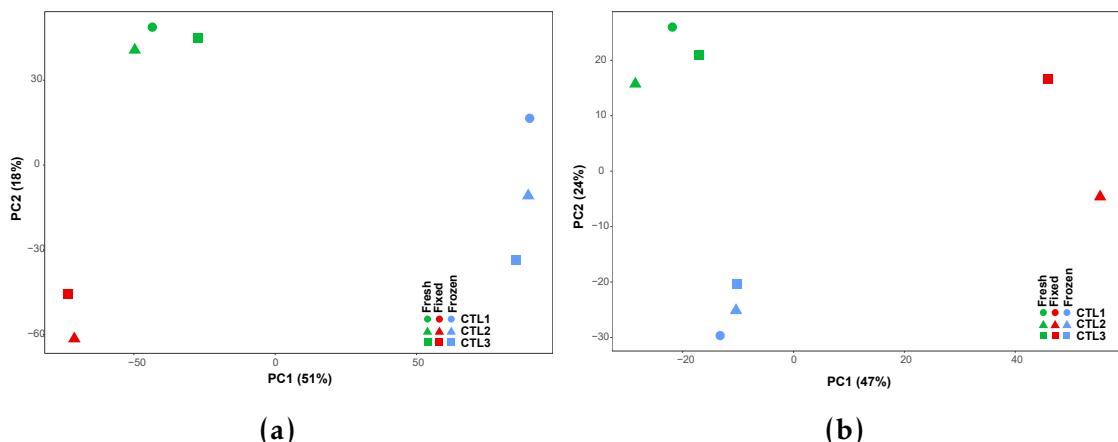


Figure 3.18: PCA analysis based on the ATAC-seq chromatin accessibility landscape in fresh, fixed and frozen samples. PCA analysis was performed using the normalised counts across the consensus master list of the combined fresh, fixed and frozen samples (ML_CD14.all_cond and ML_CD4.all_cond) in a) CD14⁺ monocytes or b) tCD4⁺ 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 master lists 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.

PCA analysis performed using the DESeq2 normalised counts for each of the two master lists demonstrated differences across the three ATAC-seq conditions in the two cell types. Plotting the first two principal components (PCs) from the PCA analysis 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⁺

Establishment of methods to assess genome-wide chromatin accessibility

monocytes (Figure 5.3 a). The second PC (16% variance) showed more moderate changes between fresh and fixed libraries. In contrast, PCA analysis in tCD4⁺ showed fixed to present the largest differences versus fresh in the genome-wide accessibility (Figure 5.3 b).

To further compare chromatin accessibility across the three conditions, the normalised read counts at each of the ML_CD14_all_cond and ML_CD4_all_cond peaks were contrasted between fresh and 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⁺ monocytes (Figure 3.19 a, b, c and d).

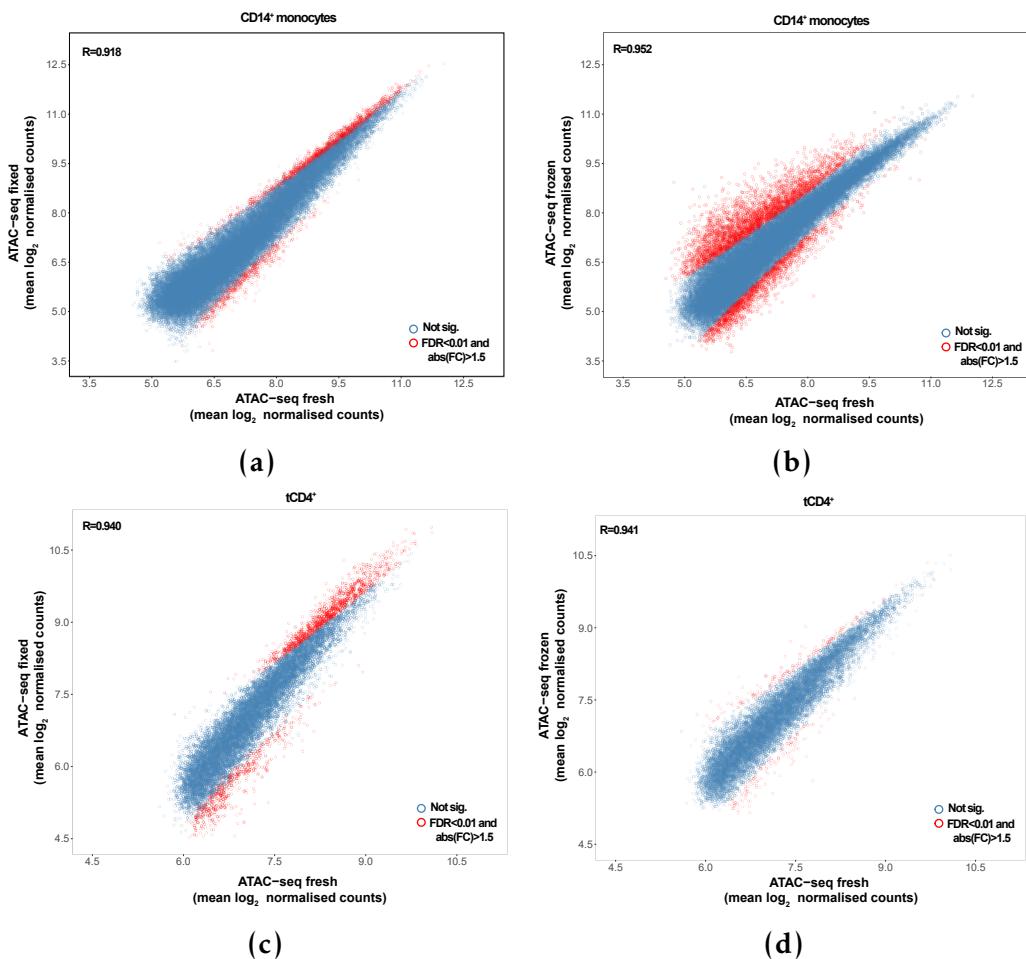


Figure 3.19: Comparison of the \log_2 normalised ATAC-seq counts at the consensus master lists peaks in fresh, fixed and frozen conditions. Each plot presents the comparison of ATAC-seq \log_2 mean normalised counts from the ML_CD14_all_cond or ML_CD4_all_cond filtered for background noise (80% empirical 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 $\text{abs}(FC)>1.5$) when performing differential chromatin accessibility analysis between ATAC-seq fresh and ATAC-seq fixed or frozen at each of the ML_CD14_all_cond and ML_CD4_all_cond regions using DESeq2. The number of significant DARs ($FDR<0.01$ and $\text{abs}(FC)\geq 1.5$) reported in each of the comparisons mirrored the PCA analysis results (Table 3.4). In $CD14^+$ monocytes, the largest number of DARs versus fresh were reported for the ATAC-seq frozen (5,269 regions in Figure 3.19 b). Conversely, in $tCD4^+$ the greatest differences in chromatin accessibility were found between fresh and

fixed ATAC-seq samples (1,564 DARs in Figure 3.19 d). 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 tCD4⁺ samples partially skewed the normalisation process, as can be observed by the location of the significant DARs in Figure 3.19 d. Altogether, the number of DARs identified in each comparison did not account for 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 ⁺ monocytes	5,269 (12.9%)	1,838 (4.5%)
tCD4 ⁺	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 abs(FC)>1.5) identified using ML_CD14_all_cond and ML_CD14_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⁺ 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.

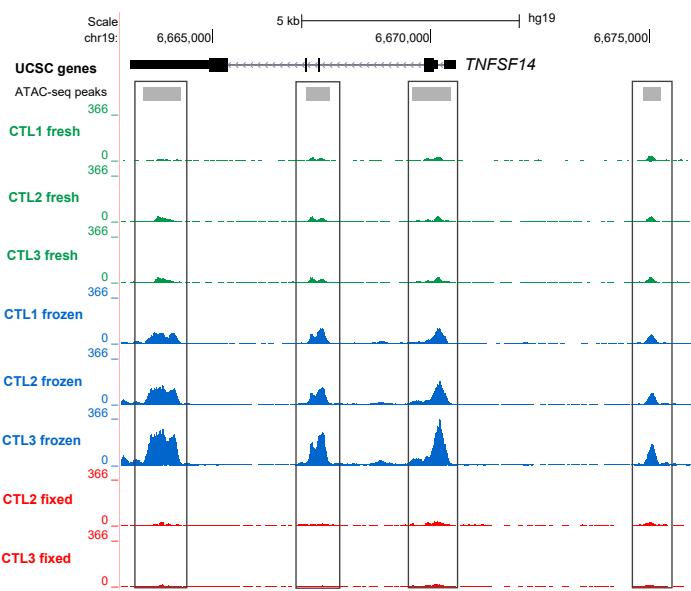


Figure 3.20: Differential chromatin accessibility at the *TNFSF14* gene between ATAC-seq fresh and ATAC-seq frozen in CD14⁺ monocytes. UCSC Genome Browser view illustrating the normalised read density (y-axis) at four significant (FDR<0.01 and abs(FC)>1.5) DARs (x-axis) within and upstream the *TNFSF14* gene in CD14⁺ monocytes. The four DARs were more accessible in ATAC-seq frozen when compared to ATAC-seq fresh. 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 novel technique and data analysis pipeline for the first time in the group. As such, an exhaustive evaluation of all possible methods was beyond scope of the project. Instead, to select appropriate methods for clinical studies, where sample availability and quality may be severely limiting, a number of alternative protocols, metrics and algorithms described in early ATAC experiments were evaluated in pilot experiments. This enabled the establishment a pipeline to be implemented in ?? and 5 for investigation of psoriasis and PsA chromatin landscape.

ATAC: technical aspects and pipeline establishment

The subsequent release of new ATAC protocols as well as alternative protocols such as THS-seq to assess chromatin accessibility in low number of cells

Establishment of methods to assess genome-wide chromatin accessibility

has confirmed some limitations of the ATAC-seq and also Fast-ATAC protocols. Quality assessment and variability across samples was difficult to detect through pre-sequencing library QC based on relative abundance of the different fragment sizes(tapestation profiles). Successful tapestation profiles showing nucleosome patters 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 QC measures beyond pre-sequencing library QC.

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 QC 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. 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 the ATAC-seq and Fast-ATAC libraries was also addressed at the peak calling level in this chapter, with the implementation of a peak filtering strategy that for each particular 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 confounder for downstream analysis. In terms of sequencing depth, analysis in this chapter showed 20 to 25 million reads after filtering to be the required minimum sufficient to identify an appropriate proportion of accessible

Establishment of methods to assess genome-wide chromatin accessibility

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 *et al.*, where IDR analysis used to evaluate consistency across replicates but not implemented for peak filtering.

Establishment of appropriate measurements for post-sequencing library QC allowed formally 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 as the most appropriate for all cell types according to pre-sequencing library QC (tapestation). Assessment of three different transposition times in the ATAC-seq protocol showed heterogenous impact on the ratio of NFF/NBF and overall no major impact on 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. Specifically, Fast-ATAC significantly reduced the percentage MT reads in all four cell types of interest for this thesis. In contrast, the improvement of signal-to-noise for hematopoietic cells claimed by the Fast-ATAC paper was not evident. In fact, Corces *et al.*, only showed improved TSS by Fast-ATAC. The publication of Omni-ATAC and a comprehensive comparison of the three ATAC protocols across a large number of cell types demonstrated that Fast-ATAC did not improve TSS fold-enrichment when compared to ATAC-seq in some of the hematopoietic cells, for example CD19⁺ cells, consistent with the finding in the systematic comparison using pilot data in this thesis.

The challenges of performing differential chromatin accessibility analysis

Until ATAC-seq release, limited research had been performed to investigate differences in chromatin accessibility, and mainly used data from cell lines (Degner *et al.* 2012). Studying chromatin accessibility in clinical samples first requires the definition of a consensus master list of accessible regions for which

Establishment of methods to assess genome-wide chromatin accessibility

no accepted method has been agreed. In this work, a master list 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 master lists or simply include all the significant peaks called in all the analysed samples (**Turner2018**; Alasoo et al. 2018).

When used for differential analysis, an additional filtering step to account for high read counts in peaks that were absent in some of the samples (background counts) has been implemented. 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 performed, 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 obtaining large number of high quality samples in a clinical setting, DEseq2 in combination with the additional filtering step to control to some extent for potential false positive appeared was chosen as an appropriately stringent method at the time study. As specific tools for ATAC analysis are developed, further comparison of the outputs will be of interest in future work.

3.3.1 Studying the chromatin landscape from psoriasis biopsies

At the time of writing, only RNA-seq studies have been performed in KCs from psoriasis skin biopsies. The relevance of KCs 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. Three different ATAC protocols (ATAC 1, ATAC2 and Fast-ATAC) performed very poorly in KCs isolated from skin biopsies and also in cultured NHEKs. The fact that similar results were obtained in KCs 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, KCs synthesis an insoluble protein structure that progressively replaces the plasma membrane, which may have been impairing appropriate cell permeabilisation and efficient transposition. Interestingly, Bao's protocol using increased Tn5 concentration to perform appropriately in NHEKS did not appear to improve ATAC quality libraries in my data. Similarly, the additional optimisation of the Fast-ATAC protocol modifying the concentration of detergent and Tn5 also failed to improve the quality of the data.

Release of the Omni-ATAC protocol and comparison of its performance with ATAC-seq and Fast-ATAC reproduced the failure to earlier two protocols to generate good quality data in KCs. However, the release of excellent performance Omni-ATAC in KCs has opened a real avenue to explore the chromatin landscape in lesional and uninvolved psoriasis biopsies for the continuation of this project.

3.3.2 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 WCHG 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⁺monocytes appeared to be more sensitive to cryopreservation than tCD4⁺ 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 immunature DCs have shown an 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.

3.3.3 Conclusions

The work described here has compared commonly used strategies from early 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 available samples whilst accounting for some quality limitations. The aim of this chapter was to establish methods that could be used (with the resources and expertise in the group available at the time) in a clinical setting. A comprehensive comparison of all possible methods recently published was therefore not performed. As ATAC has become a more commonly used technique,

Establishment of methods to assess genome-wide chromatin accessibility

new methods and updates have been introduced both in the lab and analysis side. The data presented throughout this thesis provides a resource for a representative study of the chromatin accessibility landscape using primary samples that could be revisited in the future with further optimised analytical methods. Moreover, the implementation of Omni-ATAC in future sample recruitment will further improve the quality and confidence in reported differentially open regions and is also enabling processing of historically formaldehyde-fixed chromatin recruited by the group.

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 components in 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 KCs, other circulating immune cells, such as T cells or DCs, are actively recruited to the site of inflammation contributing to disease initiation and progression (Leanne2012). 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 the impaired inhibitory function of circulating Tregs (Sugiyama2005; Kagami et al. 2010). Activated T cells isolated from psoriasis patients' blood have demonstrated their ability to induce skin lesions in xenotransplantation models of psoriasis (Wrone-Smith and clinical 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). Overall, these findings reinforce

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

there being a systemic component in 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 epigenetics field has opened an avenue to profile the epigenome of individual cell type populations in clinical samples, contributing to the interpretation and understanding of GWAS non-coding variants. ATAC-seq and ChIPm have enabled the interrogation of chromatin accessibility, histone modifications and TF binding using a few thousand cells (**Buenrostro2013; Schmid2015**). This has facilitated mapping the regulatory landscape in a wide range of cell types and tissues from clinical samples, providing details about the molecular programming of cells and the location and status of *cis*-regulatory elements in a disease-specific manner.

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 enriched for TFBS 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 global loss of chromatin accessibility in comparison to healthy tissue (Wang et al. 2018). Interestingly, treatment of retinal pigmented epithelium cells simulating cigarette smoking recapitulated the chromatin accessibility changes found in AMD, highlighting the value of investigating the chromatin landscape to characterise the effect of environmental stresses in complex diseases.

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

In addition to the study of 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 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 gene expression regulatory elements, assessing the active enhancer mark H3K27ac is of particular interest here. SNPs tagging all psoriasis GWAS loci showed enrichment for H3K27ac in twenty-six relevant cell types and/or tissues, with approximately 50% of variants in LD with those lead SNPs overlapping annotated enhancers in total CD4⁺ T cells (Lin et al. 2018). 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 the autoimmune disease juvenile idiopathic arthritis, where a disease-specific H3K27ac super-enhancer (those spanning up to 50Kb) signature has been identified in SF mCD4⁺ cells (Peeters et al. 2015). In addition to this, 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 for inflamed skin (lesional) using pre-lesional (uninvolved) skin, adjacent to the lesion, as the best internal control accounting for biological variability (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 KCs; C=control; L= psoriatic lesional skin; U=psoriatic uninvolved skin.

Author and year	Sample type and size	Technology	Description
(Jabbari2011)	SB (L=3, U=3)	RNA-seq microarray	and Technology discrepancies
(Li2014)	SB (L=92, C=82)	RNA-seq microarray	and Technology discrepancies and lncRNAs targets
(Keermann et al. 2015)	SB (L=12, U=12, C=12)	RNA-seq	co-regulation Dormant psoriasis signature and <i>IL36</i> expression in psoriasis skin
(Tsoi et al. 2015)	SB (L=97, U=29, C=90)	RNA-seq	Psoriatic skin-specific new lncRNAs
(Swindell and Genome 2015)	SB (L=14, U=14)	RNA-seq and mass-spectrometry	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

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

(Coda et al. 2012)	PBMCs (PS=6, C=5) and SB (L=5, U=5)	Microarray	Partial overlap between PBMCs and skin DEGs
(Lee et al. 2009)	PBMCs (PS=5, C=8)	Microarray	202 DEGs, circulating gene expression signature
(Mesko 2015)	PBMCs(PS=15, IBD=12,RA=12, C=18)	TaqMan customised array (96 genes)	6 psoriasis-specific DEGs
(Palau et al. 2013)	Activated CD4 ⁺ and CD8 ⁺ (PS=17, C=7)	Microarray	42 DEGs in T cell activation (<i>SPATS2L</i> and <i>KLF6</i>)
(Jung et al. 2004)	IL-10 stimulated PBMCs and CD14 ⁺ (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 in uninvolved skin (Table 4.1). Interestingly, discrepancies regarding the transcriptional similarities between normal and uninvolved skin have been identified, likely due to different FC filtering criteria (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 (Jaabari2011; Li2014). LncRNAs expression has also been proven to have a role in psoriasis pathophysiology, showing approximately 1,000 species differentially expressed between lesional and uninvolved skin (Tsoi et al. 2015). Interestingly, comparison of protein abundance and DGE in psoriatic skin has revealed that only 5% of the dysregulated transcripts present a similar trend at the protein level (Swindell and Genome 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 KCs (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.

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

Transcriptomics in circulating immune cells

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

A study conducted by Coda and colleagues explored the overlap between the DEGs in PBMCs (psoriasis versus controls) and the dysregulated genes when comparing lesional and uninvolved skin biopsies (Coda et al. 2012). The results revealed a limited overlap with more than 50% of the common genes presenting opposite directions of modulation in the two tissues. At the cell type specific level, some studies have performed *in vitro* culture and stimulation of T cells and monocytes (Palau et al. 2013; Jung et al. 2004). For instance, Palau and colleagues found forty-two DEGs enriched for cytokine and IFN (α , β and γ) signalling pathways when comparing activated CD4 $^{+}$ and CD8 $^{+}$ T cell from psoriasis patients and healthy controls. Further understanding of psoriasis-specific systemic gene dysregulation has also been approached through comparison with other chronic inflammatory diseases (Mesko2015).

4.1.4 Chromatin accessibility, gene expression and genetic variability

As previously explained, 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,

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

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

4.1.5 Fine-mapping using summary stats

The generation of cell type specific epigenetic maps can be used to inform statistical fine-mapping in the effort to identify putative causal SNPs to undergo functional validation (detailed in Chapter 1). Integration of Bayesian fine-mapping for twenty-one complex immune diseases performed by Farh and colleagues demonstrated greatest enrichment of fine-mapped causal variants in immune cell enhancer elements, particularly from activated conditions (**Farh2015**). 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 (Das et al. 2014).

Traditional Bayesian fine-mapping requires genotyping data from the GWAS cohorts to perform genotype phasing and imputation prior to association

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

analysis and calculation of 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. Since summary statistics from GWAS studies are widely available, methods like DIST have been developed to impute summary statistics instead of genotypes for the unmeasured SNPs in the study (Lee et al. 2013). In addition to this, summary statistics Bayesian fine-mapping methods using functional annotation as a prior in the model have also been developed. For example, the Risk Variant Inference using Epigenomic Reference Annotation (RiVIERA) method has been applied to perform fine-mapping for the Immunochip GWAS associated loci, incorporating in the model the forty-three ENCODE and Epigenome Roadmap annotation features showing greatest enrichment for psoriasis risk SNPs (Li and research 2016).

4.2 Aims

The aim of this chapter is to determine chromatin accessibility and gene expression differences between psoriasis patients and controls in four circulating immune cell types ($CD14^+$ monocytes, $tCD4^+$, $tCD8^+$ and $CD19^+$ cells) using ATAC, ChIPm and RNA-seq. The goal is to identify disease and cell type specific changes in putative regulatory regions and integrate them with dysregulation in genes 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 between psoriasis patients and healthy controls in immune cells isolated from PB.
2. To identify differences in the H3K27ac active enhancer mark between psoriasis patients and healthy controls in immune cells isolated from PB.

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

3. To determine changes in genes expression between psoriasis patients and healthy controls in immune cells isolated from PB.
4. To identify differentially expressed genes between lesional and uninvolved epidermis isolated from psoriatic skin biopsies.
5. To compare the dysregulation in the transcriptomic profile from circulating immune cells between patients and controls with the transcriptional differences from contrasting lesional and uninvolved epidermis.
6. To conduct fine-mapping analysis for a number of psoriasis GWAS loci using summary statistics.
7. To integrate 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 from GWAS risk loci.

4.3 Results

4.3.1 Psoriasis and healthy controls: cohort description and datasets

Peripheral blood (PB) samples were collected from a cohort of psoriasis patients and healthy individuals in order isolate four relevant immune cells types ($CD14^+$ monocytes, $tCD4^+$, $tCD8^+$ and $CD19^+$) and perform ATAC-seq, RNA-seq and ChIPm analysis. Additionally, the epidermis from paired uninvolved and lesional skin biopsies collected from three psoriasis patients were processed downstream for RNA-seq analysis.

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

Cohorts description

Psoriasis patients PB was collected as detailed in Chapter 2 from a total of eight psoriasis patients, six males and two females (Table 4.2).

Sample ID	Sex	Age at diagnosis	Disease duration (months)	PASI	Nails affected	Family history
Cohort 1A						
PS1011	Male	55	420	11	Yes	No
PS2014	Female	65	588	17	No	No
PS2015	Male	56	384	5	Yes	No
PS2016	Male	40	180	10	No	No
Cohort 1B						
PS2000	Male	61	156	10	No	Yes
PS2001	Male	56	432	10	Yes	No
PS2314	Male	42	120	6.5	Yes	No
PS2319	Female	64	372	10.2	No	Yes
Average	-	55	331.5	10	-	-

Table 4.2: Description and metadata of the psoriasis patients cohort. For each of the individuals information relating to sex, age at the time of sampling, disease duration, PASI score, affection of nails and family history has been included. Patients are divided into cohort 1A and cohort 1B based on the batch of ATAC-seq and RNA-seq processing. PASI evaluates the percentage of affected area and the severity of redness, thickness and scaling for four body locations (as detailed in Table ??). 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 as the sum of the clinical signs severity scores for that region multiplied by score of that percentage affected area and the proportion of body surface represented by that body region (0.1 for head and neck, 0.2 for upper limbs, 0.3 for trunk and 0.4 for lower limbs. 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).

The average age of the cohort was 55 years old and the average disease duration 331.5 months. All the patients presented active skin disease and none of them had reported joint affection at the time of sample collection. Disease severity was quantified using the PASI score, previously reviewed in Chapter

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

1, with the average cohort score being 10. Currently, there is no consensus on PASI thresholds to define mild and moderate-to-severe disease. A review study regarding the use of PASI as an instrument to determine disease severity of chronic-plaque psoriasis have suggested considering psoriasis as moderate when PASI ranges between 7 to 12 and severe for PASI>12 (Schmitt and Dermatology 2005). On the other hand, NICE and other studies had 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. Only two of them presented PASI<7 showing a mild psoriasis phenotype. All patients were naïve for biologics therapies. PS2319 was currently on MTX therapy and the remaining patients had only been treated occasionally with topical steroids or UVB therapy. Interestingly, PS2014 presented the most severe PASI score (17) and was a non-responder to MTX. 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 (**Moll1976; McGonagle**; Griffiths and Lancet 2007; medicine 2011). Psoriasis family history was reported by PS2000 and PS2319, which could indicate those individuals are HLA-C*06:02 positive, upon missing genotyping data. In addition to the psoriasis samples, PB blood was collected from ten sex and age-matched healthy individuals (Table 4.3)

For both patients and controls, the subdivision into cohort 1A and cohort 1B relates to the batches in which ATAC-seq and RNA-seq samples were processed and sequenced.

Datasets

For both cohorts, ATAC-seq and RNA-seq data were generated from CD14 $^{+}$ monocytes, tCD4 $^{+}$, tCD8 $^{+}$ and CD19 $^{+}$ cells (Table 4.4). For cohort 1A ATAC-seq data was generated using the standard ATAC-seq protocol from Buenrostro *et*

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

Sample ID	Sex	Age
Cohort 1A		
CTL1	Male	36
CTL2	Male	53
CTL3	Male	34
CTL4	Female	46
CTL5	Male	42
Cohort 1B		
CTL6	Male	31
CTL7	Male	57
CTL8	Female	50
CTL9	Male	50
CTL10	Male	67
Average	-	46.6

Table 4.3: Description of the healthy control cohort. Controls are divided in cohort 1A and cohort 1B based on the batch of ATAC-seq and RNA-seq processing, similarly to the psoriasis patients samples.

al., 2013, which was replaced by the FAST-ATAC method (Corces et al. 2016) in cohort 1B, due to the improvements of this protocol as explained in Chapter 3. Additionally, samples from cohort 1B were also processed to assess differences in H3K27ac modification between patients and controls using ChIPm (Table 4.4).

For three of the psoriasis patients (PS2014, PS2015 and PS2016) paired biopsies from lesional and uninvolved skin were collected and the epidermal sheets were isolated to perform RNA-seq differential analysis (Table 4.4). This should be considered as 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 and cost constraints.

In order to assess the chromatin accessibility landscape in psoriasis and control individuals, ChIPm and ATAC-seq were performed in CD14⁺ monocytes, tCD4⁺, tCD8⁺ and CD19⁺ B cells isolated from PB. ChIPm data was only

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

Technique	Cohort or samples ID	Sample size (patient/control)
ATAC-seq	Cohort 1A and 1B	8/10
RNA-seq	Cohort 1A and 1B	8/10
ChIPm	Cohort 1B	4/4
Skin RNA-seq	PS1011, PS2015 and PS2016	3/0

Table 4.4: Datasets generated for the psoriasis and control cohort samples. Cohort identity includes both, patients and controls. The skin RNA-seq samples include lesional and uninvolved paired-skin biopsies from each of the three individuals.

generated for cohort 1B whereas ATAC-seq was performed for all samples (Table 4.4).

4.3.2 Assessment of changes in the enhancer mark H3K27ac in psoriasis immune cells

Data processing and quality control

A total of 32 ChIPm libraries from four patients and four controls in four cell types were sequenced, and reads filtered as detailed in Chapter 2. 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 4.1 a). As part of the quality control, library complexity for each of the samples was measured by the NRF and the PCR bottlenecking coefficients PBC1 and PBC2. According to the ENCODE standards, most of the libraries had appropriate complexity and moderate to mild bottlenecking (Table A.2). The CD8⁺ CTL7 and the CD19⁺ PS2000 and PS2314 libraries failed the recommended complexity NRF values and also had more severe PCR bottlenecking according to the PBC1 coefficient threshold. These observations were consistent with the greater number of duplicated reads identified in these libraries compared to the rest (>50% of the total sequenced reads) and consequently lower number of reads after filtering (Figure 4.1 a).

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

Cross-correlation analysis was performed to determine the NSC and RSC coefficients, which provide a measure for the signal-to-noise ratios in the samples. All the ChIPm libraries presented appropriate signal-to-noise following ENCODE standards (Landt et al. 2012), with NSC and RSC values equal or greater than 1.05 and 0.8, respectively (Figure 4.1 b and c). Interestingly, the CD14⁺ monocytes and tCD4⁺ ChIPm libraries had lower signal enrichment compared to the tCD8⁺ and CD19⁺ libraries, which correlates with the cell type grouping during sample processing. Following QC, the tCD8⁺ CTL7 and the CD19⁺ PS2000 and PS2314 libraries were removed for downstream analysis.

PCA analysis using a combined master list of the H3K27ac enriched sites in patients and controls for all four cell types (excluding the aforementioned low quality samples) confirmed the ability of this data to recapitulate the cell type specific epigenetic landscape of this enhancer mark and reinforced the appropriate quality of the data (Figure 4.2 a). When performing PCA analysis by cell type, the PS2314 CD8⁺ library appeared as an outlier compared to the rest of CD8⁺ H3K27ac ChIPm patients and control libraries and was also removed for downstream analysis (data not shown).

Differential H3K27ac enrichment analysis

Differential H3K27ac analysis was performed between psoriasis and healthy control samples for each cell type using DiffBind. The consensus merged list of H3K27ac sites assembled by this algorithm to perform the differential analysis (as explained in Chapter 2) included a high percentage of sites annotated as heterochromatin or repetitive (Figure 4.2 b), ranging from 20.8% in CD14⁺ monocytes to 51.39% in tCD8⁺ cells. Such sites are less likely to be relevant since H3K27ac is a histone modification mainly enriched at enhancers. When restricting the differential analysis to those regions annotated as enhancers (weak and strong), CD14⁺ monocytes appeared as the cell type with the greatest

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

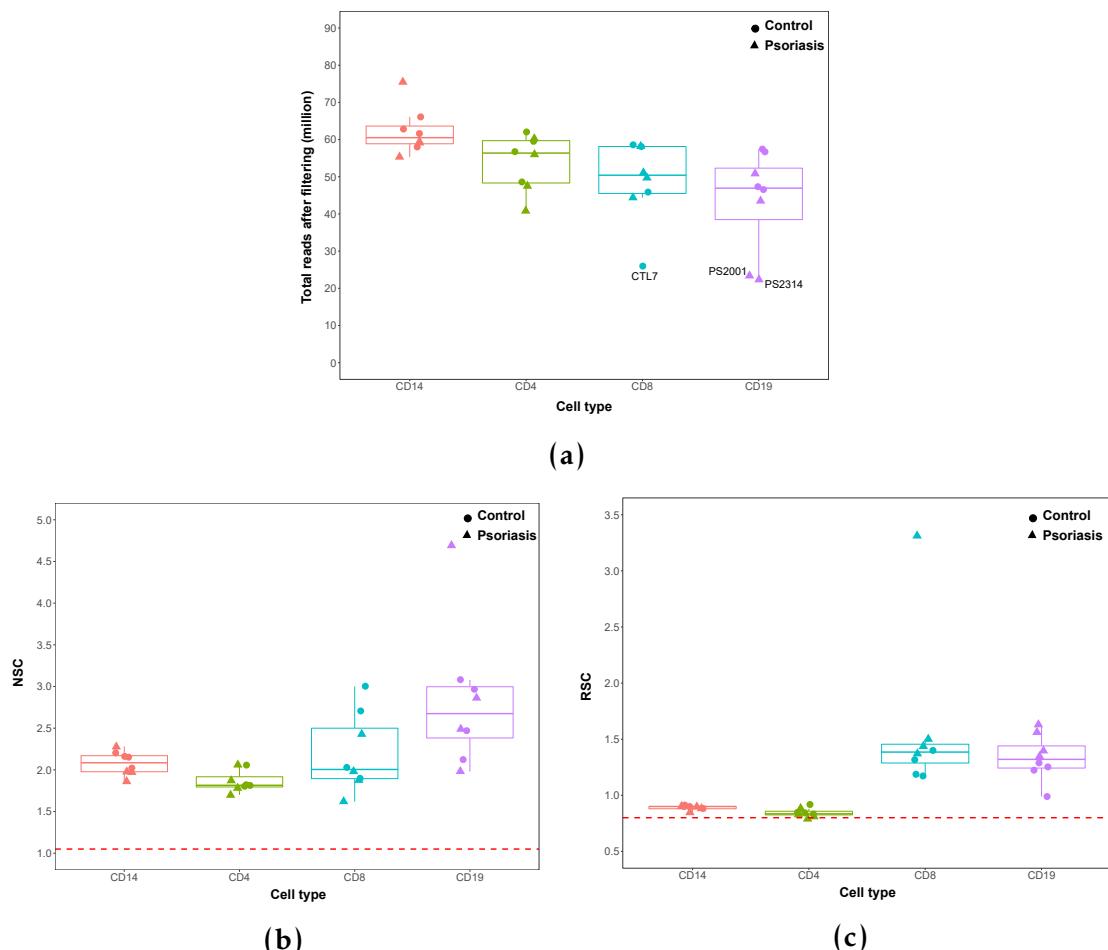


Figure 4.1: Quality control evaluation of the H3K27ac ChIPm libraries in immune cells isolated from psoriasis and control samples. For each of the cell types boxplots representing a) million of reads after filtering, b) normalised strand cross-correlation coefficient (NSC) and c) relative strand cross-correlation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

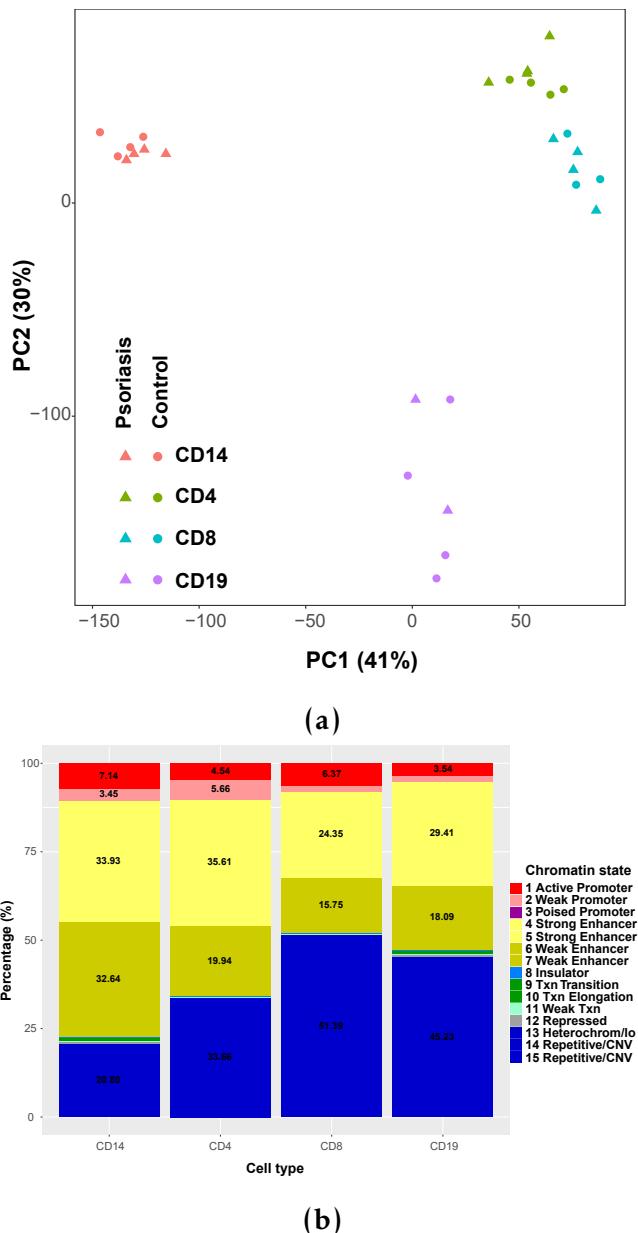


Figure 4.2: PCA analysis and chromatin annotation states of the H3K27ac enriched sites in four immune primary cell types from psoriasis and healthy control samples.

a) PCA analysis was performed using the normalised counts across a consensus master list of the combined H3K27ac enriched regions in psoriasis patients and healthy control samples across CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ cells. The first two PCs (x-axis and y-axis, respectively) for all the H3K27ac ChIPm peaks in the master list are plotted. 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 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

number of differentially modified enhancers (8 significant sites), followed by CD4⁺ (4) and CD8⁺ (1) (Table 4.5).

Cell type	Master list size genome-wide/enhancers	Differential regions genome-wide/enhancers
CD14 ⁺	99,862/60,962	15/8
tCD4 ⁺	110,353/56,282	0/4
tCD8 ⁺	137,194/51,607	8/1
CD19 ⁺	199,014/88,722	12/0

Table 4.5: Summary results from the differential H3K27ac analysis between psoriasis patients and healthy controls in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ cells. In the genome-wide analysis, the master list size refers to the number of H3K27ac enriched sites included in the consensus list built using DiffBind to perform the differential analysis. In the analysis restricted to enhancers, the size of the master list was reduced to only those sites from the genome-wide master list annotated as enhancers (weak and strong) according to the chromatin segmentation map for each particular cell type. Genome-wide significant sites in CD14⁺ monocytes and CD8⁺ also contain the sites identified in the enhancer restricted analysis. Significant differentially H3K27ac modified regions were determined using FDR<0.05 and no FC threshold.

One of the differentially H3K27ac modified regions when comparing patients versus controls in CD14⁺ monocytes was located between the *SLC15A2* and *ILDR2* genes, where *ILDR2* has recently been identified as relevant for negative regulation of T cells response in RA (Hecht et al. 2018). This region presented lower H3K27ac levels in psoriasis patients compared to controls and was annotated as enhancer by the Epigenome Roadmap chromatin segmentation map (Figure 4.3). Additionally this site was overlapping a DHS and H3Kme1 (enhancer mark) modification and a CTCF-binding site identified by ChIP-seq in K562 cells. Publicly available chromatin conformation data in T cells failed to show interactions of this region with the promoter of any of the proximal genes in monocytes and eQTL data has suggested regulation of the calmodulin-binding motif-containing protein *IQCB1* gene in monocytes and tCD4⁺ cells (GTeX; Fairfax et al. 2014; Kasela et al. 2017). Amongst the tCD4⁺ hits, all the significant regions presented minor changes with no evidence of regulating expression of any relevant gene by proximity or based on chromatin conformation data.

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

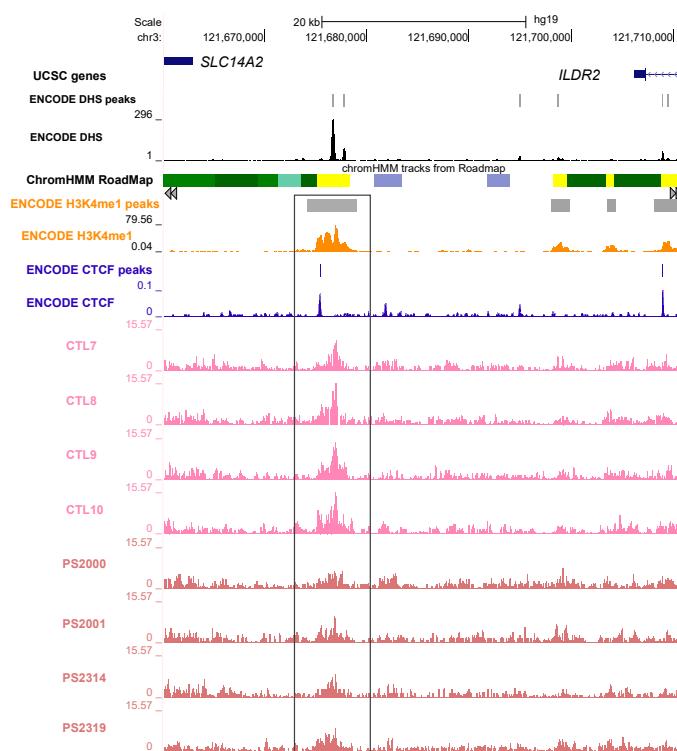


Figure 4.3: Differential H3K27ac modification at a putative intergenic enhancer region in circulating CD14⁺ monocytes between psoriasis patients and healthy controls. UCSC Genome Browser view illustrating the normalised H3K27ac fold-enrichment (y-axis) at an intergenic differentially modified region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

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

When performing genome-wide differential analysis, CD14⁺ monocytes and CD8⁺ revealed additional statistically significant differentially H3K27ac modified regions, as well as those already identified in the restricted enhancer analysis (Table 4.5). The newly identified regions in both cell types were mostly in regions lacking DHS and H3K4me1 modifications. Genome-wide analysis in CD4⁺ cells did not reveal significant differentially modified targets outside enhancers and also failed to retain significance for the four hits identified in the restricted analysis, most likely due to increase in multiple testing as a result of the larger size of the master list (Table 4.5). The differential regions between psoriasis patients and controls identified in CD19⁺ cells when performing genome-wide analysis presented considerable fold changes (absolute log₂FC ranging from 2.47 to 6.37). However, most of them (10 out of 12) did not overlap accessible chromatin and none were found to be interacting with other enhancers or distal promoters. The absence of differentially H3K27ac sites between patients and controls at CD19⁺ enhancers may reinforce the limited relevance for B cells in psoriasis when compared to the other three cell types.

Overall, restricting the differential analysis to enhancer annotated regions did not show a great increase in the number of significant differentially modified H3K27ac sites when compared to the genome-wide analysis in any of the four cell types. The results 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 Identifying global changes in immune cells chromatin accessibility between psoriasis patients and healthy controls

The cell type specific chromatin accessibility landscape is determined by the combination of histone modifications and DNA binding proteins (including

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

TF and co-regulatory proteins) in a particular locus. The previous results showing only moderate changes in the H3K27ac landscape between patients and controls are not necessarily representative of the overall changes in the chromatin accessibility landscape in disease. In order to interrogate genome-wide changes in chromatin accessibility between patients and controls, ATAC-seq was performed in the same four cell types in eight patients and ten controls (Table 4.4) giving a total of 72 libraries.

Data processing and quality control

ATAC-seq quality control first found that the median total reads after filtering ranged between 39.2 and 49.8 in CD4⁺ and CD19⁺ respectively, and was over the 15 million reads determined as appropriate an minimum (Chapter 3) in all the samples (Figure 4.4 a).

The differences in read depths across ATAC samples are mostly due to intrinsic difficulties in determining molarity of the ATAC libraries (due to the fragment size heterogeneity) as well as the number of duplicates present in each library, which also correlated with MT reads. Differences in the percentage of MT reads were noticeable between samples from cohort 1A generated with the standard ATAC-seq protocol from Buenrostro *et al.*, 2013 and the Fast-ATAC libraries from cohort 1B using the later modified (Corces *et al.* 2016) protocol (Figure ??).

