



Functional genomics of psoriasis

Alicia Lledo Lara
Hertford College
University of Oxford

*A thesis submitted in partial
fulfilment of the requirements for the degree of
Doctor of Philosophy
Trinity Term, 2018*

Abstract

Functional genomics of psoriasis

Alicia Lledo Lara, Hertford College, Trinity Term 2018

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	

Associated Publications

Title

Journal

Authors

Other Publications

Title

Journal

Authors

Contents

Abstract	i
Acknowledgements	ii
Declarations	iii
Submitted Abstracts	iv
Associated Publications	v
Contents	vi
List of Figures	ix
List of Tables	x
Abbreviations	xi
1 Cross-tissue comparison of chromatin accessibility, gene expression signature and immunophenotypes in PsA	1
1.1 Introduction	1
1.1.1 The relevance of cell type and tissue specificity in the study of PsA	1
1.1.2 Bulk transcriptomic studies and their limitations	2
1.1.3 Transcriptomics and proteomics at the single cell resolution	4
1.1.4 Using a multi-omics approach in the study of complex diseases	5
1.1.5 Genetic fine-mapping using genotyping level data	7
1.1.6 Aims	7
1.2 Results	7
1.2.1 PsA patients cohort description and datasets	7
1.2.2 Immune cellular composition of blood and synovial fluid in the PsA cohort	9
1.2.3 Differential chromatin accessibility analysis in immune cells reveals differences between SF and PB	10
1.2.4 Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility	19
1.2.5 Differential gene expression analysis in paired circulating and synovial immune cells	21
1.2.6 Characterisation of the CD14 ⁺ monocyte heterogeneity in PsA using scRNA-seq	36
1.2.7 Integration of PsA GWAS and epigenetics	42

CONTENTS

1.2.8 Mass cytometry reveals CD14 ⁺ monocytes as the most active cytokine producing cells	42
1.3 Discussion	44
Appendices	46
A Appendices	48
A.1 Additional tables	48
A.1.1 Chapter 5 Tables	48
A.1.2 Chapter 4 Tables	54
A.1.3 Chapter 5 Tables	54
A.2 Additional figures	54
A.2.1 Chapter 3 Figures	54
A.2.2 Chapter 4 Figures	54
A.2.3 Chapter 5 Figures	56

List of Figures

1.1	Comparative percentages of PB and SF immune cellular composition from the PsA cohort.	10
1.2	Quality control assessment of the ATAC-seq data generated in four immune cell types isolated from PB and SF of PsA patients samples.	12
1.3	Combined PCA analysis of all four cell types isolated from blood and SF.	13
1.4	Enrichment of eQTLs publicly available data in the combined cell type and tissue chromatin accessibility master list for the PsA cohort.	14
1.5	Annotation with genomic regions and chromatin states of the PsA DOCs from the four cell types differential analysis.	16
1.6	Enrichment of PsA DARs for the FANTOM5 eRNA dataset.	17
1.7	Differentially accessible regions located within gene bodies in CD14 ⁺ monocytes and NK cells from PsA patients.	18
1.8	Distinct enriched pathways across SF and PB in CD14 ⁺ ,CD4m ⁺ ,CD8m ⁺ and NK.	21
1.9	Representation of the FC in gene expression between SF and PB for the significant genes (pval<0.05) in at least one of the cell types.	24
1.10	Gene expression changes in immune-relevant genes between SF and PB in CD14 ⁺ monocytes, mCD4 ⁺ and mCD4 ⁺ cells.	26
1.11	Chromatin accessibility landscape at the <i>FN1</i> gene in CD14 ⁺ monocytes isolated from SF and PB.	27
1.12	Protein network analysis based on the immune qPCR array expression data.	33
1.13	Expression changes in immune-relevant genes between SF and PB in CD14 ⁺ monocytes, mCD4 ⁺ and mCD4 ⁺ cells.	34
1.14	Identification of two main CD14 ⁺ monocytes subpopulations in the SF and PB combined analysis	38
1.15	Sc-RNA-seq differential gene expression results between SF and PB in the CC-mixed and CC-IL7R CD14 ⁺ monocytes subpopulations.	40
1.16	Mean expression of IL-8 and TNF- α in SF and PB from PsA patients.	43
1.17	Mass cytometry analysis of TNF- α production in SF and PB by CD14 ⁺ monocytes after protein transport blockade with BFA.	44
A.1	FAST-ATAC and Omni-ATAC NHEK tapestation profiles.	50
A.2	Assessment of TSS enrichment from ATAC-seq and FAST-ATAC in healthy and psoriasis skin biopsies samples.	51
A.3	Percentage of MT reads in the ATAC-seq libraries generated in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ isolated from psoriasis patients and healthy controls.	54
A.4	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 ⁺	55

LIST OF FIGURES

A.5	PCA analysis illustrating batch effect in ATAC-seq and RNA-seq samples.	55
A.6	Permutation analysis SF vs PB in CD14 ⁺ ,CD4m ⁺ ,CD8m ⁺ and NK.	56
A.7	Heatmap for the top 20 marker genes of the CC-mixed and CC-IL7R CD14 ⁺ monocytes subpopulations.	57
A.8	Identification of the CD14 ⁺ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes.	57

List of Tables

1.1	Description and metadata of the PsA patients cohort.	8
1.2	Datasets generated for each sample in the PsA cohort.	9
1.3	Summary results of the differential chromatin accessibility analysis between SF and PB in PsA samples.	15
1.4	Annotation of gene body DARs in four cell types from PsA samples.	17
1.5	Immune genes with significant modulated expression in SF and proximal to a DAR in ATAC-seq.	25
1.6	Pathway enrichment analysis for the modulated genes between SF and PB in CD14 ⁺ and mCD4 ⁺	30
1.7	Most relevant enriched pathways for the DEGs between SF and PBCD14 ⁺ monocytes in CC-mixed and CC-IL7R.	42
A.1	Enrichment of ATAC-seq reads across the TSS for the CD14 ⁺ monocytes and CD4 ⁺ samples fresh, frozen and fixed.	48
A.2	Evaluation of ChiPm library complexity for the psoriasis and control chort 1B ChiPm assay.	49
A.3	Additional enriched pathways for DEGs between psoriasis and healthy controls in CD14 ⁺ monocytes and CD8 ⁺ cells.	51
A.4	Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	52
A.5	Additional enriched pathways for the DEGs between SF and PB CD14 ⁺ monocytes from the CC-mixed and CC-IL7R subpopulations.	53

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

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

1.1 Introduction

1.1.1 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 (??), 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.

Differences in the PsA inflammatory response between blood and the synovium

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 circulating IL-17⁺ and IL-22⁺ CD4⁺ T cells have been reported in PB from PsA patients compared to control individuals (**Benham2013**). Moreover, reduced percentage of pDCs and NK in PB have also been observed in PB from PsA compared to controls (**Jongbloed2008; Spadaro2004**). In terms of cytokine production, stimulated PBMCs from PsA patients released greater levels of IL-17 and IL-22 than the healthy control counterparts (**Benham2013**). Nevertheless, PsA is characterised by affection of the joints, where the local inflammatory response leads eventually to joint destruction. Oligoarticular PsA (involving four or fewer joints) is commonly managed by joint aspiration prior intra-articular steroid injection to relieve pain, facilitating the sample collection for research purposes (**Kavanaugh2006**). 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 (**Ross2000**). Additionally, elevated proportion of T cells expressing the cytokine receptors CCR6⁺ and IL-23R⁺ were found in SF compared to PB in patients (**Benham2013**).