All the samples showed the required characteristic ATAC-seq fragment size distribution recapitulating nucleosome periodicity (data not shown) (previously detailed in Chapter 3). Analysis of ATAC-seq signal enrichment across gene TSSs revealed that most of the samples had enrichment over 6 (Figure 4.4 b) and only PS2000 and PS2001 CD14+ monocytes were removed from downstream analysis due to low signal-to-noise ratios (<6). When comparing the number of peaks passing IDR filtering in each samples versus the number of reads after filtering,

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

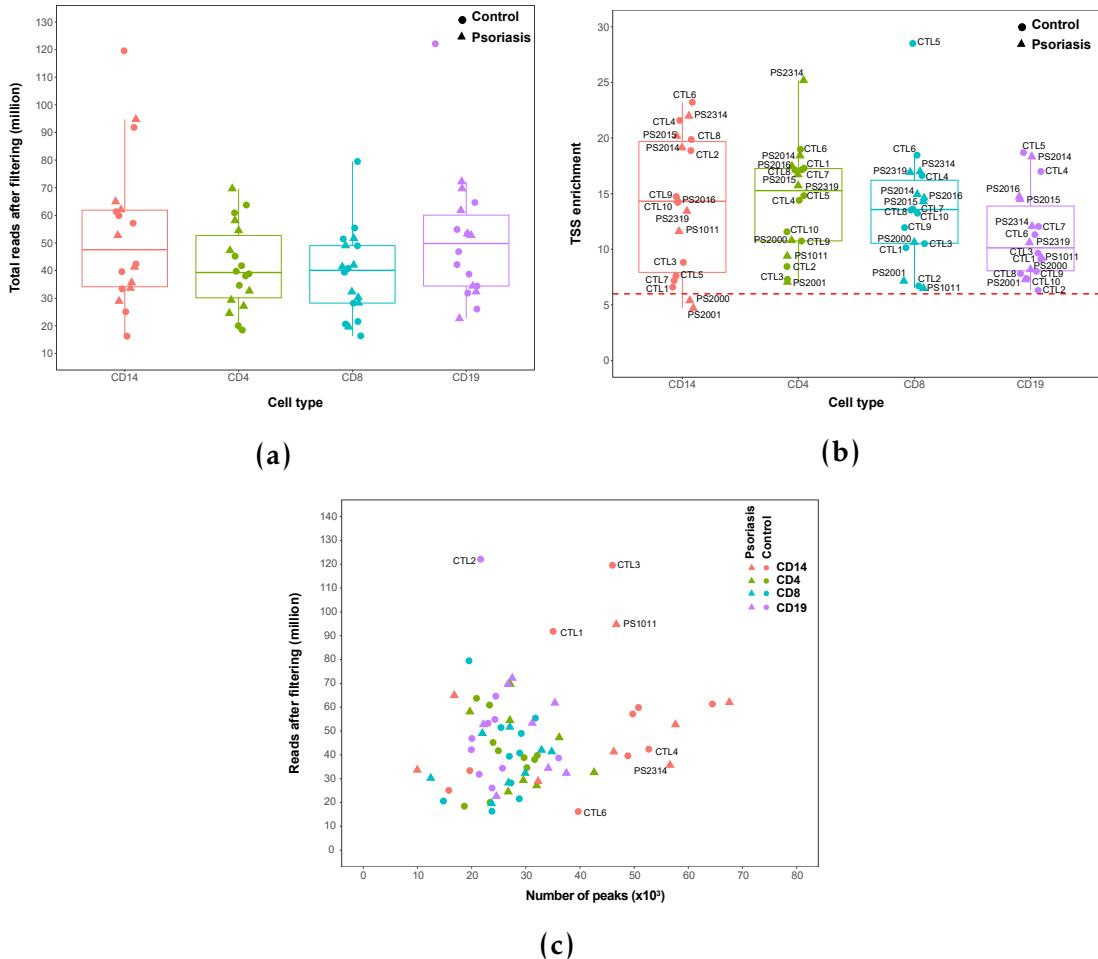


Figure 4.4: Quality control assessment of the ATAC libraries generated from circulating immune cells in psoriasis and control samples. For each of the cell types and samples, boxplots representing a) million of reads after filtering, b) values for fold-enrichment of ATAC fragments across the Ensembl annotated TSS and c) representation of the number of significant peaks based on IDR optimal pval versus the total million reads after filtering for each of the samples. 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) and c) sample IDs are included for all or some of the samples.

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

most of the samples presented between 10,000 and 35,000 peaks (Figure 4.4 c). Since the sequencing depth of most of the samples was \geq 15 million reads, most of the differences in number of called peaks were intrinsic to the cell type and the signal-to-noise differences in the samples, as previously studied in Chapter 3. For example, CD14⁺ monocytes had greater numbers of peaks when compared to the other three cell types, despite the median total number of reads after filtering being similar to the other cell types (Figure 4.4 a). CD19⁺ CTL2 appeared to be an outlier, with a noticeably lower number of peaks for its high sequencing depth (Figure 4.4 c). This observation together with its border line TSS enrichment supported removal of the CD19⁺ CTL2 sample from downstream analysis.

Functional relevance of the ATAC regions included in the differential analysis

A heatmap illustrating sample distance using the consensus master list of ATAC-seq regions across the four cell types (ML_all) showed successful separation of the samples according to the cell type into three main clusters corresponding to CD14⁺ monocytes, CD19⁺ and CD4+/CD8⁺ T cells (Figure 4.5 a). Within each of the cell type clusters, samples did not separate based on disease condition, suggesting the absence of large global differences in the chromatin accessibility landscape between psoriasis patients and control individuals. Conversely, within the cell types there was some grouping of samples by batch (Figure 4.5 a).

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

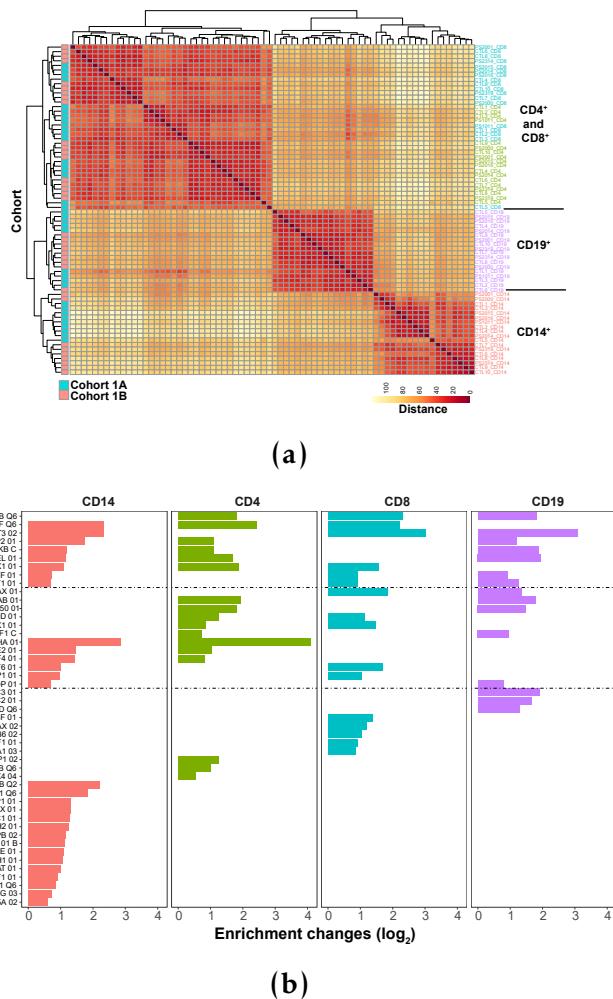


Figure 4.5: Clustered heatmap and conserved TFBS enrichment analysis in the consensus ATAC regions identified in $CD14^+$ monocytes, $CD4^+$, $CD8^+$ and $CD19^+$ cells from the patients and controls cohort. a) Distance matrix and hierarchical clustering for the 72 samples was performed based on the normalised read counts retrieved for each sample at the regions included in a consensus master list of ATAC-seq enriched sites built across all four cell types (ML_all). Clusters have been additionally annotated using cohort identity. b) Enrichment analysis for the conserved TFBS was performed for each of the ATAC-seq cell type master lists of regions used for downstream differential analysis. Enrichment was tested for 258 human conserved TFBS identified by Transfac using position-weight matrices based on experimental results in the scientific literature. Significant enrichment using FDR<0.01.

Each of the four cell type master lists of ATAC-seq peaks (ML_CD14, ML_CD4, ML_CD8, and ML_CD19) used for the downstream differential chromatin accessibility analysis (explained in Chapter 2), presented the highest percentage of regions annotated as gene promoters, intronic and intergenic, as expected for ATAC-seq (Figure ??). *cis*-eQTL SNPs from a number of immune

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

cell types, including CD14⁺ monocytes (unstimulated and stimulated), B cells, neutrophils, CD4⁺ and CD8⁺ cells (Kasela 2016; Fairfax et al. 2012; Fairfax et al. 2014), were enriched within the ATAC consensus peak list of each cell type. For example, eQTLs from unstimulated and stimulated (LPS or IFN- γ) monocytes were the most significantly enriched (FDR<0.01) in the ML_CD14 (unstimulated fold-enrichment 5.1, LPS 2h fold-enrichment 4.7 and IFN- γ fold-enrichment 5.0) when compared to the other eQTL datasets. Similarly, *cis*-eQTLs from CD4⁺ and CD8⁺ were the most enriched datasets (fold-change 8.3 and 8, respectively) in the ML_CD8. The specificity and functional relevance of each cell type master list was further reinforced by the significant enrichment (FDR<0.01) of conserved TFBS within those ATAC-seq regions (Figure 4.5 b). For example, enrichment of conserved NF κ B binding motifs(NFKB Q6, NFKB C and NFKAPPAB 01) was identified across the master lists from the different cell types. Conserved binding motifs for TF involved in T cell biology, such as AREB6 (ZEB1), ATF6 and the heat-shock transcription factor HSF1 (Guan et al. 2018; Yamazaki et al. 2009; Gandhapudi et al. 2013), were enriched in the ML_CD8. Overall, the enrichment of eQTL SNPs and conserved TFBS highlighted the potential of each cell type master list to harbour functional relevant differences in chromatin accessibility between psoriasis patients and controls.

Differential chromatin accessibility analysis

Differential chromatin accessibility analysis between patients and controls was performed on the ATAC-seq normalised read counts for the regions of each cell type master list using DESeq2. PCA analysis on these data prior to the differential analysis revealed a batch effect correlating with the different ATAC-seq protocols used in cohort 1A and cohort 1B (standard ATAC-seq and FAST-ATAC, respectively) (Figure B.6 a). Therefore, the ATAC-seq protocol was included as a covariate in the differential analysis model. Moreover, CTL5

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

appeared as a cohort 1A outlier for all the cell types (representative example Figure B.6 a) and was also removed from the differential analysis.

Genome-wide differential chromatin accessibility analysis revealed 55 significant (FDR<0.05) DARs between psoriasis patients and healthy controls in CD8⁺ cells (Table 4.6), of which 17 showed FDR<0.01. Conversely, CD14⁺ monocytes, CD4⁺ and CD19⁺ cells only presented one or no DARs.

Cell type	Number of DARs FDR<0.05
CD14 ⁺	1
tCD4 ⁺	0
tCD8 ⁺	55
CD19 ⁺	1

Table 4.6: Summary results from the differential chromatin accessibility analysis between psoriasis patients and healthy controls in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ cells. The number of differentially accessible regions (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 FC was applied in this analysis.

Annotation of the 55 CD8⁺ DARs using cell type specific Roadmap Epigenomics chromatin segmentation maps revealed the potential for some of the regions to be involved in regulation of gene expression, including 24 (44.4%) weak enhancers, 7 (12.9%) active promoters, 6 (11.1%) weak promoter and 2 (3.7%) strong enhancers. The functional relevance of the DARs in terms of regulation of gene expression was further investigated by integration of the CD8⁺ cell eRNA data from the FANTOM5 project. Only 8 of the CD8⁺ DARs overlapped significantly expressed eRNAs. These include a region at the TSS of the *TNSF11* gene and another upstream the *IL7R* promoter, which were more accessible in the psoriasis patients compared to the healthy controls (Figure 4.6 a and b). The two DARs also overlap chromatin harbouring H3K4me3, a histone mark indicating an active promoter, and H3K27ac consistent with the transcription of those regions as eRNAs in tCD8⁺ cells according to FANTOM5.

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

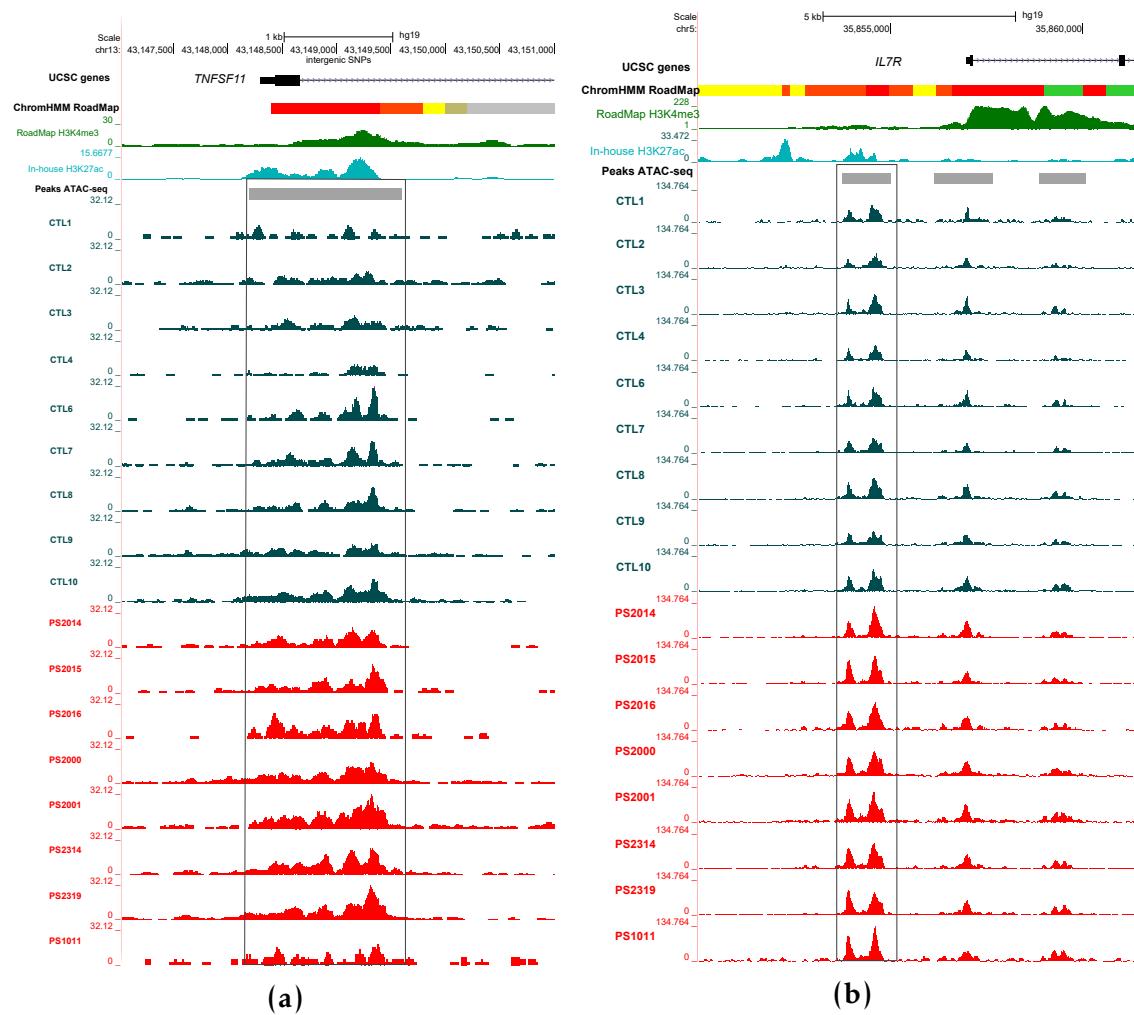


Figure 4.6: Epigenetic landscape at two ATAC differential accessible regions between patients and controls in CD8⁺ cells. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) in DARs located at a) the promoter of TNFSF11 gene and b) up-stream the IL7R gene (x-axis). Both DARs were more open in tCD8⁺ cells from psoriasis compared to controls. Tracks are colour-coded by condition: control(CTL)=dark turquoise and psoriasis (PS)=red. The Epigenome Roadmap chromatin segmentation map and H3K4me3 for tCD8⁺ 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 FC cut-off.

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

Other potentially interesting CD8⁺ DARs were found nearby genes such as the MAPK *MAP3K7CL* and *NFKB1*; However they were not at regions annotated as enhancers or overlaped with experimentally validated eRNAs.

Integration of H3K27ac ChIPm and ATAC-seq chromatin accessibility profiles

Although a very low number of differentially H3K27ac modified and DARs were found between psoriasis and control samples in the four cell types , commonalities in the disease specific changes were investigated. The the H3K27ac ChIPm and ATAC differential sites between psoriasis and control individuals only showed one overlapping region in CD8⁺ cells. The DARs and differentially H3K27ac site was located within an intron of the D-tyrosyl-tRNA deacylase 1 (*DTD1*) gene (Figure 4.7). This region presented lower levels of H3K27ac (4 patients versus 4 controls) and was less accessible (8 patients versus 9 controls) in the psoriasis patients when compared to healthy controls (Figure). This differential region was annotated as an active enhancer according to the tCD8⁺ ChromHMM segmentation map and did not interact with the promoter of any gene according to Hi-C and promoter Hi-C data in tCD8⁺ cells (Javierre2016). Conversely, SNPs within this region were eQTL for *DTD1* in whole blood (<https://gtexportal.org/home/eqtls>).

4.3.4 Gene expression analysis in psoriasis circulating immune cells

Data processing and quality control

In addition to characterising the chromatin accessibility landscape, gene expression profiles in psoriasis and healthy individuals were also analysed for the same four primary circulating immune cell types using RNA-seq. The percentage of RNA-seq reads mapping to a unique location in the genome using STAR (see Chapter 2) was appropriate (minimum recommended 70 to 80%),

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

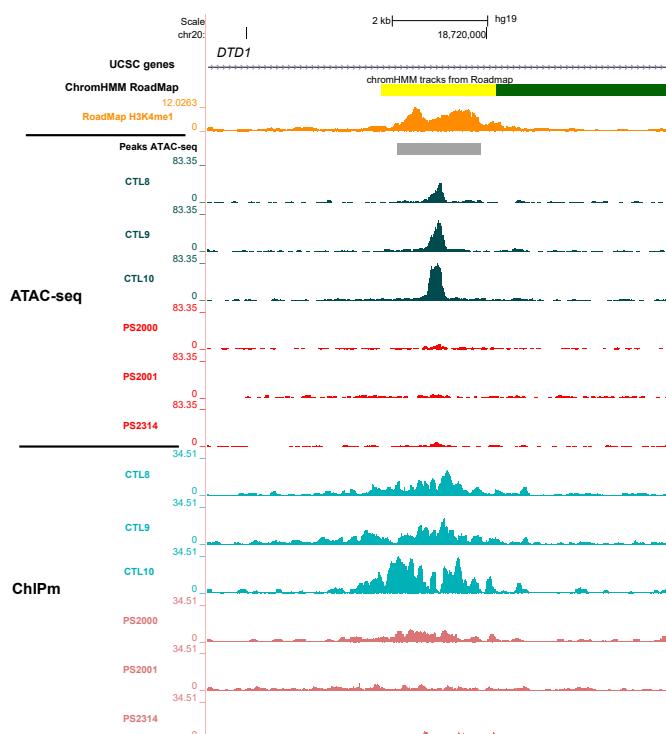


Figure 4.7: Epigenetic landscape at the only overlapping region identified as DARs and differentially H3K27ac modified between psoriasis patients and controls. UCSC Genome Browser view illustrating the normalised ATAC read density and H3K27ac normalised fold-enrichment (y-axis) at an intron of the *DTD1* gene (x-axis) in tCD8⁺ cells. This region was identified as less accessible and less enriched for H3K27ac modifications in psoriasis patients compared to healthy controls. 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 Epigenome Roadmap chromatin segmentation map and H3K4me1 for tCD8⁺ cells are also shown.

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

ranging between 79.64 and 86.19% across the 72 samples (Figure 4.8 a). After appropriate filtering, all the 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 4.8 b). The median total reads mapping to Ensembl features was greater for CD14⁺ monocytes when compared to the other three cell types. Interestingly, in all four cell types analysed, greater mapping rates and total reads mapping to Ensembl features were observed for cohort 1B samples when compared to cohort 1A. These differences were attributed to the library preparation and sequencing of each cohort in two different batches.

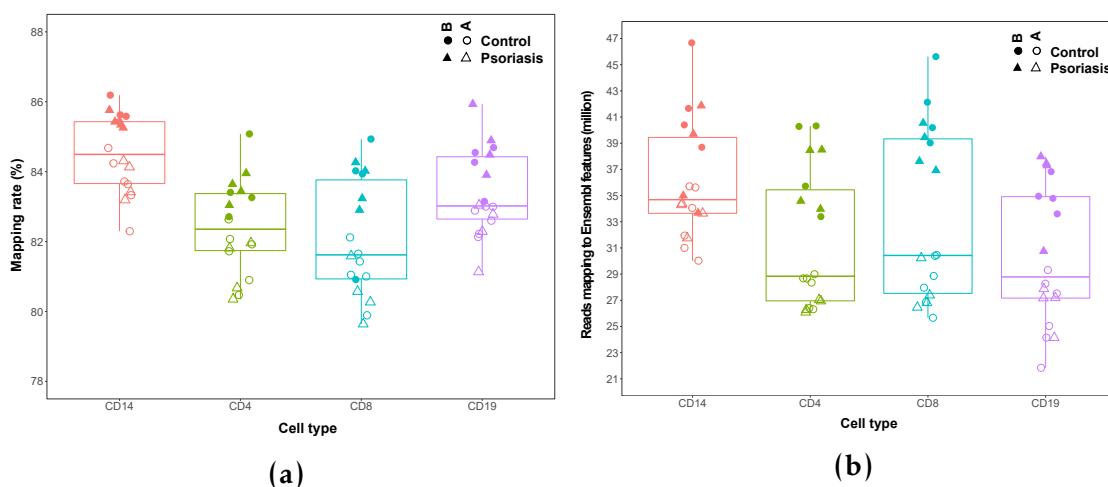


Figure 4.8: Mapping rate and total reads after filtering (million) mapping to Ensembl genes in all the RNA-seq samples from psoriasis patients and controls in four cell types. a) The mapping rate refers to the percentage of total sequenced reads from each sample that uniquely mapped to a particular site 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.

Similarly to ChIPm and ATAC-seq, the first and second PC from PCA analysis using the normalised number of reads mapping to each of the 20,493 Ensembl genes passing quality control (see Chapter 2) showed that most variability was driven by cell type differences (Figure 4.9 a). A heatmap illustrating sample distance based on the expression profile of each sample followed by hierarchical clustering revealed three main clusters corresponding

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

to CD14⁺ monocytes, CD4⁺ and CD8⁺ lymphocytes, and CD19⁺ cells (Figure 4.9 a). Within each cell type cluster, samples were further grouped by cohort (1A and 1B) and not by condition (psoriasis and control), consistent with the differences in mapping rate and total reads mapping to Ensembl genes observed across the two cohorts (Figure 4.8 a and b). Clear correlation of sample batch with PC4 from the PCA analysis led to a very clear separation of the samples into cohort 1A and 1B, explaining 3% of the total variance (Figure B.6 b). Consequently, batch was included in the DGE model as a covariate.

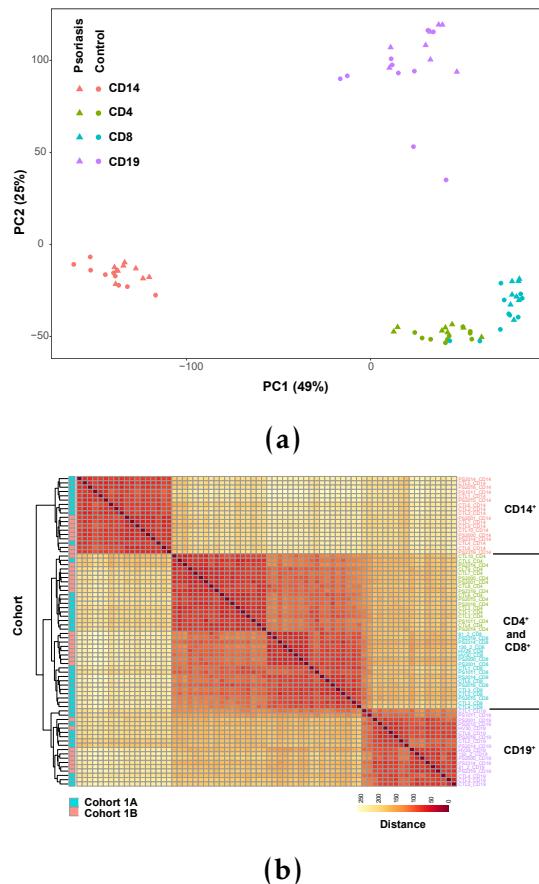


Figure 4.9: PCA analysis and sample distance heatmap with hierarchical clustering illustrating the sample variability based on the gene expression profiles. a) The first and second PCs (x-axis and y-axis, respectively) for the analysis using all the detected genes are represented to identify the main sources of variability across the 72 samples. Each point represents a sample, where the colour codes for cell type and the shape for condition. The proportion of variation explained by each principal component is indicated.b) Distance matrix clustering for the 72 samples was performed based on the normalised read counts mapping to 20,493 Ensembl featured remaining after appropriate filtering. Annotation of the clustering using cohort identity is included.

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

mRNA and lncRNA differential expression

DGE analysis between 8 psoriasis patients and 10 healthy controls in CD14⁺ monocytes, tCD4⁺, tCD8⁺ and CD19⁺ was performed using DESeq2 and including the cohort identity as a covariate to account for the batch effect previously mentioned. For each of the cell types a number of mRNAs were identified as differentially expressed at an FDR <0.05 or 0.01 (Table 4.7).

Cell type	mRNA FDR<0.05/0.01	lncRNA FDR<0.05/0.01
CD14 ⁺	671/229	28/8
tCD4 ⁺	108/40	12/4
tCD8 ⁺	656/175	31/5
CD19 ⁺	167/71	6/2

Table 4.7: Summary results from the DGE analysis between psoriasis patients and healthy controls in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ cells. The number of statistically differentially expressed mRNAs and lncRNAs are listed for two FDR threshold (FDR<0.05 and FDR<0.01). No threshold for the FC was applied in this analysis. The number and name of the lncRNAs overlapping with the Dolcino *et al.*, 2018 study comparing PBMCs between PsA patients and healthy controls are also included. (*) indicates dysregulation in the opposite direction between this data and Dolcino *et al.*.

CD14⁺ monocytes and tCD8⁺ were the two cell types presenting the largest number of mRNAs with modulated expression between psoriasis patients and controls.

The magnitude of the FC of gene expression between psoriasis patients and controls was moderate in all four cell types, with the largest changes in CD14⁺ monocytes and CD8⁺ cells. Regarding the directionality of the statistically significant modulated genes (mRNA and lncRNAs) using FDR<0.05, CD14⁺ monocytes (up 344, down 379) and CD4⁺ (up 57, down 66) presented similar numbers of genes up-regulated and down-regulated in psoriasis patients when compared to the healthy controls. In contrast, in CD8⁺ (up 278, down 429) and CD19⁺ (up 29, down 148) a larger number of modulated genes were down-regulated in patients compared to controls.

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

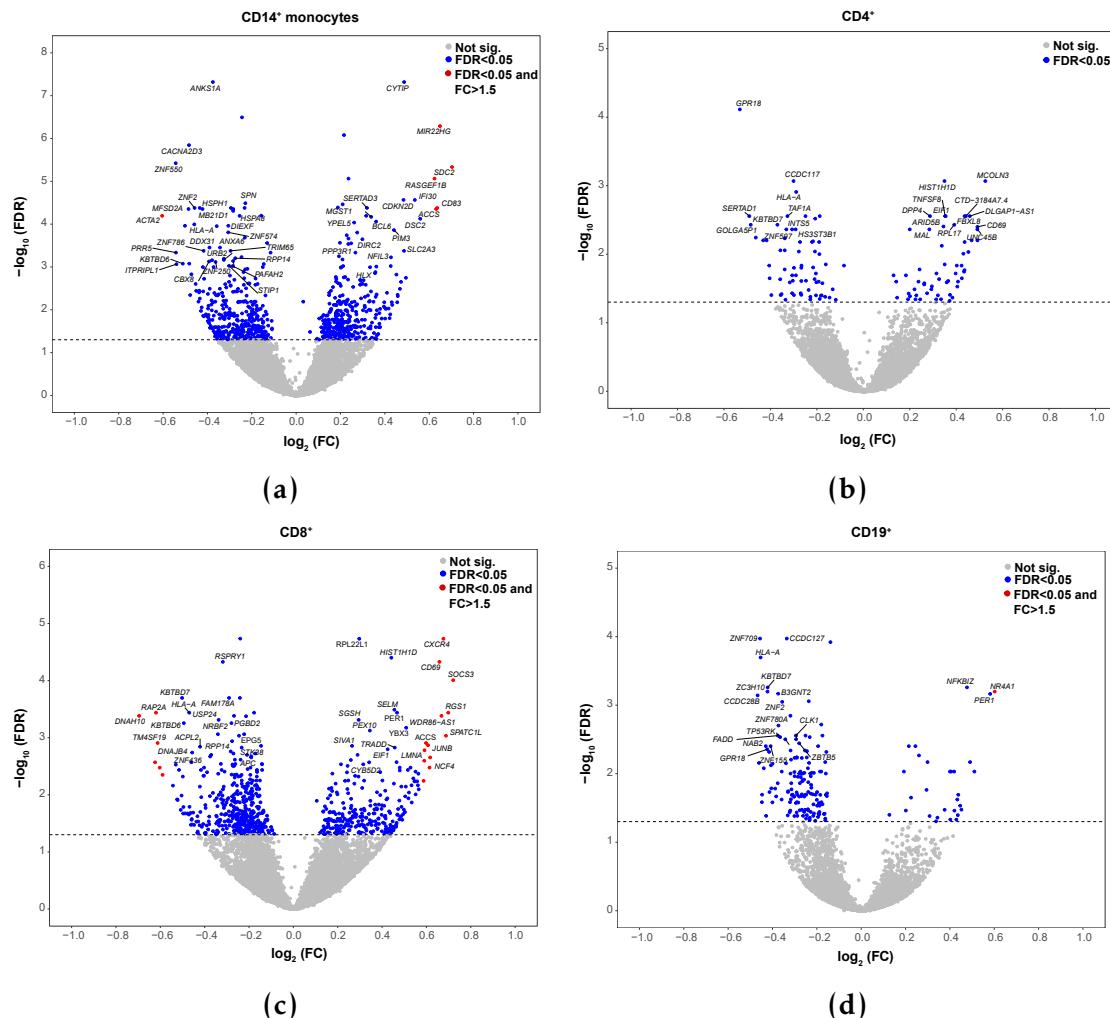


Figure 4.10: Magnitude and significance of the gene expression changes between psoriasis patients and healthy controls in four immune cell types. Volcano plots for the results of the DGE analysis in a) CD14⁺ monocytes, b) CD4⁺, c) CD8⁺ and d) CD19⁺ cells. For each gene, the $\log_2(\text{FC})$ represents the change in expression for that gene in the psoriasis group with reference to the healthy controls. Significant DEGs (FDR<0.05) in blue for FC <1.5 and red for FC >1.5. The volcano plots include mRNAs and lncRNAs species.

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

Some of the DEGs across the four cell types overlapped DEGs from the two most comprehensive studies comparing expression of PBMCs isolated from psoriasis patients and healthy controls (Lee et al. 2009; Coda et al. 2012). The greatest overlap (7 genes) was found between the DEGs in CD14⁺ monocytes and those identified by Coda *et al.*, 2012. However, 5 out of 7 presented opposite directionality. One of those genes dysregulated in the same direction was ubiquitin conjugating enzyme E2 D1 (*UBE2D1*), which mediates, for example, ubiquitination of the TNF receptor-associated factor 6 (TRAF6) protein (Gru2008). The greatest overlap with the Lee and colleagues DEGs was found for CD8⁺ cells (3 genes) in the same direction. Similarly, only one overlap (*NAMPT*) was found with the psoriasis DEGs in a study comparing PBMC transcriptional profiles of three inflammatory diseases (IBD, RA and psoriasis)(Mesko et al. 2010). The nicotinamide phosphoribosyltransferase *NAMPT*, involved in metabolism and stress response, was up-regulated in our CD14⁺ monocytes as well as in PBMCs from the three phenotypes studied by Mesko and colleagues, suggesting its role as a marker of inflammation rather than marker for psoriasis.

An overlap between the significant DEGs (FDR<0.05) across the four cell types and a list of the genes putatively associated with psoriasis GWAS hits from the NHGRI-EBI catalog (<https://www.ebi.ac.uk/gwas>) curated to include other genes from more recent studies was performed (Table 4.8). CD8⁺ was the cell type with the largest number of DEGs (7 hits) overlapping with putative GWAS genes, followed by CD14⁺ monocytes and CD4⁺ (3 hits each). Some of those genes were found in more than one cell type, including *NFKBIA*, *TNFAIP3* and *NFKBIZ*, amongst others.

The role of lncRNAs in psoriasis circulating immune cells

In addition to protein coding genes, some of the DEGs identified were classified as lncRNAs. tCD8⁺ and CD14⁺ monocytes were the two cell types

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

Cell type	Number of GWAS overlaps	Up-regulated genes	Down-regulated genes
CD14 ⁺	3	NFKBIA	IL23A, FASLG
tCD4 ⁺	3	TNFAIP3, NFKBIZ	FASLG
tCD8 ⁺	7	TNFAIP3, NFKBIA, ETS1, SOCS1, NFKBIZ	B3GNT2, FASLG
CD19 ⁺	2	NFKBIZ	B3GNT2

Table 4.8: Overlap between putative psoriasis GWAS genes and the reported significantly DEGs in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ cells. DEGs list based on FDR<0.05.

presenting the largest number of dysregulated lncRNAs between psoriasis patients and controls (Table 4.7). In contrast, CD19⁺ was the cell type showing the lowest number of lncRNAs differentially expressed. Only one lncRNA, RP11-218M22.1 appeared to be dysregulated between psoriasis and healthy controls in all the cell types.

The differentially expressed lncRNAs in this study were overlapped with the 259 lncRNAs identified as dysregulated in PBMCs when comparing PsA patients versus healthy controls by Dolcino *et al.*, 2018 (Table 4.9). The largest overlap was found in CD14⁺ monocytes, where four of the differentially expressed lncRNAs were also reported by Dolcino and colleagues. However, HOTAIRM1 and ILF3-AS1 were up-regulated in psoriasis CD14⁺ monocytes when compared to controls but appeared down-regulated in PsA PBMCs.

The number of differentially expressed lncRNAs for which a functional interaction had been experimentally found was investigated using NPInter database , which retrieves functional interactions between non-coding RNAs and biomolecules (proteins, RNAs and DNAs) which have been published (Hao et al. 2016). The majority of the differentially expressed lncRNAs (FDR<0.05) in all the cell type were found to have a functional interacting partner.

Amongst the characterised lncRNAs dysregulated between psoriasis and controls CD14⁺ monocytes were the negative regulator of antiviral response

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

Cell type	LncRNAs with functional interactions	LncRNAs overlapping Dolcino <i>et al.</i> , 2018
CD14 ⁺	24	4 (<i>HOTAIRM1*</i> , <i>ILF3-AS1*</i> , <i>MMP24-AS1</i> , <i>RP11-325F22.2</i>)
tCD4 ⁺	12	1 (<i>MMP24-AS1</i>)
tCD8 ⁺	21	1(<i>CTB-25B13.12</i>)
CD19 ⁺	5	0

Table 4.9: 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.*.

(*DYNLL1-AS1* or *NAV*), the HOXA transcript antisense RNA myeloid-specific 1 (*HOTAIRM1*) and the nuclear paraspeckle assembly transcript 1 (*NEAT1*). *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). In this data, *DYNLL1-AS1* appeared down-regulated in CD14⁺ monocytes from psoriasis patients when compared to controls but no up-regulation of *IFITM3* and *MxA* was found. Conversely, *HOTAIRM1* appeared to be up-regulated in the CD14⁺ monocytes from psoriasis patients (Figure 4.11 a). 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⁺ monocytes in psoriasis versus control samples in this data (Figure 4.11 b). Lastly, *NEAT1* was also up-regulated in psoriasis patients compared to controls and had also been found to be up-regulated in a study in SLE CD14⁺ monocytes (Zhang *et al.* 2016).

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

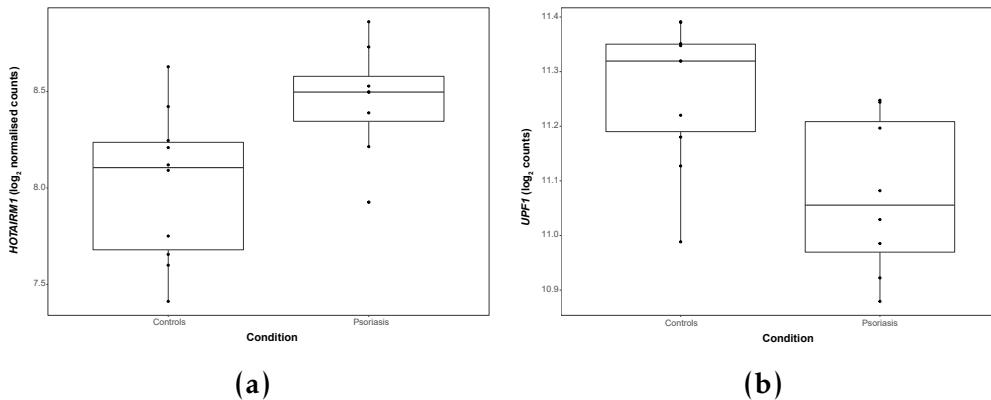


Figure 4.11: RNA-seq expression levels of the lncRNA *HOTAIRM1* and its experimentally validated target *UPF1* in psoriasis and healthy controls CD14⁺ monocytes. Expression is illustrated as the log₂ of the normalised read counts mapping to a) the lncRNA *HOTAIRM1* and b) *UPF1*, which has been experimentally identified as one of the genes regulated by this lncRNA according to NPInter database.

For tCD8⁺ cells, the most relevant non-coding RNA was the miR *MIR146A*, which was captured in the standard RNA library preparation. *MIR146A* has been characterised to have a role in negative regulation of innate immunity, inflammatory response and antiviral pathway and was found to be downregulated in psoriasis patients when compared to controls. Other lncRNAs were found to be dysregulated in more than one cell type. For example, *KCNQ1OT1* was downregulated in both, tCD4⁺ and tCD8⁺ cells. Dysregulated expression of this lncRNA has been reported in Beckwith-Wiedemann syndrome consisting of a loss-of-imprinting paediatric overgrowth disorder with some skin features such as creases or pits in the skin near the ears (Pandei2008).

Pathway enrichment analysis for the DEGs

In order to better understand the biological role of the significantly modulated genes, pathway enrichment analysis was performed for each cell type. The moderated FCs in the DGE analysis illustrated in the volcano plots suggest that the differences between patients and controls in these circulating immune cells are discrete. Nevertheless, moderate differences may have an important impact on their phenotype for infiltration and activation in the skin. Therefore,

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

exploratory pathway analysis was conducted using DEG with FDR<0.05 and no FC cut off. Biologically relevant pathways appeared to be significantly enriched (FDR<0.01) for CD14⁺ monocytes and CD8⁺ cells DEGs (Table ?? and A.3). In CD19⁺ cells, only one pathway (generic transcription) appeared to be significantly enriched for DEGs in this cell type. In contrast, in CD4⁺ cells modulated genes between psoriasis patients and controls were not enriched for any pathway.

Cell type	Pathways
CD14 ⁺ monocytes	MAPK signalling IL-12 mediated signaling events Th-1 and Th-2 cell differentiation Th-17 cell differentiation TCR signalling Platelet-derived growth factor (PDGF- β) signalling Forkhead box O (FoxO) signalling
CD8 ⁺	Osteoclast differentiation MAPK signalling TNF signalling IL-12 mediated signalling events NF- κ B signalling Chemokine signalling

Table 4.10: Most relevant pathways enriched for DEGs between psoriasis patients and healthy controls in CD14⁺ monocytes and CD8⁺ cells. The enrichment analysis was conducted using significantly DEGs FDR<0.05 and no FC threshold. Enriched pathways had FDR<0.01 and a minimum of ten gene members overlapping with DEGs for that particular cell type.

Two of the significant enriched pathways, MAPK signalling and IL-12 mediated signalling, were found to be enriched for CD14⁺ monocytes and CD8⁺ cells DEGs (Table ??). DEGs contributing to the enrichment of this pathway in both cell types included MAPK gene members such as *MAP3K4* and *MAPK14*, both down-regulated in psoriasis when compared to controls. For example, *MAP3K4* is a member of the MAPKKK family, which expression is down-regulated in LPS-stimulated PBMCs from CD patients leading to a relative immune deficiency in TLR-mediated cytokine production. Moreover, DGE of

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

members of the dual-specificity phosphatases (DUSP) family, involved in fine-tuning the immune response (Qian et al. 2009), contributed to the enrichment of the MAPK pathway in CD14⁺ monocytes and tCD8⁺ cells. For instance, *DUSP10* was down-regulated in the psoriasis CD14⁺ monocytes and its knock-out in mice led to enhanced inflammation (Qian et al. 2009). Conversely, *DUSP4* presented up-regulation in psoriasis tCD8⁺ when compared to healthy controls 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⁺ monocytes from psoriasis presented down-regulation of *STAT4* and *STAT5A* in patients compared to controls. Neither *STAT4* and *STAT5A* were dysregulated in tCD8⁺ cells between psoriasis and healthy controls. Likewise, *IFNG* expression in psoriasis patients was lower than in healthy controls in tCD8⁺ cells but changes were not found in CD14⁺ monocytes. *IL2RA* was also up-regulated in tCD8⁺ from psoriasis patients when compared to controls, which may enhance formation of the IL2-R α and the signalling by this cytokine involved in effector and regulatory T cell differentiation (Malek and Immunity 2010).

The platelet-derived growth factor (PDGF- β) signalling pathway was only enriched in CD14⁺ monocytes (Table 4.10). Within this pathway the *SLA* gene appeared to be down-regulated in psoriasis patients compared to controls. A *SLA* knock-out mouse model has shown impaired IL-12 and TNF- α 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 tCD8⁺ cells. These included TNF, NF- κ B and chemokine signalling (Table 4.10). Due to the close relationship between these three pathways, some DEGs contributed to the enrichment of more than one of them. That was the case for NF- κ B inhibitor A (*NFKBIA*) and the TNF- α induced

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

protein 3 (*TNFAIP3*) which were unexpectedly up-regulated in tCD8⁺ cells from psoriasis compared to healthy controls. *NFKBIA* up-regulation contributed to the enrichment of all three pathways (Figure 4.12 a (in green box and b) and *TNFAIP3* was a member of the TNF and NF-κB pathways (Figure 4.12 a in green box). Interestingly, *NFKBIA* and *TNFAIP3* are two of the psoriasis GWAS associated genes and were also up-regulated in psoriasis CD14⁺ monocytes and tCD4⁺ cells, respectively (Table 4.8).

Other genes with a prominent pro-inflammatory role also appeared to be down-regulated in the NF-κB or TNF signalling pathways, such as the activating transcription factor 2 (*ATF2*) and 4 (*ATF4*) members of the TNF signalling cascade and the protein kinase C beta_{PRKCB} from the NF-κB and chemokine signalling pathways. In contrast, up-regulation of pro-inflammatory genes members of these two pathways were also found. For example JunB proto-oncogene (*JUNB*) coding for one of the subunits of the TF AP-1 and three of the NF-κB subunits including *RELA*, *RELB* and *NFKB2*. Particularly, AP-1 undergoes activation following growth factors, cytokines, chemokines, hormones and multiple environmental stresses and acts as a negative regulator of cell proliferation and IL-6 production (Schonthaler and rheumatic 2011).