1.1.2 Bulk transcriptomic studies 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

Cross-tissue comparison analysis in PsA

and SF from the same PsA patients or differences with other arthritic diseases (**Stoeckman2006; Batiwalla2005; Gu2002; Dolcino2015**). Amongst the most comprehensives of theses studies, that conducted by Dolcino and colleagues revealed genes from the Th-17 axis and type-I IFN signalling to be differentially expressed between PsA and healthy controls in synovial membranes. 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 (**Ritchlin1999; Li2017**).

Studies using mixed cell populations can be influenced by the relative proportion of the different cell populations within the sample (**Whitney2003**). 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 (**Peters2016; Kasela2017; Fairfax et al. 2012; Fairfax et al. 2014; Raj et al. 2014**). 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**). 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.

1.1.3 Transcriptomics and proteomics at the single cell resolution

In addition to the study of specific cell types, evidence for heterogeneity in the transcriptome across individual 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 (**Ziegenhain2017**; 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 ??). 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-Gale2017**).

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 (**Bandura2009**). The use of isotopes enables incorporating up to 45 Abs to profile cellular populations and assess molecular functions.

1.1.4 Using a multi-omics approach in the study of complex diseases

The interaction between genetics and environmental factors in the risk to develop complex diseases, such as PsA, results in shaping the epigenetic landscape of the cells. This dynamism of the epigenome involves cell and context specific features, reinforcing the importance of studying purified cell types instead of mixed populations. Moreover, the importance of the epigenomic landscape in understanding disease state and also contextualizing the role of putative genetic risk variants has made necessary the implementation of multi-omics approaches to study complex diseases. This involves the study of chromatin accessibility, gene expression including further resolution using scRNA-seq and immunophenotypes of the relevant cell involved in disease pathophysiology.

Transcriptional profiling of discrete cell populations have also been performed in monocytes and Th-17 cells in AS, intestinal epithelial cells in CD and fibroblast-like synoviocytes in RA (**Al-Mossawi2017; Smith2008; Howell2018; Ai2016**). Moreover, recent advances have yielded the possibility to study the transcriptomic and proteomic profiles at the single-cell resolution and account for the variability at the single-cell level in gene expression and protein translation.

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 (**Jaitin2014; Villani2017; Bengsch2018**). In complex diseases such as RA, scRNA-seq has discovered heterogeneity in the synovial fibroblast population and identified a potentially pathogenic cluster highly proliferative and active in pro-inflammatory cytokine secretion (**Mizoguchi2018**). Similarly, mass cytometry analysis performed in RA identified an expanded CD4⁺ T cell population promoting B cell response (**Rao2017**).

Data integration

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. Integrating chromatin accessibility and gene expression data has allowed to better understand the correlation and specificity between open chromatin and transcription. In terms of identifying cell populations, chromatin accessibility has appeared to be more cell type specific than bulk RNA-seq (**Corces2016**). In pancreatic islets, the overlap of ATAC-seq and RNA-seq has shown chromatin accessibility to be a better proxy for gene expression in α than in β cells. Moreover, integration of both data sets reinforced the functional specificity of AMD DEGs nearby differentially accessible regions across tissues (**Ackermann2016; Wang2018**).

The most comprehensive available study integrating multi-omics (bulk RNA-seq, scRNA-seq and mass cytometry) has been performed in RA (**Zhang2018**). This recent study has identified eighteen unique subpopulations amongst the main cell types involved in RA pathophysiology by systematic correlation between transcriptional profiles and mass cytometry.

Regarding PsA, no such a comprehensive multi-omic study has been conducted to date. Characterisation of chromatin accessibility and transcriptomic landscape of the most relevant cell types in SF and PB would improve the understanding of differences across tissues as well as the relationship between chromatin accessibility and gene expression in PsA. Furthermore, implementation of single-cell transcriptomic and mass cytometry would help to

1.1.5 Genetic fine-mapping using genotyping level data

1.1.6 Aims

1.2 Results

1.2.1 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 1.1). 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 ??).

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, other 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 naïve for treatment and only PsA1505 had been on MTX therapy in the past for xxx months/years (how many years ago?). Post-visit, most of the patients qualified for TNFi biologic therapy xxxx.

Cross-tissue comparison analysis in PsA

Table 1.1: 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 assessment	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 1.2 and Chapter ??). 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	Fast-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	No
PsA1801	No	No	Yes	No
PsA1605	No	No	Yes	No

Table 1.2: 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 for bulk SFMCs and PBMCs.

1.2.2 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 ???. 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 (**Ross2000; Jongbloed2006**). Interestingly, this data also showed reduced percentage of SF NK cells percentage compared PB, in line with previous studies suggesting the role of impaired non-MHC-restricted cytotoxicity in PsA (**Spadaro2004**). Similarly, a tendency towards reduced proportions of B cells in SF compared to PB reinforced the lack of

Cross-tissue comparison analysis in PsA

contribution of 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.

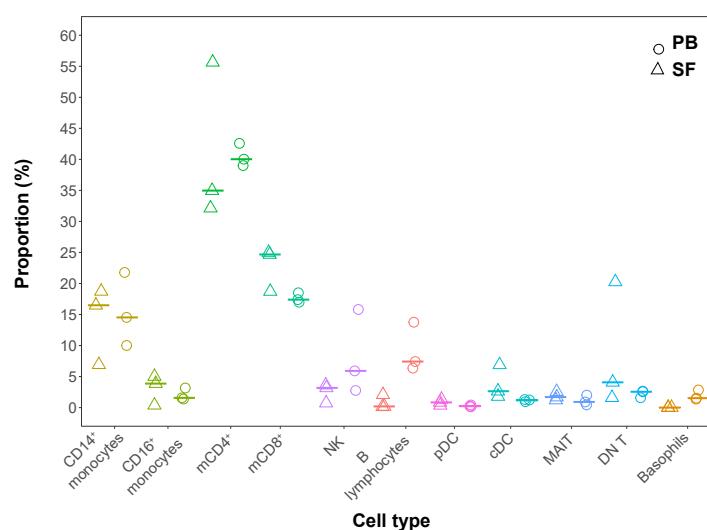


Figure 1.1: Comparative percentages of PB and SF immune cellular composition from the PsA cohort. Percentages of each of the twelve cell types identified by mass cytometry are shown by individual and tissue for PsA1718, PsA1719 and PsA1607. Horizontal line represents the median percentage for a particular 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.)

1.2.3 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 ???. 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 1.2 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 1.2 b), contributing to the loss of reads in Fast-ATAC as previously detailed in Chapter ??.

Regarding sample quality, TSS enrichment analysis showed differences in the levels of background noise across cell types and highlighted the variability of Fast-ATAC performance (Figure 1.2 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. If a larger cohort size was available these two samples would be removed from the differential analysis to prevent excessive noise to reduce the power of the differential analysis.

Cross-tissue comparison analysis in PsA

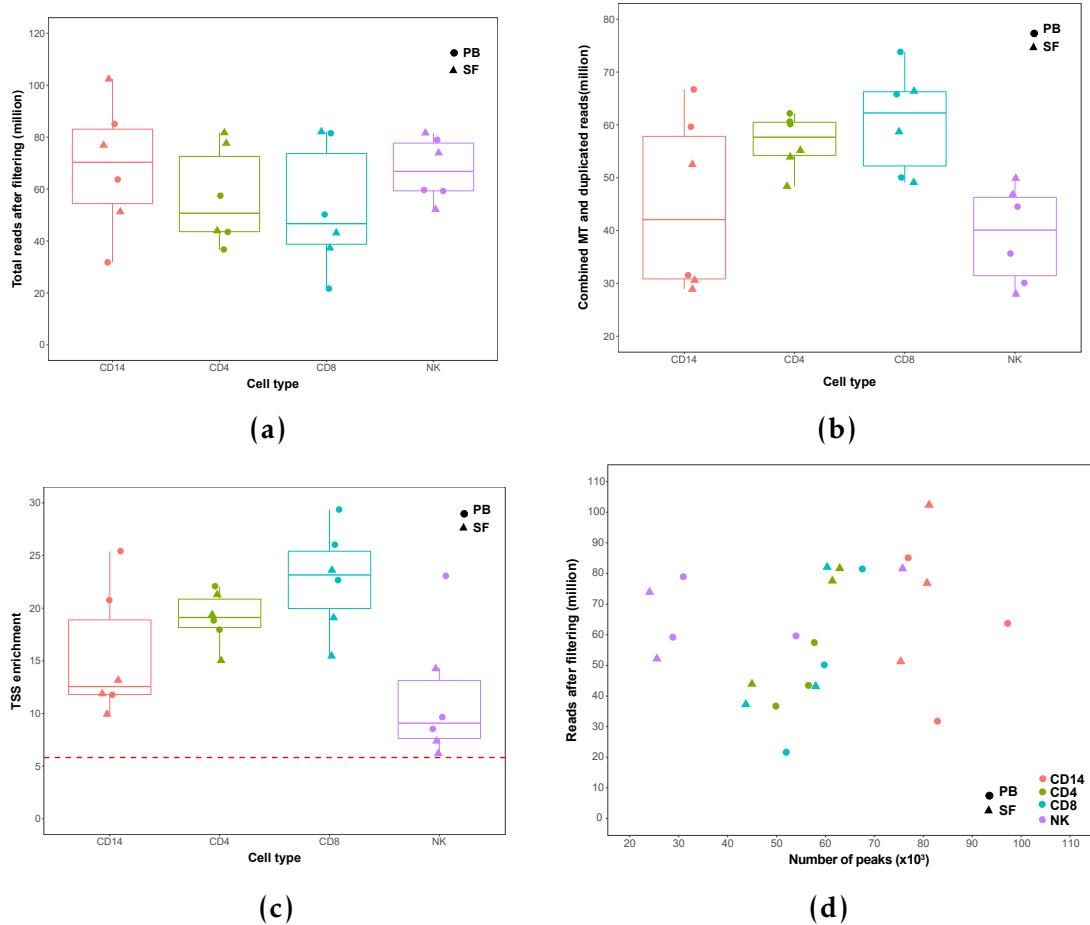


Figure 1.2: Quality control assessment of the ATAC-seq data generated in four immune cell types isolated from PB and SF of PsA patients samples. Significant differentially accessible regions for FDR<0.01 and absolute FC>1.5.

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 1.2 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

Cross-tissue comparison analysis in PsA

Chapter ?? and Chapter ??.

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 1.3). 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 1.3).

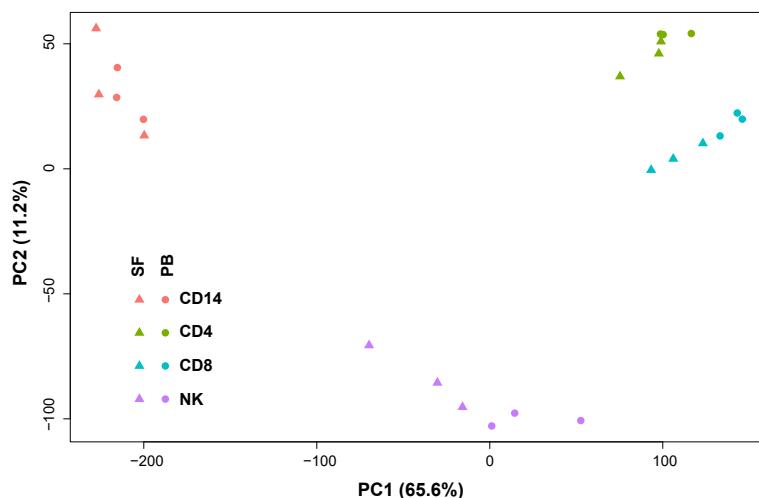


Figure 1.3: Combined PCA analysis of all four cell types isolated from blood and SF.

The ability to capture putative regulatory regions within the identified accessible chromatin regions was also explored. Enrichment analysis of different 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 1.4 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

Cross-tissue comparison analysis in PsA

unstimulated, LPS 2h and IFN- γ 24h) followed by mCD8 $^{+}$ T cells (Figure 1.4 b).

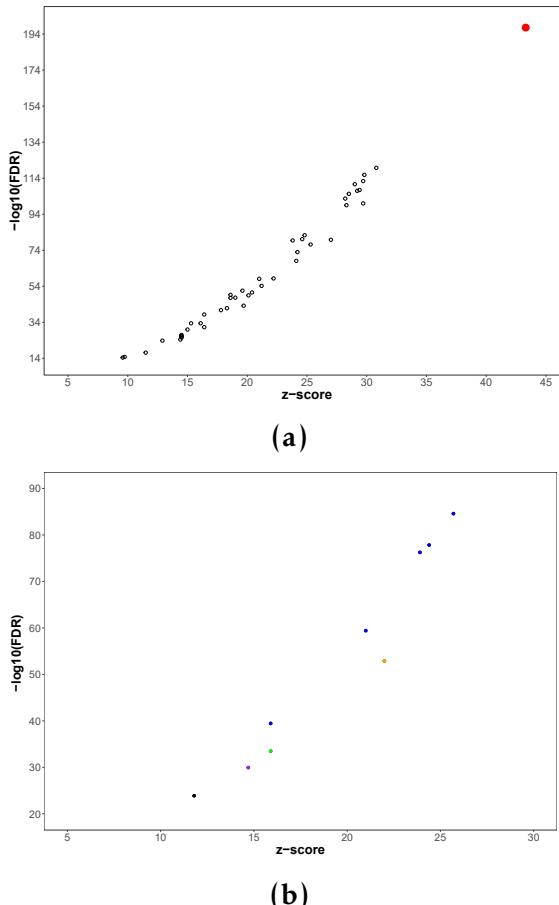


Figure 1.4: Enrichment of eQTLs publicly available data in the combined cell type and tissue chromatin accessibility master list for the PsA cohort. The dot plots showed the z-score values of the enrichment analysis in the x-axis and the significance (-log₁₀FDR) in the y-axis for a) GTEx eQTL datasets and b) non-GTEx immune-related cell types including CD14 $^{+}$ monocytes (unstimulated, 2 or 24h LPS stimulated and 24h IFN γ stimulated) in blue, B cells in black, total CD4 $^{+}$ in green, total CD8 $^{+}$ in orange and neutrophils in purple.

Characterisation of the differential accessible chromatin regions

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 DESeq2 and a paired design (Table 1.3). A 80% cut-off for background noise

Cross-tissue comparison analysis in PsA

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 1.3). 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 1.3: Summary results of the differential chromatin accessibility analysis between SF and PB in PsA samples.

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 A.6).

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 1.5 a). Universal promoter regions was the third most represented genomic feature, accounting for the annotation of 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 1.5 b). For all four cell types, between 44.96 and 72.11% of the DARs were annotated as weak enhancers, which represented the most

Cross-tissue comparison analysis in PsA

prominent category and the most significantly enriched (data not shown). This 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 1.5 b).