Regarding dysregulation of chemokines, a mix of up-regulation and down-regulation of members of this pathway was found in CD8⁺ cells from psoriasis patients when compared to healthy controls (Figure 4.12 b). One of the most relevant dysregulated cytokine genes was *CCR10*, the receptor for the chemotactic skin-associated chemokine CCL27. In this data, tCD8⁺ cells but not tCD4⁺ cells from psoriasis patients presented up-regulated expression of the *CCR10*. Other up-regulated chemokine receptors in tCD8⁺ circulating psoriatic cells 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- α , IL-17 and IL-6,

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

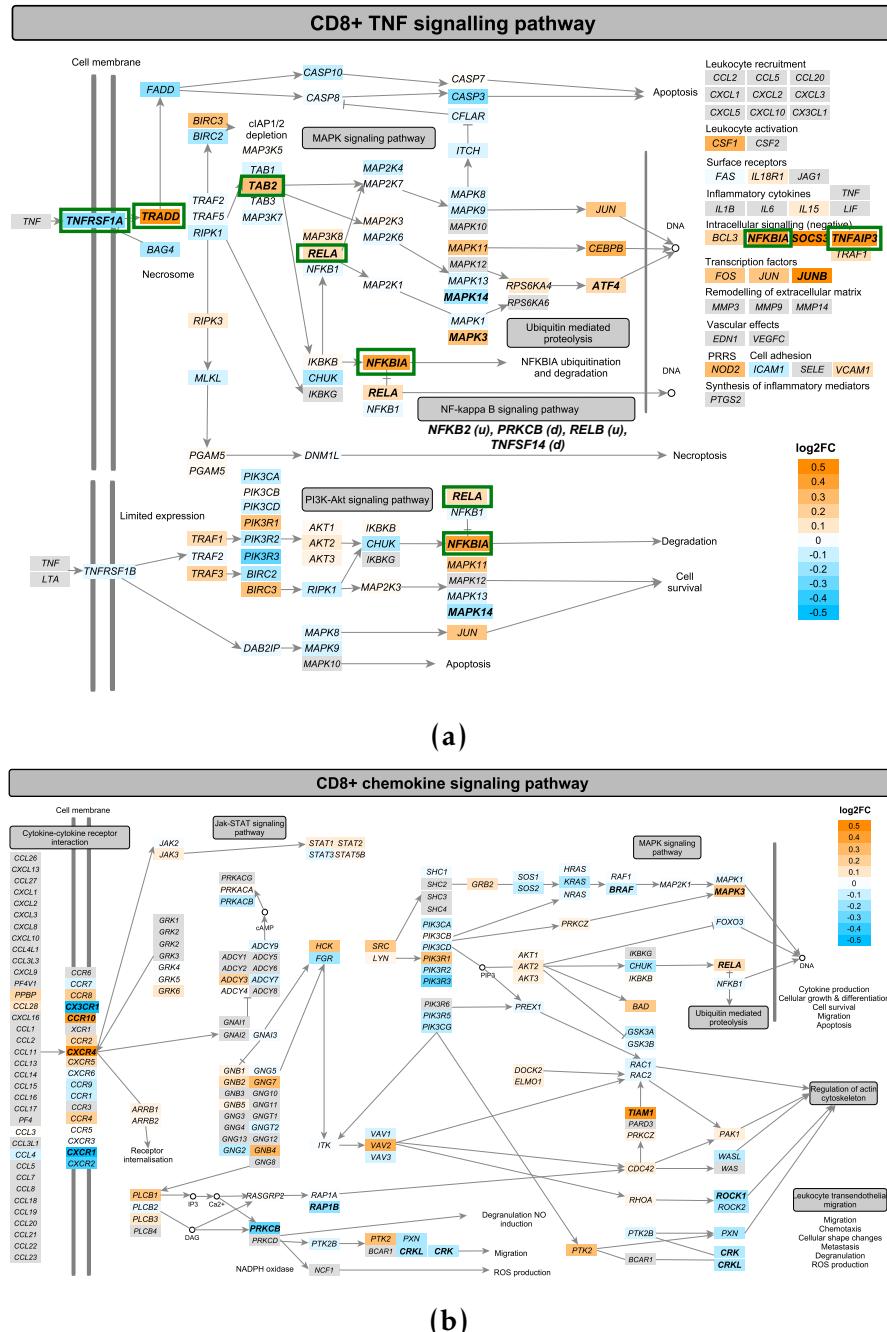


Figure 4.12: Mapping of the DEGs identified in CD8⁺ cells between psoriasis patients and healthy controls onto the TNF- α and the chemokine signalling pathways. The a) TNF- α and b) chemokine pathways were sourced from KEGG, manually curated in a way that all member genes are maximised visually and then automatically color-coded by the log₂FC expression between psoriasis patients and healthy controls CD8⁺ cells isolated from PB. Significant DEGs (FDR<0.05) are highlighted in bold. In a), members of the TNF- α pathway shared with the NF- κ B are highlighted with a green box. Additional members of the NF- κ B pathway differentially regulated in CD8⁺ cells have also been indicated in brackets. Enrichment for a) and b) was identified by using only the CD8⁺ DEGs (FDR<0.05).

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

appeared to be up-regulated in any of the four cells types from psoriasis patients compared to healthy controls.

4.3.5 RNA-seq in epidermis from psoriasis patients

Data processing and quality control

For the three paired uninvolved-lesional samples (Table 4.4), all presented a mapping rate greater than 80%, with the rate moderately greater in the lesional compared to the uninvolved samples in all three patients (Figure ?? a). The number of reads after filtering that were mapped to Ensembl genes ranged between 29.5 and 33.2 million in PS1011 uninvolved and PS1011 lesional, respectively (Figure 4.13 b). Similarly to the mapping rate, the final number of reads mapping to genes was greater in the lesional samples compared to the controls.

PCA analysis using the normalised number of reads mapping to the genes after filtering (see Chapter 2) revealed separation of the lesional samples from the uninvolved by the first PC, which explained 37% of the variance (Figure 4.13 c). 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 DGE analysis accounted.

Summary of the DGE results

DGE analysis revealed a total of 1,227 (FDR<0.05) and 702 (FDR<0.01) genes dysregulated between the 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 in the changes of gene

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

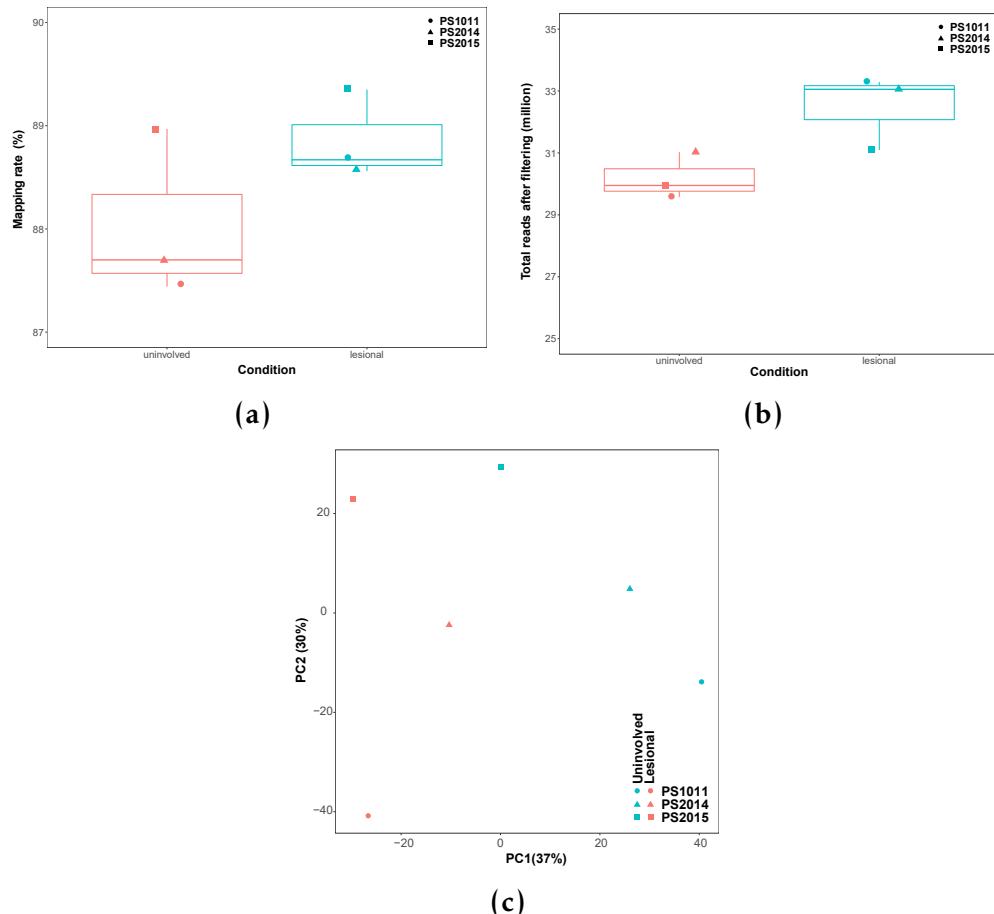


Figure 4.13: Mapping quality control and PCA analysis for the RNA-seq data in the uninvolved and lesional epidermis from psoriasis patients. a) Mapping rate calculated as the proportion of sequencing reads mapping uniquely to a particular region of the genome. b) The total number of reads mapping to an Ensembl feature (including protein coding genes and lncRNAs) after removing the non-uniquely mapped and duplicated reads. c) 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 this study. Dots colour corresponds to condition (lesional or uninvolved) and shape refers to the patient ID.

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

expression between lesional and uninvolved skin were notably larger when compared to the changes in expression from analysis in circulating immune cells, with 874 out of 1,227 genes showing $FC \geq 1.5$.

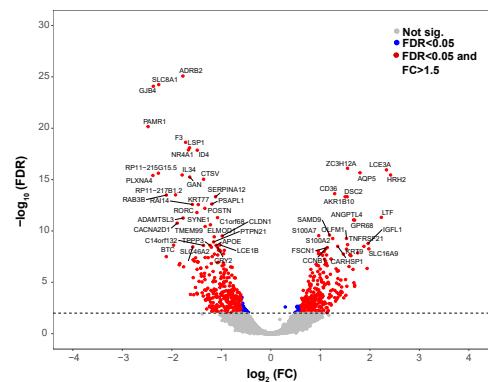


Figure 4.14: Magnitude and significance of the gene expression changes between matched lesional and uninvolved epidermal biopsies from three psoriasis patients. The volcano plot represents for each gene the significance ($-\log_{10}FDR$) of the $\log_2(FC)$ in expression for that gene in the lesional skin group with reference to the uninvolved skin. Significant DEGs ($FDR < 0.05$) in blue for $FC < 1.5$ and red for $FC > 1.5$. 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* showed the opposite behavior.

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 4.11: Summary results of the DGE 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.

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

Overall comparison with other skin transcriptomic studies

As detailed in Chapter 2, the approach to study DGE in skin is different from most of the previously published studies using whole punch biopsies to compare lesional and uninvolved skin from psoriasis patients. During the course of this project a study published by Tervaniemi and colleagues also aimed to characterise the transcriptional profiles of the epidermis from psoriasis patients lesional and uninvolved skin in a more elegant way than the previous studies using full thickness skin biopsies (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 reported a total of 2,589 DEGs passing their filtering criteria ($FC < 0.75$ or $FC > 1.5$ and $FDR < 0.05$) and showing overall a larger number of differentially expressed genes between the two types of biopsies compared this study. The number of genes up-regulated in lesional epidermis compared to uninvolved (2,330) was larger than the number of down-regulated targets (261), contrasting to the in-house results where similar numbers of up- and down-regulated genes were found (Table ?? and Figure 4.15 bottom panel). Regarding overlap, 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. Amongst the up-regulated genes in both studies TFs such as *STAT1*, genes from the *S100* family (e.g *S100A9* and *S100A12*) and genes nearby psoriasis GWAS loci such as *STAT3* and *IFIH1*. The direction of change in 45 out of the 359 shared genes appeared to be opposite across the two datasets. For example, *SERPINB2* gene, a serine protease inhibitor of the serpin superfamily, presented down-regulation in the in-house data and up-regulation in the Tervaniemi results. Interestingly, a study demonstrated a defective stratum corneum in *SERPINB2* deficient mice as well as greater susceptibility

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

to developing inflammatory lesions upon chemically induced atopic dermatitis compared to wild type controls (Schroder et al. 2016).

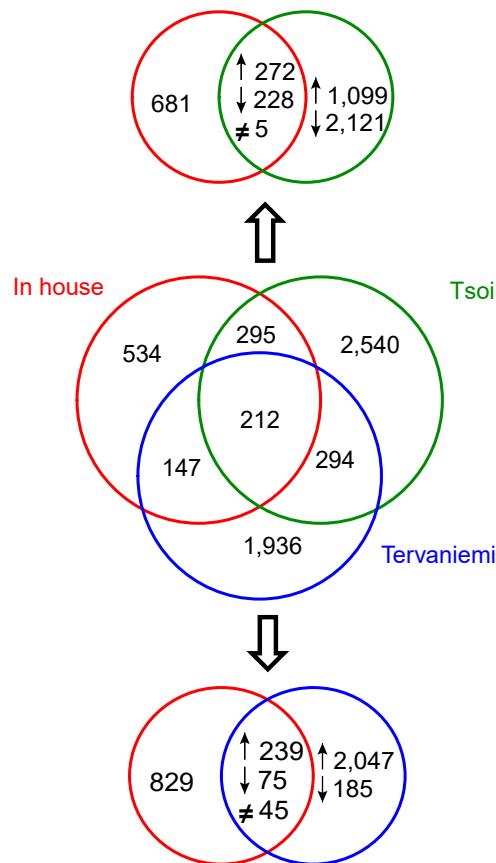


Figure 4.15: Overlap of the significantly differentially expressed genes between lesional and unininvolved epidermal sheets, split epidermis and full-thickness skin biopsies. The central venn diagram illustrated the DEGs overlapping between this study (in house), Tervaniemi *et al.*, 2016 split epidermis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

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

the two datasets, 272 were up-regulated, 228 down-regulated and 7 presented opposite direction of change (Figure 4.15 top panel). Some of the genes found dysregulated in the same direction between the in-house and Tervaniemi's study were also consistently dysregulated in Tsoi analysis, including *STAT1*, *S100* and the GWAS genes *STAT3* and *IFIH1*. Moreover, the GWAS gene *NOS2* was also a shared up-regulated hit with Tsoi's data, which was found to be dysregulated in Tervaniemi's analysis. The genes showing dysregulation in the opposite direction included *ALOX15B*, *ARG2*, *LCE6A*, *MGST1*, *PNLIPRP3*, *TLDC1* and *UBL3*. For example, *LCE6A*, a member of the *LCE* family involved in the synthesis of the later cornified envelope layer, was down-regulated in lesional skin in our study, in contrast to the up-regulation found in the Tsoi analysis. Notably, down-regulation of genes from the *LCE* family, including *LCE1B*, *LCE1F* and *LCE2A*, showed opposite direction of change in both, Tsoi and Tervaniemi's analyses. A study performing qPCR quantification of *LCE* genes from groups 1, 2, 5 and 6 demonstrated increased expression in psoriasis lesional skin, in line with the in-house results (Bergboer et al. 2011). In contrast to the other genes from the *LCE* family, all three datasets presented up-regulation of the GWAS risk associated gene *LCE3B*.

Overlap across the three studies only identified 212 DEGs shared by the three datasets. Despite having a 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. This may suggest, amongst other things, some of the DEGs being specific to the type of biopsy used on each approach.

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

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

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 KC 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 commented in the DGE 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), and was also shown to have increased expression when comparing lesional skin versus healthy biopsies (Li2014). One of the predicted miR-146a targets by Target Rank software in a study conducted by Jazdzewski and colleagues revealed *NFAT5*, also down-regulated in lesional skin compared to uninvolved in the in-house data, as the 11th most confidently predicted target (Jazdzewski et al. 2009). Interestingly, a negative correlation ($R=-0.981$, $pval=5.3\times 10^{-4}$) between the normalised counts of the two genes was found in the three lesional-uninvolved paired samples (Figure 4.16 a).

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. In our data, *HIF-1A* showed up-regulation in lesional skin compared to uninvolved, and a trend for positive

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

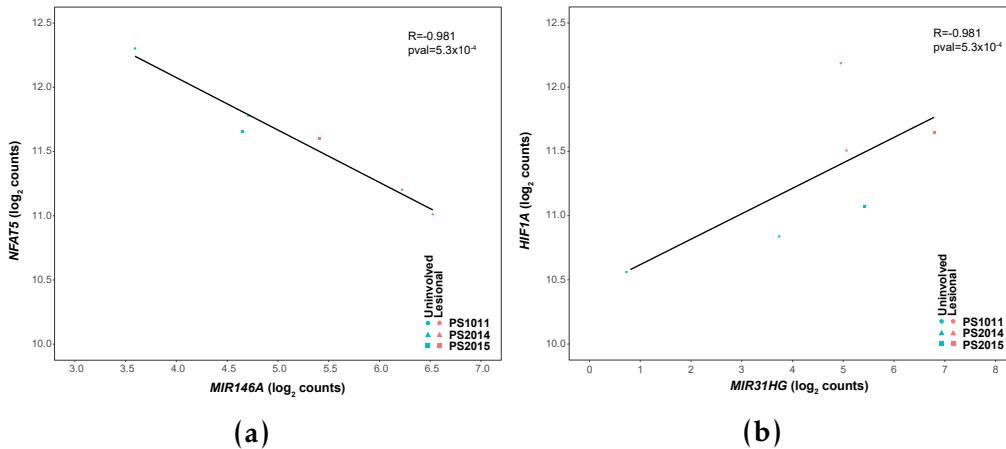


Figure 4.16: Correlation in gene expression between two dysregulated miRs in lesional skin and their putative target genes. Plots showing the correlation in \log_2 normalised read counts for a) *MIR146A* and its putative target genes *NFAT5* and b) *MIR31HG* and its putative target genes *HIF-1A*. Pearson correlation values (*R*) and significance (*pval*) are included. Each of the dots represents one sample, where colour represents condition (lesional or uninvolved) and shape corresponds to the patient ID.

correlation ($R=0.690$, $pval=0.12$) between normalised counts of this gene and the putative regulator *MIR31HG* was also observed (Figure 4.16 b).

Pathway enrichment analysis

In order to better understand the functional role of the DEGs ($FDR<0.05$) between lesional and uninvolved epidermis from psoriasis patient skin biopsies, pathways enrichment analysis was performed. A considerable number of pathways were significantly enriched ($FDR<0.005$) for DEGs found in our analysis (Table 4.12 and A.4).

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 amino acids and derivatives. HIF-1 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 the GWAS gene *NOS2*, amongst others, contributed to the enrichment of this pathway (Figure 4.18).

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

Lesional versus uninvolved epidermis enriched pathways	
IFN- α/β /signalling	
Peroxisome proliferator-activated receptors (PPAR) signalling	
NOD-like receptor signaling pathway	
IL-17 signalling	
IL2-mediated signalling	
G protein coupled receptor (GPCR) ligand binding	
Hypoxia-inducible factor 1 (HIF-1) signalling	
Cytokine signalling in immune system	
Cell cycle	
Apoptosis	
Arginine and proline metabolism	

Table 4.12: Most relevant pathways enriched for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies. Significant pathways for FDR<0.005. The analysis was performed using significantly DEGs FDR<0.05 and no FC threshold. Enriched pathways had a minimum of ten members overlapping with DEGs.

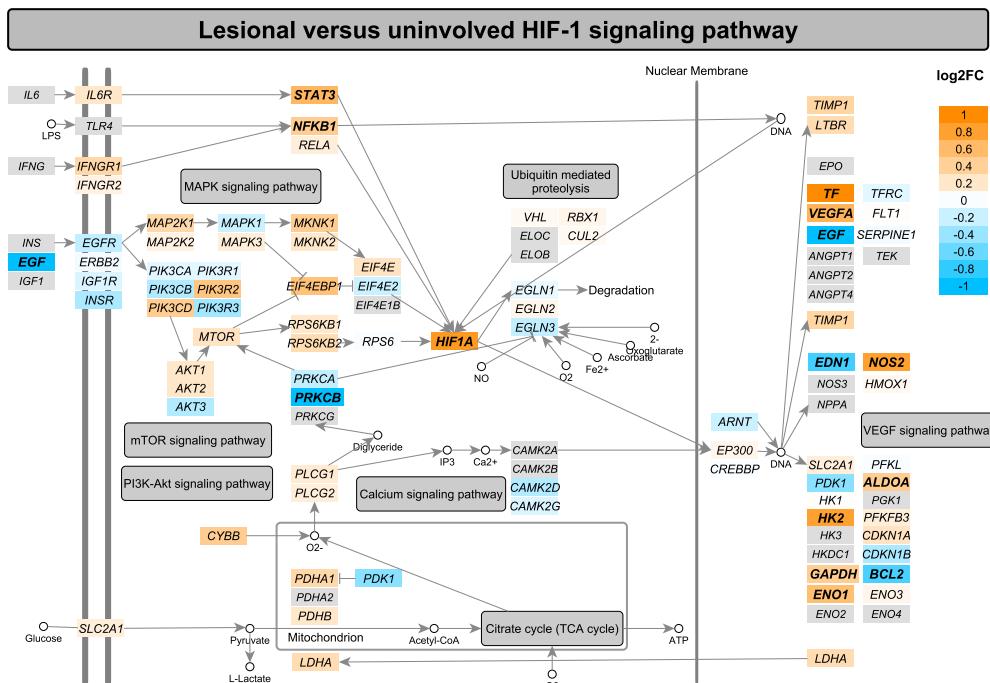


Figure 4.17: Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the HIF-1 signalling pathway. This pathway was sourced from KEGG, manually curated in a way that all member genes are maximised visually and then automatically color-coded by the log₂FC expression 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

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- κ B and MAPK activation, appeared enriched with 23 significantly DEGs (Figure 4.18 in orange and bold). Some of the most up-regulated genes contributing to the enrichment included *NOD2*, *CARD6* or *IFI16*, amongst others, and they were also up-regulated in Tervaniemi's data, where 42 DEGs mapped to this pathway. Amongst the down-regulated genes contributing to this pathway were *TNFAIP3* and *BCL-2* (Figure 4.18 in blue and bold). Performing pathway enrichment analysis using the DEGs from Tsoi and colleagues failed to show significant enrichment for NOD-like receptors signalling (19 DEGs in the NOD-like pathway). Nevertheless, NOD-like receptors signalling remained significantly enriched (13 DEGs mapping to this pathway) when the analysis was conducted using only the DEGs from the in-house data not overlapping the Tsoi and colleagues results.

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

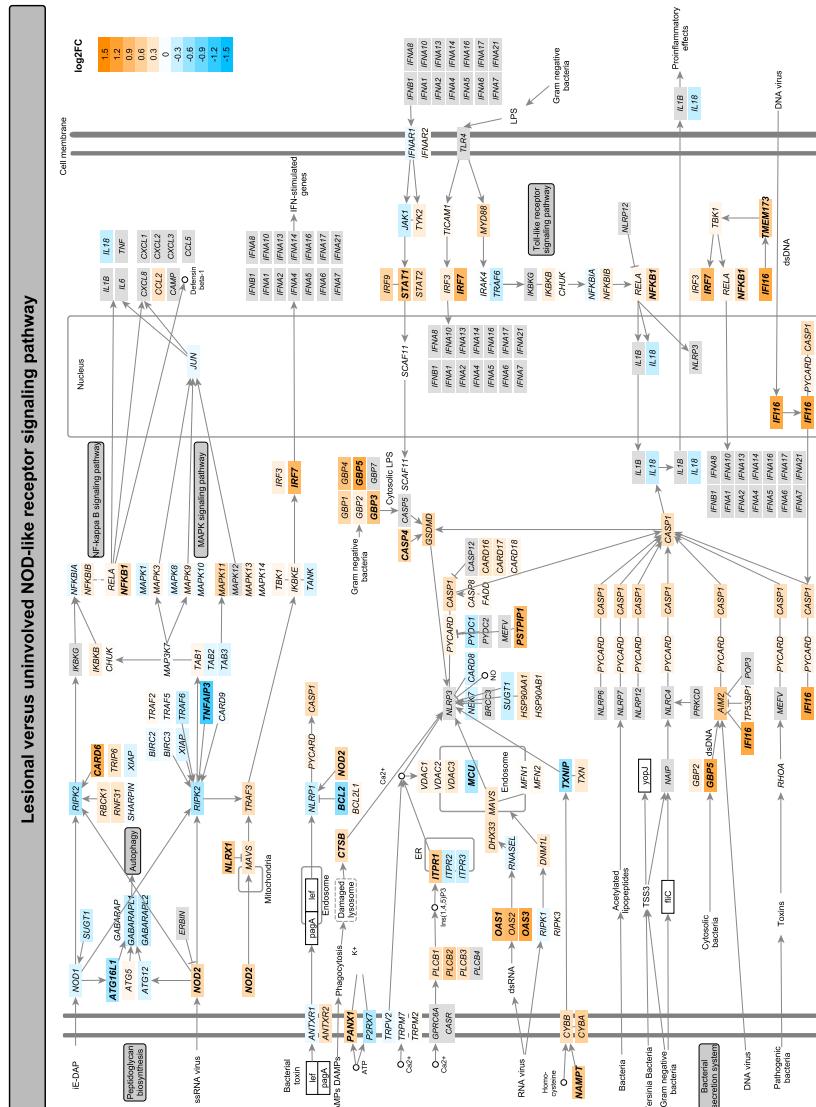


Figure 4.18: Mapping of the DEGs between lesional and uninvolved epidermis from psoriasis patients onto the NOD-like signalling pathway. This pathway was sourced from KEGG, manually curated in a way that all member genes are maximised visually and then automatically color-coded by the \log_2FC expression 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 the NOD-I 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, similarly to Tsoi and Tervaniemi's data, which found up-regulation for a number of these genes. Dysregulation for other IL-17 ligands and receptors or IL-23A was neither detected, in contrast to the observations from Tsoi and Tervaniemi. Moreover, 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 δ (*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 the systemic and tissue-specific gene expression signatures in psoriasis

In order to describe 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 dysregulated genes in lesional skin compared to uninvolved and the DEGs identified in circulating immune cells, with CD14 $^{+}$ monocytes and tCD8 $^{+}$ cells showing the greatest overlap. A similar or larger proportion of the total overlapping DEGs presented opposite direction of change in circulating immune cells and in skin from psoriasis patients. An example was *TNFAIP3* gene, which was up-regulated in psoriasis tCD4 $^{+}$ and tCD8 $^{+}$ cells

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

compared to controls and down-regulated in lesional epidermis when compared to uninvolved.

Another two relevant transcript showing opposite dysregulation were the early growth response 1 and 3 (*EGR1* and *EGR3*) genes, two genes involved in maintenance of the homeostasis of the adaptive immune response (Li et al. 2012). Both genes were up-regulated in CD14⁺ monocytes in psoriasis compared to controls and *EGR2* was also up-regulated in tCD4⁺ cells. Conversely, great up-regulation of *EGR2* and *EGR3* ($\log_2\text{FC}$ -0.74 and -0.53) was observed in lesional skin when compared to uninvolved. Interestingly, no overlap was observed for genes of the *S100* family, only found to be up-regulated in lesional skin. Lastly, the previously described *MIR146A* also appeared dysregulated in opposite direction in tCD8⁺ cells and in the skin analysis, presenting down- and up-regulation, respectively.

DEGs overlapping with skin	Total overlap	Same direction	Opposite direction
CD14 ⁺ monocytes	37	19	18
tCD4 ⁺	10	6	4
tCD8 ⁺	37	24	13
CD19 ⁺	16	5	11

Table 4.13: 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⁺ and tCD8⁺ DEGs were mostly immune-related pathways, including TCR , IL-12 , TNF and NF- κ B signalling. Moreover,some of the pro-inflammatory genes contributing to those pathways appeared down-regulated in psoriasis when compared to controls, as previously commented.In skin, the DEGs in lesional epidermis when compared to uninvolved were not only enriched in immune-related pathways but also for pathways involved in metabolism,

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

oxidative stress and cell cycle. In contrast to the systemic observation, the dysregulation of genes contributing to the enrichment of immune-related pathways appeared to present a more clear pro-inflammatory signature, which would be 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 in circulating immune cells

The characterisation of the chromatin accessibility landscape and the transcriptome in circulating immune cells from psoriasis patients has revealed a greater effect of disease in gene expression than in chromatin accessibility. In order to assess to some extent the relationship between the two, overlap between DEGs and the genes proximal to DARs ($\leq 5\text{Kb}$) was performed. 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*).

An example was T Cell lymphoma invasion and metastasis 1 (*TIAM1*), which activates IL-17 expression and T cell transendothelial migration during inflammation (Kurdi et al. 2016; Grard et al. 2009). This gene showed an increased expression ($\log_2\text{FC} 0.44$) in psoriasis patients tCD8 $^{+}$ cells (Figure 4.19 left). Likewise, psoriasis CD8 $^{+}$ cells presented greater chromatin accessibility compared to healthy controls ($\log_2\text{FC} 0.41$) in a region located at an intron of the *TIAM1* gene and annotated as an active enhancer according to the Roadmap chromatin segmentation data in this cell type (Figure 4.19 right). Common SNPs within this peak did not appear to be an eQTL regulating expression of any gene in tCD8 $^{+}$ 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

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

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). Both genes presented up-regulated expression and increased chromatin accessibility in psoriasis patients tCD8⁺ cells compared to healthy controls.

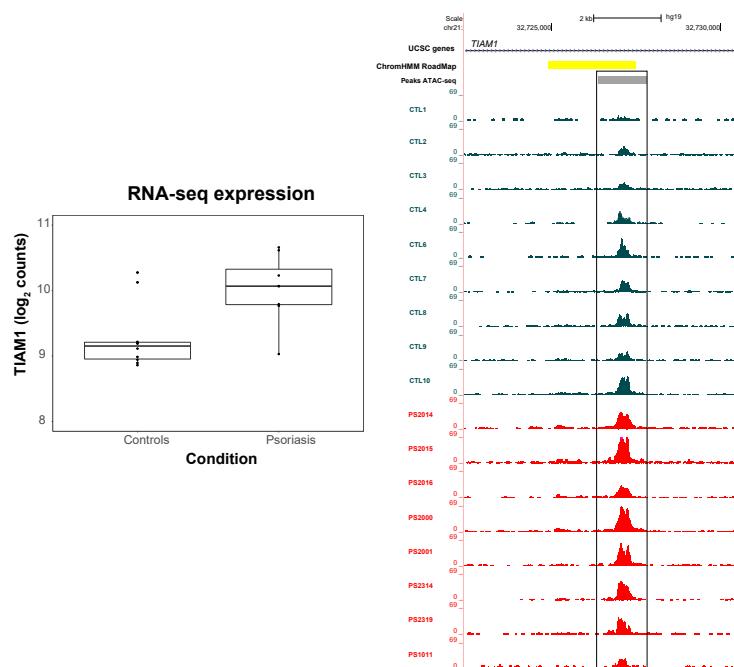


Figure 4.19: Differential gene expression and chromatin accessibility landscape for *TIAM1* gene in tCD8⁺ cells. Boxplot on the left represents RNA-seq log₂ normalised counts for the *TIAM1* gene in psoriasis and healthy controls tCD8⁺ cells. The right panel is the UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) at an intron of the *TIAM1* gene (x-axis) in tCD8⁺ cells. This region was 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 Epigenome Roadmap chromatin segmentation map for tCD8⁺ cells is included.

4.3.8 Fine-mapping of psoriasis GWAS loci and functional interpretation

Fine-mapping using summary statistics data

In absence of permission to access 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 cohort genotyped using the Immunochip in Tsoi and colleagues' publication (PAGE cohort) were not publicly available through ImmunoBase at the time of this analysis. As explained in Chapter 2, fine-mapping from summary statistics with DIST uses the statistics z-score of each of the genotyped SNPs to impute 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, association analysis is performed and ABF, PP and credible set of SNPs are built for each of the signals.

Fine-mapping was performed for 26 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 with highly confident predicted damaging effects. Out of the 26 regions, 9 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 A.5). Of the 17 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, which was not included in the credible set (Table 4.14 with *). This is likely due to the effect of reduced sample size, (only GACP cohort) compared to Tsoi *et al.*, 2012, on the ability to identify association signals (Bunt *et al.* 2015).

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

Table 4.14: Summary table of the psoriasis Immunochip GWAS loci presenting $\log_10ABF > 3$ for the fine-mapping lead SNP. For 16 of the Immunochip psoriasis GWAS loci presented $\log_10ABF > 3$ the table reports the closer gene(s), FM lead SNP, MAF, OR for the GAPC cohort, \log_10ABF for the FM lead SNP, PP, number of SNPs included in the 90% credible set and the Tsoi *et al.*, 2012 GWAS lead SNP. In 10 of those loci (*) the fine-mapping lead SNP was in low LD ($r^2 < 0.5$) with the psoriasis GWAS SNP and did not contain it in the credible set. FM=fine-mapping; MAF=minor allele frequency; OR=odds ratio; ABF=approximate Bayes factor; PP=posterior probability.

chr	Closer gene	FM lead SNP	MAF	OR	\log_10ABF	PP set	90% credible set	Tsoi lead SNP
1	<i>IL28RA</i>	rs61774731*	0.10	1.21	11.5	0.99	1	rs7552167
2	<i>FLJ16341/REL</i>	rs6714339*	0.14	1.17	9.0	0.99	1	rs62149416
2	<i>IFIH1</i>	rs2111485*	0.59	1.27	6.7	0.50	2	rs17716942
5	<i>TNIP1</i>	rs17728338	0.06	1.59	10.6	0.40	6	rs2233278
5	<i>IL12B/ADRA1B</i>	rs12188300	0.05	1.58	11.2	0.18	9	rs12188300
6	<i>TNFAIP3</i>	rs1416173*	0.85	1.23	7.7	0.15	10	rs582757
14	<i>NFKBIA</i>	rs74243591	0.21	1.16	5.0	0.30	12	rs8016947
17	<i>NOS2</i>	rs117094752*	0.02	1.22	7.3	0.94	1	rs28998802
1	<i>SLC45A1/TNFRSF9</i>	rs425371	0.25	1.13	5.6	0.14	22	rs11121129
1	<i>RUNX3</i>	rs61774731 *	0.10	1.13	11.5	0.99	1	rs7536201
2	<i>B3GNT2/TMEM17</i>	rs9309343*	0.33	1.12	4.3	0.66	34	rs10865331

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

7	<i>ELMO1</i>	rs77840275*	0.10	1.11	7.5	0.99	1	rs2700987
11	<i>ZC3H12C</i>	rs11213274	0.40	1.14	4.0	0.05	69	rs4561177
16	<i>PRM3/SOCS1</i>	rs111251548*	0.02	1.13	9.4	0.97	1	rs367569
17	<i>PTRF/STAT3</i>	rs963986	0.17	1.15	4.1	0.18	8	rs963986
19	<i>ILF3/CARM1</i>	rs34536443*	0.03	1.17	9.7	0.93	1	rs892085

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

The 6 loci including the Tsoi *et al.*,2012 lead GWAS SNP presented a number of SNPs in the 90% credible set of SNPs ranging from 6 to 69 (Table 4.14). Of those 6 loci, *TNIP1*, *IL12B/ADRA1B* and *PTRF/STAT3* had 90% credible sets which included 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.

Integration with functional data

A total of 126 unique SNPs formed the union of 90% credible sets from the 6 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⁺ monocytes, tCD4⁺, tCD8⁺ and CD19⁺ cells. Conversely, overlap with the consensus master list ATAC peaks from each of the cell types revealed a total of 16 SNPs from 5 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)). However, no overlap was found for the 9 SNPs of the *IL12B/ADRA1B* 90% credible set. CD14⁺ monocytes appeared as the cell type showing the largest proportion of accessible chromatin regions containing SNPs from the credible (2.3%), followed by CD19⁺, tCD4⁺ and tCD8⁺ cells (1.7, 1.6 and 1.05%, respectively) (Table 4.15). Altogether, integration of the SNPs from the credible set with ATAC accessible regions in four cell types allowed to further refine the number of genetic variants with a putative functional role in psoriasis for 5 of the 6 analysed loci. Moreover, in the *PTRF/STAT3* and

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

SLC45A1/TNFRSF9 loci all the SNPs from the credible set overlapping accessible chromatin appeared to be cell type specific (Table 4.14 and 4.15). Of the 8 SNPs in the *PTRF/STAT3* credible set, only 1 overlapped accessible chromatin in CD14⁺ 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⁺ monocytes-specific.

ATAC cell type master list	90% credible set overlapping SNPs (number)	Cell type specific overlap
CD14 ⁺ monocytes	13	<i>PTRF/STAT3</i> (1), <i>SLC45A1/TNFRSF9</i> (4)
tCD4 ⁺	5	None
tCD8 ⁺	3	<i>TNIP1</i> (2)
CD19 ⁺	5	None

Table 4.15: 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 126 SNPs) from the 6 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⁺ monocyte ATAC-specific peaks was the fine-mapping lead SNP rs425371, which is located at an intergenic region, approximately 269.3Kb downstream *TNFRSF9* gene. The other three SNPs overlapping CD14⁺ monocytes accessible chromatin included rs11121131, rs12745477 and rs417065.

4.3.9 Maximising genetic commonalities across chronic inflammatory diseases: 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 the chronic inflammatory diseases AS and CD. Interestingly, Reveille's AS GWAS study reported the same lead SNP as the psoriasis Immunochip (Reveille et al. 2010). The later AS Immnochip study identified rs6759298 (OR=1.29) as the tag SNP for this region, which is in high LD ($r^2=0.84$) with rs10865331 (Cortes 2012). A recent GWAS meta-analysis combining data for five chronic inflammatory diseases also identified association for the chr2p15 locus, confirming the same association and direction to be shared by the three out of five phenotypes (Ellinghaus et al. 2016). Therefore the AS and psoriasis associations are considered to be the same signal and likely to share the functional mechanism increasing the risk of a dysregulated inflammatory response.

As previously presented, fine-mapping using summary statistics from the psoriasis GACP Immunochip cohort failed to successfully fine-map this locus ($\log_{10}BF_i3$). In contrast, genotype-based fine-mapping analysis performed in collaboration with Dr Anna Sanniti in the UK Immunochip AS data successfully identified an independent signal at chr2p15 ($\log_{10}BF=18.43$) tagged by the SNP rs4672505. The refinement of this association signal in AS yielded a 95% credible set containing only three SNPs. Out of the three identified SNPs, rs4672505 accounted for 40% of the association in chr2p15 locus whereas the additional two SNPs (rs6759298 and rs6759003) explained together 60% of this association. Interestingly, the SNP rs4672505 was also the lead SNP for the chr2p15 signal identified in the multi-disease meta-analysis from Ellinghaus and colleagues, where the risk allele was found to be A, similar to the results in the AS GWAS association analysis performed for fine-mapping (Ellinghaus et al. 2016).

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

Integration of ATAC and publicly available epigenetic data

rs4672505 overlaps an accessible ATAC region included in the ML_CD8 and not present in CD14⁺ monocytes, tCD4⁺ and CD19⁺ cells (Figure 4.20 a). The tCD8⁺ accessible region harbouring rs4672505 was not a DAR between patients and controls and appeared to have variability across individuals unrelated to disease status (Figure 4.20 b).

For example, PS2314 and CTL1 showed no ATAC signal at this location when compared to PS2000 and CTL4. Integration with publicly available ENCODE DHS data also revealed accessible chromatin at rs4672505 in Th-1, Th-2 and Th-17 cells from ENCODE (Figure 4.20 a). Although this region was tagged as quiescent according to the tCD8⁺ chromatin segmentation, Epigenome Roadmap primary tCD8⁺ DNase-I digital genomic footprinting signal was found to overlap with the location of rs4672505 and the in-house ATAC peak (Figure 4.20 b), and may suggest a putative *cis*-regulatory role for this region. The absence of histone marks and chromatin accessibility leading to annotation of this chromatin segment as quiescent by the Epigenome Roadmap could be explained by the variability across individuals found at this location. For example, H3K27ac data generated in cohort 1B demonstrated modest signal enrichment in CTL7, which also presented accessible chromatin at rs4672505; however no enrichment was detected for PS2314 (Figure 4.20 b). Regarding TFs, ENCODE ChIP-seq experimental data from GM12878 showed binding for RUNX3, a psoriasis and AS GWAS associated gene, with nominal association also in the PsA Immunochip GWAS study. Importantly, RUNX3 together with a number of TFs is involved in CD8⁺ cell differentiation (Wong et al. 2011). Further investigation using *in silico* TFBS prediction, such as 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

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

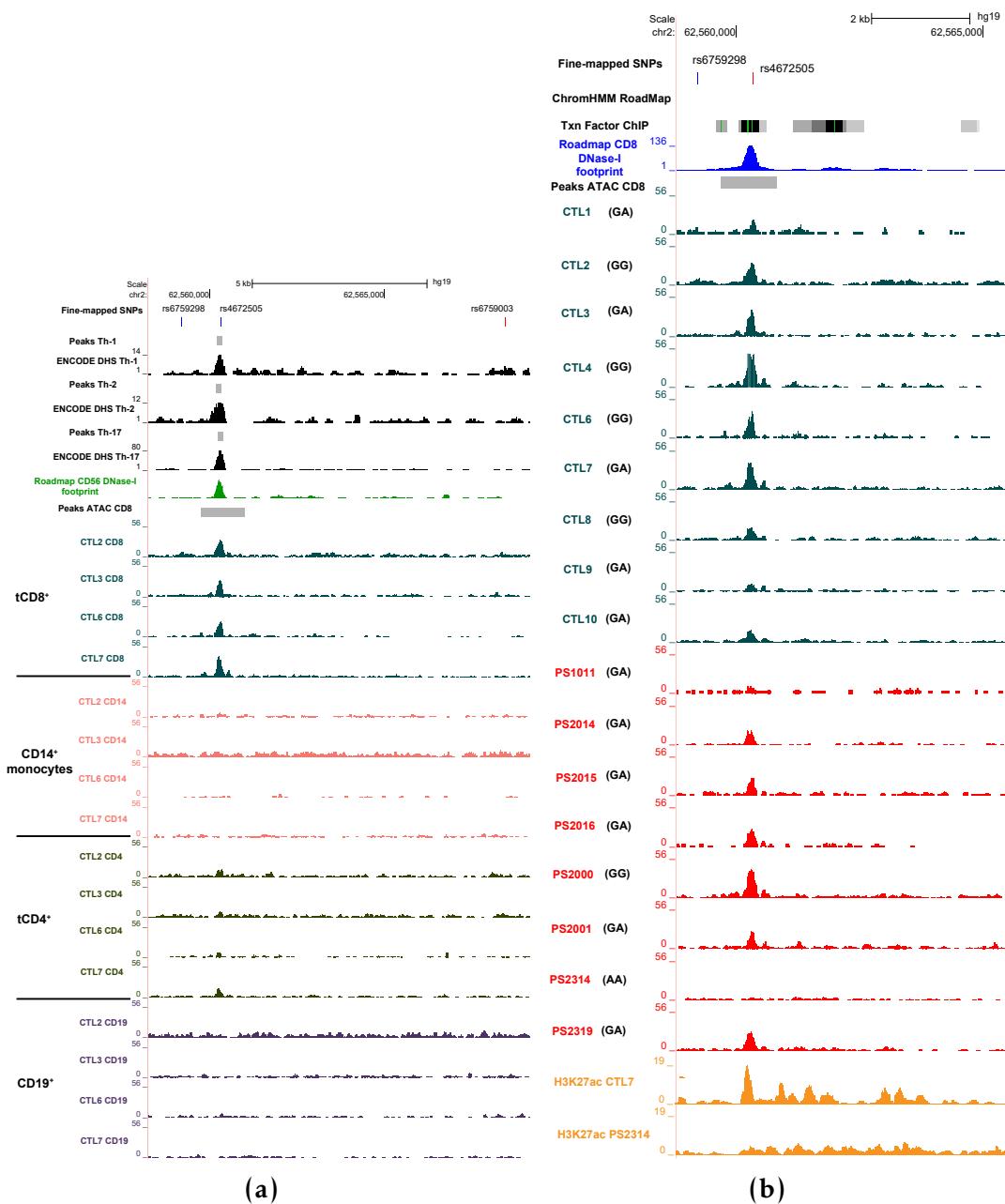


Figure 4.20: Epigenetic landscape at the location of the SNPs in the 95% credible set for the chr2p15 GWAS psoriasis locus. a) UCSC Genome Browser view illustrating normalised read density for in-house ATAC and a number of other publicly available epigenetic data (DHS, DNase-I footprint, chromatin segmentation map) (y-axis) at the location of the three SNPs (rs6759298, rs4672505 and rs6759003) (x-axis) from the 95% credible set obtained in the fine-mapping analysis of the chr2p15 GWAS association in AS. Representative ATAC data from the same four controls in the cohort and the four cell types included in this study are shown. b) UCSC Genome Browser view illustrating the normalised read density for tCD8⁺ 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

the three fine-mapped SNPs included in the 95% credible set, to have a functional role explaining the association of chr2p15 with psoriasis risk.