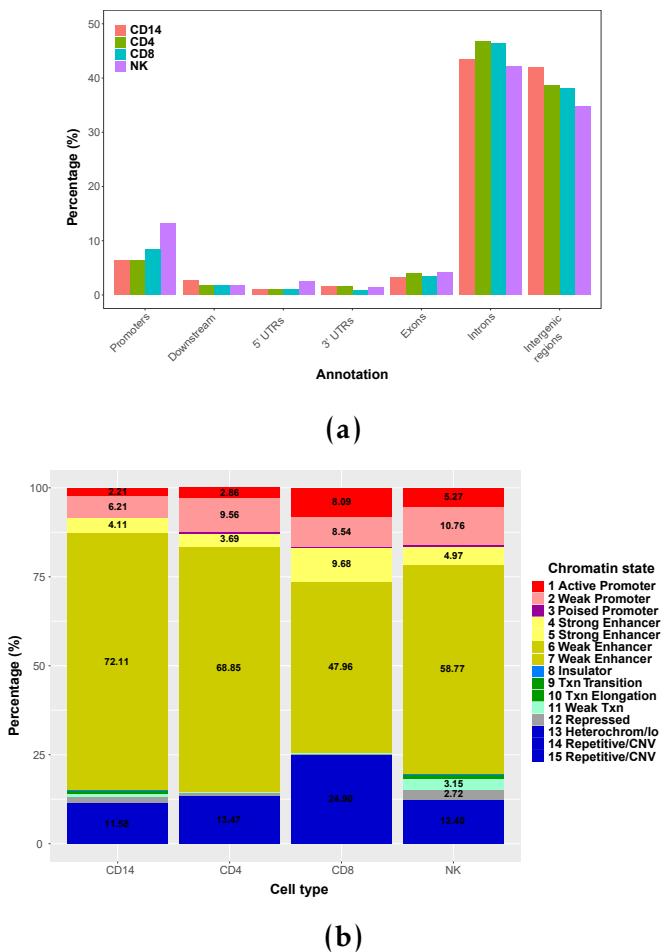


Figure 1.5: Annotation with genomic regions and chromatin states of the PsA DOCs from the four cell types differential analysis. xxxx

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

Cross-tissue comparison analysis in PsA

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 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.

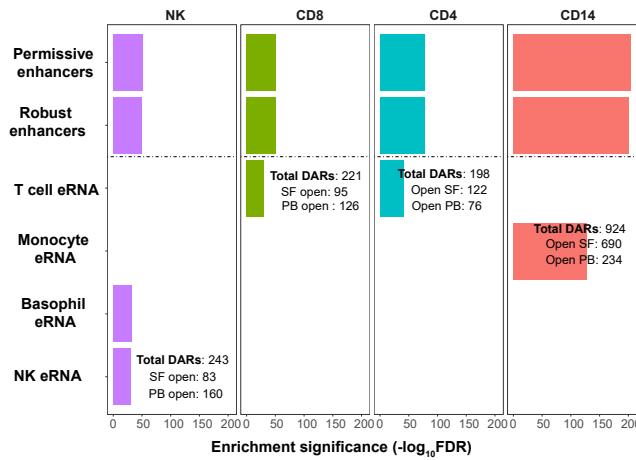


Figure 1.6: Enrichment of PsA DOCS for the FANTOM5 eRNA dataset. Robust enhancers have been defined as those detected at the genome-wide significant level in at least one primary cell type or tissue. Permissive enhancers are all detected eRNAs but not passing genome-wide filtering criteria (Andersson2014). Robust enhancers represent a subset of the permissive.

From the differential analysis between SF and PB, a number of DARs were overlapping a gene body (Table 1.4). Interestingly, the majority were located within introns instead of untranslated regions (UTRs) and have also been annotated as weak or strong enhancers according to the cell type specific chromatin segmentation map.

Cell type	DARs in gene body	Gene with > 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 1.4: Annotation of gene body DARs in four cell types from PsA samples.xxxx

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 1.7 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 1.7 b).

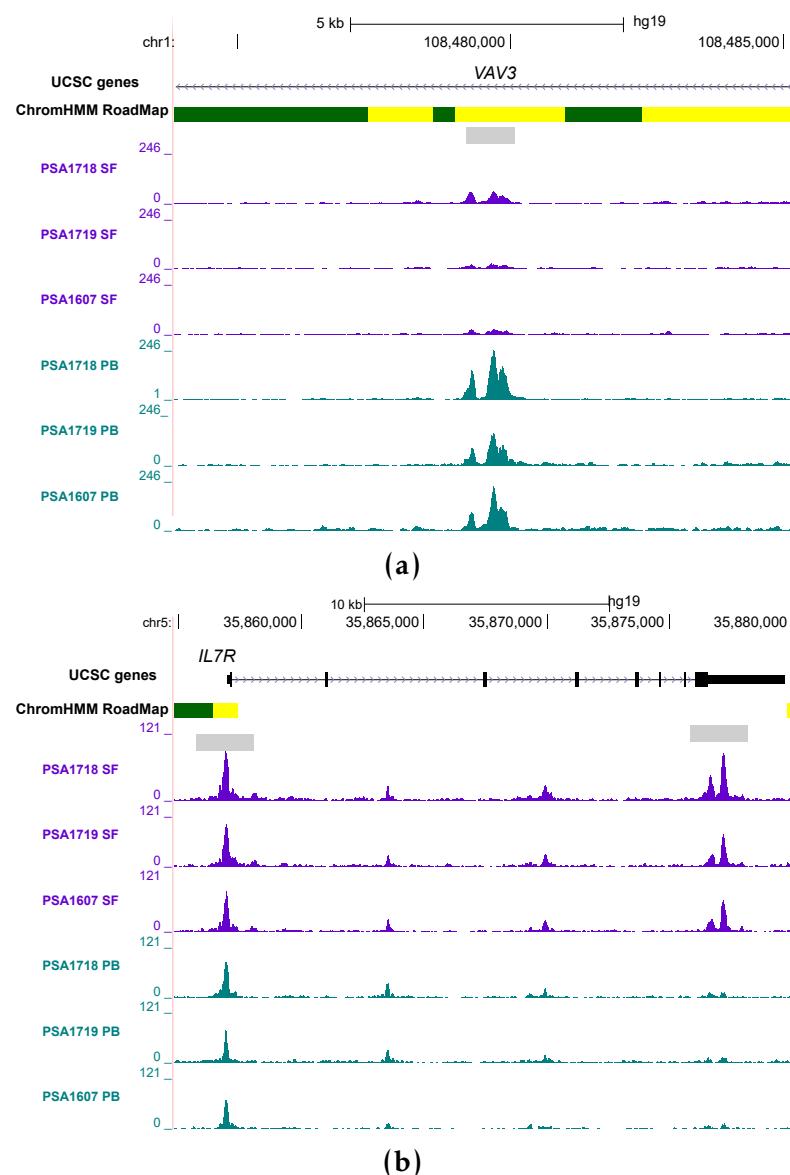


Figure 1.7: Differentially accessible regions located within gene bodies in CD14⁺ monocytes and NK cells from PsA patients. xxxx

The relevance of the differences in chromatin accessibility in the context of psoriasis and PsA GWAS hits was also addressed. Enrichment analysis of psoriasis and PsA GWAS hits for the DARs in each cell type was performed using XGR co-localisation and permutation analysis. At the SNP level, no significant enrichment was reported between DARs and GWAS lead SNPs and those in LD $r^2 \geq 8$. When the enrichment analysis was performed for the psoriasis and PsA LD blocks, significant enrichment (2-fold enrichment and empirical p-val 0.043) was observed only for the CD14⁺ DARs.

1.2.4 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 ???. Despite commonalities, differences in significant enriched pathways (FDR<0.01 or 0.05) were also identified within the same cell type between SF and PB open DARs (Figure 1.8). 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 1.8 a).

mCD4⁺ SF open DARs compared to PB open DARs showed enrichment for TCR signalling as well as chemokine signalling, which included DARs in proximity to IFN- γ , IL-2 receptor alpha (*IL2RA*) and IL-5 receptor alpha (*IL5RA*), amongst others (Figure 1.8 b). The SF open DARs in *IL2R* and *IL5R* may be related to the the IL-2, IL-3 and IL-5 pathway enrichment in CD14⁺ SF open DARs. Although the T cell signaling pathway appears only enriched for SF open DARs in mCD4⁺, PB open DARs in this cell type were also enriched for focal

Cross-tissue comparison analysis in PsA

adhesion members, also involved in the T cell activation (**Dustin2001**). Enriched pathways for SF or PB open DARs in mCD8⁺ were only significant when using an FDR<0.05 threshold. (Figure 1.8 c). Interestingly, the G protein coupled receptor (GPCR) signalling was enriched for mCD8⁺ PB open DARs, consistent with the role of this pathway in mediating the chemotactic recruitment of T cells to the inflamed tissue. 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”-e”2014**).

NK SF open DARs presented enrichment for Fc-gamma receptor (FC γ R)-mediated phagocytosis that could be triggered by occurrence of monoclonal gammopathy of undetermined significance (MGUS) in PsA patients and consequently induce NK activation (Figure 1.8 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 NK cell mediated toxicity genes was unveiled. According to FACS analysis, the proportion of NK CD56^{bright} was greater in PB compared to SF in this sample cohort (data not shown). This is consistent with the observed enrichment for NK cytotoxicity in PB open DARs and previous studies demonstrating that CD56^{bright} NK cells are preferentially cytokine producers compared to the tissue resident ones.

Cross-tissue comparison analysis in PsA

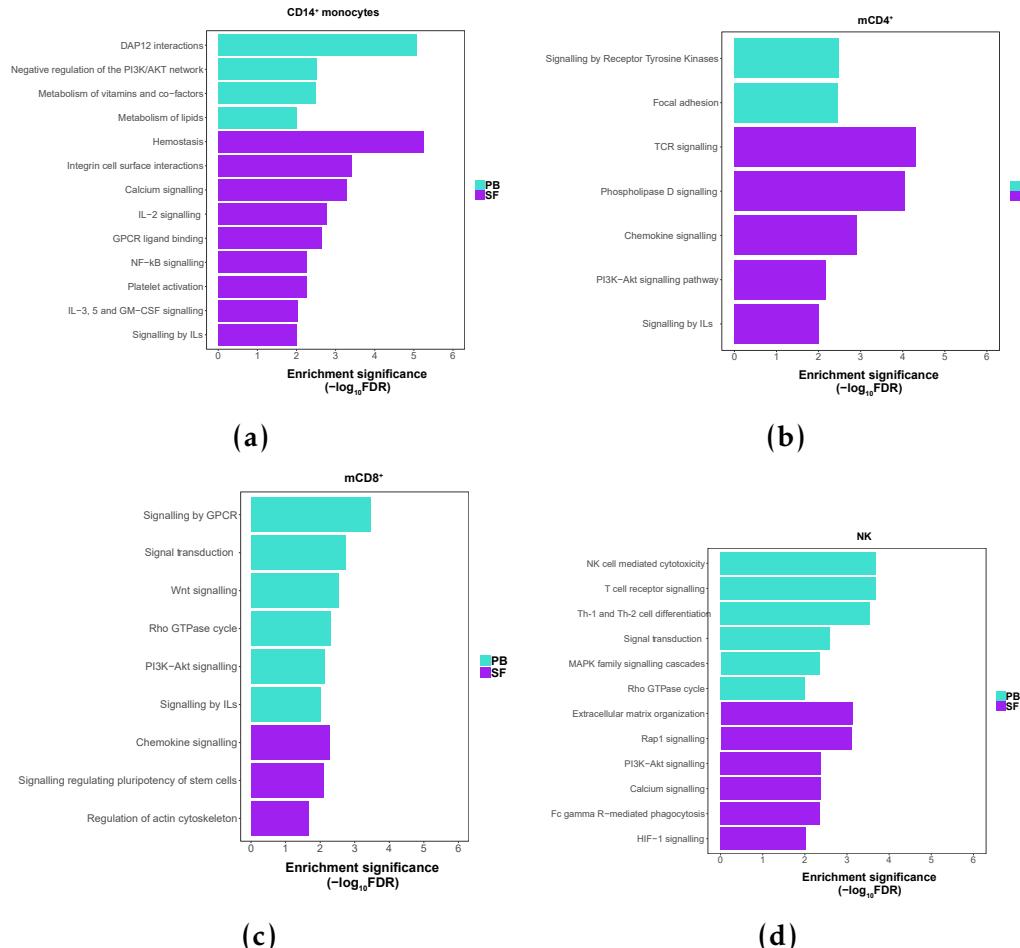


Figure 1.8: Distinct enriched pathways across SF and PB in CD14⁺, CD4m⁺, CD8m⁺ and NK. All pathways shown have an FDR <0.01.

1.2.5 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 of the chromatin landscape also presents some limitations, including the

Cross-tissue comparison analysis in PsA

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 1.2). Those appeared as the most abundant cell types in PB and SF from patients and particularly for 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 pathophysiological characteristics of 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 ??). 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 1.9). 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 1.9 orange box). On the other hand, a number of genes presented reduced expression in SF (FC<1) in at least one of the three cell types, including *FOS*, *IL16*, *PPBP* and *TPST1* (Figure 1.9 purple box). Also, a number of genes were only consistently modulated in the three CD14⁺ monocyte samples but not in T cells (Figure 1.9 dark blue box). For example, *CCR7* and *IL7R* were up-regulated in SF CD14⁺ monocytes compared

Cross-tissue comparison analysis in PsA

to PB; however the FCs between SF and PB were largely variable across the three patients in mCD4⁺ and mCD8⁺. Moreover, differences in the magnitude of FCs across the three cell types were observed for some of the genes modulated in the same direction, for instance *VEGFB* and *CXCR6* (Figure ?? green box).

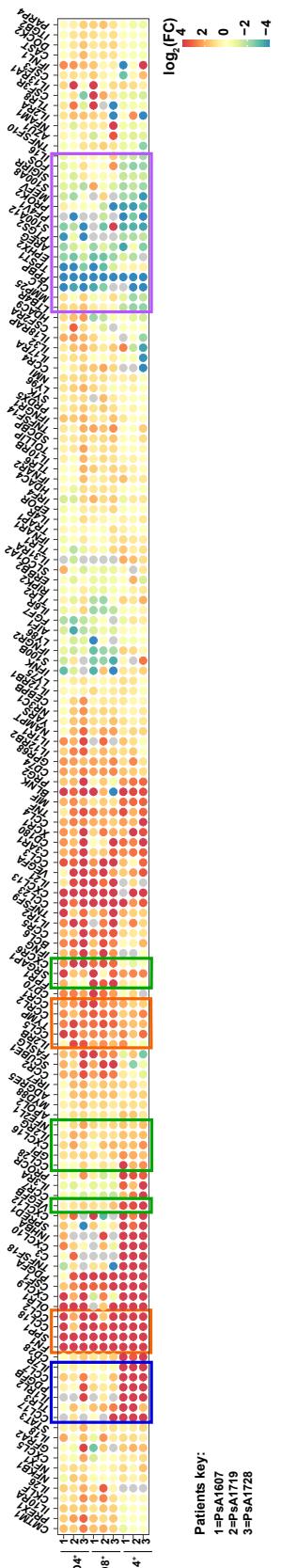


Figure 1.9: Representation of the FC in gene expression between SF and PB for the significant genes ($pval < 0.05$) in at least one of the cell types. XXXX

Filtering of all the genes tested for expression in the qPCR array based on both, statistical significance ($p\text{val}<0.05$) and FCs ($\text{abs FC}>1.5$), revealed that CD14^+ monocytes and mCD8^+ presented greater number of significantly modulated genes (70 and 73, respectively) compared to mCD4^+ cells (46 genes) (Figure 1.10 a, b and c). For the three analysed cell types, the majority of modulated immune genes presented up-regulation in the SF (Figure 1.10 a, b and c). For example, 56 out of the 70 significantly modulated genes in CD14^+ monocytes showed a mean $\text{FC}>1.5$ versus the 14 genes with mean $\text{FC}<1.5$ (Figure 1.10 a).