Allele-dependent chromatin accessibility and allelic imbalance using ATAC reads at rs4672505

The genotype of each individual at the rs4672505 SNP was characterised using Sanger sequencing. Amongst the eighteen samples (ten controls and eight psoriasis patients), one (PS2314) was homozygous for the risk allele (A, MAF=0.43), eleven were heterozygous and six were homozygous for the protective allele (G) (Figure 4.20 b genotypes in brackets). Interestingly, PS2314, the only homozygous individual for the risk allele, showed complete absence of the peak at rs4672505. In order to further investigate the role of rs4672505 genotype in the variability of chromatin accessibility across individuals, the normalised read counts retrieved at the chr2:62,559,749-62,561,442 ML_CD8 peak were used as the dependent variable in linear model analysis based on rs4672505 genotype, using batch as a covariate.

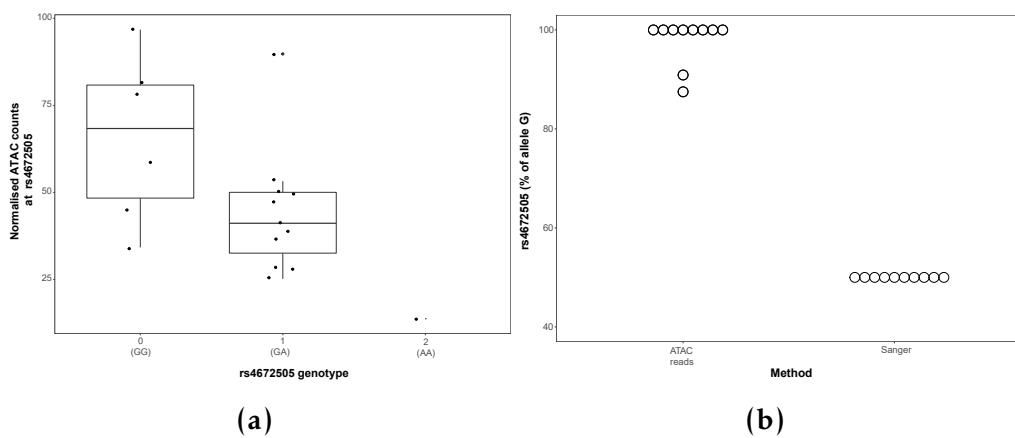


Figure 4.21: Effect of rs4672505 genotype in tCD8⁺ cells chromatin accessibility at chr2:62,559,749-62,561,442. a) Boxplot illustrating the effect of the rs4672505 genotype in chromatin accessibility at the chr2:62,559,749-62,561,442 ATAC peak. Log₂ normalised ATAC counts adjusted for batch effect, also 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.

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

Significant negative correlation ($pval=0.035$) 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 a tendency towards greater chromatin accessibility in presence of the fine-mapped protective allele rs4672505(G) at the chr2p15 locus.

Potential regulatory role of rs4672505 in the expression of B3GNT2

As previously mentioned, one of the issues with intergenic GWAs signals is the difficulty in determining the gene they may be having an effect on through regulation of gene expression. rs4672505 is located 140Kb downstream of the *B3GNT2* gene and 150Kb upstream of *TMEM1*. Publicly available promoter capture data from Javierre *et al.* 2016 in tCD8⁺ revealed a genome-wide significant interaction (CHiCAGO score=7.67) between a region containing rs6759298 and rs4672505 and the promoter of the *B3GNT2* gene (Figure 4.22 a).

This interaction was not found in any of the additional sixteen human primary hematopoietic cell included in the study. Moreover, no upstream interaction with *TMEM1* promoter was identified. Investigation of the publicly available T cell eQTL dataset from Kasela *et al.* and Raj *et al.* did not show a significant eQTL for this SNPs or SNPs in high LD ($r^2>0.8$) either in tCD8⁺ or tCD4⁺ (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). Conversely, whole blood eQTL study from Jansen and colleagues revealed a

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

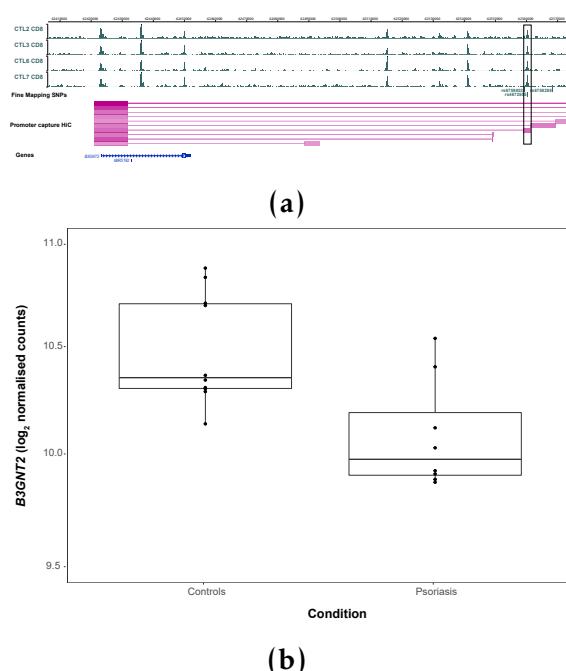


Figure 4.22: Potential role of rs4672505 in regulating B3GNT2 gene expression. a) WASHU Genome Browser track showing in tCD8⁺ cells normalised ATAC read density in four of the healthy controls, the location of the three SNPs of the 95% credible set and the promoter capture HiC data depicting the regions interacting with the bait at the B3GNT2 promoter. b) Boxplot illustrating the B3GNT2 \log_2 normalised RNA-seq counts adjusted for batch effect in the psoriasis and healthy control groups.

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

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, significant (FDR<0.05) moderate down-regulation was observed in psoriasis patients when compared to controls in tCD8⁺ (Figure 4.22 b) and CD19⁺ cells (-log₂FC=-0.317 and -log₂FC=-0.317, respectively). *B3GNT2* was not significantly dysregulated in lesional skin when compared to uninvolved. This data confirms dysregulation of *B3GNT2* expression in psoriasis and may suggest a role for rs4672505 in the regulation of *B3GNT2* expression in tCD8⁺ only under certain conditions.

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 tCD8⁺ cells had the largest number of discrete changes. In ATAC, greater accessibility in tCD8⁺ 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 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* gene, which participates in initiation of DNA replication and is associated with aspirine-intolerance in

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

asthmatics(Pasaje et al. 2011). However, evidence of *DTD1* involvement in chronic inflammation has yet not been reported. 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 the studied circulating immune cells. Recent similar studies performing ATAC in B cells from SLE patients and in AMD retina and retinal pigmented epithelium have revealed larger differences in the chromatin accessibility landscape between patients and controls (Scharer et al. 2016; Wang et al. 2018). Similarly, H3K27ac mapping in mCD4⁺ cells isolated from juvenile idiopathic arthritis SF found approximately one thousand differential enhancers when compared to healthy control circulating cells (Peeters et al. 2015). Conversely only small differences were found when comparing mCD4⁺ from PB of patients and controls, highlighting the specificity of the disease signature at the site of inflammation. Importantly, direct interrogation of the main cell type or tissue affected by inflammation in those studies would partly explain the more profound changes observed in the chromatin landscape compared to my study. Additionally, some differences will be driven by genotype, and by the nature of complex diseases different patients have different genetic backgrounds, with some variants also shared with control individuals. Thus it may be necessary to study changes in chromatin accessibility in the context of genotype and under exogenous inflammatory stimuli that may manifest those differences (Calderon2018; Alasoo et al. 2018).

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⁺ monocytes and tCD8⁺ 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 tCD8⁺ compared to tCD4⁺ may suggest that, as in skin, CD8⁺ are the main effector cells upon induced-activation by CD4⁺ 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- α production contributes towards maintenance of inflammation (Nickoloff2000; Wang et al. 2006).

The overlap of DEGs with previous studies comparing PBMCs from psoriasis patients was limited, probably due to many differences identified in my cell type specific analysis being masked by the admixture of cells as well as those studies using microarrays instead of RNA-seq (Lee et al. 2009; Coda et al. 2012). Although those studies did not find specific enrichment for any pathway, Coda and colleagues identified some genes were associated with pathophysiological processes such as immune response, oxidative stress or apoptosis (Coda2012.).

The cell type specific analysis conducted in my thesis identified significant enrichment of relevant biological processes, including MAPK and IL-12 signalling, in the CD14⁺ monocytes and tCD8⁺ cells contrast. Interestingly, some of the well-known pro-inflammatory genes contributing to the enrichment of these pathways were down-regulated in psoriasis compared to controls. For example, *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 α . In the IL-12 signalling pathway, leading to T cell

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

proliferation and IFN- γ production through activation of TFs from the STAT family, CD14 $^{+}$ cells presented down-regulation of *STAT4* and *STAT5A* in psoriasis versus controls. Other member of the STAT family, such as *STAT2*, was found to be down-regulated in psoriasis PBMCs and in AS monocyte-derived macrophages when compared to controls (Coda et al. 2012; Smith et al. 2008). Monocytes do not express *STAT4* in basal conditions but up-regulation follows IL-12, IL-18 and IFN- α stimulation (Frucht et al. 2000; Schindler et al. 2001). STAT5 phosphorylation in monocytes is mainly induced by granulocyte macrophage-colony stimulating factor (GM-CSF) and promotes differentiation into macrophages. Interestingly, STAT5 downstream gene targets such as prostaglandin synthase 2 (*COX2*) and *IL-10* were not dysregulated in CD14 $^{+}$ monocytes in my data. In other chronic inflammatory diseases such as T2D, persistent STAT5 phosphorylation has been found in circulating monocytes isolated from T2D upon GM-CSF (Litherland et al. 2005). Further investigation to determine phosphorylated STAT4 and STAT5 protein abundance will be required to determine if the down-regulation at the transcript level observed in psoriasis CD14 $^{+}$ monocytes is biologically relevant. In tCD8 $^{+}$, expression of *IFNG* a gene activated by the IL-12 signalling pathway, was down-regulated when compared to healthy controls. Down-regulation of *IFNG* has previously been reported in unstimulated and stimulated macrophages derived from AS patients, in SF from SpA patients compared to RA and in a SpA rat model (Smith et al. 2008; Fert et al. 2014). This down-regulation was accompanied by an overall inverse transcriptional response of IFN-regulated genes, which was not seen in my data. Moreover, the 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 (Cañete et al. 2000; Chu et al. 2007). Therefore, down-regulation of *IFNG* may actually result in a pro-inflammatory effect.

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

In tCD8⁺ 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⁺ monocytes and CD4⁺ cells respectively, and both associated with psoriasis GWAS signals. Polymorphisms within *TNFAIP3* or in its vicinity have also been associated with a number of chronic inflammatory diseases including MS, RA, SLE and T1D (Vereecke et al. 2011). *NFKBIA* codes for IκBα which inhibits NF-κB by binding it and preventing translocation to the nucleus. *TNFAIP3* codes for the zinc finger protein and ubiquitin-editing enzyme A20, and is up-regulated in the presence of inflammation and NF-κB activation in order both to inhibit the NF-κB TNF-mediated response and promote return to homeostasis. Either *NFKBIA* or *TNFAIP3* were found to be 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 (PASI<4.84) and severe (PASI>4.84) psoriasis vulgaris revealed a significant negative correlation between *TNFAIP3* expression and disease severity (Jiang et al. 2012). Furthermore, this study also demonstrated that in the mild group of patients but not in the severe *TNFAIP3* expression was down-regulated when compared to healthy control PBMCs. This is in line with my findings with the caveat that all patients from my cohort would be classified as severe by Jian *et al.*. Altogether, the up-regulated expression of *TNFAIP3* and *NFKBIA* compared to healthy controls may not be unexpected as it reflects a persistent inflammatory stimuli in psoriasis PB and a

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

mechanism that limits the systemic inflammatory response to some extent (Idel et al. 2003).

Of interest was also the up-regulation of the chemokine receptor *CCR10* in tCD8⁺ cells from psoriasis patients. 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 CCR10⁺ infiltrated T lymphocytes in psoriatic skin, where KCs 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⁺ CCR10⁺ cells ready to migrate into the skin lesions. Moreover, correlation between the frequency of CTLA⁺ CD8⁺ cells and disease severity measured by PASI score has been found (Sigmundsdttir and 2001). Overall, these data have revealed dysregulation between psoriasis patients and controls for relevant immune genes showing pro- and anti-inflammatory effects in circulating immune cells. Although down-regulation of pro-inflammatory genes and up-regulation of anti-inflammatory genes has been detected, understanding the overall effect of those interactions in the inflammatory response requires further investigation.

4.4.3 Correlation between changes in chromatin accessibility and gene expression

In this chapter, greater changes in gene expression have been identified compared to chromatin accessibility. Strikingly, in tCD8⁺ 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. Correlation between chromatin accessibility measured by ATAC and gene expression has been reported to some extent in a number of studies, with limitations in

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

establishing relationships between enhancer regions and the regulated target (Ackermann et al. 2016; Wang et al. 2018).

Chromatin accessibility shows the current functional landscape of the genome and will have some correlation with the current transcriptional state of the cell. The RNA transcripts however will be both a view of the current transcription and previous changes in transcription as well as other events related to RNA turnover and lncRNAs and microRNAs interactions and regulatory processes at the RNA level. As such, the larger changes in gene expression compared to chromatin accessibility here could be due to a non-direct relationship between chromatin accessibility and transcripts. Also, the discrepancy may also be due to different sensitivities from the ATAC assay to identify changes in chromatin accessibility and that of the RNA-seq analysis. An example of a relevant DEG nearby a DAR was *TIAM1*, with increased chromatin accessibility and gene expression in psoriasis tCD8⁺ cells compared to healthy controls. *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.

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 in my cohort. Most previous transcriptional studies in psoriasis have used full thickness skin biopsies, formed of a mix of cell types including fibroblast, adipocytes, KCs from the epidermis and dermis and infiltrated immune cells. A study from Ahn and colleagues demonstrated large differences in gene expression between whole biopsies and FACS-isolated KCs, which may be

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

masking KC-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 KCs 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- γ 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). This suggests that the three psoriasis patients in the study are heterozygous for the deletion.

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 DGE 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). The enrichment of the HIF-I pathway in psoriasis is the result of an increased rate of cell proliferation leading to hypoxia and angiogenesis. Up-regulation of the hypoxia-inducible TFs HIF-1 α and HIF-2 α has been found in lesional skin, correlated 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 my data, 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 2013).

Immune-related pathway enrichment were also found in my analysis, including Th-17, IL-12, cytokine-cytokine 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 KCs. NOD-like signalling involves signal transduction by NOD-like receptors, a type of pattern-recognition receptors, which can recruit and activate caspases

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

into the inflammasomes or trigger inflammation through NF- κ B and MAPK. Amongst the genes contributing to 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 a pathway linking 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). Similar to my study, *PPARD* has been found to be up-regulated in psoriatic lesional skin, and molecular studies have demonstrated a role of this PPAR in KC hyperproliferation through induction of heparin-binding EGF-like growth factor (HB-EGF) (Romanowska et al. 2008).

4.4.5 LncRNAs in psoriasis

In addition to protein coding genes, the transcriptomic comparisons conducted in my study revealed dysregulation of a number of lncRNAs. The role of lncRNAs has been studied in RA, SLE, AS, and PsA (Muller2014; Shi et al. 2014; Zhang et al. 2017; Dolcino et al. 2018) but no study has been conducted to identify differentially lncRNAs in a cell type-specific manner in psoriasis PB. Conversely, several studies have contrasted lncRNAs in lesional compared to uninvolved or healthy skin (Li2014; Gupta2016; Ahn et al. 2016; Tsoi et al. 2015). My analysis in circulating immune cells revealed the largest number of differentially expressed lncRNAs in CD14 $^{+}$ monocytes and tCD8 $^{+}$ (28 and 31, respectively for FDR<0.05) when comparing psoriasis patients to healthy controls. In skin, 46 lncRNAs showed dysregulated expression between lesional

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

and uninvolved skin ($\text{FDR} < 0.05$), and 24 had also been reported by Tsoi *et al.*, 2015.

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. For example, in psoriasis CD14⁺ monocytes the up-regulation of *HOTAIRM1* corresponded to down-regulation of the predicted target gene *USP1*, unlike in Dolcino *et al.* 2018. *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⁺ 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⁺ monocytes was *NEAT1*, which has also shown up-regulation in SLE CD14⁺ monocytes. Knock-down demonstrated impairment of TLR4 signalling and down-regulation of inflammatory genes including IL-6 and CXCL10 (Zhang et al. 2016). However, neither of those two genes was dysregulated in my study in this cell type.

Interestingly, *MIR146A* was differentially expressed between lesional and uninvolved skin but also when comparing psoriasis tCD8⁺ 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). *TRAF6* and *IRAK1* are adaptor molecules involved in the activation of kinases that eventually lead to translocation of NF- κ B and AP-1 into the nucleus. Opposite direction of change was observed in the two comparisons here. The up-regulation of *MIR146A* in psoriasis tCD8⁺ compared to controls is in line with findings in serum from SLE and early RA patients (Filkova; Wang et al. 2012). In contrast,

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

transcriptomic studies using PBMCs from plaque psoriasis and also similar studies in RA (including PBMCs, SFMCs and CD4⁺ isolated from both tissues, amongst others) have reported increased levels of miR-146a in patients when compared to controls (Ele-Refaei et al. 2015; Churov et al. 2015). Conversely, 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 (Lerman2014; Li2014; Tsoi et al. 2015). One of the predicted gene targets of miR-146a, the TF *NFAT5* was down-regulated in lesional skin and showed significant negative correlation with *MIR146A* expression. Interestingly, up-regulation of *NFAT5* has been reported in RA SF and has a role in mediating angiogenesis and proliferation of synoviocytes (Han et al. 2017). Although in disagreement with the observations in PBMCs, the down-regulation of miR-146a levels in tCD8⁺ cells would support failure of one of the check-points controlling sustained inflammation and the subsequent pathophysiological implications. In contrast, the up-regulation observed in lesional skin would not fit directly with the dysregulated inflammatory response in skin.

Other dysregulated non-coding RNAs in lesional epidermis relevant to psoriasis pathophysiology were *HG19* and *MIR31HG*, down- and up-regulated respectively in my data. Both non-coding RNAs were also differentially expressed in Tsoi et al., 2015 and are involved in KC differentiation. In particular, silencing miR-31hg in the KC immortal cell line HaCaT induced cell cycle arrest and inhibited cell proliferation, consistent with two characteristic aspects dysregulated in psoriatic KCs (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

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

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 blood and skin

Comparison of the dysregulated genes in circulating immune cells and in psoriatic skin revealed very limited overlap. Overall, FCs in the epidermis appeared to be larger than in circulating immune cells, likely due to a more profound pro-inflammatory environment driving gene expression changes in skin. CD14⁺ monocytes and tCD8⁺ cells showed the greatest overlap of DEGs with genes dysregulated between lesional and uninvolved epidermis (37 genes), consistently with the larger number of DEGs detected. However, almost half showed opposite directionality across the two comparisons. This is in line with the finding of Coda and colleagues when comparing DEGS genes in psoriasis patients and controls PBMCs to genes dysregulated between lesional and uninvolved skin biopsies. Genes showing opposite change in circulation and in skin included the GWAS gene *TNFAIP3*, *EGR2*, and *EGR3*. As previously mentioned, *TNFAIP3* down-regulation in lesional skin may reflect complete loss of an NF-*kappaB* 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. Importantly, loss of *EGR2* and *EGR3* expression leads to hyperactive STAT1 and STAT3 signalling, associated with SLE pathophysiology (Li *et al.* 2012). Down-regulation of *EGR2* and *EGR3* was not observed in Tsois study, and in Tervaniemis data *EGR3* appeared to be up-regulated in lesional skin. In addition to its role in

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

regulating inflammation, down-regulation of *EGR2* in skin may also increase KC proliferation as has been shown in certain types of cancer (Wu et al. 2010).

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 a metabolic dysregulation that appears to be characteristic of this site of inflammation. Moreover, immune related pathway such as NOD-like signalling also seemed to be specific for the dysregulated gene signature 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 and the chr2p15 locus

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, I performed fine mapping for x of the genome-wide significant loci. Integration of the credible set of SNPs from each fine-mapped loci with DARs and differentially H3K27ac modified regions did not reveal any overlap. I therefore considered overlap with all ATAC peaks blab la bla . Similar approaches integrating fine-mapping SNPs and tissue-specific chromatin accessibility maps have led to successful prioritisation of putative causal variants in other diseases. For example, in T2D 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

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

chromatin and increased enhancer activity (Gaulton et al. 2010; Stefan et al. 2014).

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. Although fine-mapping at chr2p15 failed in psoriasis (probably due to lack of power), fine-mapping analysis using AS Immunochip GWAS genotyping data yielded a credible set of three SNPs, of which only rs4672505 overlapped a tCD8⁺-specific ATAC peak. Chromatin accessibility varied across individuals unrelated to disease status, with complete ablation of the ATAC peak in some individuals. Integration of genotyping data with ATAC revealed allele-dependent chromatin accessibility, with ATAC reads negatively correlated with the risk allele. Promoter capture-C data linked rs4672505 to the *B3GNT2* promoter only in tCD8⁺ cells, suggesting the accessible chromatin at rs4672505 may be highlighting an enhancer element interacting with *B3GNT2* promoter as a priming event (Javiere2016). This regulation may only occur under pro-inflammatory stimuli through recruitment of TFs such as STAT1 or RUNX3, found to be binding at this location in lymphocytic cell lines. *B3GNT2* is a major polylactosamine synthase involved in the post-translational modifications of carbohydrate chains, which are essential for cellcell, receptorligand and carbohydratecarbohydrate 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 tCD8⁺ activation. Under this scenario, presence of the risk allele (A) at this stimulus-specific enhancer could increase risk of disease

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

by reducing chromatin accessibility, in both homozygous and heterozygous individuals. In fact, the down-regulation of *B3GNT2* expression in tCD8⁺ cells from psoriasis patients compared to controls may be the result of the majority of the patients being AA homozygous (1 out of 8) or heterozygous (6 out of 8). Nevertheless, the observation of heterozygous individuals with either presence or absence of the peak in both phenotypic groups may suggest additional mechanisms influence the epigenetic landscape at this location, such as other environmental cues interacting with the genotype. Formulation of a more comprehensive and accurate model to further explain the functional role of rs4672505 in psoriasis susceptibility will require additional work, such as increasing the sample size to acquire more homozygous individuals for the risk allele, studying chromatin accessibility and *B3GNT2* expression in relation to rs4672505 genotype in stimulated tCD8⁺ and performing EMSA for relevant TFs.

4.4.8 Limitations in the approach and future work

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 KCs was not achieved. This may have revealed larger differences in chromatin accessibility between psoriasis patients and controls compared to circulating immune cells, as in other studies performed in affected tissues (Scharer et al. 2016; Wang et al. 2018). Additionally, chromatin and transcriptomic profiles from skin infiltrated cells could be generated using FACS or single-cell technologies to better understand the changes in chromatin accessibility and gene expression driven by the inflammatory stimuli at the site

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

of inflammation. Moreover, generating this data would also allow comparison to the profiles obtained in blood to better understand disease pathophysiology.

Other limitations in this study include its relatively small sample size, lack of genotyping data and skin biopsies only being available for three patients in the cohort. These limitations are intrinsic to time and project budget constraints and will be addressed as the study continues. Recruitment of additional patients would allow validation of the findings described in this chapter. Genotyping data would permit the study of chromatin accessibility in a genotype-specific manner, using the current samples with prospective integration of chromatin conformation data (Kumasaka et al. 2018). Importantly, this will enable exploration of changes in chromatin accessibility at GWAS loci in combination with fine-mapping, similarly to the chr2p15 locus analysed in this chapter. Furthermore, new sample recruitment could be used to study chromatin accessibility and gene expression in additional cell populations sorted by FACS and also to include *in vitro* stimulations. Overall, this strategy would allow better characterisation of the differences and similarities between patients and controls in context-specific regulatory elements *in vivo* and *in vitro* (Peeters et al. 2015).

Finally, improvements in analytical methods will also be required to ascribe chromatin accessibility changes in enhancers to target genes potentially regulated by these regions. These could involve a more systematic integration of available chromatin conformation data, eRNA FANTOM data and also use of analytical models and tools currently available or that may be further developed in the future to specifically address this challenge (Wang et al. 2016; Cao et al. 2018).

4.4.9 Conclusions

In this chapter, use of the latest epigenetic methodologies (as established in the previous chapter) together with gene expression profiling has allowed

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

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 immunophenotypes in PsA

5.1 Introduction

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

Consideration of cell and type specificity in the study of complex diseases is fundamental for the understanding of the disease pathophysiology. As previously reviewed (1), the dysregulated immune response in PsA is the results 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 increasing disease risk and identify drugs with optimal efficacy and specificity.

PsA is considered a systemic disease where studies in PBMCs have demonstrated changes in cell type composition and cytokine production when compared to healthy individuals. For example, increased frequencies of

Cross-tissue comparison analysis in PsA

circulating IL-17⁺ and IL-22⁺ CD4⁺ T cells have been reported in PB from PsA patients compared to control individuals (Benham et al. 2013). Moreover, reduced percentage of pDCs and NK in PB have also been observed in PB from 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 eventually 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 PB and SF in PsA patients. For example, expansion mCD8⁺ but not mCD4⁺ T cells was observed in SF when compared to PB in PsA paired samples (Ross et al. 2000). Additionally, elevated proportion of T cells expressing the cytokine receptors CCR6⁺ and IL-23R⁺ were found in SF compared to PB in patients (Benham et al. 2013). Moreover, the elevated TNF- α , IL-1, IL-6 and IL-18 production in PsA SF 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 SFMCs samples. Several studies have been conducted to better understand gene expression differences in blood between PsA and controls and also specific differences between PB and SF from the same PsA patients or differences with other arthritic diseases (**Batiwalla2005**; Stoeckman et al. 2006; Gu et al. 2002; Dolcino et al. 2015). Amongst the most comprehensives of these studies, that conducted by Dolcino

Cross-tissue comparison analysis in PsA

and colleagues revealed genes from the Th-17 axis and type-I IFN signalling to be differentially expressed between PsA and healthy controls in synovial membranes. Moreover, the overlap of genes that were differentially expressed between patients and controls in each of the compartments, highlighted differences and commonalities in the systemic and synovial immune response in PsA. Cytokines production measurements have also been conducted in serum and SF, revealing increased levels of TNF- α , for example, in both tissues when compared to controls and osteoarthritis patientst, 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). For instance, 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). These studies have highlighted the regulatory role of some genetic variants only in particular cell types and conditions, previously masked when considering mixed population of cells such as PBMCs. In this respect, 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 PB and SF of PsA patients (**Antoniv2006; Leipe2010**). 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 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, including Drop-seq, SmartSeq2 and 10X Chromium amongst the most widely used (Ziegenhain 2017; Picelli et al. 2014). 10X Chromium technology is based on microfluidics where cells in suspension get directly encapsulated into nanoL droplets that incorporate cell and transcript barcode identifiers (see 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 fourty-five Abs to profile cellular populations and assess molecular functions.

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

As previously highlighted in Chapter ??, the methodological advances in the epigenetics field have allowed to map the regulatory landscape from clinical samples. This approach has facilitated the characterisation of the closest regulatory landscape to disease conditions using cell populations directly isolated from patients, instead of cultured cell lines or primary cells with additional stimulus, and the integration with the transcriptional profiles from paired samples. 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

Cross-tissue comparison analysis in PsA

in RA identified an expanded CD4⁺ T cell population promoting B cell response (Rao et al. 2017).

One of the most challenging aspects of using a multi-omics approach is the appropriate integration of the data in order to maximise the amount of information extracted and also the reliability of the findings. The power of this integration is increased by generating paired data for all the omics across all the individuals in the cohort, which cannot always be achieved due to sample availability and cost. 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 eighteen 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

As already explained in Chapters 1 and ??, fine-mapping of the 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 conditional analysis to identify independent secondary signals, prior to calculate PP and credible sets for each of them (Bunts2015; Maller et al. 2012). In some cases, this enables to reduce the size of the region associated to disease and thus the number of putative causal SNPs that could be functionally relevant for the disease pathophysiology. Although fine-mapping reduces the number of putative causal SNPs from thousands to tens, additional integration of epigenetics and functional

Cross-tissue comparison analysis in PsA

data as well as molecular assays are required to pinpoint the genetic variant and the mechanisms driving the association with disease.

The PsA GWAS study conducted by Bowes and colleagues successfully performed fine-mapping for seventeen of the associated regions (Bowes 2016). 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 more relevant cell type where they may have an effect. Bowes *et al.* 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.

Leveraging epigenetic data to further refine the candidate causal SNPs from fine-mapping studies could 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 develop a framework for the integration of a multi-omic dataset in PsA blood and synovial immune cells. Specifically, using the chromatin accessibility landscape and the transcriptomic profile in a number of immune relevant genes in four cell populations isolated from SF and PB to improve the understanding of cell and tissue-specific differences and the relationship between chromatin accessibility and gene expression in PsA.

Cross-tissue comparison analysis in PsA

Moreover, single-cell transcriptomic and mass cytometry will attempt to identify common and tissue-associated cell subsets contributing to pathophysiological relevant pathways in PsA and chronic inflammation. Lastly, fine-mapping of PsA GWAS loci and integration with chromatin accessibility maps of relevant immune cells from PsA samples will be performed to further narrow down the putative causal variants driving such associations.

Specifically, the aims of this chapter are:

1. To identify differences in the chromatin accessibility landscape between SF and PB in CD14⁺ monocytes, mCD4⁺, mCD8⁺ and NK cells isolated from PsA patients.
2. To characterise differences in the transcriptomic profiles between SF and PB for relevant immune genes in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ from the same PsA patients.
3. To integrate differences in chromatin accessibility and changes in genes expression from SF and PB.
4. To further explore transcriptional differences at the single-cell level in cell types of interest and perform a basic integration with mass cytometry data.
5. To conduct fine-mapping for a number of PsA GWAS loci using genotype data.
6. To integrate the credible set of SNPs identified by fine-mapping with cell and tissue-specific PsA chromatin accessibility maps and 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 (PB) and SF 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 affection and had been first 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 including joints and global assessment compose the PsARC disease activity scores, used by clinicians as the main indicator of response to treatment by recommendation of the National Institute for Health and Care Excellence (NICE) (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. Interestingly, PsA1728 was diagnosed at a later age compared to the other patients in the cohort (late PsA onset clinical significance??). Moreover, C-reactive protein (CRP) levels, a further marker of inflammation, presented an average of 17.45 mg/L and was particularly higher in PsA1719 and PsA1728 compared to the other patients. At the time of sample recruitment all the PsA patients were not on active immunosuppressive therapy. Post-visit, most of the patients qualified for TNFi biologic therapy xxxx.

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 self-patient and physician global assessment (5 point score). Joint pain and global assessment use a likert scale based on questionnaire answers that measure the level of agreement with each of statements included. C-reactive protein (CRP).

Sample ID	Sex	Age at diagnosis	Disease duration (months)	Type	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
Average	-	44.3	106	-	1.5/1.5	3	3.2	17.4

Cross-tissue comparison analysis in PsA

For each of the patients, paired PB and SF 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 SF and PB from the same individual. The available datasets vary between individuals due to project constraints. Fast-ATAC data was generated for CD14⁺, mCD4⁺, mCD8⁺ and NK cells. RNA expression by PCR array was performed only for CD14⁺, mCD4⁺ and mCD8⁺ cells. scRNA-seq data was generated using 10X 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 SF and PB using the ICS mass cytometry panel in Chapter 2. For both tissues, mCD4⁺ (between 32.1 and 55.6%) constituted the most abundant cell type followed by mCD8⁺ (between 16.9 and 24.9%) and CD14⁺ "non-classical" monocytes (between 6.9 and 21.7%). Consistently with previous studies, a trend of increased percentage of mCD8⁺ pDCs and cDCs was observed in SF compared to PB (Ross et al. 2000; Jongbloed et al. 2006). This data also showed reduced percentage of SF NK cells percentage compared to PB, in line with previous studies suggesting the role of impaired non-MHC-restricted cytotoxicity in PsA (Spadaro et al. 2004). Similarly, a tendency towards reduced proportions of B cells in SF compared to PB reinforced the lack of contribution of

Cross-tissue comparison analysis in PsA

the humoral immune response in PsA pathophysiology. The observed differences in cell composition between SF and PB were not statistically significant for any of the twelve analysed populations likely due to the small samples size ($n=3$) available for the analysis. Further increase in the sample size will probably prove statistical significance for the observed differences in immune cell composition between the two tissues reproducing the results published by other studies.

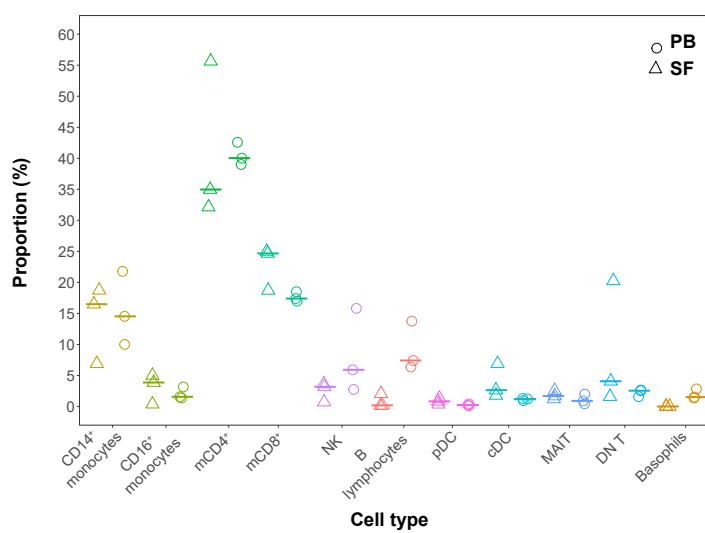


Figure 5.1: Comparative percentages of PB and SF immune cellular composition from the PsA cohort. Percentages of each of the twelve cell types identified by mass cytometry are shown by individual and tissue for PsA1718, PsA1719 and PsA1607. Horizontal line represents the median percentage for a particular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

5.3.3 Differential chromatin accessibility analysis in immune cells reveals differences between SF and PB

Quality control of Fast-ATAC data

Twenty four Fast-ATAC PsA samples from four different cell types and two tissues (PB and SF) were sequenced and processes using the in-house pipeline as previously detailed in Chapter 2. After filtering for low quality mapping, duplicates and MT reads, the median of total number of reads ranged

Cross-tissue comparison analysis in PsA

between 46.6 and 70.2 millions (Figure 5.2 a). Overall, MT and duplicated reads accounted for a median of 40 to 62.2% from the total number of unfiltered reads depending on cell type (Figure 5.2 b), contributing to the loss of reads in ATAC as previously detailed in Chapter 3.

Regarding sample quality, TSS enrichment analysis showed differences in the levels of background noise across cell types and highlighted the variability of ATAC performance (Figure 5.2 c). A general trend towards greater TSS enrichment in PB samples compared to SF was observed. mCD4⁺ and mCD8⁺ presented the best signal-to-noise ratios, with median of 19.1 and 23.1 fold enrichment, respectively. In contrast, NK was the cell type with the lowest TSS enrichment values. Particularly, the fold enrichment for PsA1719 and PsA1607 in NK were close to the 6 fold enrichment considered by ENCODE as acceptable. 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.

Cross-tissue comparison analysis in PsA

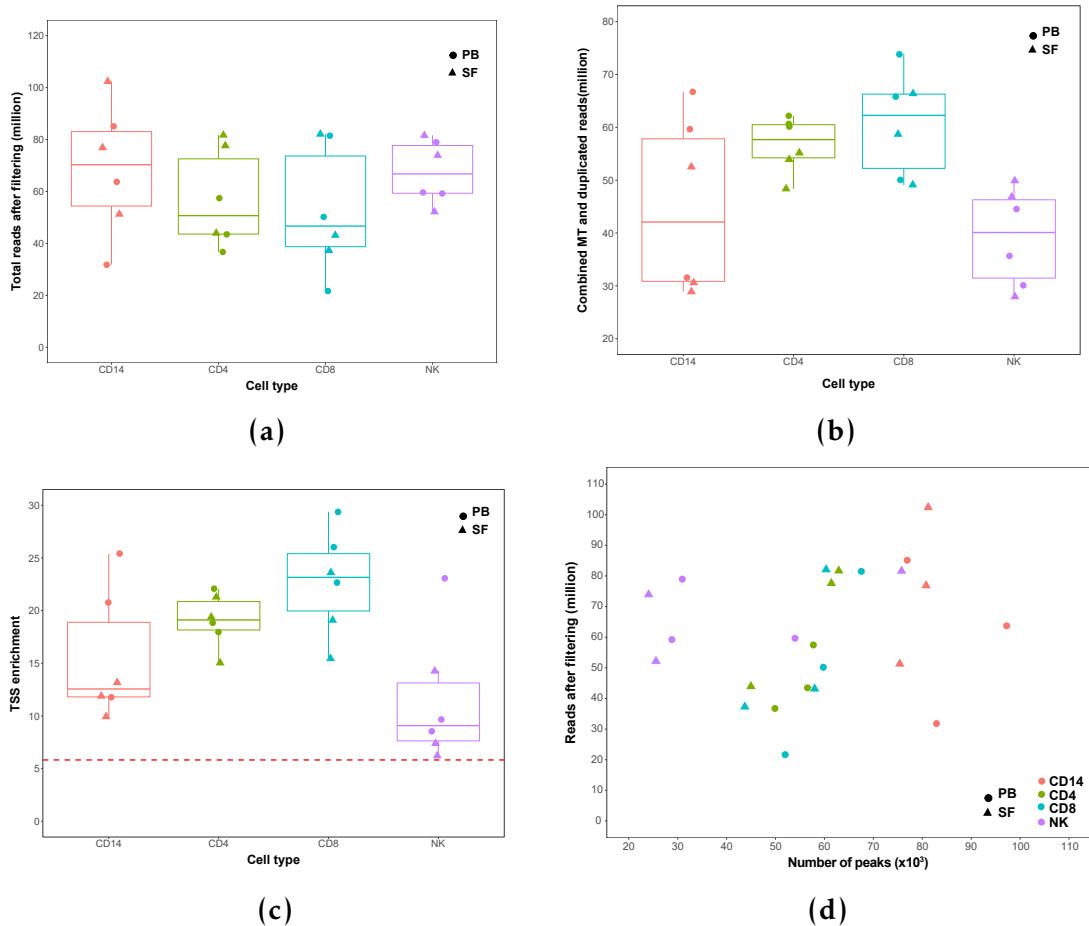


Figure 5.2: Quality control assessment of ATAC data generated in four immune cell types isolated from PB and SF of PsA patients samples. For each of the cell types and samples, boxplots representing a) million of reads after filtering, b) million of duplicated and MT reads combined and c) values for fold-enrichment of ATAC fragments across the Ensembl annotated TSS. In c) the dashed red line indicates the recommended Encode threshold for TSS enrichment values. d) Representation of the number of significant peaks based on IDR optimal pval 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 or PB).

When identifying open chromatin regions by peak calling followed by pval filtering based on IDR analysis, the number of accessible regions per sample ranged approximately between 24×10^3 and 97×10^3 (Figure 5.2 d). The total number of called peaks passing filtering varied across cell types and was influenced by the quality sample, as previously demonstrated in Chapter ???. Overall, appropriate number of peaks were called in all the samples and no concerning outliers were identified.

Accessible chromatin reflects cell type specificity and functional relevance

A consensus master list of accessible chromatin regions identified across all the samples and cell types (ML_all) was built, as previously explained in Chapter 2 and Chapter 3. PCA analysis based on the normalised counts for each region of the ML_all showed that most of the variability (PC1 65.6% of the variability) in the chromatin landscape correlated with cell type, leading to sample separation in four cluster (Figure 5.3). The myeloid ($CD14^+$ monocytes) and lymphoid ($mCD4^+$ and $mCD8^+$) clusters appeared as the most different between them based on the Fast-ATAC profile. Conversely, the $mCD4^+$ and $mCD8^+$ clusters were the most similar between them, altogether supporting the ability of Fast-ATAC to capture cell type chromatin accessibility features. In addition to this, modest separation between SF and PB samples was also found in the $mCD4^+$, $mCD8^+$ and NK clusters (Figure 5.3).

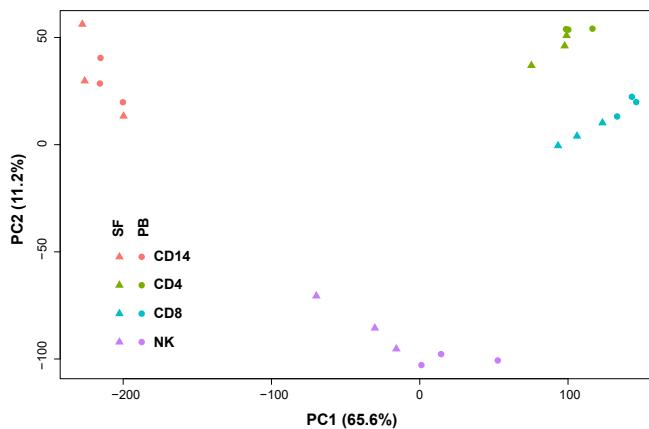


Figure 5.3: PCA analysis based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and SF. PCA analysis was performed using the normalised counts from the combined consensus master list (ML_all) across the four cell types ($CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells) and two tissues (SF and PB) of interest. 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 and PB). The proportion of variation explained by each principal component is indicated.

The ability to capture putative regulatory regions within the identified accessible chromatin regions was also explored. Enrichment analysis of different

Cross-tissue comparison analysis in PsA

eQTL publicly available datasets for the regions contained in the MASTER_ALL list was performed. Amongst the GTEx eQTL data, the largest (z-score) and most significant (-log₁₀FDR) enrichment was found for the venous blood data set (red dot), consistent with the cell types included in the study (Figure 5.4 a). In terms of publicly available eQTLs studies in immune cells, the strongest enrichment for the MASTER_ALL regions were found for CD14⁺ monocytes (importantly unstimulated, LPS 2h and IFN- γ 24h) followed by mCD8⁺ T cells (Figure 5.4 b). eQTLs in B cell appeared as the least enriched when compared to the other datasets, consistently with the absence of this cell type in the ATAC experiments, and reinforcing the cell specificity captured by this assay.

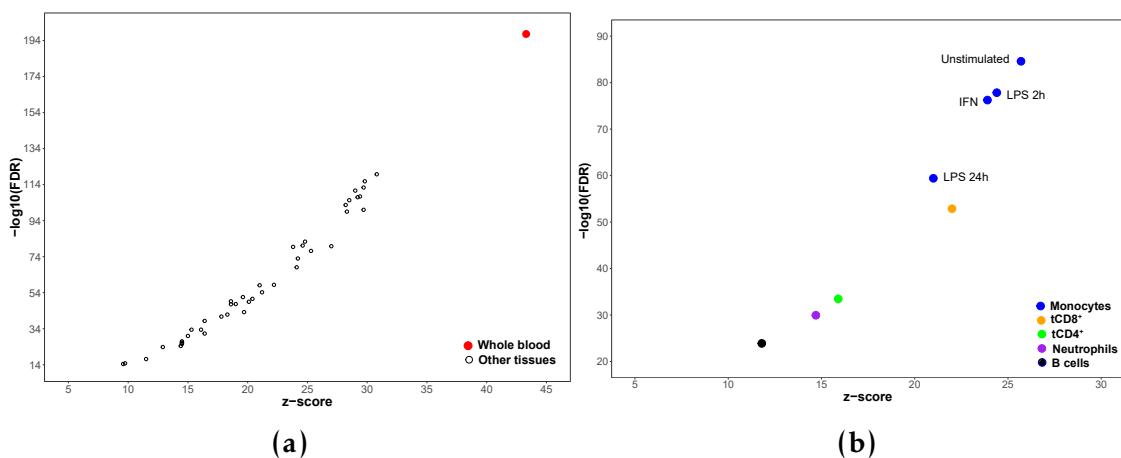


Figure 5.4: Enrichment of eQTLs publicly available data in the combined cell type and tissue chromatin accessibility master list for the PsA cohort. The dot plots showed the z-score values of the enrichment analysis in the x-axis and the significance (-log₁₀FDR) in the y-axis for a) GTEx eQTL datasets and b) non-GTEx immune-related cell types including CD14⁺ monocytes (unstimulated, LPS 2h, LPS 24h and 24h IFN γ stimulated), B cells, tCD4⁺, tCD8⁺ and neutrophils. Dots are colour-coded by cell type.

CD14⁺ monocytes present the greatest proportion of changes in chromatin accessibility

A consensus master list of chromatin accessible regions was built for each of the four cell types of interest (ML_CD14, ML_CD4, ML_CD8 and ML_NK). Differential chromatin accessibility analysis between SF and PB was performed on the normalised counts retrieved for each of the cell type master lists using

Cross-tissue comparison analysis in PsA

DESeq2 and a paired design (Table 5.3). A 80% cut-off for background noise was used to filter the count matrix, as previously explained in Chapter ???. The CD14⁺ monocytes and NK were the two cell types presenting the greatest total number (5,285 and 2,314, respectively) and proportion of DARs (23.3 and 8.9%, respectively). For each cell type, DARs were divided in DARs more open in SF compared to PB (SF open DARs) and DARs less open in SF compared to PB (PB open DARs). In CD14⁺ monocytes the number of SF open DARS were notably larger than the number of PB open DARs (3,779 and 1,506DARs, respectively) (Table 5.3). Conversely, the number of SF and PB open DARs were similar for the other three cell types.

Cell type	Total DARs	Proportion DARs (%)	SF open DARs	PB open DARs
CD14 ⁺	5,285	23.3	3,779	1,506
CD4 ⁺	1,329	4.3	621	708
CD8 ⁺	1,570	4.5	807	763
NK	2,314	8.9	1,223	1,091

Table 5.3: Summary results of the differential chromatin accessibility analysis between SF and PB 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 SF (DARs open in SF) when compared to PB and those less accessible in SF when compared to PB (DARs open in PB).

Permutation analysis was used to determine if the large number of DARs (particularly by comparison to limited finding in the psoriasis analysis) were more than would be expected by chance. None of the ten possible permutations demonstrated a greater number of DARs than the ones identified for the true groups, reinforcing the robustness of the differential analysis results (Figure B.7).

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 5.5 a). Universal promoter regions was the third most represented genomic feature, accounting for the annotation of

Cross-tissue comparison analysis in PsA

approximately between 5 to 15% of the DARs in each cell type. In addition to this, the chromatin states from the Roadmap Epigenomics maps were also used for annotation (Figure 5.5 b). For all four cell types, between 44.96 and 72.11% of the DARs were annotated as weak enhancers, which represented the most prominent category and the most significantly enriched (data not shown). 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 annotated as heterochromatin and repetitive regions but not significant enrichment for these two chromatin states was found for any of the four cell types (Figure 5.5 b).

Cross-tissue comparison analysis in PsA

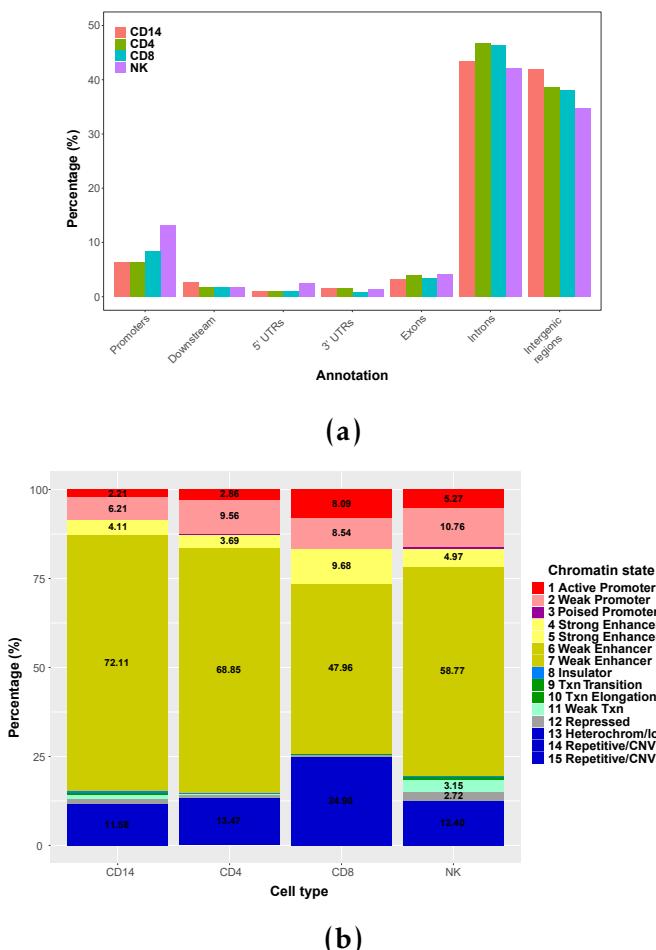


Figure 5.5: Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states. a) Barplot illustrating the percentage of nucleotides within DARs for each cell type that are annotated as promoters, downstream (regions at $\leq 1,000$ bp to a promoter), exons, introns, 5' or 3'UTR and intergenic regions. b) Stacked barplot representing the percentage of DARs annotated for each of the fifteen chromatin states defined in each of the four relevant cell types by Epigenome Roadmap chromatin segmentation maps (CD14⁺ PB isolated monocytes, mCD4⁺, mCD8⁺ and NK 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 ??). Moreover, DARs from all four cell types also presented significant enrichment for the corresponding cell type eRNA set. The proportion of DARs overlapping the appropriate cell type set of expressed eRNAs ranged between

Cross-tissue comparison analysis in PsA

19.8% (83 open in SF and 160 open in PB) in NK and 31.8% (83 open in SF and 160 open in PB) in CD4⁺ cells.

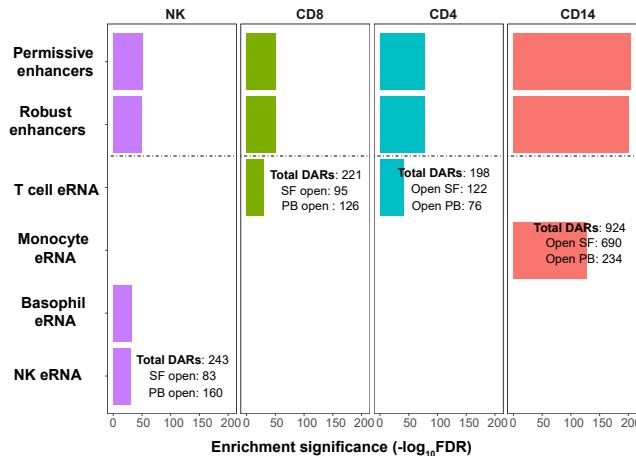


Figure 5.6: Enrichment of PsA DOCs for the FANTOM5 eRNA dataset. Robust enhancers have been defined as those detected at the genome-wide significant level in at least one primary cell type or tissue. Permissive enhancers are all detected eRNAs but not passing genome-wide filtering criteria (Andersson et al. 2014). Robust enhancers represent a subset of the permissive enhancers. Significance is considered for FDR<0.01.

From the differential analysis between SF and PB, 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.

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

Table 5.4: Characterisation of the DARs located within genes in each of the four cell types from PsA samples. The number of DARs 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 Epigenome Roadmap chromatin segmentation maps of each appropriate cell type.

Cross-tissue comparison analysis in PsA

For example, NK analysis identified a PB open DAR located in an intron of the *VAV3* gene and also significantly expressed as eRNA (Figure 5.7 a). Additionally, a number of gene entities contained more than one DAR, showing the same direction of chromatin accessibility between SF and PB. For example, in CD14⁺ two DARs located at the 5' and 3' UTRs of *IL7R* gene were more accessible in SF compared to PB (Figure 5.7 b).

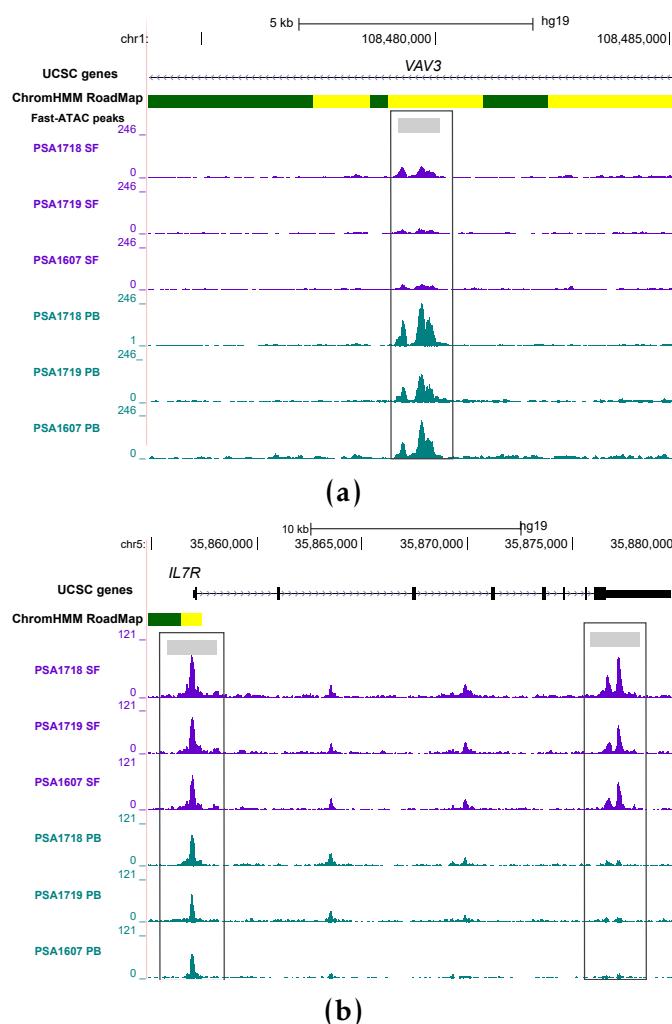


Figure 5.7: Differentially accessible regions located within gene bodies in CD14⁺ monocytes and NK cells from PsA patients. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) in a) DAR located at an intron of *VAV3* gene (x-axis) in NK (less accessible in SF compared to PB) and b) two DARs mapping to the 5' and 3'UTR of the *IL7R*, respectively, in CD14⁺ monocytes (both more accessible in SF compared to PB). Tracks are colour-coded by tissue (SF=purple and PB=turquoise). The Epigenome Roadmap chromatin segmentation track for the appropriate cell type are also shown. All DARs were significant based on FDR<0.01 and FC>1.5.

Cross-tissue comparison analysis in PsA

In all four cell types, overlap between putative psoriasis and PsA GWAS genes and proximal ($\leq 5\text{Kb}$) DARs were found. CD14^+ monocytes presented the largest number of overlaps (12), followed by mCD8^+ (9), NK^+ (8) and mCD4^+ (4). For example, DARs proximal to *ELMO1* and *RUNX3* were found for all the cell types. Co-localisation and permutation analysis to explore the enrichment of DARs for PsA and psoriasis GWAS LD blocks only found significance in CD14^+ monocytes (empirical pval=0.043).

5.3.4 Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility

Pathway enrichment analysis was conducted separately for SF open DARs and PB open DARs in each cell type. Gene annotation of the DARs was performed by physical proximity, as detailed in Chapter 2. Despite commonalities, differences in significant enriched pathways ($\text{FDR}<0.01$ or 0.05) were also identified within the same cell type between SF and PB open DARs (Figure 5.8). In CD14^+ monocytes SF open DARs presented enrichment for pathways involved in regulation of immunity, inflammation and cell survival such as the NF- κ B pathway and cytokine related pathways, including IL-2 and IL-3, 5 and granulocyte-macrophage colonystimulating factor (GM-CSF) signalling (Figure 5.8 a).

mCD4^+ SF open DARs compared to PB open DARs showed enrichment for TCR signalling as well as chemokine signalling, which included, amongst others, DARs in proximity to IFN- γ or *CXCL13* and *CXCR6*, respectively (Figure 5.8 b). T cell signaling pathway appeared only enriched for SF open DARs in mCD4^+ ; however PB open DARs in this cell type were also enriched for focal adhesion members, also involved in the T cell activation (Dustin 2001). Enriched pathways for SF or PB open DARs in mCD8^+ were only significant when using an $\text{FDR}<0.05$ threshold. (Figure 5.8 c). The G protein coupled receptor (GPCR) signalling,

Cross-tissue comparison analysis in PsA

with varied roles in regulation of inflammation and mediation of the chemotactic recruitment of T cells to the inflamed tissue, was enriched for mCD8⁺ PB open DARs. mCD8⁺ SF open DARs showed enrichment for the Wnt signaling pathway involved in the production of memory cells with enhanced proliferative potential and stronger protective capacity (Boudousqui et al. 2014).

NK SF open DARs presented enrichment for Fc-gamma receptor (FC γ R)-mediated phagocytosis (Figure 5.8 d). Moreover, members of the HIF-1 pathway involved in oxygen homeostasis were also enriched in NK SF open DARs, in line with the hypoxic environment found in joint inflammation. Interestingly, enrichment of open PB DARs in the proximity of genes involved in NK-mediated toxicity was unveiled.

Cross-tissue comparison analysis in PsA

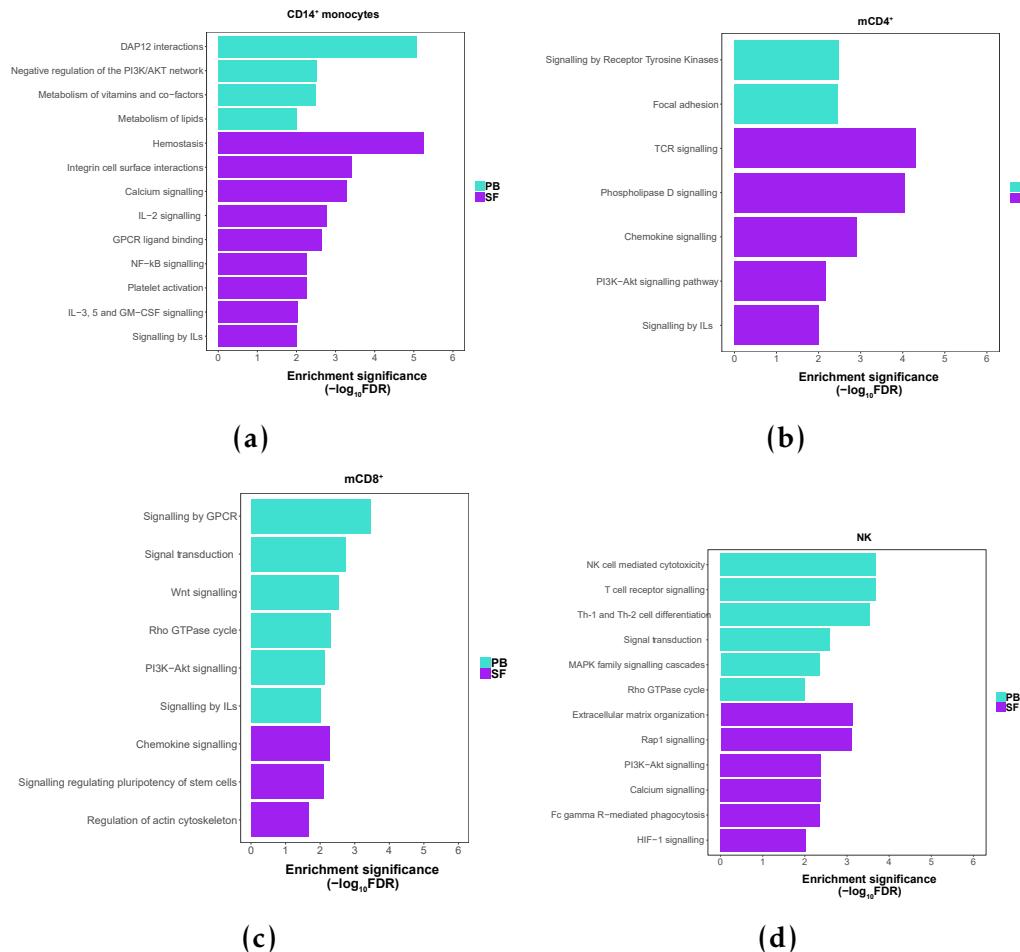


Figure 5.8: Distinct functional pathways enriched for DARs open in SF or open in PB in CD14⁺ monocytes, CD4m⁺, CD8m⁺ and NK. Enrichment analysis was performed separately for the DARs open in SF and DARs open in PB separately in a) CD14⁺ monocytes, b) mCD4⁺, c) mCD8⁺ and d) NK. The pathways presenting significant enrichment (FDR < 0.01) only in open SF or open PB DARs are shown here.

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

Immune-relevant gene expression by qPCR

Mapping chromatin accessibility represents an informative tool to identify regulatory elements undergoing histone modifications, DNA methylation and TF binding, as previously explained. All those elements are involved in the regulation of gene expression, making the study of chromatin accessibility a good proxy for the inference of gene expression. Nevertheless, the characterisation

Cross-tissue comparison analysis in PsA

of the chromatin landscape also presents some limitations, including the discordance between open chromatin and functionality of the regulatory element, shown by CAGE studies, as well as the identification of the target gene regulated by a particular element.

In order to contextualise the ATAC-seq data, qPCR gene expression analysis for 370 key genes in the inflammatory and autoimmune response was conducted in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells isolated from SF and PB of three PsA patients (Table 5.2). Those appeared as the most abundant cell types in PB and SF from patients, and particularly mCD4⁺ and mCD8⁺ cells have been shown to expand in PsA inflamed synovium, as previously mentioned. The PCR array represented a cost-effective approach to study gene expression between PB and SF focusing in a relevant subset of genes of notably importance, given the inflammatory component in PsA. For each cell types, FC in expression was calculated pair-wise for SF respect to PB within each sample for each individual genes (detailed in Chapter 2). Likely due to the small sample size, the majority of the modulated genes between SF and PB lacked of significance (FDR<0.05) after multiple testing correction. Therefore, to explore the biological relevance of this data, a less stringent pval<0.05 was used as the filtering threshold.

When considering the significantly modulated genes (pval<0.05) in at least one cell type, differences in magnitude and reproducibility in FCs were observed across samples and cell types (Figure 5.9). Some of the modulated genes showed up-regulation (FC>1.5) in SF compared to PB across the three cell types, for example *FN1*, *SPP1* or *CCL2*, amongst others (Figure 5.9 orange box). On the other hand, a number of genes presented reduced expression in SF (FC<1.5) in at least one of the three cell types, including *FOS*, *IL16*, *PPBP* and *TPST1* (Figure 5.9 purple box). Also, a number of genes were only consistently modulated in the three CD14⁺ monocyte samples but not in T cells (Figure 5.9 dark blue box).

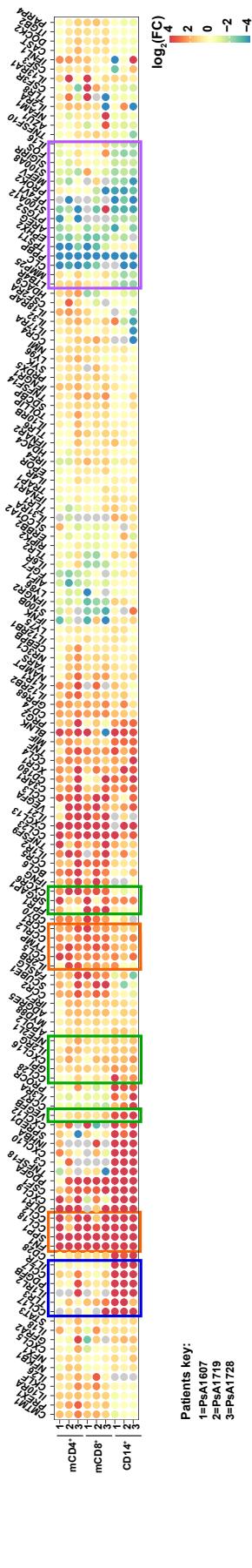


Figure 5.9: Heatmap of gene expression FCs between SF and PB for those gene significantly modulated ($pval<0.05$) in at least one cell type. Amongst the 370 genes measured by qPCR, the FC in gene expression between the SF and PB for each pair of samples has been represented only those genes which were consistently modulated across the three PsA samples ($pval<0.05$) in at least one of the cell types. Each column represents a genes and each row a pair of SF-PB PsA samples. The \log_2FC in gene expression between SF and PB is colour-coded. Overall, the heatmap allows to observe the change in gene expression as well as the magnitude between SF and PB for each gene in each of the three pairs of PsA samples in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells.

Cross-tissue comparison analysis in PsA

For example, *CCR7* and *IL7R* were up-regulated in SF CD14⁺ monocytes compared to PB; however changes between SF and PB were not consistent across patients in mCD4⁺ and mCD8⁺. Moreover, differences in the magnitude of FCs were also observed for some of the genes modulated in the same direction across the three cell types, for instance *VEGFB* and *CXCR6* (Figure ?? green box).

Filtering of all the genes tested for expression in the qPCR array based statistical significance ($pval < 0.05$) and mean $FC > 1.5$ revealed CD14⁺ monocytes and mCD8⁺ presenting greater number of significantly modulated genes (72 and 77, respectively) compared to mCD4⁺ cells (46 genes) (Figure 5.10 a, b and c). For the three analysed cell types, the majority of modulated immune genes showed up-regulation in the SF (Figure 5.10 a, b and c). For example, 56 out of the 70 significantly modulated genes in CD14⁺ monocytes showed mean $FC > 1.5$ versus the 14 genes with mean $FC < 1.5$ (Figure 5.10 a).

Correlation between gene expression and chromatin accessibility

Overlap between differentially modulated genes and DARs in the proximity were observed in the three cell types (Table 5.5). The overlap in CD14⁺ monocytes revealed significant enrichment of modulated genes between SF and PB for DARs with the same direction of change (Fisher exact test $pval = 0.028$). In contrast to CD14⁺ monocytes, the observed overlap between gene expression and chromatin accessibility did not appear to be significant in mCD4⁺ and mCD8⁺ cells (Fisher exact test $pval = 0.466$ and 0.173, respectively).

In CD14⁺ monocytes, 13 out of the 56 significantly up-regulated genes in SF overlapped with SF open DARs. For example, the increased expression of *IL7R* in SF correlated with increased chromatin accessibility at the 5' and 3' UTR of this gene, previously shown (Figure 5.7 b). Another relevant example was the *FN1* gene, which up-regulated expression in synovial biopsies compared to PB has already been reported by others (Dolcino et al. 2015). In this

Cross-tissue comparison analysis in PsA

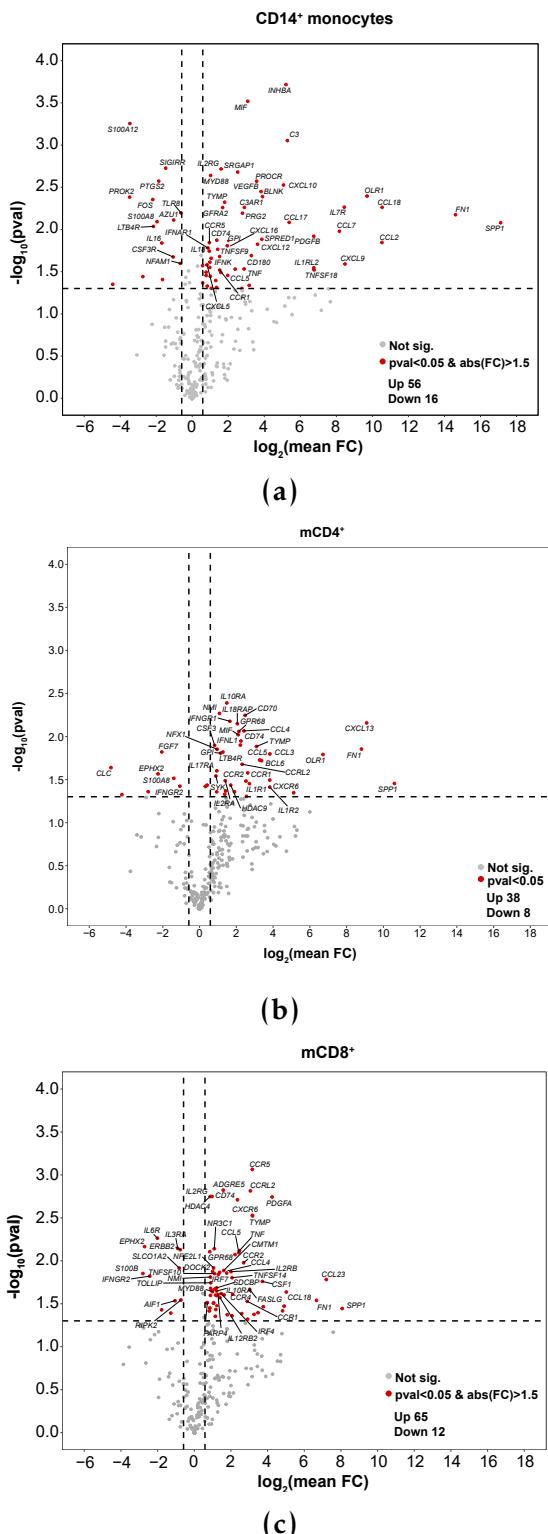


Figure 5.10: Gene expression changes in immune-relevant genes between SF and PB in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells. Volcano plots showing differences in gene expression measured by qPCR array between SF and PB for a) CD14⁺ monocytes, b) mCD4⁺ and c) mCD8⁺ cells. The significance ($\log_{10}p\text{val}$) of the modulation in gene expression between the two tissues (y-axis) is plotted against the \log_2 of the mean FC across the three PsA patients. Positive FC indicates higher expression in SF. Genes showing $p\text{val}<0.05$ and mean FC>1.5 are coloured in red, with the most significant genes labelled.

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 ⁺ monocytes	13 (<i>BLNK, CCL2*, CCR1*, CD180, CXCL10, FN1, IL18, IL31RA*, IL7R*, NFKB1*, PRG2, SRGAP1, STAT3</i>)	2 (<i>FOS, PROK2*</i>)
mCD4 ⁺	3 (<i>CXCL13, CXCR6*, IL2RA</i>)	0
mCD8 ⁺	6 (<i>CCL3, CCR2, CCR5 ,IRF4 TNFSF10, YARS</i>)	1 (<i>EPHX2</i>)

Table 5.5: Immune genes with significant modulated expression in SF and proximal to a DAR in Fast-ATAC. An overlap is defined by significant change in expression ($pval < 0.05$) 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).

cohort, *FN1* expression was up-regulated in SF for all three cell types with the greater FC found in CD14⁺ monocytes (Figure 5.10 a), concomitantly with more accessible chromatin at the promoter and downstream the 3' UTR of the gene (Figure 5.11). Lesser overlap between up-regulated gene expression and open chromatin in SF compared to PB was observed in mCD4⁺ and mCD8⁺ (6 and 3 hits, respectively). Only CD14⁺ monocytes and mCD8⁺ cells presented overlap between SF down-regulated genes and proximal less accessible in SF (2 and 1, respectively). Notably, none or very few genes presented opposite direction of change in gene expression and chromatin accessibility on a proximal DAR (4 in CD14⁺ monocytes, 0 in mCD4⁺ and 2 in mCD8⁺), reinforcing the biological relevance of the observed overlaps.

Pathway enrichment and network analysis highlights the role of synovial CD14⁺ monocytes in cytokine and chemokine production

To identify relevant pathways amongst the modulated genes between SF and PB, enrichment analysis was performed for each individual cell type. Up-

Cross-tissue comparison analysis in PsA

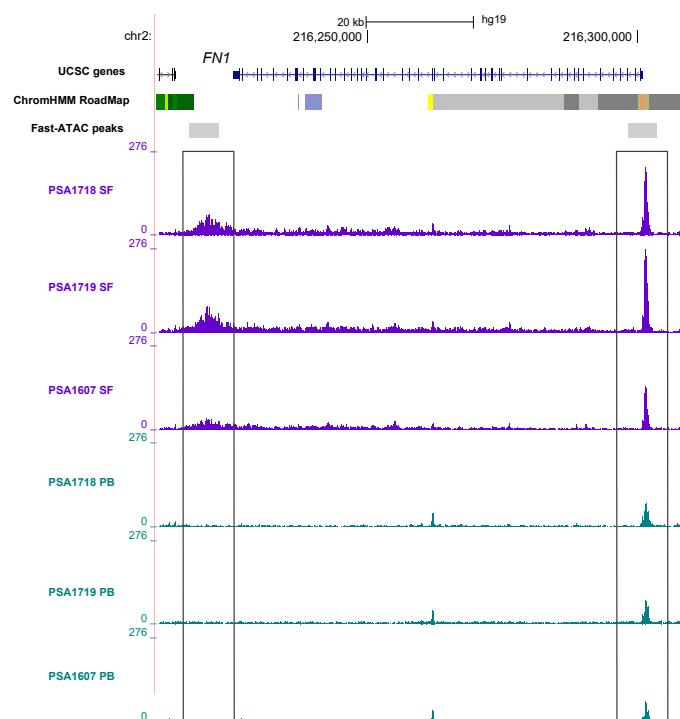


Figure 5.11: Chromatin accessibility landscape at the qPCR differentially expressed *FN1* gene in CD14⁺ monocytes. UCSC Genome Browser view illustrating the ATAC normalised read density (y-axis) in two DARs located at the promoter and downstream the 3' UTR of the *FN1* gene (x-axis) in CD14⁺ monocytes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

Cross-tissue comparison analysis in PsA

regulated and down-regulated genes showing abs mean FC>1.5 and pval<0.05 were used as input for the enrichment analysis. Interestingly, the modulated genes between SF and PB in CD14⁺ monocytes were enriched for chemokine, NOD-like signalling and TLR signalling pathways (Table 5.6). All three pathways are involved in the activation of cytokines and chemokines gene expression, leading to T cell recruitment and inflammatory response.

The TLR signalling pathways enrichment involved the *FN1* (previously mentioned) and *SPP1*, two of the top differentially expressed genes found in this pilot study (Table 5.6). Together with *FN1*, *SPP1* was highly up-regulated (mean FC>16) in the three cell types, showing the greatest FC in monocytes (Figure 5.10 a). Moreover, some of the genes driving enrichment, such as *CCL5* and *NFKB*, were also shared across the three pathways of interest. Others genes, including *TNF*, *IRF7* and *MYD88*, highlighted the cross-link between the NOD-like and the TLR signalling pathways.

Accordingly, the enrichment of SF open DARs in CD14⁺ monocytes for the NFκB pathway is closely related to the enrichment for TLR and NOD-like signalling pathways at the transcriptomic level, since both pathways lead to the activation of the NFκB TF (Figure 5.8 a).

The enrichment for the chemokine pathway in CD14⁺ monocytes (Table 5.6) included modulated genes highly up-regulated (mean FC>16) in SF compared to PB (e.g *CCL18* and *CCL2*) for all three cell types (Figure 5.10) as well as genes only consistently modulated between SF and PB in CD14⁺ monocytes (e.g *CCL28*, Figure 5.9 green box).

At the transcriptional level, significantly modulated genes between SF and PB in mCD4⁺ T cells were enriched for the IL-10 signalling pathway (Table 5.6), in lines with the enrichment for IL signalling of open chromatin in SF cells (Figure 5.8 b).

Cross-tissue comparison analysis in PsA

Table 5.6: Pathway enrichment analysis for the modulated genes between SF and PB in CD14⁺ and mCD4⁺. The analysis was performed using only those genes showing pval<0.05 and mean FC>1.5. Reported enriched pathways were significant at an FDR <0.05.

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

Cross-tissue comparison analysis in PsA

In addition to pathway enrichment, network analysis was performed to understand the interaction and relationship of the genes with modulated expression between SF and PB. A gene subnetwork was identified from the STRING functional interaction database using as input all the genes from the qPCR array (regardless pval significance and FC) and ranking them based on the best pval across the three analysed cell types. The identified subnetwork predominantly included significant modulated genes between SF and PB in at least one of the cell types. Amongst the most interesting nodes was the single Ig and Toll-interleukine domain containing gene (*SIGIRR*), which is a negative regulator of the TLR signalling pathway (Figure 5.12). *SIGIRR* was significantly down-regulated in SF CD14⁺ monocytes only and it is consistent with the significant up-regulation (pval<0.05) of the *TLR1*, *TLR2*, *MYD88* genes in SF as well as the enrichment for the TLR pathway in this cell type (Figure 5.9 and 5.6). Moreover, the significant up-regulation of *NFKB* and *TNF* in the SF CD14⁺ monocytes appeared as a downstream result of the functional connection with TLR pathway members such as *MyD88*, previously mentioned. Conversely, in mCD4⁺ and mCD8⁺ the modulation of these members did not appear to be significant between SF and PB.

Another interesting part of the network is the connection of the TLR pathway and the chemokine production through *NFκB*, *TNF* and *CCL2* (Figure 5.12). *CCL2* is connected to *CXCL10* and subsequently with *CCL18* and *CCR5*, all chemokines and chemokine receptors regulating migration and infiltration of monocytes and memory T cells at the sites of inflammation. This network analysis also highlighted relationship between *IL7R* and *IL2RG* coding for the two chains of the IL-7R.

Overall, the integration of the chromatin accessibility and immune transcriptional data reinforced a relevant role of synovial CD14⁺ monocytes in the production of cytokines and chemokines, likely leading to activation

Cross-tissue comparison analysis in PsA

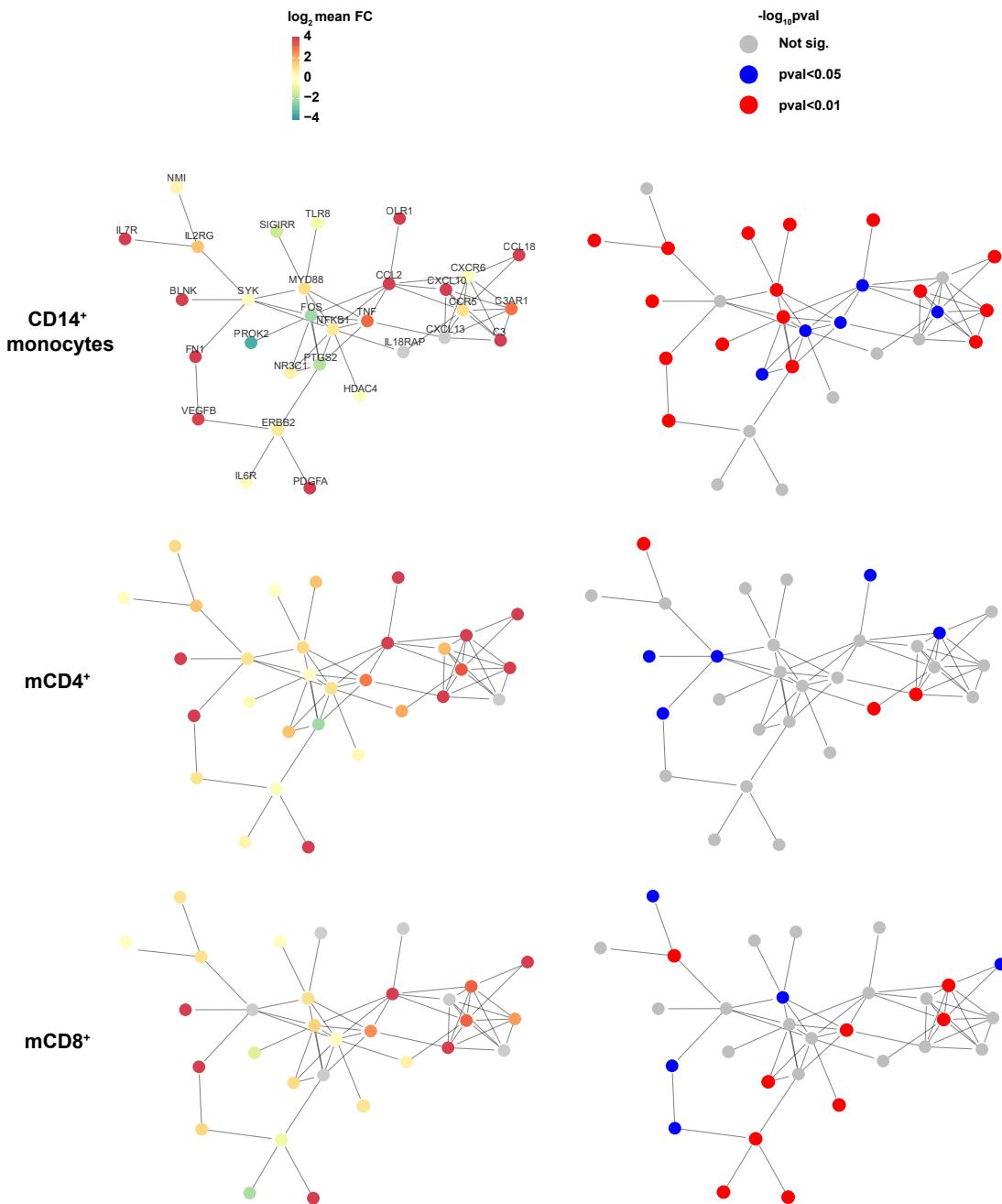


Figure 5.12: Protein network analysis based on the immune qPCR array expression data. The list of all the genes quantified in the qPCR array genes together with the best pval for significance of the mean FC across the three cell types was used to perform gene network analysis. STRING interaction network (including known and predicted proteinprotein interactions) was used to superpose the aforementioned list of genes and obtain a 30 gene size subnetwork common for all three cell types. This included maximal number of significant genes ($pval < 0.05$) in at least one cell type and minimal presence of non-significant genes as linkers in the network. In the left hand panel, for each cell type each of the nodes (proteins) of the identified subnetwork (the same for each cell type, as previously explained) are colour-coded by the change of expression (\log_2 mean FC) for the corresponding gene in the qPCR analysis. On the right hand panel, each of the nodes in the same subnetwork are colour-coded by the level of significance (pval) for the reported modulation in gene expression (\log_2 mean FC) in the qPCR array.

Cross-tissue comparison analysis in PsA

of the innate immune response and the recruitment of T cells to this site of inflammation.

Tissue and disease specificity in gene expression modulation and relevant biological pathways

In order to better understand the disease and tissue specificity of the prior transcriptomic results, gene expression was analysed in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ isolated from PB in three healthy controls using the same qPCR array. In each of the cell types, the FC in was calculated for the mean PB expression across the three PsA patients compared to the mean expression of the three healthy controls (as detailed in Chapter 2). Similar to the previous analysis, pvals for the FC significance were calculated for each particular genes. Integration of the previous results of modulated gene expression between SF and PB in PsA (see Immune-relevant gene expression by qPCR) with this analysis allowed the identification of three group of genes (Figure 5.13). First, the genes only significantly modulated (based on pval and FC threshold criteria) in PB between controls and PsA were designated as systemic genes (Figure 5.13 green dots). Those genes were not significantly modulated in the prior analysis comparing SF versus PB within PsA patients and could then be considered as the circulating disease "footprint". In this respect, CD14⁺ monocytes was the cell type with lower number of systemic modulated genes (14), compared to mCD8⁺ (23) and mCD4⁺ (42) (Figure 5.13a, b and c). Amongst these genes were *CCL24* and *CCL27* in CD14⁺ monocytes, *CCR7* and *TLR4* in mCD4⁺ cells and *CCR10* in mCD8⁺ cells.

A second group of genes were designated as tissue-specific, since they were significantly modulated between SF and PB in PsA patients but did not show significant changes between controls and PsA at the circulating level (Figure 5.13 red dots). Interestingly, in CD14⁺ monocytes the tissue specific modulated genes

Cross-tissue comparison analysis in PsA

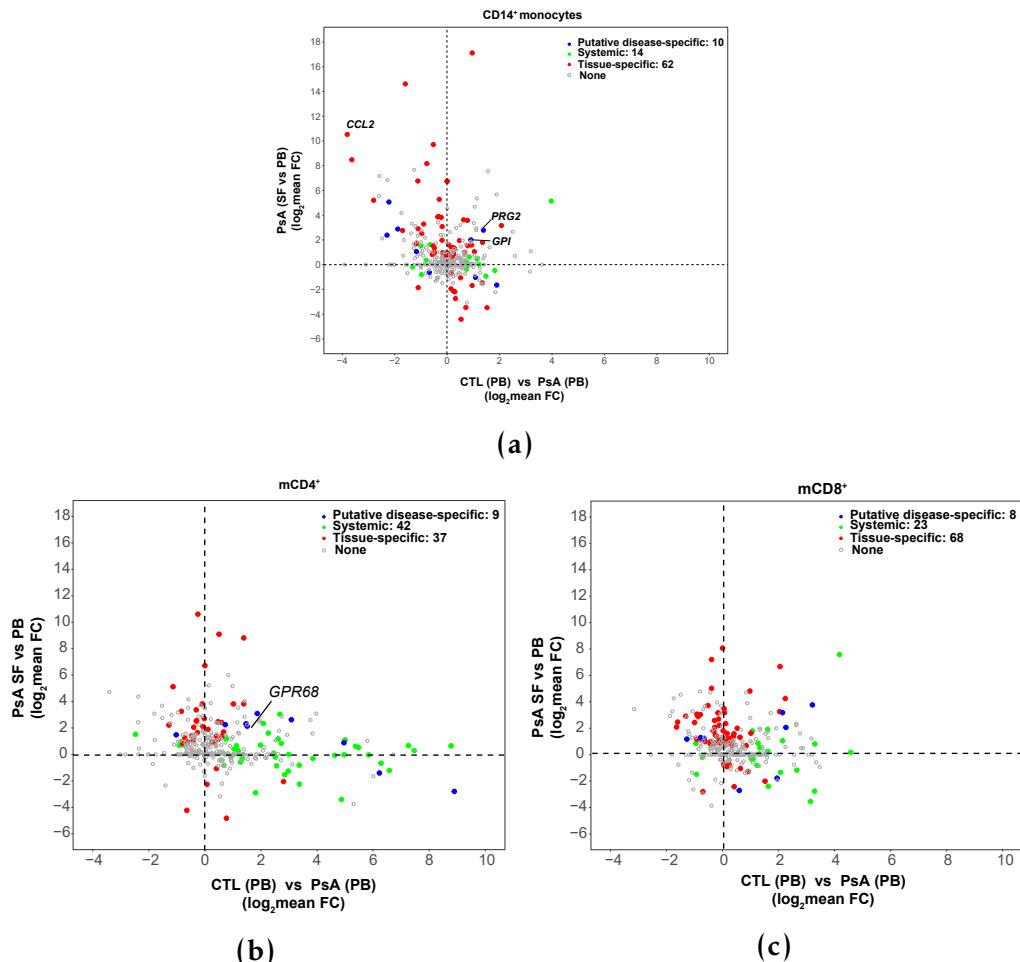


Figure 5.13: Comparison of immune-relevant gene expression modulation across PsA tissues (SF vs PB) and in PsA patients verus healthy controls. The qPCR \log_2 mean FC for each of the genes in the PsA SF vs PB contrast are plotted against the \log_2 mean FC for the same genes in the PsA PB vs healthy control PB contrast in a) CD14⁺ monocytes, b) mCD4⁺ and c) mCD8⁺ cells. The genes are colour-coded based three categories of genes built according by comparison of changes in gene modulation between the two contrasts: only significantly modulated in PB between controls and PsA (systemic genes), only significantly modulated between SF and PB in PsA patients (tissue-specific) and significantly modulated between controls and PsA patients in PB as well as between SF and PB in PsA patients (putative disease-specific).