Correlation between gene expression and chromatin accessibility

Amongst the significantly modulated genes (based on $p\text{val}$ and FC cut-offs) between SF and PB, overlap with DARs were observed for the three cell types (Table 1.5). In CD14^+ monocytes, 13 out of the 56 significantly up-regulated genes in SF overlapped with SF open DARs.

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

Table 1.5: Immune genes with significant modulated expression in SF and proximal to a DAR in ATAC-seq. An overlap is defined by significant change in expression ($p\text{val}<0.05$) 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).

Cross-tissue comparison analysis in PsA

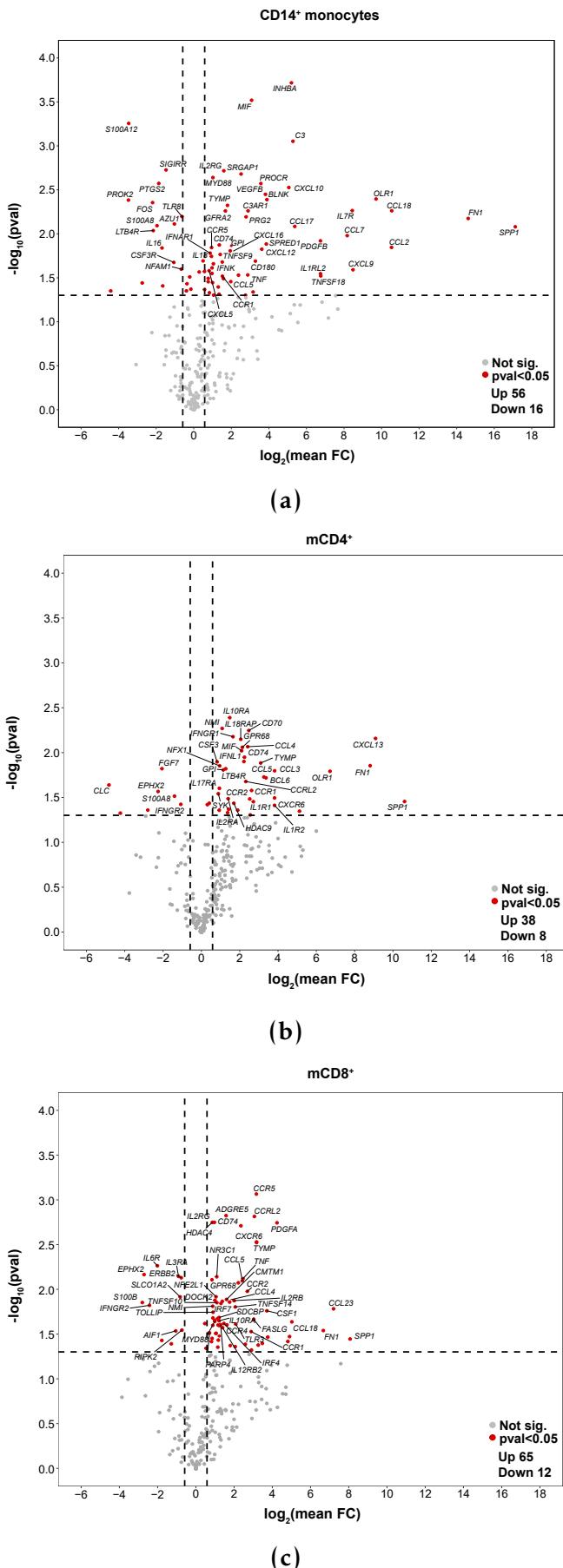


Figure 1.10: Gene expression changes in immune-relevant genes between SF and PB
 26
 in CD14⁺ monocytes, mCD4⁺ and mCD4⁺ cells.

Cross-tissue comparison analysis in PsA

For example, the greater chromatin accessibility at the *IL7R* 5' and 3' UTR, previously shown (Figure 1.7 b) correlated with greater mRNA expression in SF CD14⁺ monocytes compared with PB. Another relevant example is the *FN1* gene, involved in cell adhesion, migration and osteoblast biology. Up-regulated expression in synovial biopsies compared to PB has already been reported (**Dolcino2015**). In this cohort, *FN1* expression was up-regulated in SF for all three cell types with the greater FC found in CD14⁺ monocytes (Figure 1.10 a), concomitantly with more accessible chromatin at the promoter and 3' UTR of the gene (Figure 1.11). 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). Regarding significantly down-regulated genes, only CD14⁺ monocytes and mCD8⁺ presented overlap with proximal less accessible chromatin regions in SF (2 and 1, respectively).

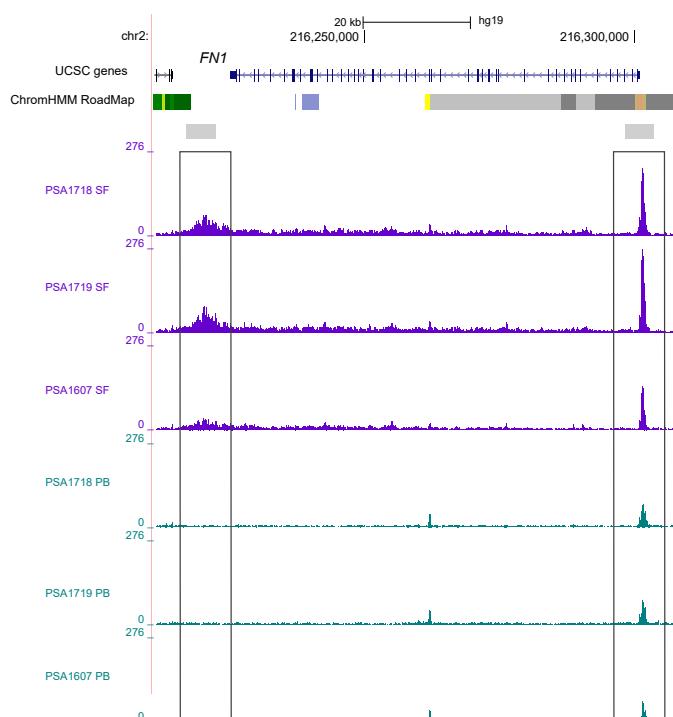


Figure 1.11: Chromatin accessibility at the *FN1* gene in CD14⁺ monocytes isolated from SF and PB. xxxx

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-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 1.6). 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 three most differentially expressed genes reported by Dolcino and colleagues in a study comparing synovial biopsies between healthy and PsA individuals (**Dolcino2015**) (Table 1.6). Together with *FN1*, *SPP1* was also one of the genes highly up-regulated (mean FC>16) in the three cell types (Figure 1.9 orange box), showing the greatest FC in monocytes (Figure 1.10 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 (Figure 1.8 a). Both, TLR and NOD-like pathways lead to the activation of the NFκB TF, which induces transcriptional activation of pro-inflammatory cytokines, further supported by the enrichment of SF open DARs for IL-2, IL-3, IL-5 and GM-CSF pathways (Figure 1.8 a). Moreover, the pivotal role of NFκB in the immune transcriptional profile of SF CD14⁺ monocytes is additionally sustained at the chromatin

Cross-tissue comparison analysis in PsA

accessibility level by the enrichment of SF accessible chromatin sites for this TF (Figure ??).

The enrichment for the chemokine pathway in CD14⁺ monocytes (Table 1.6) 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 1.9 orange box) as well as genes only consistently modulated between SF and PB in CD14⁺ monocytes (e.g *CCL28*, Figure ?? green box). The chemokine pathway includes production of chemottractant molecules involved in the recruitment of leukocytes to the site of inflammation and reactive oxygen species (ROS) through Ca²⁺ mobilisation, a pathway that has previously presented to be enriched in SF open DARs in CD14⁺ monocytes (Figure 1.8 a).