Cross-tissue comparison analysis in PsA

considerably outnumbered the systemic ones (62 versus 14), showing a more pronounced change in the expression profile of immune genes across patients' tissues than between healthy and diseased PB (Figure 5.13 a red dots). For example, the aforementioned *NFKB* and *MYD88*, *TLR2* genes were only up-regulated in PsA SF CD14⁺ monocytes and their expression was not significantly modulated between healthy controls and PsA circulating CD14⁺ monocytes. Similarly to CD14⁺ monocytes, mCD8⁺ cells also presented greater disease tissue-specific modulation than genes differentially expressed when compared to controls in PB (Figure 5.13 c red dots). For all three cell types, the two genes presenting the greatest FC between SF and PB *SPP1* and *FN1* appeared to be tissue-specific genes and no significant changes in PB between PsA and healthy controls were identified.

The third category comprised genes significantly modulated for each cell type between controls and PsA patients in PB as well as between SF and PB in PsA patients. These genes defined as putative disease-specific genes presented similar numbers across CD14⁺ monocytes, mCD4⁺ and mCD8⁺ (10, 9 and 8, respectively) (Figure 5.13 blue dots in a, b and c). In CD14⁺ monocytes two of those genes, *GPI* and *PRG2*, were up-regulated in both comparisons, with further exacerbation in SF (Figure 5.13 a). Evidence of the glucose-6-phosphate isomerase *GPI* up-regulation in disease has been found in RA synovial fibroblasts and linked to increased levels of TNF- α and IL-1 β in the synovium (Zhong2015). Another example of exacerbated up-regulation in SF was the expression of *GPR68* in mCD4⁺. This gene was up-regulated in PsA PB mCD4⁺ when compared to the control counterparts and further up-regulated in SF when compared to PB in PsA individuals (Figure 5.13 b). *GPR68* was also up-regulated in SF compared to PB in mCD8⁺ cells, reinforcing the relevance of this gene in the synovial pathophysiological aspect of PsA. Amongst the genes presenting an opposite behavior is the epidermal growth factor-like amphiregulin (*AREG*),

Cross-tissue comparison analysis in PsA

which in mCD8⁺ is significantly up-regulated in PsA PB compared to the controls but is down-regulated in PsA individuals when comparing SF versus PB (Figure 5.13 c). Despite the interesting aforementioned findings, the identification of disease-specific and disease tissue-specific genes is clearly limited by the impossibility of obtaining healthy controls SF to include in the experimental design.

When performing pathway enrichment analysis using the significantly modulated genes between healthy controls and PsA patients PB in the qPCR array, only the Reactome immune system pathway appeared as significant for CD14⁺ monocytes and mCD4⁺ cells. This result reinforced the tissue-specificity of the pathways enriched for the modulated genes between SF and PB in CD14⁺ monocytes PsA patients and clearly suggest a more pronounced inflammatory phenotype of the pathological CD14⁺ monocytes in SF compared to PB.

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

According to the analysis of chromatin accessibility and immune-related gene expression in this pilot cohort, the CD14⁺ monocytes showed the greatest changes in chromatin accessibility and the most reliable modulation of expression for pro-inflammatory chemokines and cytokines between PB and SF. 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.

Cross-tissue comparison analysis in PsA

scRNA-seq reveals two main subpopulations in SF and PB combined CD14⁺ monocytes

ScRNA-seq 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 filtered as explained in 2 and CD14⁺ monocytes were subset from the rest of cell populations by expression of *CD14* and *LYZ*, two of the most accurate expression markers defining this cell population (Figure B.9 a and b). Across all six samples (three SFMCs and three PBMCs), 2,459 cells were CD14⁺ monocytes cells, representing approximately 17% of the bulk SFMCs and PBMCs cells included in the analysis and in line with the proportion of CD14⁺ monocytes previously reported using cell surface markers by mass cytometry (Figure 5.1). The CD14⁺ monocytes identified in each of the three paired PBMCs-SFMCs PsA samples were combined using CCA to correct for intrinsic batch effect, unavoidable due to patient samples recruitment in different days and generation of SFMCs and PBMCs 10X libraries separately. CCA alignment of the six CD14⁺ monocytes populations was followed by conservative unsupervised clustering (using resolution 0.1) and t-SNE visualisation.

Cross-tissue comparison analysis in PsA

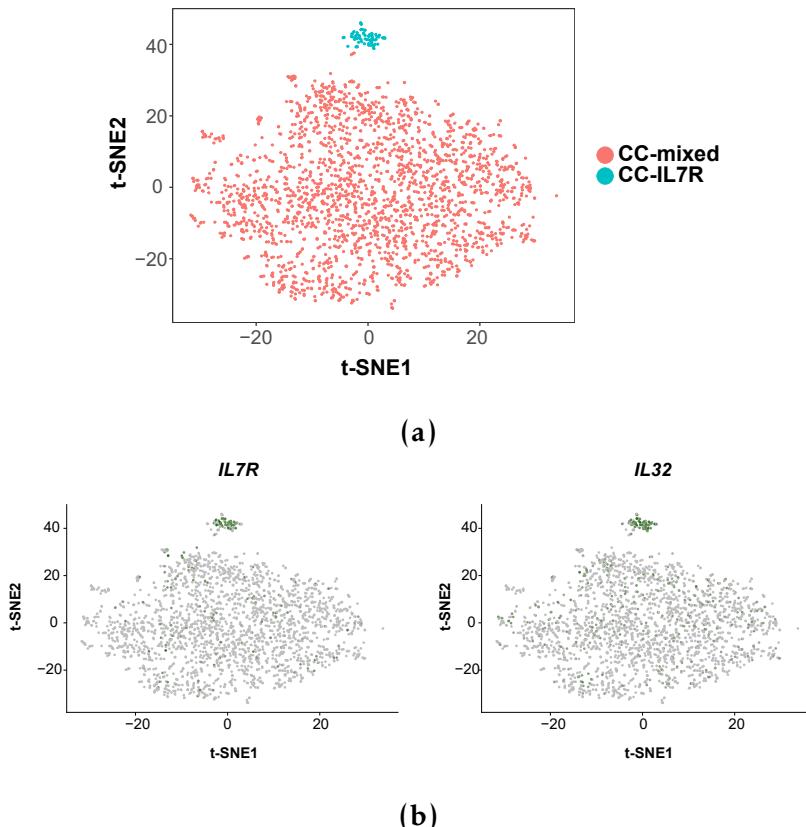


Figure 5.14: Identification of two main $CD14^+$ monocytes subpopulations in the SF and PB combined analysis. a) Visualisation using t-SNE dimensional reduction of the two cluster (CC-mixed and CC-IL7R) identified in the combined SF and PB $CD14^+$ monocyte cells using a very conservative resolution ($\text{res}=0.1$) for the unsupervised clustering analysis. Each of the dots represents a cell, colour-coded by the cluster membership (pink=CC-mixed and turquoise=CC-IL7R). b) Overlap of *IL7R* and *IL32* expression intensities (green) on the t-SNE representation of the SF and PB $CD14^+$ monocytes. *IL32* and *IL7R* gene expression appeared as markers for the $CD14^+$ monocytes from the CC-IL7R cluster.

Using this conservative approach for cluster definition, two robust clusters were identified (Figure B.9 a). The smallest cluster, named a CC-IL7R, was characterised by the expression of *IL7R*, *IL32* and *CCL5*, amongst others, and was formed by a total of 72 cells (43 from SF versus 29 from PB) (Figure ?? b and B.8) ((Al-Mossawi et al. 2018), in revision). The proportion of $IL7R^+CD14^+$ monocytes when compared to the total $CD14^+$ monocyte population was very similar in SF and PB (3 and 2.7%, respectively) in this data. The largest cluster, named as CC-mixed, consisted of 2,387 (1,356 SF and 1,031 PB). CC-mixed

Cross-tissue comparison analysis in PsA

was an heterogeneous cluster, without consistent expression pattern for those genes identified as cluster markers (Figure B.8). When using a less conservative approach for cluster definition by increasing the resolution (resolution 0.4, 0.6 and 0.8), additional clusters were identified. Similarly to the observation in the most conservative approach, no consistency was found in the expression of the top genes defined as markers by cells of the same cluster (data not shown). Due to the moderate cohort size, limitation in accounting for batch effect for cluster identification and the complexity in the definition and identification of stable clusters this analysis could only yield limited information about monocytes subpopulations. Increasing cohort size and a implementation of more exhaustive analysis and alternative methods could lead to identification of additional subpopulations within the combined SF and PB CD14⁺ monocytes in the CC-mixed cluster (see 5.4.6). For the scope of this project, downstream analysis, including DGE between the two tissues, was performed in the CC-mixed and CC-IL7R clusters identified by the most conservative approach (resolution 0.1).

Differential gene expression between SF and PB CD14⁺ monocytes in CC-mixed and CC-IL7R

DGE analysis was performed in order to explore differences between SF and PB within each of these two main CD14⁺ monocyte subpopulations. For the CC-mixed cluster, a total of 251 genes were differentially expressed at an FDR<0.01 and FC>1.5 between SF and PB, of which 149 and 102 presented up- and down-regulation, respectively (Figure 5.15 a). Differential analysis within the CC-IL7R cluster revealed a total of 37 modulated genes, with the majority (35 out of 37) up-regulated in SF compared to PB (Figure 5.15 a). Due to the low number of cells in the CC-IL7R cluster and the limited sample size (n=3), the analysis only identified as significantly differentially expressed (FDR<0.01) genes presenting FC>1.5. Out of the 37 DEGs in the CC-IL7R cluster between the two tissues, 30

Cross-tissue comparison analysis in PsA

were also shared by the CC-mixed cluster. The seven distinctly modulated genes in the CC-IL7R cluster included *CD44* (receptor of the protein product of *SPP1*), *MT-CO2* or S-ribosomal protein (RPS) genes (*RPS29* and *RPS27*).

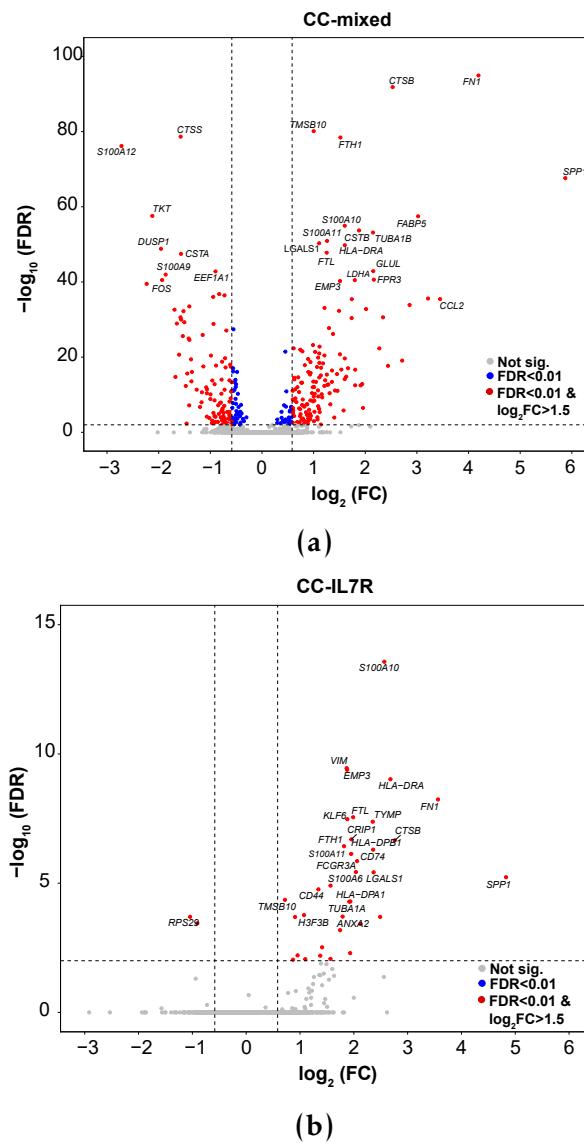


Figure 5.15: Sc-RNA-seq differential gene expression results between SF and PB in the CC-mixed and CC-IL7R CD14⁺ monocytes subpopulations. Vulcano plots showing differences in gene expression between SF and PB a) CC-mixed and b) CC-IL7R CD14⁺ monocyte cluster. In a and b, the significance ($\log_{10}\text{FDR}$) of the differential expression (y-axis) is plotted against the $\log_2\text{FC}$. A positive FC indicates higher expression in CD14⁺ monocytes from SF compared to PB. Genes showing $\text{FDR} < 0.01$ are coloured in blue and genes presenting $\text{FDR} < 0.01$ and $\log_2\text{FC} > 1.5$ are coloured in red. The most significant genes are labelled.

Cross-tissue comparison analysis in PsA

Comparison with the qPCR expression analysis revealed a modest overlap between the two assays, particularly for the DEGs in the CC-IL7R. Amongst the 72 DEGs ($p\text{val}<0.05$) detected by qPCR between SF and PB in CD14⁺ monocytes, only 12 and 4 genes were also differentially expressed ($\text{FDR}<0.01$ and no FC threshold) in the CC-mixed and CC-IL7R clusters, respectively (Figure ?? a). Genes with reproducible differential expression between SF and PB by the two approaches included genes with the largest FCs in both assays, such as *SSP1*, *FN1*, *OLR1* and *S100A12*, being the direction of change also consistent for all of them.

Cross-tissue comparison analysis in PsA

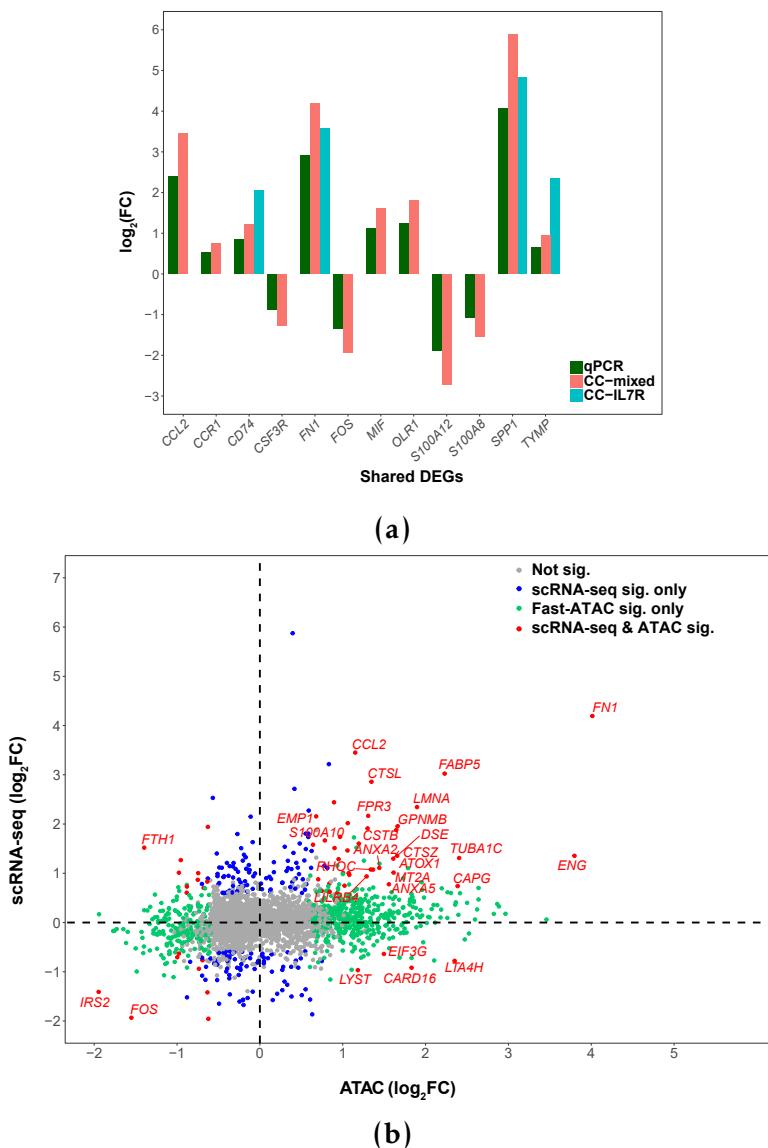


Figure 5.16: Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14⁺ monocytes. a) Overlap between the scRNA-seq DEGs (significant based only on $\text{FDR} < 0.01$) from the CC-mixed and CC-IL7R CD14⁺ monocyte cluster in SF versus PB and the corresponding DEGs ($p\text{val} < 0.05$) detected by qPCR in the bulk CD14⁺ monocytes population. For each of the shared DEGs, the $\log_2\text{FCs}$ from the qPCR and the 10X scRNA-seq analysis are represented. b) Correlation plot comparing SF and PB differences in scRNA-seq expression of the CC-mixed CD14⁺ monocytes and ATAC chromatin accessibility in total CD14⁺ monocytes. The $\log_2\text{FCs}$ for scRNA-seq differential expression of all transcripts in the CC-mixed CD14⁺ monocytes are plotted against the $\log_2\text{FC}$ for total CD14⁺ 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⁺ monocytes. Significance is based on FDR and FC thresholds ($\text{FDR} < 0.01$ and $\text{FC} > 1.5$) in both assays.

Cross-tissue comparison analysis in PsA

Pathway enrichment analysis was performed for the significantly DEGs between SF and PB in the CC-mixed and CC-IL7R subpopulations. The DEGs in the CC-mixed cluster were significantly enriched (FDR<0.01) a number of pathways (Table 5.7 and A.6), some of them of particular pathological relevance (Table 5.7).

Cluster	Pathways
CC-mixed	Ag processing and presentation via MHC-II
	Extracellular matrix and extracellular matrix-associated proteins
	Phagosome and lysosome formation
	IFN signalling
	Cytokine signalling *
	Apoptosis
CC-IL7R	Innate immunity
	Adaptive immunity
	Ag processing and presentation
	Phagosome
	Extracellular matrix and extracellular matrix-associated proteins

Table 5.7: Most relevant enriched pathways for the DEGs between SF and PBCD14⁺ monocytes in CC-mixed and CC-IL7R. Significantly DEGs based on the FDR and FC threshold were used for the analysis. Most relevant enriched pathways based on FDR<0.01. (*) Enrichment for FDR<0.05. Additional pathways included in Table A.6.

One of those pathways was for the Ag processing and presentation pathway, contributed by up-regulated expression in SF CD14⁺ monocytes of *CD74* and genes from the *HLA-D* family such as *HLA-DQA1* and *HLA-DRB1* (Table 5.7). Enrichment for IFN signalling was driven differential expression of genes such as *IFI6*, *IFITM3*, *ISG15* and the TF *STAT1*, all of them up-regulated in SF when compared to PB CD14⁺ monocytes. Another relevant enriched pathway was the extracellular matrix and extracellular matrix-associated proteins, which involve genes of the S100 family (including *S100A8*, *S100A9*, *S100A10*, *S100A11* and *S100A12*, that interact with the receptor for advance glycosylation end products (RAGE) and induce production of matrix-degrading enzymes. The phagosome

Cross-tissue comparison analysis in PsA

and lysosome formation pathway also appeared to be more active in SF CD14⁺ monocytes, with up-regulation of genes such as *CTSL*, which is involved in protein degradation in lisosomes and phagocytosis of apoptotic cells. Lastly, enrichment for cytokine signalling was not driven by differential expression of cytokines but was contributed by up-regulation of pro-inflammatory TFs such as *STAT1*, amongst other genes. The most functionally relevant significantly enriched pathways identified for DEGs in the CC-IL7R subpopulation were common to the ones found in the CC-mixed cluster (Table 5.7).

Moderate genome-wide correlation between chromatin accessibility and scRNA-seq expression in the CC-mixed cluster

In order to determine the overall correlation between scRNA-seq expression and chromatin accessibility, comparison between the log₂FCs for all the expressed genes in the CC-mixed CD14⁺ monocytes cluster and all the accessible chromatin regions in bulk CD14⁺ monocytes between SF and PB was conducted . Changes in expression and chromatin accessibility only presented a moderate correlation in this data ($R=0.214$, $pval=2\times 10^{-16}$) (Figure 5.16 b). In the CC-mixed cluster, 64 genes out of the 251 DEGs ($FDR<0.01$ and $FC>1.5$) were proximal ($\leq 5\text{Kb}$) to one or more ATAC DARs (Table 5.8). This overlap was significant and highlighted the enrichment (Fisher exact test $pval=1.5\times 10^{-3}$) of DEGs in the CC-mixed cluster for proximal DARs identified by ATAC in bulk CD14⁺ monocytes. The majority of the overlaps corresponded to matched increase or decrease (40 and 12 genes, respectively) of gene expression and chromatin accessibility in SF vs PB. However, 14 DEGs in the CC-mixed cluster showed opposite direction of change between expression and chromatin accessibility (Table 5.8).

Amongst the DEGs in the CC-mixed cluster overlapping a proximal DAR were *CCL2* and *FN1* (Figure 5.16 b). Both genes were up-regulated in SF compared to PB in the CC-mixed cluster, proximal to a SF open DAR, and found

Cross-tissue comparison analysis in PsA

Cluster	Up-regulated genes with proximal SF open DAR	Up-regulated genes with proximal PB open DAR	Opposite direction in expression and DAR
CC-mixed	40	10	14
CC-IL7R	9	0	4

Table 5.8: scRNA-seq DEGs in SF versus PB CD14⁺ monocytes proximal to an ATAC DAR.. For each of the two CD14⁺ monocytes cluster identified by scRNA-seq analysis, an overlap is defined when a gene is differentially expressed (FDR<0.01 and FC>1.5) between SF and PB and a proximal significant DAR ($\leq 5\text{Kb}$) showing same or opposite direction of change is also found.

up-regulated in the same direction by the qPCR expression analysis in SF bulk CD14⁺ monocytes compared to PB (Figure 5.16 a).

Similarly to the CC-mixed results, the DEGs in the CC-IL7R cluster between SF and PB were also enriched for those genes with at least one DARs nearby (Fisher exact test $pval=1.85\times 10^{-9}$). Amongst the 22 DEGs overlapping a proximal DAR, 13 of them had correlated dysregulation of expression and chromatin accessibility only 4 presented opposite directionality in the variation of the two features (Table 5.8). The CC-IL7R cluster and CD44 data not shown.

Overall, this results have shown only moderate correlation between gene expression and proximal chromatin accessibility, which may highlight causality to some extent for the dysregulation of the chromatin landscape in the alteration of gene expression between CD14⁺ monocytes in the two tissues.

5.3.7 Mass cytometry reveals differences in protein expression consistent with the chromatin accessibility and transcriptomic profile in PsA CD14⁺ monocytes.

To determine the effect of chromatin accessibility and genes expression differences in CD14⁺ monocytes between SF and PB at the protein level mass cytometry was conducted in collaboration with Dr Hussein Al-Mossawi and Dr Nicole Yager. For those samples with paired ATAC and/or qPCR expression data,

Cross-tissue comparison analysis in PsA

intra-cellular staining for relevant cytokines was performed before and after treatment with BFA, which blocks cytokine release and enables identification of cells actively producing cytokines in absence of additional inflammatory stimuli (see Chapter 2). Due to technical problems, only a limited number of intra-cellular cytokine staining passed QC, being TNF- α amongst the more relevant ones.

Mass cytometry expression of CD14 $^{+}$ versus TNF- α demonstrated greater intensity of TNF α staining after 6h of BFA treatment in SF (Figure 5.17 a blue dots bottom panel) when compared to PB CD14 $^{+}$ monocytes in all three PsA patients (Figure 5.17 a blue dots top panel). Furthermore, the percentage of TNF- α producing cells in each tissue was quantified as the difference of TNF- α $^{+}$ cells before and after BFA treatment for 6h (6h minus 0h). In PB, the resulting percentage of CD14 $^{+}$ monocytes TNF- α $^{+}$ did not surpass the 1%, indicating a very low increment of TNF α producing cells after BFA treatment. Conversely, SF CD14 $^{+}$ cells showed a larger percentage of cells actively producing TNF- α , ranging between 1.5 and 11.8% (Figure 5.17 b). This observation was consistent with increased chromatin accessibility nearby genes of the NF- κ B signalling pathway as well as with increased expression of a number of genes in the TLR and NOD-like signalling in SF compared to PB, which were hypothesised to result into enhanced NF κ B activation and cytokine production. Nevertheless, this trend of increased CD14 $^{+}$ cells producing TNF- α in SF compared to PB did not reach significance ($pval=0.25$), likely due to the small sample size.

Cross-tissue comparison analysis in PsA

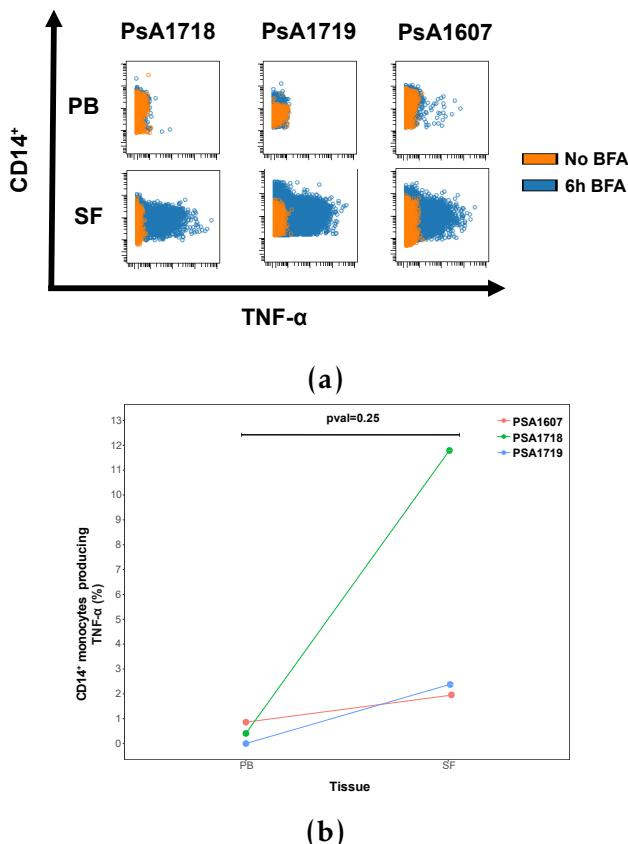


Figure 5.17: Comparison of TNF- α expression by SF and PB CD14 $^{+}$ monocytes before and after protein transport blockade with BFA using mass cytometry. a) For each of the three PsA patients, representation of CD14 $^{+}$ (y-axis) versus TNF- α (x-axis) intensity of expression in matched SF and PB without protein transport blockade (blue dots) or after 6h treatment with BFA (orange dots). b) Percentage of CD14 $^{+}$ monocytes producing TNF- α in SF and PB. For each tissue and sample shown in a, this percentage is calculated as the increment in cells producing TNF- α before and after protein transport blockade with BFA (6h minus 0h).

In order to validate this observation and also assess other cytokines of interest, mass cytometry was conducted in PB and SF from another ten PsA patients. This validation cohort included the patients for which scRNA-seq data was presented in this chapter (Table 5.2). As previously, percentage of CD14 $^{+}$ monocytes producing TNF- α , MCP-1 (protein product of CCL2 and osteopontin (protein product of SPP1) was computed as the difference between percentage of cells producing each of the cytokines before and after BFA treatment. The percentage of CD14 $^{+}$ monocytes producing TNF- α was greater in SF compared to PB (pval=0.048) (Figure 5.18 a), reinforcing the results from the previous cohort.

Cross-tissue comparison analysis in PsA

Likewise, a larger percentage of CD14⁺ monocytes producing osteopontin and MCP-1 (pval=0.001 and pval=0.003, respectively) were also observed in SF compared to PB (Figure 5.18 b and c), in line with the up-regulated expression of these two genes in SF compared to PB in bulk CD14⁺ monocytes and in the CC-mixed scRNA-seq cluster. The percentage of cells producing cytokines in SF was particularly moderate for TNF- α and MCP-1, not exceeding of 1.8 and 3, respectively, for the majority of the PsA samples (Figure 5.18 a and c). However, one of the patients (PSA1505) appeared to have particularly higher percentage of SF CD14⁺ monocytes producing all three cytokines when compared to the rest of the patients in the cohort. Although no justification was found to remove this sample, the statistical significance (pval<0.05) in the differences found between percentage of cell producing TNF- α , MCP-1 and osteopontin in SF versus PB remained when repeating the analysis in absence of PSA1505.

Cross-tissue comparison analysis in PsA

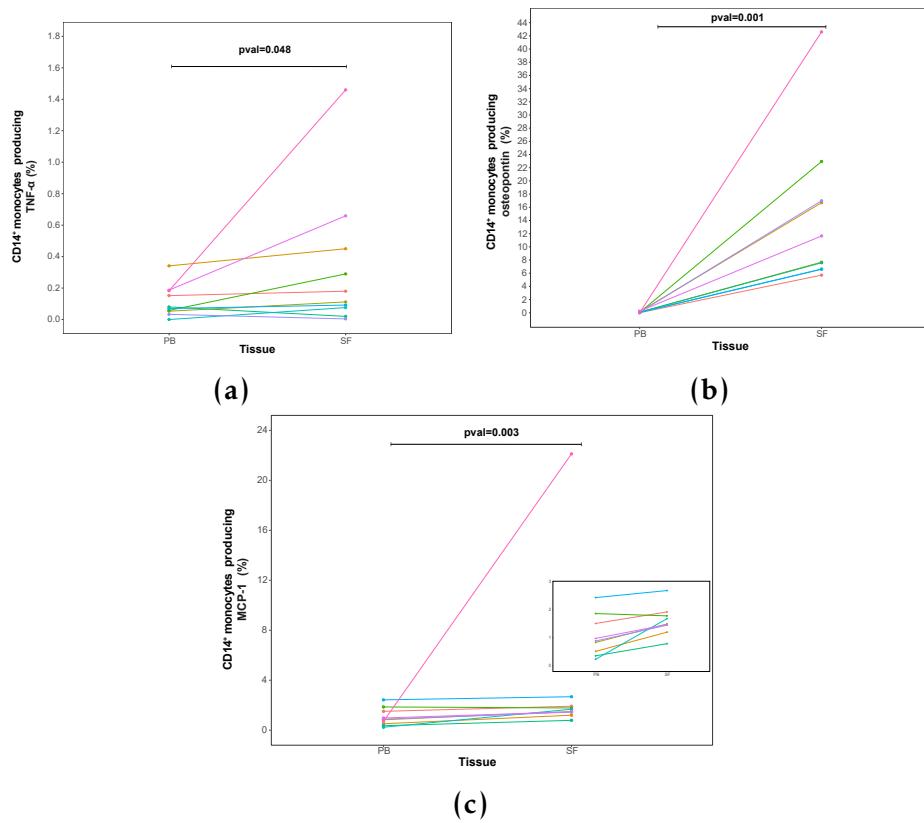


Figure 5.18: Percentage of CD14⁺ monocytes producing TNF- α , osteopontin and MCP-1 in SF and PB in a validation cohort of ten PsA samples. The percentage of CD14⁺ monocytes producing a) TNF- α , b) osteopontin and c) MCP-1 in SF and PB are shown for each of the ten samples in a PsA cohort used to validate cytokine production by mass cytometry. In each sample and tissue, this percentage is calculated as the increment in cells producing the relevant cytokine before and after protein transport blockade with BFA (6h minus 0h). In c), a zoom in for the percentage of CD14⁺ monocytes producing MCP-1 in all patients minus PSA1505 is included for further detail on the differences across SF and PB for these samples.

CCL2-CCR2 signalling: an example of multi-omics correlation

The differences in percentage of CD14⁺ monocytes producing MCP-1 between SF and PB represent an example of a putative correlation between changes observed in the chromatin accessibility landscape, bulk and scRNA-seq expression and protein production. Differential chromatin accessibility analysis between SF and PB identified a statistically significant cell type-specific SF open DAR upstream *CCL2* gene (Figure 5.19). This SF open DAR region is annotated as enhancer according to Epigenome Roadmap chromatin segmentation and overlaps a eRNA reported by FANTOM5 in CD14⁺ monocytes. The expression

Cross-tissue comparison analysis in PsA

of *CCL2* was shown to be significantly modulated ($p\text{val}<0.05$ and mean FC >1.5) between SF and PB by qPCR in CD14 $^{+}$ monocytes only, whereas no significant changes were observed for mCD4 $^{+}$ and mCD8 $^{+}$ in this data from the same patients (Figure 5.10 a, b, c and Table 5.5). Up-regulation of *CCL2* was not found in PB CD14 $^{+}$ monocytes compared to PsA patients and healthy controls, being defined as one of the tissue-specific genes in the previous analysis (Figure ?? a). Furthermore, *CCL2* was also identified by scRNA-seq as one of the up-regulated genes in the CC-mixed cluster (Figure 5.15 a and 5.16 a and b). Expression of *CCR2*, the receptor for the chemokine MCP-1 (protein product of *CCL2*), appeared up-regulated by qPCR in SF mCD4 $^{+}$ and mCD8 $^{+}$ cells in the same individuals, which could suggest increased chemotaxis driven by CD14 $^{+}$ monocytes and leading to T cell infiltration in the synovium. Interestingly, in this data no significant up-regulation of *CCR2* was observed in PsA PB when compared to healthy controls in any of the three cell types.

Cross-tissue comparison analysis in PsA

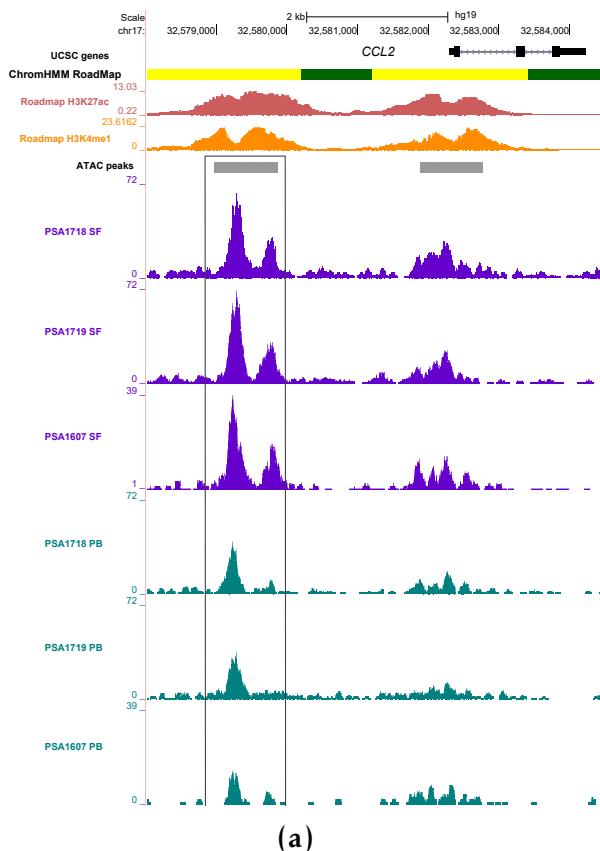


Figure 5.19: Chromatin landscape in CD14⁺ monocytes upstream the differentially expressed gene CCL2. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) for two ATAC peaks at the promoter and upstream the CCL2 gene (x-axis) in SF and PB CD14⁺ monocytes from three PsA patients. The ATAC peak upstream CCL2 (black rectangle) appeared as a significant DAR (FDR<0.01 and FC>1.5) in the differential analysis, being more accessible in SF when compared to PB. Tracks are colour-coded by tissue (SF=purple and PB=turquoise). Additional epigenetic tracks of PB isolated CD14⁺ monocytes from the Epigenome Roadmap, including chromatin segmentation map and the enhancer marks H3Kme1 and H3K27ac, are also included.

In addition to the mass cytometry data, collaborators in Basel have performed measurements of cytokine levels in PB and SF from same ten patients in the validation cohort. High levels of MCP-1 (approximately 1,000 pg/mL) were also reported in SF whereas this cytokine was below the lower detection level in the assay in plasma, supporting additional evidence of CCL2 up-regulated expression at the protein level (Figure B.10). Overall, the data here presented suggests a tissue and cell type specific dysregulation of CCL2 expression in the CC-mixed CD14⁺ monocytes cluster that may be related to alterations in the chromatin accessibility of an enhancer in the proximity to this gene.

5.3.8 Prioritisation and interpretation of PsA GWAS SNPs

The generation of epigenetic and expression data from different cell types isolated from SF and PB aims to contribute to the general understanding of disease pathophysiology and differences between affected and non-affected tissue. Furthermore, overlapping this data derived from clinical samples with fine-mapped credible sets of SNPs may be more informative for refining the number of putative functional causal variants in non-coding or intergenic GWAS associations, compared to the only integration of epigenetic data from cell lines or healthy controls.

Bayesian fine-mapping using genotyping data

In order to further refine the PsA Immunochip GWAS signals identified by Bowes and colleagues, Bayesian fine-mapping was conducted using genotype data from 1,103 patients and 8,900 controls, which corresponds to the PsA Immunochip UK cohort from Bowes *et al.*, 2015. Bowes and colleagues performed fine-mapping for some of the loci and reported the number of independent signals for each locus as well as the number of SNPs in the credible set (Bowes et al. 2015) accounting for 99% of the association in that regions. Nevertheless, *de novo* fine-mapping was conducted here to obtain the exact identity of the SNPs in the credible set and integrate them with the chromatin accessibility data generated in this chapter.

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 $\log^{10}ABF$ under 3 (cut-off used in (Bunt et al. 2015)) for the lead SNP in the fine-mapping association analysis (Table A.7). The majority corresponded to GWAS signals with the lowest significance for the lead SNP ($pval > 10^{-4}$) in Bowes *et al.* study, such as *RSPH3/TAGAP* or *ELMO1* (Table A.7) and they were not fine-mapped either by Bowes *et al.*, 2015. Some of those regions with $pval < 10^{-4}$, for

Cross-tissue comparison analysis in PsA

example *DDX58*, were also discarded by Bowes and colleagues for fine-mapping based on marker density being lower than 100 SNPs.

Cross-tissue comparison analysis in PsA

Table 5.9: Summary table of the PsA GWAS loci presenting $-\log_{10}ABF > 3$ for the fine-mapping lead SNP. For 12 PsA loci $\log_{10}ABF$ of the fine-mapping lead SNP was 3 or greater. In 4 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. FM=fine-mapping; MAF=minor allele frequency; OR=odds ratio; ABF=approximate Bayes factor; PP=posterior probability.

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

Cross-tissue comparison analysis in PsA

Cross-tissue comparison analysis in PsA

Amongst the loci that failed to be fine-mapped in this analysis, 4 (*B3GNT2*, *NOS2A*, *REL* and *TNIP1*) had been successfully fine-mapped by Bowes and colleagues (Table A.7). This is consistent with the dependence of fine-mapping success on sample size (Bunt et al. 2015) and highlighted the reduced power of the analysis presented here, limited by a smaller sample size (only UK cohort as previously mentioned).

Of the 12 loci passing the $\log_{10}ABF \geq 3$ cut-off, 5 showed 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 5.9 labelled with *). In some cases this was the result of the association analysis identifying spurious signals or signals from other loci nearby. For example, the association analysis performed for the fine-mapping of the *IL13* locus was confounded by the *TYK2* signal. The additional 7 signals successfully fine-mapped in this study contained the lead GWAS SNP and Bowes and colleagues fine-mapped lead SNPs in the credible sets and ranged between 2 and 110 SNPs in the 90% credible set (Table 5.9). Moreover, two of those loci, *IL23R* and *TRAF3IP2*, presented the same fine-mapping lead SNPs as the reported by Bowes *et al.*, 2015. Regarding additional independent signals, a secondary signal reported by Bowes and colleagues in the *IL12B* locus was also identified here by step-wise conditional analysis.

Integrating fine-mapped SNPs and chromatin accessibility from PsA samples

The union of the 90% credible sets for the 7 successfully fine-mapped loci comprised a total of 292 unique SNPs. These SNPs were used to perform overlap with the significant (FDR<0.01 and FC>1.5) DARs identified between SF and PB in CD14⁺ monocytes, mCD4⁺, mCD8⁺ and NK cells. Unfortunately, none of the 292 SNPs were contained by a DAR in any of these cell types. Additional overlap was performed between these SNPs and the accessible chromatin regions (consensus peaks without filtering based on the differential analysis) in each

Cross-tissue comparison analysis in PsA

of the four cell types assayed by ATAC. The largest number of SNPs (17) was found to overlap accessible chromatin in CD14⁺ monocytes, followed by mCD8⁺, mCD4⁺ and NK cells (Table 5.10). The 43 unique SNPs from the 90% credible set overlapping ATAC accessible chromatin were distributed across the *CSF2* (8), *IL12B* (3), *IL23R* (4), *RUNX3* (6), *STAT2* (14), *TRAF3IP2* (1) and *TYK2* (7) loci. A number of these SNPs were found to only overlap accessible chromatin in one particular cell type. Interestingly, for the *TRAF3IP2* locus, which was fine-mapped with the greatest resolution, none of the two SNPs of the 90% credible set overlapped accessible chromatin in any of the four studied cell types.

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

Table 5.10: 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 the eight fine-mapped loci overlapping each cell type ATAC master list are reported. Furthermore, 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.

SNPs from the fine-mapping credible set were significantly enriched in ATAC peaks when compared to all other GWAS Catalog SNPs and SNPs in LD ($r^2=8$) in all four cell types (Fisher exact test: CD14⁺ monocytes $pval=7.12\times 10^{-8}$, mCD4⁺ $pval=1.69\times 10^{-10}$, mCD8⁺ $pval=6.40\times 10^{-9}$ and NK $pval=1.86\times 10^{-5}$). Notably, the GWAS Catalog SNPs overlapping ATAC accessible regions were significantly enriched (FDR<0.001) for particular terms from the Experimental Factor Ontology (EFO) (Figure 5.20). The EFO is a hierarchical tree-like ontology where each term represents a (disease) trait or group of related (disease) traits

Cross-tissue comparison analysis in PsA

with which disease-risk SNPs may be annotated (Figure 5.20). Enrichment for general terms (towards the root of the tree) including autoimmune diseases, rheumatic diseases and skin diseases were found across all four cell types. Disease-specific terms (amongst the branches of the tree) related to these general terms, such as CD and IBD, were also found to be enriched for SNPs overlapping ATAC in all four cell types. Conversely, other "branches" from more general terms, including psoriasis and MS, presented significant enrichment (FDR<0.001) only in CD14⁺ monocytes and mCD4⁺ cells, respectively. Overall, this reinforced the specificity of the overlap between GWAS Catalog genetic variants not included in the fine-mapping credible set with accessible chromatin across the immune cell types investigated in this chapter.

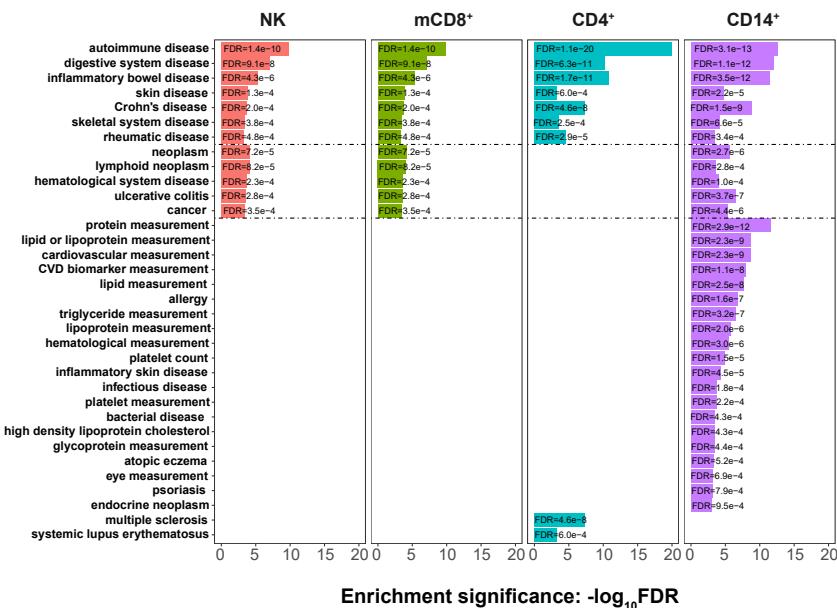


Figure 5.20: Experimental Ontology Factor terms enriched for GWAS Catalog SNPs overlapping ATAC regions in four cell types. Each term annotates a set of risk SNPs associated with a disease trait or a group of related disease traits. Enrichment analysis was performed using as input data the GWAS Catalog SNPs overlapping ATAC accessible chromatin regions. A minimum of ten SNPs overlap and FDR<0.001 was required for enrichment to be considered significant.

Cross-tissue comparison analysis in PsA

Further investigating the PsA-specific 5q31 locus

Following fine-mapping, integration of in-house ATAC and additional functional data with the 90% set of SNPs was conducted to further investigate the 5q31 locus, harbouring the only PsA GWAS association not shared with psoriasis. Out of the 58 SNPs in the 90% credible set, only 8 overlapped ATAC accessible chromatin in at least one of the four cell types included in this study (Figure ?? top panel). Amongst those SNPs were three (rs10065787, rs27437 and rs7721882) of the four variants highlighted by Bowes and colleagues as the most functionally relevant according to ENCODE epigenetics data.

SNP	Cell type ATAC overlap	Top eGene , cell type and condition
rs10065787	CD14 ⁺ , mCD4 ⁺	<i>P4HA2</i> (monocytes LPS2, LPS24, IFN- γ) <i>SLC22A5</i> (monocytes UT)
rs11242104	All	NA
rs11242105	All	NA
rs2069803	All	<i>SLC22A5</i> (CD4 ^{+(*)} , CD8 ⁺)
rs27437	CD14 ⁺ , mCD4 ⁺	<i>SLC22A5</i> (CD4 ⁺ , CD8 ⁺)
rs4705908	All	<i>SLC22A5</i> (CD4 ^{+(*)} , CD8 ⁺)
rs2089855	All	<i>P4HA2</i> (monocytes LPS2, LPS24, IFN- γ), <i>SLC22A5</i> (monocytes UT, IFN- γ , CD4 ^{+(*)} , CD8 ^{+(*)})
rs7721882	mCD4 ⁺	<i>SLC22A5</i> (CD4 ^{+(*)} , CD8 ⁺)

Table 5.11: 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 monocytes (UT, LPS 2h, LPS 24h, IFN- γ) (Fairfax et al. 2014), B cells (Fairfax et al. 2012), NK untreated (Naranbhai et al. 2015), neutrophils untreated (unpublished), tCD4⁺ and tCD8⁺ (Kasela et al. 2017) and whole blood (Jansen et al. 2017). (*) for eQTLs extremely significant (FDR<2.2x10⁻³⁰⁸).

The SNP rs10065787, highlighted by Bowes *et al.* for overlapping with ENCODE clusters of occupancy for TFs relevant in CD4⁺ and CD8⁺ biology, presented accessible chromatin in CD14⁺ monocytes and mCD4⁺ cells, showing also moderate enrichment for the the enhancer histone mark H3K4me1 in

Cross-tissue comparison analysis in PsA

mCD4⁺ (Figure ?? right hand side panel red line). Similarly, the nearby SNP rs27437 overlapped an ATAC peak in CD14⁺ monocytes and mCD4⁺ and the same TFBS site cluster as rs10065787 (Table 5.11 and Figure ?? right hand side panel green line). rs10065787 appeared to be part of an eQTL signal for *SLC22A5* and *P4HA2* expression in unstimulated and stimulated monocytes (LPS2, LPS24, IFN- γ), respectively (Table 5.11). However, no *cis* eQTL for tCD4⁺ was reported for this SNP in the Kasela and colleagues dataset (Table 5.11). In contrast, rs27437 was part of a *cis*-eQTL in tCD4⁺ and tCD8⁺ for *SLC22A5*, the same eGene reported by Bowes and colleagues in their pilot eQTL study. Chromatin conformation data using promoter capture-HiC (**Javierre2016**) in unstimulated monocytes does not clearly reveal interaction for rs10065787 with the promoter of *P4HA2* or *SLC22A5*. Conversely, rs27437 is relatively close to the bait in the *IL3* promoter, which interacts with *SLC22A5*, potentially bringing this SNP in proximity with the promoter of *SLC22A5*.

Other SNPs such as rs2089855 also appeared in an eQTL signal for *P4HA2* and *SLC22A5* in untreated and stimulated monocytes, and were associated with *SLC22A5* expression in tCD4⁺ cells (Table 5.11). This SNP is proximal to rs11955347, which in Bowes *et al.* presented the most significant correlation with expression of *SLC22A5* in tCD4⁺ and tCD8⁺, and has also shown to be within the *cis*-eQTL signal for *SLC22A5* expression in unstimulated and IFN- γ stimulated monocytes and for *P4HA2* in monocytes stimulated with LPS (2 and 24h) (Fairfax *et al.* 2014). The effect of rs11955347 genotype in modulation of *SLC22A5* expression was also confirmed in the larger data set of Kasela *et al.* in both tCD4⁺ and tCD8⁺ cells. These observations suggest 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 conditions.

Another two relevant SNPs from the 90% credible set reported here were rs2069803 and rs4705908, both overlapping ATAC peaks in all four cell types

Cross-tissue comparison analysis in PsA

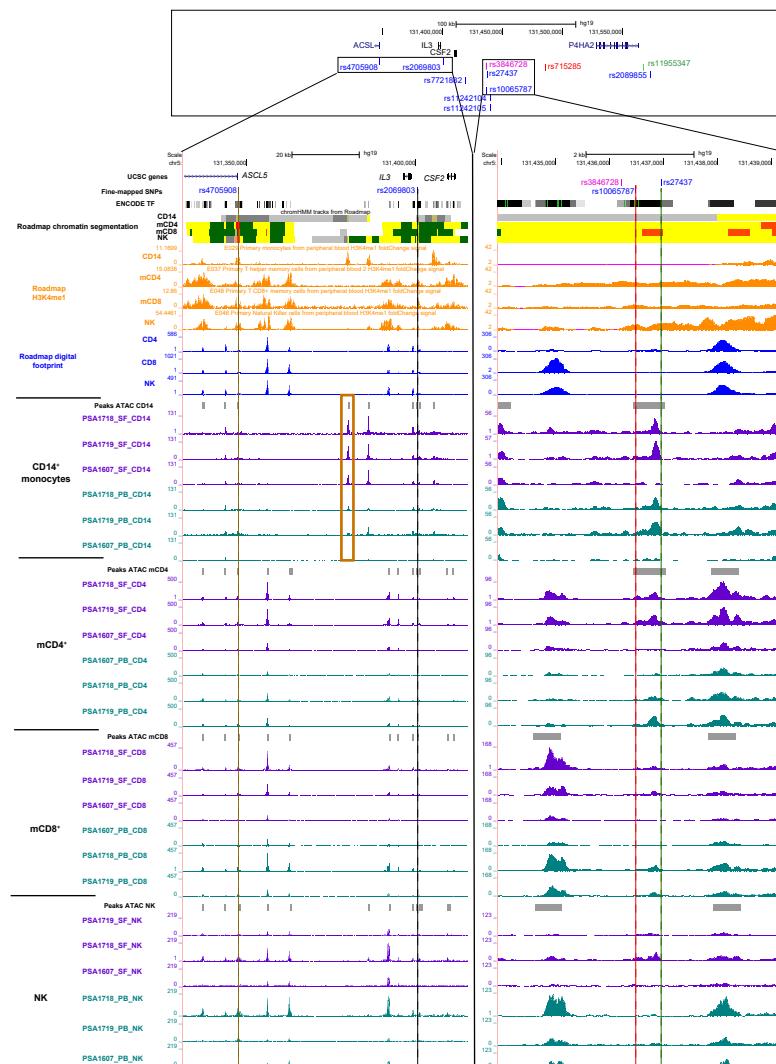


Figure 5.21: Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal. The top panel shows the genomic location of the six SNPs in the 5q31 fine-mapping 90% credible set overlapping PsA ATAC accessible regions in at least one of the four cell types (in blue). The schema also includes the location of relevant SNPs for the 5q31 PsA-specific GWAS region from the Bowes *et al.*, study, including the GWAS lead SNP (in red), the eQTL SNP showing the best correlation with the GWAS lead SNP (in green) and one SNPs from the credible set overlapping several ENCODE annotation features with no overlap for my PsA ATAC data (in purple). The left and right hand side panels are the UCSC visualisation of the epigenetic landscape for rs4705908 and rs2069803 (left brown and black lines, respectively) and rs1006587 and rs27437 (right, red and green lines, respectively), which represent four of the most relevant fine-mapped SNPs at the 5q31 in terms of overlap with PsA ATAC data signal for eQTLs. For each panel, all the ATAC tracks from three PsA patients and four cell types isolated from SF and PB are included. ATAC tracks are colour-coded by tissue (SF=purple and PB=turquoise). Additionally, publicly available epigenetic data (H3K4me1, chromatin segmentation maps, digital footprint and ENCODE cluster for TF binding) generated in the same cell types as the in-house ATAC are included. The yellow box highlights a DAR found in CD14⁺ monocytes near rs2069803 and rs4705908. H3K4me1 relative fold-enrichment signal and ATAC normalised counts within each cell type are shown (y-axis).

Cross-tissue comparison analysis in PsA

(Figure ?? left hand side panel black and brown lines, respectively). rs4705908 is located upstream from the promoter of the *ACSL6* gene in a region showing H3K4me1 enrichment, supporting a regulatory role. Notably, rs4705908 maps to a CTCF binding site reported in GM12878 and LCLs cell lines. Likewise, rs2069803 also overlaps moderate H3K4me1 signal in mCD4⁺, mCD8⁺ and NK cells (Figure ?? left hand side panel). The region has also been annotated as an enhancer and weakly transcribed in mCD4⁺, mCD8⁺ and NK cells by the Epigenome Roadmap chromatin segmentation maps (yellow and light green in ?? left hand side panel). Although accessible chromatin has been identified at rs2069803 and rs4705908 for all cell types, *cis*-eQTL for these two SNPs have only been reported in CD4⁺ or CD8⁺, with the genotypes of both SNPs correlating with regulation of *SLC22A5* expression with extremely high significance in tCD4⁺ cells (Table 5.11). Promoter capture Hi-C data in naïve and total CD4⁺ CD8⁺ cells revealed interaction of the *IL3* promoter bait containing rs2069803 with the promoter of *SLC22A5*. Interestingly, rs4705908 also appeared within the bait of the *ACSL6* promoter interacting with the *IL3* promoter bait, which also includes the previously mentioned rs2069803 variant. Overall, promoter-capture HiC data revealed potential physical interactions between rs27437, rs2069803 and rs4705908 in CD4⁺ and CD8⁺ cells with potential functional relevance in regulation of *SLC22A5* expression.

Lastly, rs7721882 was the only SNP showing a very significant eQTL effect in tCD4⁺ for *SLC22A5* expression and overlapping with a putative mCD4⁺ cell type specific ATAC peak (Table 5.11). However, additional replicates and greater sequencing depth would be required in order to confirm the robustness of this peak as well as the cell type specificity in mCD4⁺ cells.

In addition to epigenetic evidence, further investigation of the suitability of *SLC22A5* and *P4HA2*, two of the genes showing eQTLs for some of the SNPs in the 5q31 credible set, as drug targets for PsA was investigated. Dr Hai Fang

has developed an algorithm named Priority Index (Pi) based on random forest to leverage genetic information in the prioritisation of putative drug targets for particular complex diseases. Interestingly, *SLC22A5* appeared as the third gene in the psoriasis rank supported as eGene in several immune cell eQTLs studies, annotation by the disease ontology with related inflammatory disease terms (including CD, IBD and RA, among others) as well as prediction for high druggability based on protein structure data (http://galahad.well.ox.ac.uk:3010/pidb/discovery/PS0/SLC22A5#bookmark_details_genomic). In contrast, *P4HA2* appeared 4,172 in the rank for psoriasis, not being such a suitable putative drug target in this disease.

5.4 Discussion

5.4.1 Characterising chromatin accessibility in PsA samples

Technological advances has enabled characterization of the epigenetic landscape in immune cells isolated from PB and SF in PsA patients. The cell types included in this study ($CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK) represent key players in the innate and adaptive immune response dysregulated in PsA pathophysiology. In particular, expansion of $mCD4^+$ and $mCD8^+$ cells in the synovium of PsA patients has been described, and GWAS have highlighted significant association with MHC-I and other pivotal genes involved in T cell immune response. (Taams2018).

In this chapter I have used ATAC to identify genome-wide differences between PB and SF in four disease-relevant cell types from PsA patients. $CD14^+$ monocytes demonstrated the largest number of DARs between SF and PB (23.3% of the investigated regions). For all the cell types, the majority of the DARs were located at intergenic and intronic regions annotated as enhancers (weak or strong), and were also enriched for eRNAs identified in each cell type by the

Cross-tissue comparison analysis in PsA

FANTOM project (Fantom et al. 2014). Altogether these findings may suggest a role for the differential regions in long-range regulation of gene expression (Qu et al. 2015). Pathway enrichment analysis of the genes proximal to the DARs revealed both commonalities and differences across cell types. In SF, enrichment for genes in the NF- κ B pathway was found in CD14 $^{+}$ monocytes, consistent with the TNF- α signature and the efficacy of anti-TNF therapies already reported for PsA (**Ahil2016**). Interestingly, in mCD8 $^{+}$ T cells enrichment of DARs open in PB was found proximal to genes in the Wnt signalling 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 PB 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 accessibility proximal to gene members of the NK-mediated cytotoxicity pathway was found in PB compared to SF cells. NK CD56 bright cells resident at sites of inflammation are more specialised for cytokine production than cytotoxicity (Michel et al. 2016). Notably, matched mass cytometry data in these patients have shown a reduced proportion of CD56 bright cells in PB compared to SF which would be consistent with the NK-mediated cytotoxicity enrichment.

Overall, this approach has identified robust differences in chromatin accessibility between SF and PB 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. Moreover, accessible chromatin is not a definitive marker for regulatory regions and mapping of histone marks such as H3K4me1 and H3K27ac together with eRNA quantification will help to refine the functional relevance of the identified DARs.

5.4.2 Bulk gene expression profiling and integration with 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 SF and PB from the same patient have yet to be reported.

In this chapter, I have presented a pilot study characterising expression of relevant immune genes in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells isolated from SF and PB and integrated this with paired chromatin accessibility data from the same samples. CD14⁺ monocytes and mCD8⁺ presented the largest number of consistently modulated genes between SF and PB in this pilot analysis. The most significantly dysregulated genes between SF and PB in the four cell types were *SPP1* and *FN1*, the same as reported by Dolcino *et al.*. Other highly DEGs in Dolcino *et al.*s study, including *TNFA*, *CXCL13* or *CCL18*, were also found to be modulated between SF and PB in at least one of the cell types in this pilot data. Consistent with their role in Th17 cell biology, *CXCL13* and *IL26* appeared significantly up-regulated in SF mCD4⁺ and/or mCD8⁺ cells but not in CD14⁺ monocytes (Takagi et al. 2008).

Cross-tissue comparison analysis in PsA

The subsequent integration of transcriptome profiling with paired-ATAC data in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ revealed that genes with differentially modulated expression in SF and PB corresponded with nearby DARs showing changes in the same direction. Some of those DARs had also been identified as eRNAs in CD14⁺ monocytes (e.g *NFKB1*) and mCD4⁺ cells (e.g *CCR6*) by the FANTOM project. Although the characterization of those DARs as eRNAs evidences the downstream role of those regions in active transcriptional regulation, it does not unequivocally link this regulatory role to the proximal gene found to be differentially expressed between tissues by qPCR.

The integration of differences in PB gene expression between PsA patients and controls with the cross-tissue comparison within patients led to identification of systemic, tissue-specific and putative-disease specific modulated genes in this pilot data. According to my data, more profound transcriptional changes across PsA tissues (tissue-specific genes) were identified when compared to changes in expression between PsA patients and controls in PB for CD14⁺ monocytes and mCD8⁺ cells. Systemic genes for mCD8⁺ cells included for example *CCR10*, a chemokine receptor co-expressed by a subset of memory cells that preferentially migrate to skin, which has also been identified as an up-regulated gene in tCD8⁺ cells in the psoriasis cohort (Chapter ?? and in patients with atopic dermatitis (Hijnen et al. 2005). Interestingly, *SPP1* and *FN1* appeared in the tissue-specific category in the three cell types. Dolcino and colleagues had reported these genes as the two most dysregulated when comparing bulk synovial membrane transcriptome from PsA patients and healthy individuals (Dolcino et al. 2015). My data suggests that CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells may all contribute to the up-regulation of *SPP1* and *FN1* in the PsA synovium membranes. The *SPP1* protein product, osteopontin, is a cytokine and chemokine expressed by many cell types, including monocytes/macrophages and T cells. It is involved in cell migration, adhesion and cell-mediated immune

Cross-tissue comparison analysis in PsA

response through regulation of T cells, importantly in Th-17 (Morimoto et al. 2010). *SPP1* also plays a role in other chronic inflammatory and autoimmune diseases (Rittling and research 2015). *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). Moreover, *FN1* has been shown to induce bone resorption mediated by pro-inflammatory mediators, such as nitric oxyde and IL-1 β (Gramoun et al. 2010). In Dolcinos data *FN1* up-regulation was only found when comparing synovial membranes of PsA versus controls, and no changes were reported in PB samples, altogether suggesting the tissue-specificity of this dysregulation in PsA. Furthermore, the identification of DARs open in SF at the promoter and 3 downstream of *FN1* in CD14 $^{+}$ monocytes may suggest a link between changes in modulation of gene expression and chromatin accessibility in this particular cell type.

In contrast to the observation in my pilot study, Dolcino also showed moderate up-regulation of *SPP1* expression in PB from PsA patients compared to controls. This could be explained by the fact that Dolcino's study was performed in bulk PBMCs and dysregulation of *SPP1* in CD14 $^{+}$ monocytes, mCD4 $^{+}$ and mCD8 $^{+}$ cells may be tissue-specific. However, this could also be the result of the small number of qPCR replicates and *SPP1* dysregulation in PB between PsA and healthy samples failing to reach significance in my study. Overall, up-regulation of these two genes in SF cells reflect the activation of chemotaxis and infiltration of circulating monocytes and T cells, activation of the Th-17 immune response and dysregulation of osteoclast bone remodeling , all of which are pivotal in PsA pathophysiology, particularly at the inflamed tissue (**Mensah2017**; Durham et al. 2015).

An example of an interesting putatively disease-specific gene identified by this study was *GPR68*, up-regulated in mCD4 $^{+}$ PsA PB compared to controls

Cross-tissue comparison analysis in PsA

with expression further increased in SF. *GPR68* is a G protein-coupled receptor (GPCR), expressed in T cells and others, that undergoes activation through pH acidification, characteristic of synovial tissues under inflammation (Saxena 2011; Biniecka et al. 2016). Interestingly, *GPR65*, another member of the acid-sensing GPCR family, has been associated with a number of immune mediated diseases, including AS, CD and MS (Cortes et al. 2013; Lassen et al. 2016; Wirasinha et al. 2018). *GPR65* was found to be a marker of pathogenic Th17 cells in the murine and human systems (Al-Mossawi 2017; Gaublomme et al. 2015). Unfortunately, *GPR65* was not included in the gene array used in this study. Indeed, the use of a gene array rather than a transcriptomics approach using RNA sequencing is a limitation of this study.

In terms of relevant biological processes, pathway analysis using consistently modulated genes between SF and PB in this data revealed enrichment for TLR and NOD-like signalling pathways in CD14⁺ monocytes. This was consistent with the relevance of TLR and NOD-like receptors for rheumatic diseases (WJ and Arthritis 2009). Up-regulated expression of *TLR1* and *TLR2* was significant in SF CD14⁺ monocytes compared to PB and a similar trend in *TLR4* expression was also observed but failed to reach significance in this pilot study. These finding were in line with some studies that have identified increased *TLR-2* and *TLR-4* expression in SFMCs compared to PBMCs in patients with juvenile idiopathic arthritis (Myles and Aggarwal 2011). The relevance of NOD-like signaling has also been highlighted in the genome-wide transcriptomic analysis in lesional and uninvolved psoriatic skin presented in Chapter ??, reinforcing the role of NOD-like signalling in the inflammatory response at the site of inflammation. The cross-talk between the TLR and NOD-like signalling pathways was further evidenced by network-based analysis in this data highlighting increased activation of the NFκB TF in SF, particularly in CD14⁺ monocytes. In my data, enrichment of DARs open in SF in the proximity

Cross-tissue comparison analysis in PsA

of genes within the NF- κ B pathway was also found and further supported transcriptionally by the consistent up-regulation of downstream genes such as *TNFA*, *textitCCL2* and *CCL5* in SF CD14 $^{+}$ monocytes. Moreover, analysis for conserved TFBS motifs in the ATAC data revealed enrichment for NF- κ B binding motifs within DARs open in SF CD14 $^{+}$ monocytes but not in DARs open in PB (data not shown).

The enrichment of differentially modulated genes between SF and PB in mCD4 $^{+}$ for the IL-10 signalling pathway was particularly interesting in the context of flow cytometry from the same three patients revealing expansion of Tregs in SF compared to PB (data not shown). Tregs are characterised by the expression of anti-inflammatory cytokines, including IL-10 (OGarra2004). The qPCR transcriptomic data showed in SF mCD4 $^{+}$ and mCD8 $^{+}$ a significant increase in expression of the IL-10 receptor subunit α *IL10RA* and a similar trend for *IL10* up-regulation ($pval=0.14$ and 0.07, respectively) compared to PB. Altogether, this could suggest that inflammation in PsA is refractory to the immunomodulatory effects for IL-10 signalling in SF or counterbalanced by the immunostimulatory properties of this cytokine, which could be one of the reason for failure of IL-10 agonist therapy in CD and psoriasis (Marlow et al. 2013; Kimball et al. 2002).

5.4.3 The relevance of monocytes and investigation of other cell types

In this exploratory data, CD14 $^{+}$ monocytes showed the most DARs and confidently modulated genes between SF and PB as well as functionally relevant pathway enrichment. The work presented is part of a collaborative multi-omics PsA pilot study. In light of this thesis, and 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 SF and PB using PsA samples included in this thesis. Interestingly, a significant number of mCD8⁺ TCR clones with potentially shared antigen recognition across patients were found to be enriched in SF compared to PB (manuscript in preparation).

5.4.4 Characterisation of monocytes by scRNA-seq

Monocytes are very plastic cell populations that undergo cell differentiation at the site of inflammation, with differences that may be better captured at the single-cell level. In this pilot experiment, cluster identification in scRNA-seq from combined SF and PB monocytes using a conservative approach (see 5.4.6) revealed two main subpopulations (CC-mixed and CC-IL7R). CC-mixed appeared as a large heterogenous cluster in contrast to the CC-IL7R subpopulation, characterised by cells consistently expressing *IL7R*, *IL32* and *textit{CCL5}*, amongst other markers. According to the scRNA-seq data, the CC-IL7R cluster represented around 3% of the total monocytes and had approximately the same number of cells from SF and PB (3 and 2.7%, respectively). Conversely, flow cytometry data in PsA and AS patients have shown up to 35% of the total SF CD14⁺ monocyte to be IL-7R⁺ versus approximately 1% in PB (in revision (Al-Mossawi et al. 2018)). This could be due to discrepancies between gene expression and protein translation acknowledged in the literature but may also be a consequence of lower sensitivity of scRNA-seq in quantifying gene expression (Liu et al. 2016; Islam et al. 2014). This may be leading to underestimate the CC-IL7R cluster size and the predominance of SF monocytes IL-7R⁺ in the contribution towards this cluster. Differences in sensitivity between qPCR and scRNA-seq may also partly explain the limited

Cross-tissue comparison analysis in PsA

overlap of differentially expressed genes between SF and PB when using qPCR and scRNA-seq. Although the same top dysregulated genes (including *SPP1*, *FN1* or *OLR1*) were identified by both techniques, a modest number of significantly modulated genes found by qPCR were reproduced by the scRNA-seq analysis in the CC-mixed cluster. In terms of chromatin accessibility, the comparison of FCs from contrasting SF and PB chromatin accessibility in CD14⁺ monocytes and scRNA-seq expression in the CC-mixed cluster only showed moderate correlation. This limited correspondence between chromatin accessibility and gene expression has also been reported by other studies and may be also the result of aforementioned limitations in annotating accessible chromatin with a putative target gene (Ackermann et al. 2016; Wang et al. 2018).

The identification of sub-populations of monocytes expressing IL-7R⁺ is of biological interest as *IL7R* polymorphisms are associated with a number of chronic inflammatory diseases, including AS and MS (Cortes 2007; Gregory et al. 2007). Although the role of IL-7 and IL-7R in mediating the immune response was only characterised in T cells, the relevance of IL-7R in CD14⁺ monocytes under LPS stimulation has been demonstrated in eQTL studies and also at the protein level in a manuscript under review, to which I have contributed (Fairfax et al. 2014; Al-Mossawi et al. 2018). Al-Mossawi and colleagues identified a distinct transcriptional profile of the IL7R cluster in PsA SF monocytes very similar to the gene expression profile from IL7R⁺ *in vitro* stimulated monocytes. Interestingly the *IL-7R* locus showed differentially accessible chromatin in PsA SF and was one of the top DEGs in the CD14⁺ monocyte qPCR array. Moreover, my analysis *CD44* appeared to be up-regulated in SF compared to PB in the CC-IL7R cluster. Interestingly, *CD44* is the receptor for osteopontin and this observation may suggest that the SF CC-IL7R cells may be more responsive to this cytokine. Taken together, this data supports a role for IL-7 signalling in PsA circulating and tissue monocytes in chronic inflammation.

Cross-tissue comparison analysis in PsA

Pathway enrichment analysis using scRNA-seq DEGs between SF and PB in the CC-mixed cluster identified biologically relevant processes, including MHC-II Ag processing, IFN signaling and extracellular-matrix components, amongst others. Interestingly, up-regulation of *IFI6* and *IFITM3*, two of the genes contributing to the enrichment of this pathway, have also been identified as markers of a subpopulation of IFN- γ activated monocytes in from RA synovial tissue using scRNA-seq (Zhang et al. 2018). In addition to the IFN- γ activated, this study identified other three cluster within RA and osteoarthritis (OA) patients monocytes; however IL-7R $^+$ were not explicitly mentioned. Genes enriched for the extracellular-matrix pathway included members of the S100 protein family, previously reported to be dysregulated in lesional skin from psoriasis patients (Chapter ??), which are also involved in joint erosion and development of arthritis (Raghunatha 2012). Two genes of this family, *S100A8* and *S100A9*, were up-regulated in lesional skin compared to uninvolved, but down-regulated in SF CC-mixed monocytes in this data. The lack of overlap between the pathways identified for the DEGs in the CC-mixed and those found by qPCR in bulk CD14 $^+$ monocytes could be due the result of the qPCR array being biased to a small number of genes versus the unbiased approach of the scRNA-seq.

5.4.5 Mass cytometry in CD14 $^+$ monocytes and the integration with chromatin accessibility and transcriptomic data

Single-cell mass cytometry in matched PB and SF was first conducted in the same three PsA with available paired ATAC and/or qPCR data. Despite technical limitations, mass cytometry in this samples identified a greater percentage of CD14 $^+$ monocytes producing TNF- α in the SF compared to PB, in lines with the chromatin accessibility and transcriptomic data suggesting increased NF- κ B activation in this tissue. A cohort with additional ten SF and PB PsA samples

Cross-tissue comparison analysis in PsA

validated with statistical significance the TNF- α observation. Moreover, the expanded cohort also demonstrated increased percentage of CD14 $^{+}$ monocytes actively producing MCP-1 and osteopontin in SF compared to PB, consistently with the up-regulation in *CCL2* and *SPP1* expression in SF. A study using quantitative mass cytometry comparing SF from PsA and OA as control did not identify any of these three proteins to be up-regulated (Cretu et al. 2014). Conversely, another study using enzyme-linked immunosorbent assay (ELISA) reported an increased production of TNF- α , amongst other cytokines, in PsA SF compared to OA (Partsch et al. 1997).

Notably, *CCL2/MCP-1* represented a good example of correlating differences between SF and PB monocytes across chromatin accessibility, gene expression and protein production data. Dolcinos study did not identify up-regulation of *CCL2* expression in PsA synovial membranes when compared to controls, which could be due to these differences being masked by the mix of cell populations in this tissue. Interestingly, increased levels of MCP-1 in SF, similarly to the observation made by our collaborators in Basle, were previously reported, and correlation with the infiltrated levels of T cells was also demonstrated (Ross et al. 2000). Regarding the open DAR in SF proximal to *CCL2*, no eQTL or chromatin conformation data in CD14 $^{+}$ monocytes has revealed direct evidence for a relationship between differential expression of *CCL2* and changes in chromatin accessibility in this nearby region.

5.4.6 Challenges and future perspectives in multi-omic approaches

The work presented in this chapter is an exploratory study and a proof of principle for the implementation of a multi-omics approach, which represents a very powerful strategy to dissect disease pathophysiology in a cell type specific manner. Nevertheless, a number of limitations and challenges were

Cross-tissue comparison analysis in PsA

encountered and need to be taken into account to contextualise these results. One limitation is the small sample size ($n=3$) and the lack of paired data across all the techniques presented. This results from difficulties of recruiting PsA patients naïve for any treatment, the logistical difficulties to coordinate all of the techniques from the same sample, and the high cost of this approach. A further limitation in this study is the lack of PB from healthy controls or SF from another autoimmune or non-inflammatory joint disease, as included by other studies (Fumitaka2018; Dolcino et al. 2015; Zhang et al. 2018). The definition and categorisation of the qPCR significantly modulated genes into systemic, tissue-specific and putative disease-specific was particularly limited by a lack of control samples to compare to the SF cells and the use of a biased transcriptomic analysis using a qPCR array. Another challenge in this study relates to the analysis and integration of scRNA-seq and mass cytometry data. Both techniques still represent emerging fields were no consensus has been reached on the best strategy to combine samples across patients and experiments, accounting for batch effect. In this exploratory study, monocytes were identified from each SFMCs and PBMCs scRNA-seq sample and combined using CCA for further subpopulation identification. However standard resolution for cluster identification yielded potentially spurious subpopulations, leading to the adoption of a more conservative approach to define clusters in this particular analysis. This may be the consequence of remaining batch effects, and alternative methods of combining samples from the different experiments should be investigated. In this respect, 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). In addition to this, incorporation of bulk RNA-seq data from CD14⁺ monocytes will help interpretation and validation of the scRNA-seq results. In

Cross-tissue comparison analysis in PsA

mass cytometry, to reduce batch effects patient samples are undergoing *ex-vivo* fixation and cryopreservation followed by simultaneous staining and barcoding. Moreover, different methodologies for cluster identification and annotation are also being explored and so far no clear clusters have confidently been found.

In terms of data integration, this pilot study used relatively simplistic multi-omics data integration limited by sample size, technical aspects and time scale, and provides a platform for future validation studies. A more systematic integrative approach should be implemented for the expanded cohort to establish appropriate correlation across datasets. 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. Additionally, the correlation between bulk ATAC and scRNA-seq is clearly limited by the different scales of the two approaches. Therefore, generation of scATAC-seq data, identification of clusters based on chromatin profiling and appropriate methods for the overlap with scRNA-seq populations should be used to have a better understanding of the correlation between chromatin accessibility and gene expression at the single-cell level (Duren et al. 2018).

5.4.7 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 to be 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. 2014). Although DARs between SF and PB in four cell types did not show any overlap

Cross-tissue comparison analysis in PsA

with SNPs from the credible set, significant enrichment of fine-mapped SNPs for accessible chromatin was demonstrated in all four cell types. A number of SNPs from the 5q31 credible set in my analysis overlap accessible chromatin and eQTL signals in the same cell type. Integration of tCD4⁺ and tCD8⁺ eQTL from Kasela *et al.* confirmed the association between SNPs from the 5q31 GWAS signal and T cells *SLC22A5* expression previously reported in a smaller study by Bowes *et al.*, and also suggested a potential role for the 5q31 PsA-specific GWAS association in regulating *P4HA2* and *SLC22A5* expression in unstimulated and stimulated monocytes. *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. Supporting evidence highlights the implications of this gene in other inflammatory conditions such as CD and the potential as a druggable therapeutic target in PsA (Leung *et al.* 2006).

Contrary to the initial hypothesis, the integration of fine-mapping data and DARs between SF and PB in PsA relevant cell types failed to show overlap in any of the four studied cell types. For non-coding signals hypothesised to have a regulatory role, these results may suggest that fine-mapped SNPs from PsA GWAS loci do not have a tissue specific effect in chromatin accessibility changes for any of these cell types. These results may also be biased by the small size in my differential chromatin accessibility analysis and the limited power of the fine-mapping analysis using only a subset of the PsA GWAS cohort. Studying variation in chromatin accessibility upon genotype of the fine-mapped SNPs, similar to the example of chr2p15 presented in Chapter ??, may be more informative when integrating epigenetics at a GWAS associated locus.

5.4.8 Conclusions

The strategy and analysis presented in this chapter is a proof of principle for conducting a multi-omics approach on clinical samples in a cell type and tissue-specific manner. The study of chromatin accessibility in immune cells from SF and PB of PsA patients has demonstrated differences across the two tissues and shown specific enrichment for relevant pathophysiological functional pathways. Transcriptional analysis conducted on the same samples for a number of genes involved in the immune response has also revealed differential expression between the two tissues in all the cell type and shown some of those genes to be proximal to regions presenting changes in chromatin accessibility in the same direction. In this study, both data types highlighted CD14⁺ monocytes as the cell type presenting the largest number of significant changes in chromatin accessibility and consistent modulation of gene expression between SF and PB, with enrichment for pathways leading to inflammation through NF-κB activation and subsequent cytokine and chemokine production. Implementation of scRNA-seq has shown the ability of this approach to identify sub-populations within SF and PB CD14⁺ monocytes. Lastly, relatively basic integration of mass cytometry data has confirmed increased ability of SF CD14⁺ monocytes to produce a number of cytokines and chemokines and supported differences in chromatin accessibility and gene expression between the two tissues at the protein level. Overall, this chapter has shown that in PsA the pro-inflammatory environment at the site of inflammation drive changes in chromatin accessibility, gene expression and protein production that distinguish circulating cells from those infiltrated homing the involved tissue.

Appendices

A Tables	294
A.1 Chapter 3 Tables	294
A.2 Chapter 4 Tables	294
A.3 Chapter 5 Tables	299
B Additional figures	302
B.1 Chapter 3 Figures	302
B.2 Chapter 4 Figures	305
B.3 Chapter 5 Figures	308
Bibliography	345

List of Tables

A.1 Enrichment of ATAC-seq reads across the TSS for the CD14 ⁺ monocytes and CD4 ⁺ samples fresh, frozen and fixed.	294
A.2 Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChiPm assay.	295
A.3 Additional enriched pathways for DEGs between psoriasis and healthy controls in CD14 ⁺ monocytes and CD8 ⁺ cells.	296
A.4 Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	297
A.5 Loci from the psoriasis GWAS Immunochip presenting log ₁₀ ABF< 3 for the fine-mapping lead SNP in the association analysis.	298
A.6 Additional enriched pathways for the DEGs between SF and PB CD14 ⁺ monocytes from the CC-mixed and CC-IL7R subpopulations.	299
A.7 PsA GWAS Immunochip loci presenting log ₁₀ ABF< 3 for the fine-mapping lead SNP in the association analysis.	300

List of Figures

B.1	Distribution of the background read counts from all the master list peaks absent in each sample.	302
B.2	Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples.	303
B.3	Fast-ATAC and Omni-ATAC NHEK tapestation profiles.	304
B.4	Percentage of MT reads in the ATAC-seq libraries generated in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ isolated from psoriasis patients and healthy controls.	305
B.5	Genomic annotation of the consensus master list of ATAC-seq enriched sites built for downstream differential chromatin accessibility analysis in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺	306
B.6	PCA analysis illustrating batch effect in ATAC and RNA-seq samples.	307
B.7	Permutation analysis SF vs PB in CD14 ⁺ monocytes, CD4m ⁺ , CD8m ⁺ and NK.	308
B.8	Heatmap for the top 20 marker genes of the CC-mixed and CC-IL7R CD14 ⁺ monocytes subpopulations.	309
B.9	Identification of the CD14 ⁺ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes.	310
B.10	Quantification of cytokine levels in SF from ten PsA patients.	311

Appendix A

Tables

A.1 Chapter 3 Tables

Cell type	Condition	TSS enrichment		
		CTL1	CTL2	CTL3
CD14	Fresh	17.4	19.6	14.11
	Frozen	26.3	25.2	27.1
	Fixed	2.5	16.5	22.4
CD4	Fresh	5.3	5.6	7.7
	Frozen	17.9	14.1	16.1
	Fixed	7.9	23.0	14.3

Table A.1: Enrichment of ATAC-seq reads across the TSS for the CD14⁺ monocytes and CD4⁺ samples fresh, frozen and fixed.

A.2 Chapter 4 Tables

Tables

Sample ID	NRF	PBC1/PBC2
PS2000 CD14	77.6	0.60/2.5
PS2001 CD14	84.9	0.70/3.0
PS2314 CD14	81.1	0.60/1.8
PS2319 CD14	79.9	0.60/2.2
CTL7 CD14	81.1	0.65/2.2
CTL8 CD14	83.9	0.66/2.3
CTL9 CD14	80.7	0.60/2.3
CTL10 CD14	83.1	0.65/2.1
PS2000 CD4	84.8	0.75/3.4
PS2001 CD4	82.0	0.72/2.9
PS2314 CD4	82.9	0.71/2.8
PS2319 CD4	82.4	0.73/3.2
CTL7 CD4	78.6	0.68/2.5
CTL8 CD4	81.8	0.71/2.9
CTL9 CD4	81.6	0.74/3.3
CTL10 CD4	77.6	0.61/1.9
PS2000 CD8	77.0	0.76/4.5
PS2001 CD8	74.7	0.74/4.0
PS2314 CD8	74.2	0.75/4.1
PS2319 CD8	72.2	0.75/4.0
CTL7 CD8	32.7	0.32/1.5
CTL8 CD8	70.1	0.70/3.3
CTL9 CD8	73.9	0.73/3.7
CTL10 CD8	68.2	0.65/2.9
PS2000 CD19	38.0	0.42/1.9
PS2001 CD19	71.4	0.71/3.7
PS2314 CD19	29.4	0.34/1.8
PS2319 CD19	76.1	0.78/4.8
CTL7 CD19	74.2	0.69/3.1
CTL8 CD19	68.4	0.67/3.2
CTL9 CD19	75.1	0.76/4.6
CTL10 CD19	61.7	0.59/2.6

Table A.2: 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.

Tables

CD14⁺ monocytes additional enriched pathways in psoriasis
Generic transcription
RNA transport
GnRH signalling
Ribosome biogenesis in eukaryotes
Neurotrophin signaling
Spliceosome
Autophagy
Protein processing in endoplasmic reticulum

CD8⁺ additional enriched pathways in psoriasis
Epstein-Barr virus infection
RNA Polymerase I and III, and mitochondrial transcription
Apoptosis

Table A.3: Additional enriched pathways DEGs between psoriasis and healthy controls in CD14⁺ monocytes and CD8⁺ cells. Significant pathways for FDR<0.01. All the enriched pathways contained a minimum of ten DEGs FDR<0.05 from the analysis.

Tables

Lesional versus uninvolved epidermis additional enriched pathways	
Genes encoding extracellular matrix and extracellular matrix-associated proteins	
Serine/threonine-protein kinase (PLK1) signalling	
Genes encoding secreted soluble factors	
Glycolysis/gluconeogenesis	
FOXM1 transcription factor network	
Phase 1 functionalization of compounds	
Biological oxidations	
G2/M Checkpoints	
Biological oxidations	
Aurora B signaling	
Chemical carcinogenesis	
Serotonergic synapse	
Drug metabolism-cytochrome P450	
Mitotic M-M/G1 phases	
DNA Replication	
MicroRNAs in cancer	
Metabolism of amino acids and derivatives	
Metabolism of carbohydrates	
Glycosaminoglycan metabolism	
E2F transcription factor network	
p73 transcription factor network	
Genes encoding structural ECM glycoproteins	
Transmembrane transport of small molecules	
Fc-epsilon receptor I signaling in mast cells	
Tight junction	
Origin recognition complex subunit 1 (Orc1) removal from chromatin	

Table A.4: 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

Table A.5: Loci from the psoriasis GWAS Immunochip presenting $\log_{10}\text{ABF} < 3$ for the fine-mapping lead SNP in the association analysis.
 For each of the locus the closer gene, FM lead SNP, $\log_{10}\text{ABF}$, Tsoi *et al.*, 2012 GWAS lead SNP, the OR in the GAPC cohort and the number of SNPs in the 90% credible set are reported. FM=fine-mapping; ABF=approximated Bayesian factor; OR=odd ratio.

chr	Closer gene	FM lead SNP	$\log_{10}\text{ABF}$	FM lead SNP	GWAS lead SNP	GAPC OR	90% credible set
10	<i>ZMIZ1</i>	rs1316431	2.4		rs1250546	1.09	401
11	<i>RPS6KA4/PRDX5</i>	rs58779949	0.3		rs645078	1.06	334
20	<i>RNF114</i>	rs13041638	3.1		rs1056198	1.11	116
6	<i>EXOC2/IRF4</i>	rs113866081	2.4		rs9504361	1.14	400
6	<i>TAGAP</i>	rs62431928	2.2		rs2451258	1.11	853
9	<i>DDX58</i>	rs7045087	0.4		rs11795343	1.05	167
9	<i>KLF4</i>	rs6477612	2.1		rs10979182	1.12	80
11	<i>ETS1</i>	rs10893884	3.5		rs3802826	1.15	19
18	<i>POL1/STARD6</i>	rs11661229	1.6		rs545979	1.11	121

Tables

A.3 Chapter 5 Tables

CC-mixed CD14 ⁺ monocytes additional enriched pathways
SLE
Translation
3'-UTR-mediated translational regulation
Th-1 and Th-2 cell differentiation
Peptide chain elongation
Rheumatoid arthritis
Metabolism of proteins
Cell adhesion molecules (CAMs)
Th-17 cell differentiation
Nonsense mediated decay enhanced by the exon junction complex
SRP-dependent co-translational protein targeting to membrane
Hemostasis
Metabolism of mRNA
Platelet activation, signalling and aggregation
HTLV-I infection
Innate immune system
Adaptive immune system
CC-IL7R CD14 ⁺ monocytes additional enriched pathways
SLE
Tuberculosis
Epstein-Barr virus infection
Immune System

Table A.6: Additional enriched pathways for the DEGs between SF and PB CD14⁺ monocytes from the CC-mixed and CC-IL7R subpopulations. All the enriched pathways contained a minimum of ten DEGs from the analysis and were significant at an FDR<0.01.