At the transcriptional level, significantly modulated genes between SF and PB in mCD4⁺ T cells were enriched for the IL-10 signalling pathway (Table 1.6), in lines with the enrichment for IL signalling of open chromatin in SF cells (Figure 1.8 b).

Cross-tissue comparison analysis in PsA

Table 1.6: 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, 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
mCD4 ⁺	TLR signalling	CCL5, CXCL10, CXCL9, IFNAR1, IRF7, MYD88, NFKB1, SPP1, FOS, TLR1, TLR2, TLR8, TNF
	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 significant modulation based on $p\text{val}<0.05$) and ranking them based on the best $p\text{val}$ 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 1.12). *SIGIRR* was significantly down-regulated in SF CD14 $^{+}$ monocytes only and it is consistent with the significant up-regulation ($p\text{val}>0.05$) of the *TLR1*, *TLR2*, *MYD88* genes in SF as well as the enrichment for the TLR pathway in this cell type (Figure 1.9 and 1.6). 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; however, in mCD8 $^{+}$, *TNF* expression is also significantly up-regulated in the synovium compared to PB.

Another interesting part of the network is the connection of the TLR pathway and the chemokine production through *NFKB*, *TNF* and *CCL2* (Figure 1.12). *CCL2* is connected to *CXCL10* and subsequently with *CCL18* and *CCR5*, all chemokines 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. Interestingly, these two nodes were only significantly up-regulated in SF CD14 $^{+}$ monocytes when compared to PB, supporting the novel cell and context specific

Cross-tissue comparison analysis in PsA

role of IL-7R and IL-7R polymorphism under inflammatory conditions in CD14⁺ monocytes(Al-Mossawi2018).

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 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 ??). 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 1.13). 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 1.13 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 1.13a, b and c).

Cross-tissue comparison analysis in PsA

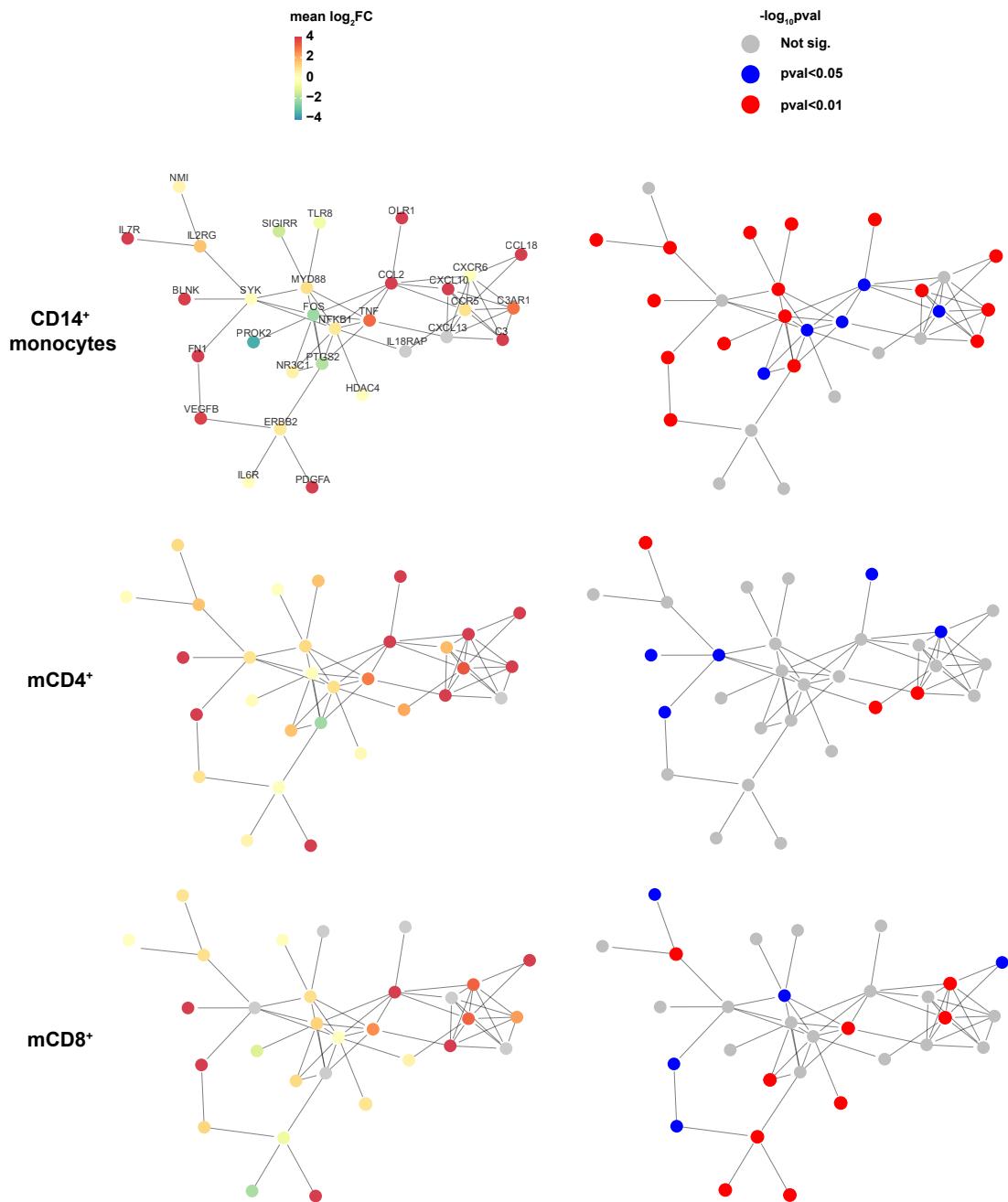


Figure 1.12: Protein network analysis based on the immune qPCR array expression data. xxxx

Cross-tissue comparison analysis in PsA

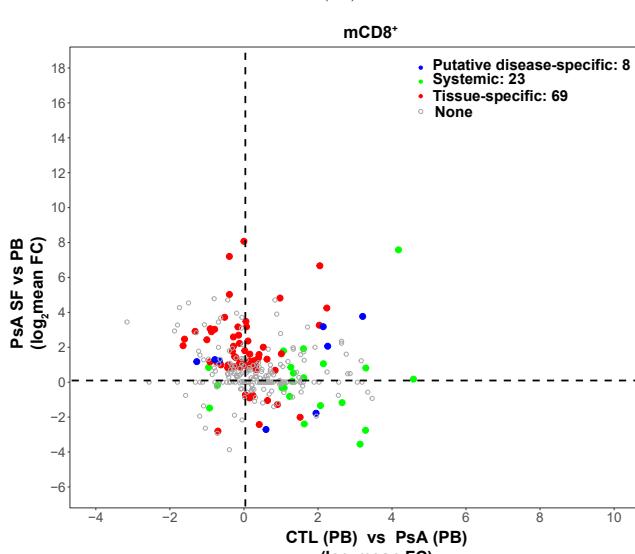
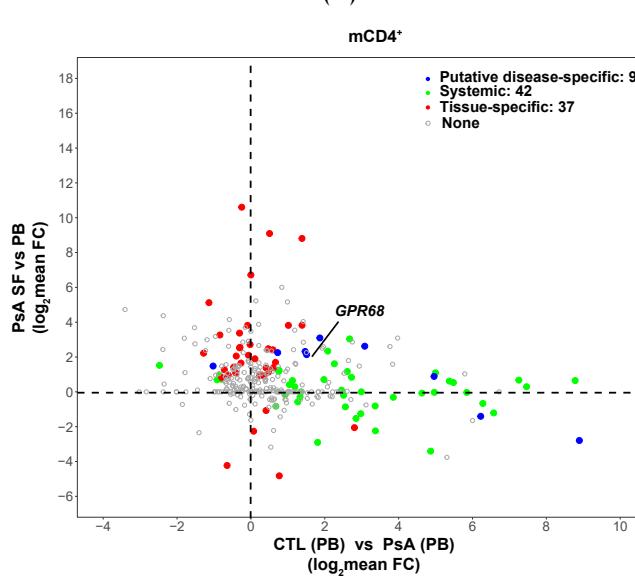
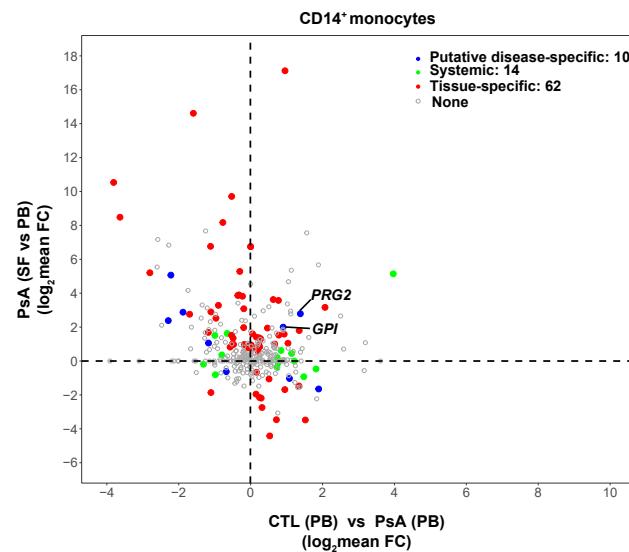