Tables

Table A.7: 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 pval 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 (pval)	OR	lead SNP	Bowes FM	Bowes 99% credible set
2	<i>B3GNT2/TMEM17</i>	2:62501912(INS)	1.8	rs6713082 (4.59x10 ⁻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 (3.36x10 ⁻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 (4.38x10 ⁻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 (1.94x10 ⁻⁷)		1.2	rs4795067	2	

Tables

2	<i>PAPOLG/REL</i>	rs60685986	2.0	rs1306395 (2.99×10^{-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 (9.65×10^{-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 (4.98×10^{-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 (2.90×10^{-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

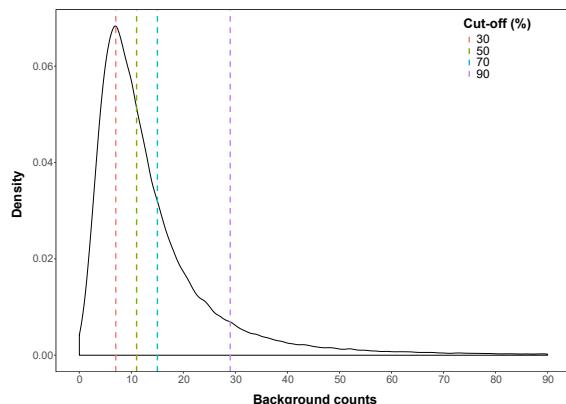


Figure B.1: Distribution of the background read counts from all the master list peaks absent peaks in each sample. Each cut-off corresponds to the number of background counts showed by a particular percentage of the total number of absent peaks.

Additional figures

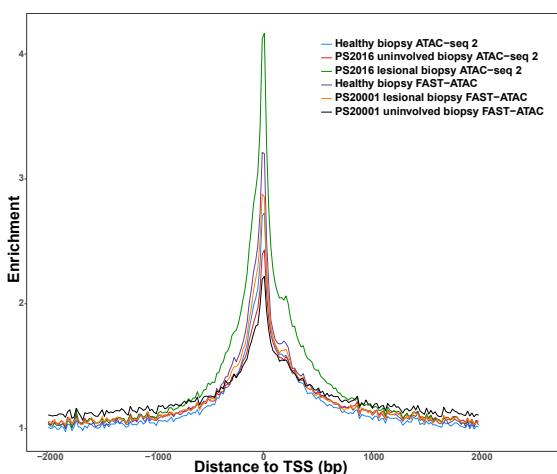


Figure B.2: Assessment of TSS enrichment from ATAC 1 and Fast-ATAC in healthy and psoriasis KCs isolated from skin biopsy samples. Fold-enrichment of ATAC fragments across the Ensembl annotated TSS from the different ATAC libraries.

Additional figures

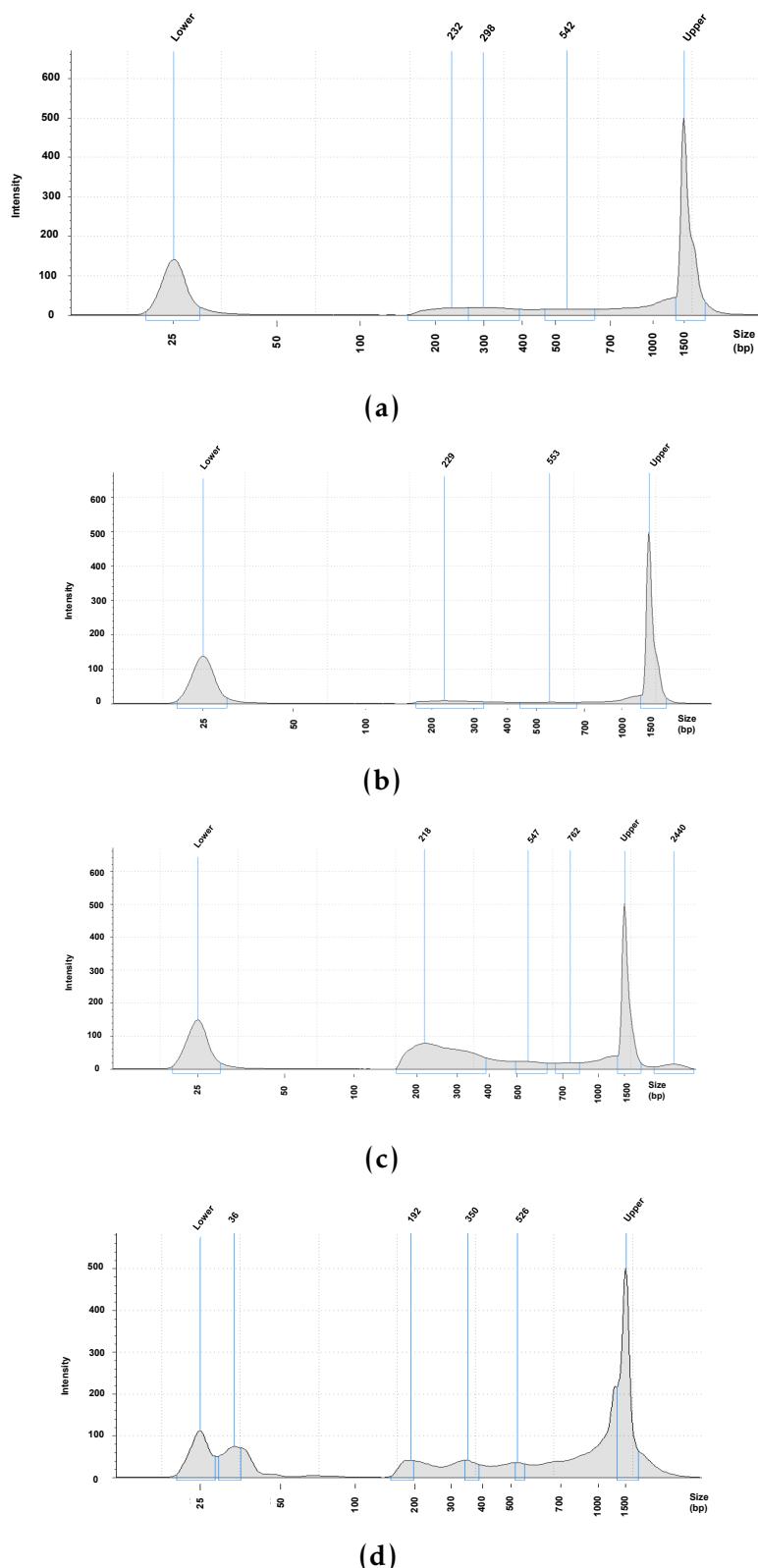


Figure B.3: Fast-ATAC and Omni-ATAC NHEK tapestation profiles. Pre-sequencing quantification of DNA fragment sizes from the libraries generated using the a) C2, b) C3, and c) C4 versions of the Fast-ATAC protocol based on modifications in the detergent and Tn5 concentration and d) Omni-ATAC. C2, C3 and C4 detergent and Tn5 concentrations are detailed in Table 3.3.

Additional figures

B.2 Chapter 4 Figures

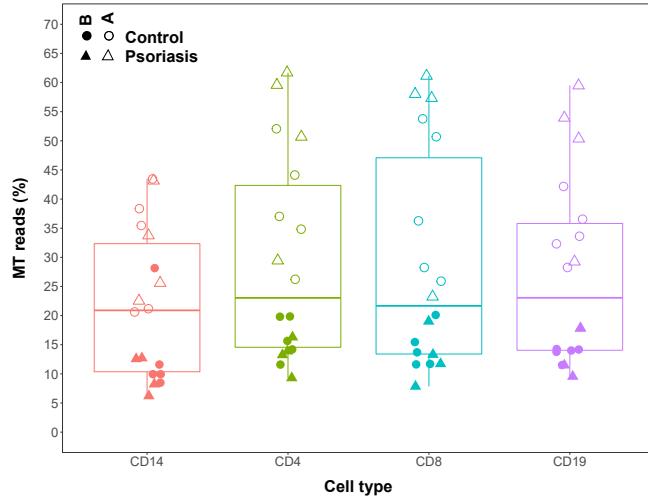


Figure B.4: Percentage of MT reads in the ATAC-seq samples generated in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ isolated from psoriasis patients and healthy controls. Samples from cohort 1A (open circles and triangles) were generated with the standard ATAC-seq protocol from Buenrostro *et al.*, 2013 whereas samples from cohort 1B (filled circles and triangles) were processed using FAST-ATAC (Corces *et al.* 2016).

Additional figures

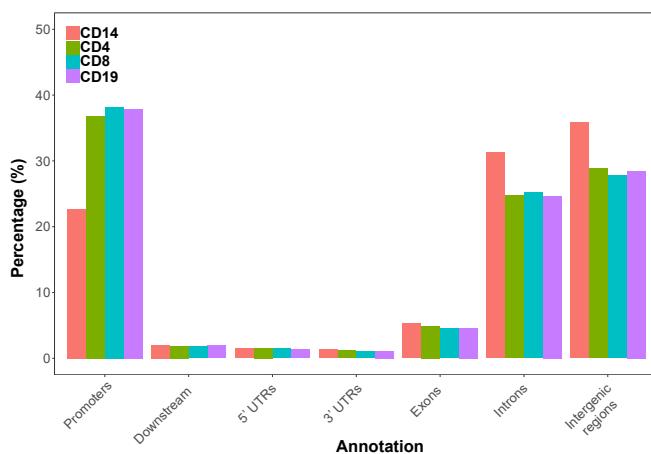


Figure B.5: Genomic annotation of the consensus master list of ATAC-seq enriched sites built for downstream differential chromatin accessibility analysis in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺. Annotation is expressed in percentage over the total number of ATAC-seq sites included in each particular cell type master list.

Additional figures

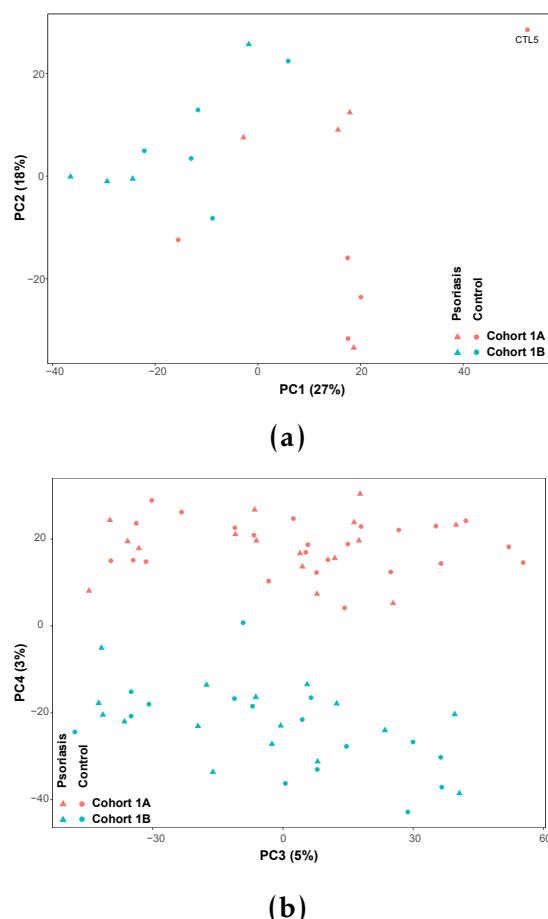


Figure B.6: PCA analysis illustrating batch effect in ATAC and RNA-seq samples. a) First and second component of the PCA analysis performed using the normalised ATAC counts in a master list of consensus regions across all the combined tCD8⁺ samples from psoriasis patients and healthy controls. b) Third 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 tCD8⁺ cells from psoriasis patients and healthy controls. For each dot colour corresponds to cohort ID (batch) and shape to condition (psoriasis or control).

Additional figures

B.3 Chapter 5 Figures

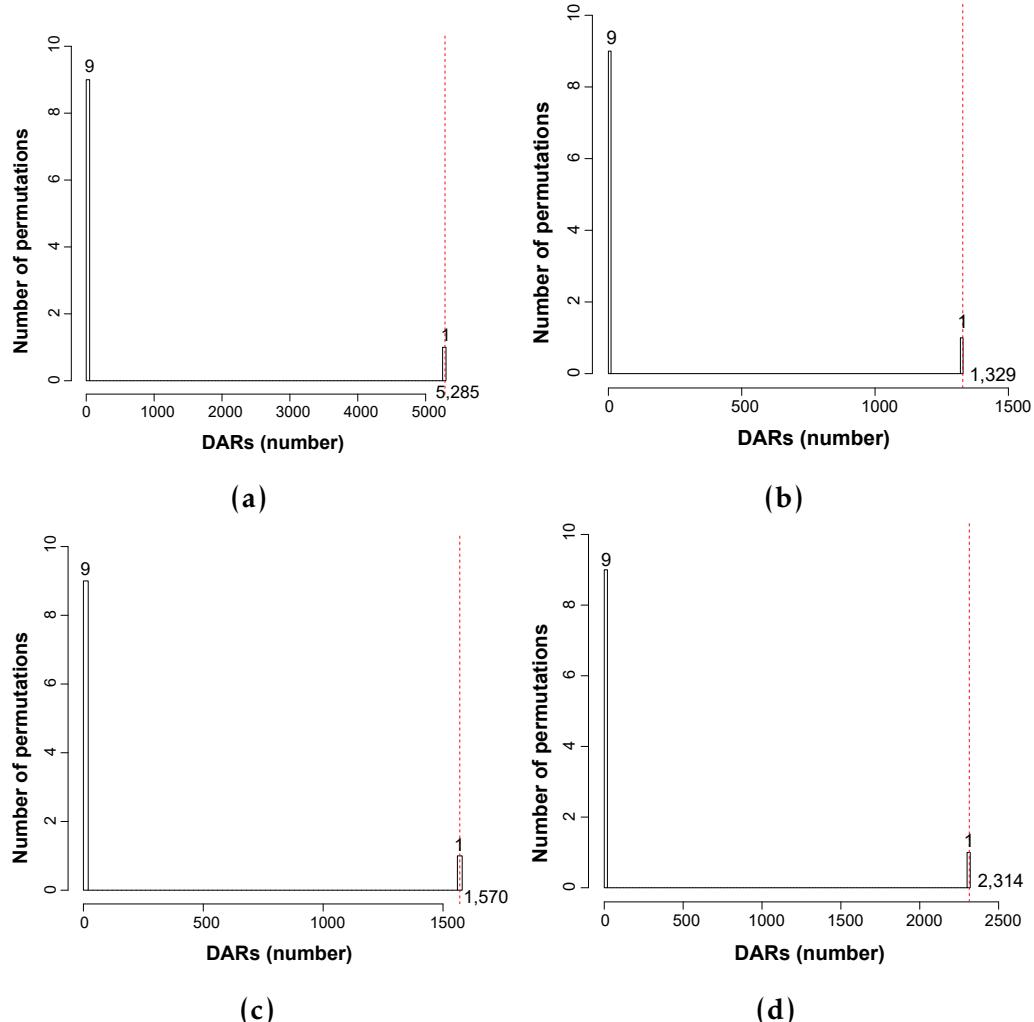


Figure B.7: Permutation analysis SF vs PB in CD14⁺ monocytes, CD4m⁺, CD8m⁺ and NK. Sample labels were permuted within each cell type to achieve the ten unique possible combinations and differential analysis was performed. The number of significant DARs (FDR<0.01 and no abs(FC)>1.5) across all permutations is plotted for a) CD14⁺ monocytes, b) CD4m⁺, c) CD8m⁺ and d) NK, demonstrating that the true observation (dashed red line) is significantly more than expected by chance ($pval<0.1$, the lowest $pval$ for the maximum number of permutations that can be conducted with this sample size) in all four cell types.

Additional figures

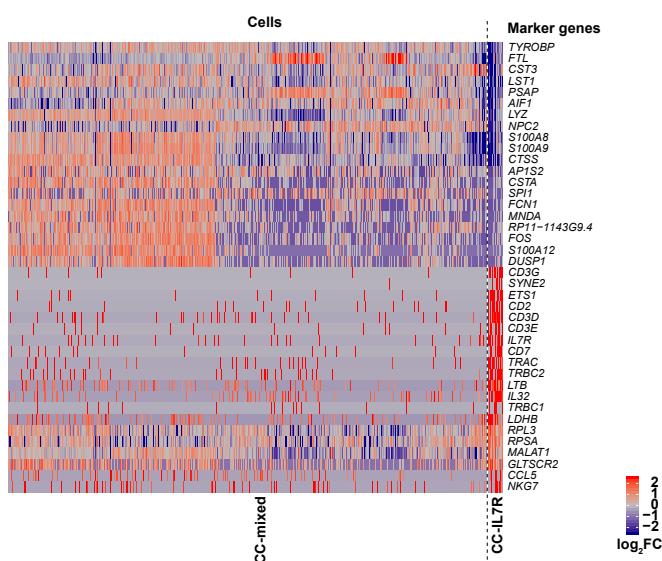


Figure B.8: Heatmap for the top 20 marker genes of the CC-mixed and CC-IL7R CD14⁺ monocytes subpopulations. Rows are the top 20 marker genes for each of the two subpopulations (total of 40 genes). The columns represent each of the cells members of the CC-mixed (left) or CC-IL7R (right) clusters. The colour scale represents the log₂FC in the expression of the marker gene in a particular cell of the cluster compared to the average expression of all the cells from the other cluster.

Additional figures

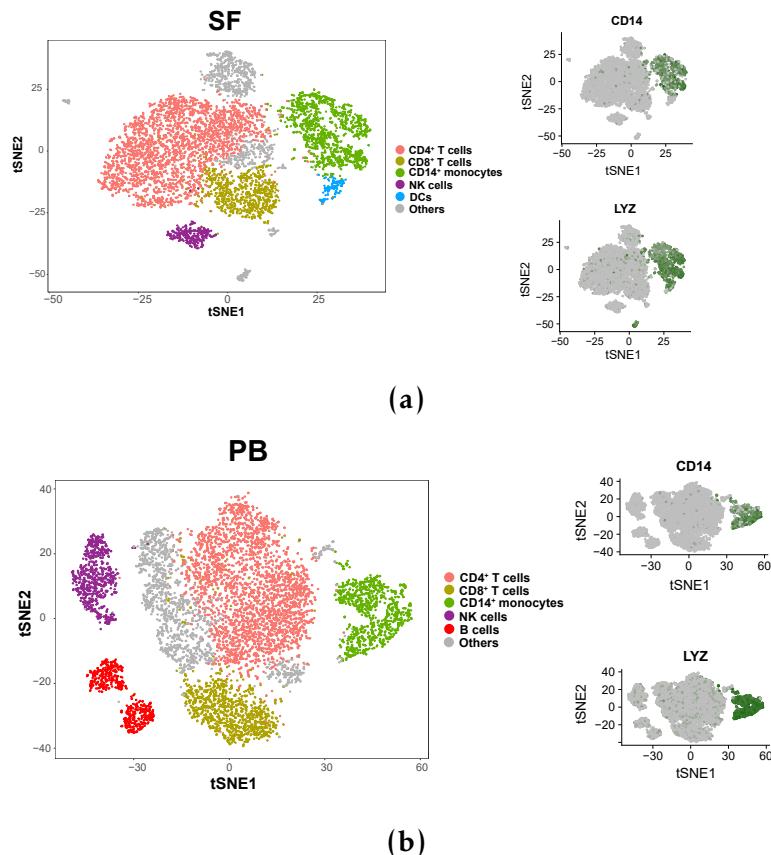


Figure B.9: Identification of the CD14⁺ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. Visualisation using t-SNE dimensional reduction of the cell subpopulations identified in a) SFMCs and b) PBMCs and the overlay of CD14⁺ monocytes characteristic markers (left hand side panel of a and b) for a representative PsA sample. Clustering performed using recommended resolution (res=0.6) allowed to identify CD4⁺ (pink), CD8⁺ (khaki), CD14⁺ monocytes (green), NK (purple), DCs (blue), B cells (red) and others (grey). On the left hand side panel, expression for two characteristics CD14⁺ monocytes markers (CD14 and LYZ) used to subset this population (dark green dots) is overlaid on the t-SNE visual representation of all the cells in each of the tissues.

Additional figures

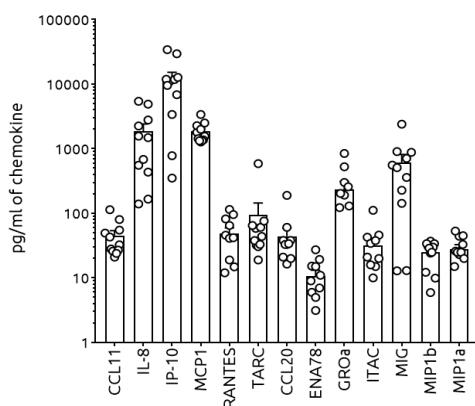


Figure B.10: Quantification of cytokine levels in SF from ten PsA patients. Barplot graph illustrating pg/mL (x-axis) for a number of cytokines measured from SF of ten PsA patients by collaborators at University Basle using enzyme-linked immunosorbent assay (ELISA). Each circle represents a patients and error bars represent the standard deviation (SD) of the mean from all patients combined. Measurement for the same cytokines was performed in matched plasma from the same patients and all of them failed to be detected.

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.
- Abraham, Gad and Inouye, Michael (2014). "Fast principal component analysis of large-scale genome-wide data". *PloS one* 9.
- Abramson, S. R. and and, Khattri-S. of (2016). "A Review of Current Evidence for TNF Inhibitor Switching in Psoriatic Arthritis". *Journal of Psoriasis and*.
- 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*.
- 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*.
- 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). "HTSeq Python framework to work with high-throughput sequencing data". *Bioinformatics*.
- 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.

BIBLIOGRAPHY

- 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*.
- 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”. *Arthritis*.
- Arend, W. P., Palmer, G., and reviews, Gabay-C. (2008). “IL1, IL18, and IL33 families of cytokines”. *Immunological reviews*.
- 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*.
- Austin, Lisa M., Ozawa, Maki, Kikuchi, Toyoko, Walters, Ian B., and Krueger, James G. (1999). “The majority of epidermal T cells in psoriasis vulgaris lesions can produce type 1 cytokines, interferon-, interleukin-2, and tumor necrosis factor-, defining TC1 (Cytotoxic T Lymphocyte) and TH1 effector populations: 1 a type 1 differentiation bias is also measured in circulating blood T cells in psoriatic patients”. *Journal of Investigative Dermatology* 113: pp. 752–759.
- B, Uszczynska-Ratajczak, Lagarde, J, and resource, Frankish A (2018). “Towards a complete map of the human long non-coding RNA transcriptome”. *resource*.
- 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 &*.
- 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 research, Kouzarides T (2011). “Regulation of chromatin by histone modifications”. *Cell research* 21: pp. 381–395.

BIBLIOGRAPHY

- 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". *Cell*. 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 &*.
- Bell, A. C. and Nature, Felsenfeld-G. (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene". *Nature*.
- Bengsch, 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*.
- 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. and Nature, Stamatoyannopoulos-J. A. (2010). "The NIH roadmap epigenomics mapping consortium". *Nature*.
- 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 of, ArdernJones-M. R. journal (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*.

BIBLIOGRAPHY

- Blonska, M. and research, Lin-X. (2011). "NF-B signaling pathways regulated by CARMA family of scaffold proteins". *Cell research*.
- 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.
- 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.
- Braun, J. E., Huntzinger, E., Fauser, 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., 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 Dermo, 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*.
- Butler, Andrew, Hoffman, Paul, Smibert, Peter, Papalexi, Efthymia, and Satija, Rahul (2018). "Integrating single-cell transcriptomic data across different conditions, technologies, and species". *Nature biotechnology* 36: p. 411.
- Cai, Yihua, Fleming, Chris, and Yan, Jun (2012). "New insights of T cells in the pathogenesis of psoriasis". 9: p. 302.
- Cameron, A. L. and of, 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*.
- Cao, J, Cusanovich, DA, and, Ramani V (2018). "Joint profiling of chromatin accessibility and gene expression in thousands of single cells" ..
- Capon, Francesca (2017). "The Genetic Basis of Psoriasis". *International Journal of Molecular Sciences* 18: p. 2526.
- 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" ..
- Chang, X, Yamada, R, Suzuki, A, and, Kochi Y (2005). "Citrullination of fibronectin in rheumatoid arthritis synovial tissue" ..
- Chen, Xingqi, Shen, Ying, Draper, Will, Buenrostro, Jason D., Litzenburger, Ulrike, et al. (2016). "ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing". *Nature methods* 13: p. 1013.

BIBLIOGRAPHY

- 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 &*
- Churov, A. V., Oleinik, E. K., and reviews, Knip-M. (2015). “MicroRNAs in rheumatoid arthritis: altered expression and diagnostic potential”. *Autoimmunity reviews*.
- 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*.
- Clark, R. A., Chong, B., Mirchandani, N., Brinster, N. K., Yamanaka, K., et al. (2006). “The vast majority of CLA+ T cells are resident in normal skin”. *J Immunol* 176: pp. 4431–9.
- 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 (2016). “Psoriasis, psoriatic arthritis, and rheumatoid arthritis: Is all inflammation the same?” *Seminars in arthritis and*.
- 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*.
- Cong, Le, Ran, F. A., Cox, David, Lin, Shuailiang, Barretto, Robert, et al. (2013). “Multiplex genome engineering using CRISPR/Cas systems”. *Science (New York, N.Y.)* 339: pp. 819–823.
- 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, Beijing, 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, Adrian and Brown, Matthew A. (2011). “Promise and pitfalls of the Immunochip”. *Arthritis research & therapy* 13: p. 101.

BIBLIOGRAPHY

- 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 cycle and developmentally regulated promoter". *Cell*.
- Costello, P., Bresnihan, B., and O'Farrelly-C. (1999). "Predominance of CD8+ T lymphocytes in psoriatic arthritis". *The Journal of*.
- Cretu, Daniela, Prassas, Ioannis, Saraon, Punit, Batruch, Ihor, Gandhi, Rajiv, et al. (2014). "Identification of psoriatic arthritis mediators in synovial fluid by quantitative mass spectrometry". 11: p. 27.
- Creyghton, M. P. and 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 Wu-R. in Biology (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.
- 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*.
- 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*.
- 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.
- Degner, Jacob F., Pai, Athma A., Pique-Regi, Roger, Veyrieras, Jean-Baptiste, Gaffney, Daniel J., et al. (2012). "DNase I sensitivity QTLs are a major determinant of human expression variation". *Nature* 482: p. 390.
- 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.
- Diani, M., Altomare, G., and reviews, Reali-E. (2015). "T cell responses in psoriasis and psoriatic arthritis". *Autoimmunity reviews*.

BIBLIOGRAPHY

- 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.
- Doбин, 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, Ottria, 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 and Arthritis, Collins ES (2012). "New insight into the functions of the interleukin-17 receptor adaptor protein Act1 in psoriatic arthritis". *Arthritis*.
- Dozmanov, 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*.
- 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.
- Durham, LE, Kirkham, BW, and Taams, LS (2015). "Contribution of the IL-17 Pathway to Psoriasis and Psoriatic Arthritis." *Curr Rheumatol Rep* 17: p. 55.
- 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*.

BIBLIOGRAPHY

- 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.
- Ele-Refaei, A. M., research, and practice, El-Esawy-F. M. (2015). "Effect of narrow-band ultraviolet B phototherapy and methotrexate on MicroRNA (146a) levels in blood of psoriatic patients". *Dermatology research and practice*.
- 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.
- Ermann, Joerg and Glimcher, Laurie H. (2012). "After GWAS: mice to the rescue?" *Current opinion in immunology* 24: pp. 564–570.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., and Shoresh, 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.
- 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*.
- Eyerich, S, Eyerich, K, and, Pennino D of (2009). "Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling". *The Journal of*.
- Fabregat, Antonio, Jupe, Steven, Matthews, Lisa, 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*.

BIBLIOGRAPHY

- 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.
- Fantom, D. G. T., Forrest, Alistair R. R., Kawaji, Hideya, Rehli, Michael, Baillie, Kenneth J., et al. (2014). "A promoter-level mammalian expression atlas". *Nature* 507.
- Farh, Kyle, Carrasco-Alfonso, J., Dita, Mayer, Luckey, C., Nikolaos, A., et al. (2014). "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*.
- 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.
- 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.
- Fredriksson, T. and Dermatology, Pettersson-U. (1978). "Severe psoriasisoral 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*.
- Frucht, David M, Aringer, Martin, Galon, Jérôme, Danning, Carol, Brown, Martin, et al. (2000). "Stat4 is expressed in activated peripheral blood monocytes, dendritic cells, and macrophages at sites of Th1-mediated inflammation". 164: pp. 4659–4664.
- 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)*.

BIBLIOGRAPHY

- 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*.
- 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*.
- Gelfand, Joel M., Troxel, Andrea B., Lewis, James D., Kurd, Shanu, Shin, Daniel B., et al. (2007). "The Risk of Mortality in Patients With Psoriasis: Results From a Population-Based Study". *Archives of Dermatology* 143: pp. 1493–1499.
- 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.
- 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*.
- 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" ..
- 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*.

BIBLIOGRAPHY

- 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.
- 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*.
- Gregory, Adam P., Dendrou, Calliope A., Attfield, Kathrine 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.
- Griffiths, C. E. M. and Lancet, Jnwn (2007). "Pathogenesis and clinical features of psoriasis". *The Lancet*.
- 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.
- Gudjonsson, J. E. and of, Karason-A. journal (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.
- 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.
- Haddad, Amir and Chandran, Vinod (2013). "Arthritis Mutilans". *Current Rheumatology Reports* 15.

BIBLIOGRAPHY

- 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*.
- Han, EJ, Kim, HY, Lee, N, Kim, NH, Yoo, SA, et al. (2017). "Suppression of NFAT5-mediated inflammation and chronic arthritis by novel B-binding inhibitors". *EBioMedicine*.
- 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, 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 Medicine, Baylin-S. B. of (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.

BIBLIOGRAPHY

- 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.
- 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*.
- 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 &*.
- 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

- Islam, S., Zeisel, A., Joost, S., Manno, La G., and Nature, Zajac P (2014). “Quantitative single-cell RNA-seq with unique molecular identifiers”. *Nature*.
- 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.
- Jackson, B., Tilli, Cmlj, Hardman, M. J., and Investigative, Avilion-A. A. of (2005). “Late cornified envelope family in differentiating epithelial response to calcium and ultraviolet irradiation”. *Journal of Investigative*.
- 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*.
- Jazdzewski, Krystian, Liyanarachchi, Sandya, Swierniak, Michal, Pachucki, Janusz, Ringel, Matthew D, et al. (2009). “Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer”. 106: pp. 1502–1505.
- Jenuwein, T. and Science, Allis-C. D. (2001). “Translating the histone code”. *Science*.
- Ji, Jong, Cheon, Hyeonjoo, 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*.
- 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, L. M., McNutt, N. S., and clinical, Krueger-J. G. of (2009). “Cytokine-producing dendritic cells in the pathogenesis of inflammatory skin diseases”. *Journal of clinical*.

BIBLIOGRAPHY

- 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*.
- Jordan, C. T., Cao, L., Roberson, E. D. O., and of, Duan-S. (2012a). "Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis". *The American Journal of*.
- Jordan, CT, Cao, L, Roberson, EDO, and, Duan S of (2012b). "Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis". *The American Journal of*.
- 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, CM and, SandrinGarcia - P (2009). "Differential gene expression of peripheral blood mononuclear cells from rheumatoid arthritis patients may discriminate immunogenetic, pathogenic and treatment ..".
- 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.

BIBLIOGRAPHY

- Kent, L. (2009). "Freezing and thawing human embryonic stem cells". *Journal of visualized experiments: JoVE*.
- Kent, W. J., Zweig, A. S., Barber, G., and, Hinrichs-A. S. (2010). "BigWig and BigBed: enabling browsing of large distributed datasets" ..
- Kheradpour, Pouya, Ernst, Jason, Melnikov, Alexandre, Rogov, Peter, Wang, Li, et al. (2013). "Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay". *Genome research* 23: pp. 800–811.
- Kichaev, G. and Genetics, Pasaniuc-B. of (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.
- Knight, J. C. (2014). "Approaches for establishing the function of regulatory genetic variants involved in disease". *Genome medicine*.
- 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*.
- Lai, A. Y., Mav, D., Shah, R., Grimm, S. A., and Genome, Phadke-D. (2013). "DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B cell activation". *Genome*.
- 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.
- 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, Heekyoung, Qian, Kun, Toerne, Christine von, Hoerburger, Lena, Claussnitzer, Melina, et al. (2017). "Allele-specific quantitative proteomics unravels molecular mechanisms modulated by cis-regulatory PPARG locus variation". *Nucleic acids research* 45: pp. 3266–3279.
- 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*.
- Leijten, Emmerik F. A., Kempen, Tessa S., Boes, Marianne, Amelsfort, Jocea M. R., Hijnen, Dirkjan, et al. (2015). "Brief report: enrichment of activated group 3 innate lymphoid cells in psoriatic arthritis synovial fluid". *Arthritis & Rheumatology* 67: pp. 2673–2678.
- 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, 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.

BIBLIOGRAPHY

- 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, Suling, Miao, Tizong, Sebastian, Meera, Bhullar, Punam dip, Ghaffari, Emma, et al. (2012). "The transcription factors Egr2 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells". 37: pp. 685–696.
- Li, Y. and research, Kellis-M. acids (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*.
- 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., Beyer, A., and Cell, Aebersold R (2016). "On the dependency of cellular protein levels on mRNA abundance". *Cell*.
- 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*.
- 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*.

BIBLIOGRAPHY

- 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*.
- Mahil, Satveer K., Capon, Francesca, and Barker, Jonathan N. (2016). "Update on psoriasis immunopathogenesis and targeted immunotherapy". *Seminars in Immunopathology* 38: pp. 11–27.
- Maine, G. N., Mao, X., and journal, Komarck-C. M. (2007). "COMMD1 promotes the ubiquitination of NFB subunits through a cullincontaining ubiquitin ligase". *The EMBO journal*.
- 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.
- 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*.
- Maurano, M. T., Humbert, R., Rynes, E., and, Thurman-R. E. (2012). "Systematic localization of common disease-associated variation in regulatory DNA" ..
- 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.

BIBLIOGRAPHY

- medicine, Kong HH in molecular (2011). "Skin microbiome: genomics-based insights into the diversity and role of skin microbes". *Trends in molecular medicine*.
- Meglio, Paola, Villanova, Federica, and Nestle, Frank O. (2014). "Psoriasis". *Cold Spring Harbor Perspectives in Medicine* 4.
- 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 medical, 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 Gastroenterology*.
- Michel, T., Poli, A., Cuapio, A, and, Briquemont B of (2016). "Human CD56bright NK cells: an update". *The Journal of Immunology*.
- 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.
- Moll, J. M. H., in, arthritis, and rheumatism, Wright-V. (1973). "Psoriatic arthritis". *Seminars in arthritis and rheumatism*.

BIBLIOGRAPHY

- 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.
- 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*.
- Moura, Carlos, Assis, Luiz, Ges, Paulo, Rosa, Fabiana, Nunes, Victor, et al. (2015). “A Case of Acute Generalized Pustular Psoriasis of von Zumbusch Triggered by Hypocalcemia”. *Case reports in dermatology* 7: pp. 345–351.
- Mumbach, Maxwell R., Rubin, Adam J., Flynn, Ryan A., Dai, Chao, Khavari, Paul A., et al. (2016). “HiChIP: efficient and sensitive analysis of protein-directed genome architecture”. *Nature methods* 13: pp. 919–922.
- Myles, Arpita and Aggarwal, Amita (2011). “Expression of Toll-like receptors 2 and 4 is increased in peripheral blood and synovial fluid monocytes of patients with enthesitis-related arthritis subtype of juvenile idiopathic arthritis.” *Rheumatology (Oxford)* 50: pp. 481–8.
- 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.
- 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.
- Nickoloff, B. and Nestle, F. (2004). “Recent insights into the immunopathogenesis of psoriasis provide new therapeutic opportunities”. *J Clin Invest* 113.

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.
- Niimi, Tomoaki, Munakata, Mitsuru, Keck-Waggner, Catherine L., Popescu, Nicholas C., Levitt, Roy C., et al. (2002). "A polymorphism in the human UGRP1 gene promoter that regulates transcription is associated with an increased risk of asthma". *American journal of human genetics* 70: pp. 718–725.
- Noordenbos, T., Yeremenko, N., and, Gofita-I. (2012). "Interleukin17positive mast cells contribute to synovial inflammation in spondylarthritis". *Arthritis &*
- 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.
- 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.
- 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*.
- 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.
- 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*.

BIBLIOGRAPHY

- Pasaje, Charisse A., Bae, Joon, Park, ByUNG-L. A. E., Jang, An-Soo, Uh, Soo-Taek, et al. (2011). "Association analysis of DTD1 gene variations with aspirin-intolerance in asthmatics". *International journal of molecular medicine* 28: pp. 129–137.
- 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*.
- 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.
- Picard, C., Dogniaux, S., and of, Chemin-K. journal (2009). "Hypomorphic mutation of ZAP70 in human results in a late onset immunodeficiency and no autoimmunity". *European journal of*.
- 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, Bagnyukova-T. V. (2009). "Hepatic epigenetic phenotype predetermines individual susceptibility to hepatic steatosis in mice fed a lipogenic methyl-deficient diet". *Journal of*.
- 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*.
- Polachek, 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*.

BIBLIOGRAPHY

- 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 Bioinformatics, Hall-I. M. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". *Bioinformatics*.
- Raghunatha, R. (2012). "S100 proteins in cartilage: role in arthritis". *Biochimica et Biophysica Acta (BBA)-Molecular Basis*.
- 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.
- 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*.
- 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*.
- 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*.
- 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*.

BIBLIOGRAPHY

- 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.
- 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.
- S., Andrews (2010). "FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>".
- 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.
- 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.
- 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*.
- Schindler, Heike, Lutz, Manfred B, Röllinghoff, Martin, and Bogdan, Christian (2001). "The production of IFN- by IL-12/IL-18-activated macrophages requires STAT4 signaling and is inhibited by IL-4". 166: pp. 3075–3082.
- Schmid, P., Itin, P., and interferon, Cox-D. of (1994). "The type I interferon system is locally activated in psoriatic lesions". *Journal of interferon*.
- Schmitt, J. and Dermatology, Wozel-G. (2005). "The psoriasis area and severity index is the adequate criterion to define severity in chronic plaque-type psoriasis". *Dermatology*.
- Schmitt, J. and of, Rosumeck-S. (2014). "Efficacy and safety of systemic treatments for moderate to severe psoriasis: metaanalysis of randomized controlled trials". *British Journal of*.
- Schonthaler, H. B. and rheumatic, Guinea-Viniegra-J. of the (2011). "Targeting inflammation by modulating the Jun/AP-1 pathway". *Annals of the rheumatic*.

BIBLIOGRAPHY

- 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.
- Schroder, Wayne A., Anraku, Itaru, Le, Thuy T., Hirata, Thiago D. C., Nakaya, Helder I., et al. (2016). "SerpineB2 deficiency results in a stratum corneum defect and increased sensitivity to topically applied inflammatory agents". *The American journal of pathology* 186: pp. 1511–1523.
- Sellars, M. L., Huh, J. R., Day, K., Issuree, P. D., and Nature, Galan-C. (2015). "Regulation of DNA methylation dictates Cd4 expression during the development of helper and cytotoxic T cell lineages". *Nature*.
- 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.
- Sigmundsdottir, H. and, Gudjonsson-J. E. (2001). "The frequency of CLA+CD8+T cells in the blood of psoriasis patients correlates closely with the severity of their disease". *Clinical &*.
- 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.

BIBLIOGRAPHY

- 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.
- Solomon, M. J., Larsen, P. L., and Cell, Varshavsky-A. (1988). "Mapping protein-DNA 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 genetics, Barrett-J. C. molecular (2015). "Strategies for fine-mapping complex traits". *Human molecular genetics*.
- 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.
- 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.

BIBLIOGRAPHY

- Swindell, W. R. and Genome, 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*.
- 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, Siitonen, 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.
- 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*.
- Tiilikainen, A., Lassus, A., Karvonen, J., Vartiainen, P., and Julin, M. (1980). "Psoriasis and HLA Cw6". *British Journal of Dermatology* 102: pp. 179–184.
- 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.
- 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.

BIBLIOGRAPHY

- Tomfohrde, J., Silverman, A., and Barnes-R. (1994). "Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q" ..
- Tonel, Giulia, Conrad, Curdin, Laggner, Ute, Di Meglio, Paola, Grys, Katarzyna, et al. (2010). "Cutting edge: A critical functional role for IL-23 in psoriasis." *J. Immunol.* 185: pp. 5688–91.
- 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.
- 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*.
- Valdimarsson, H. and in, Thorleifsdottir-R. H. (2009). "Psoriasisas an autoimmune disease caused by molecular mimicry". *Trends in*.
- 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*.
- 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*.

BIBLIOGRAPHY

- 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*.
- 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, H., Peng, W., Ouyang, X., Li, W., and Research, Dai-Y. (2012). "Circulating microRNAs as candidate biomarkers in patients with systemic lupus erythematosus". *Translational Research*.
- 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.
- Wang, Y, Jiang, R, and review, Wong WH science (2016). "Modeling the causal regulatory network by integrating chromatin accessibility and transcriptome data". *National science review*.
- Ward, L. D. and biotechnology, Kellis-M. (2012). "Interpreting noncoding genetic variation in complex traits and human disease". *Nature biotechnology*.
- Weiss, G., Shemer, A., and Academy, Trau-H. of the (2002). "The Koebner phenomenon: review of the literature". *Journal of the European Academy*.
- Welter, D., MacArthur, J., Morales, J., and acids, Burdett-T. (2013). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations". *Nucleic acids*.
- 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.

BIBLIOGRAPHY

- 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.*
- WJ, McCormack, and Arthritis, Parker AE (2009). "Toll-like receptors and NOD-like receptors in rheumatic diseases". *Arthritis.*
- 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.*
- Wray, Naomi R. (2005). "Allele frequencies and the r² 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 clinical, Nickoloff-B. J. of (1996). "Dermal injection of immunocytes induces psoriasis". *The Journal of clinical.*
- 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*
- 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.*
- Yamamoto, Mayuko, Nakajima, Kimiko, Takaishi, Mikiro, Kitaba, Shun, Magata, Yasuhiro, et al. (2015). "Psoriatic inflammation facilitates the onset of arthritis in a mouse model". 135: pp. 445–453.
- Yamashita, Riu, Sathira, Nuankanya P., Kanai, Akinori, Tanimoto, Kousuke, Arauchi, Takako, et al. (2011). "Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis". *Genome research* 21: pp. 775–789.
- 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.*
- 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.

BIBLIOGRAPHY

- 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". *Nat Genet* 42: pp. 565–9.
- Yen, J. H., Moore, B. E., Nakajima, T., and of, Scholl-D. (2001). "Major histocompatibility complex class Irecognizing receptors are disease risk genes in rheumatoid arthritis". *Journal of*.
- 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.
- Yosef, N and Regev, A (2016). "Writ large: Genomic dissection of the effect of cellular environment on immune response". 354: pp. 64–68.
- Zgraggen, S., Huggerberger, 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, 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

- Zhou, F., Wang, W., Shen, C., Li, H., Zuo, X., et al. (2016). "Epigenome-wide association analysis identified nine skin DNA methylation loci for psoriasis". *Journal of Investigative Dermatology*.
- 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*.
- Zuk, O., Hechter, E., and the, Sunyaev-S. R. of (2012). "The mystery of missing heritability: Genetic interactions create phantom heritability". *Proceedings of the National Academy of Sciences of the United States of America*.
- 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.