Figure 1.13: Expression changes in immune-relevant genes between SF and PB 34
CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells.

Cross-tissue comparison analysis in PsA

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 1.13 red dots). Interestingly, in CD14⁺ monocytes the tissue specific modulated genes 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 1.13 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 1.13 c red dots).

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 1.13 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 1.13 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 1.13 b). *GPR68* is a G protein-coupled receptor, expressed in T cells, amongst other cells, that undergoes activation through pH acidification, characteristic of synovial tissues under inflammation

(**Biniecka2016**). *GPR68* activation leads to an increase of Ca^{2+} levels and subsequent activation of immune-related pathways (**Saxena2011**). *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 behaviour is the epidermal growth factor-like amphiregulin (*AREG*), 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 1.13 c). *AREG* deficiency in mouse models have shown an impaired immunosuppressive response by Treg cells (**Zaiiss2013**), which could be contributing to the exacerbated immune response in the synovium. 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.

1.2.6 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 appeared as the studied population showing the greatest changes in chromatin accessibility and the most reliable modulation of expression for pro-inflammatory chemokines and cytokines between PB and SF. Since monocytes are very plastic cells

Cross-tissue comparison analysis in PsA

which differentiate into macrophages at the site of inflammation, exploring differences at the single-cell level and integrating chromatin accessibility and gene expression data could highlight subpopulations of particular interest and help to better understand the differences in the immune response between the blood and the inflamed tissue.

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 1.2). scRNA-seq data from each of the PBMCs and SFMCs samples, were filtered as explained in ?? and CD14⁺ monocytes were subset from the rest of the cells by expression of *CD14* and *LYZ*, the two appropriate markers defining this cell population (Figure A.8 a and b). These cells represented a total of 2,459 cells, approximately 17% of all the SFMCs and PBMCs cells, consistently with the proportion of CD14⁺ monocytes previously reported using mass cytometry (Figure 1.1). All the CD14⁺ monocytes identified in each of the three PsA paired PBMCs-SFMCs samples were combined using canonical-correlation analysis and followed by conservative unpervised clustering (using resolution 0.1) tSNE visualisation.

Using this conservative approach two clear clusters were identified (Figure A.8 a). The smallest cluster, named a CC-IL7R, was characterised by the expression of *IL7R*, *IL23* and *CCL5*, amongst others, and was formed by a total of 72 cells (43 cells from SF versus 29 SF) (Figure ?? b and ??). This data has been included in a publication from Al-Mossawi and colleagues in current revision (**Al-Mossawi2018**). Interestingly, the proportion of IL7R⁺CD14⁺ monocytes when compared to the total number CD14⁺ monocytes was very similar in SF and PB (3 and 2.7%, respectively) in this data. The largest cluster, named as CC-mixed, was formed by 2,387 (1,356 SF and 1,031 PB). It was quite heterogeneous,

Cross-tissue comparison analysis in PsA

without consistent expression of the genes identified as markers (Figure ??). In future increased cohort size and more exhaustive analysis could be used to identify additional subpopulations within the CD14⁺ monocytes from SF and PB in the CC-mixed cluster. For the scope of performing DGE between the two tissues, CC-mixed was considered as a unique cluster in this study.

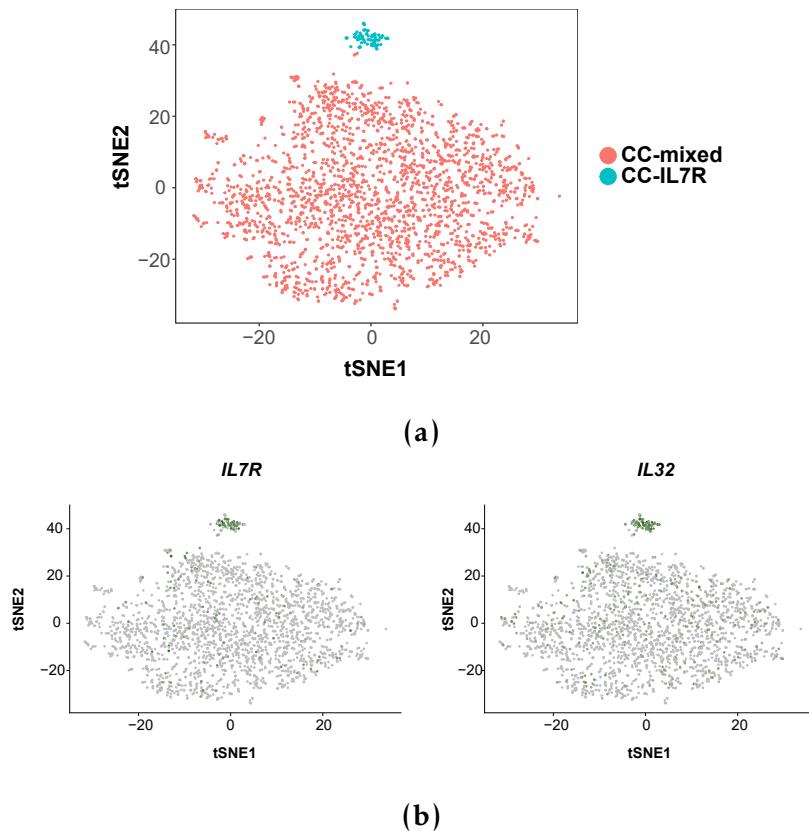


Figure 1.14: Identification of two main CD14⁺ monocytes subpopulations in the SF and PB combined analysis. a) Visualisation using tSNE dimensional reduction of the two clusters, CC-mixed and CC-IL7R, identified from the combined SF and PB CD14⁺ monocyte cells using a very conservative resolution for the unsupervised clustering analysis. b) Overlap of IL7R and IL32 expression intensities by the CC-mixed and CC-IL7R CD14⁺ monocytes.

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

In order to explore the differences between SF and PB at a genome-wide level within each of these two main CD14⁺ monocytes subpopulations, DGE

Cross-tissue comparison analysis in PsA

was performed. In the CC-mixed cluster a total of 251 genes were differentially expressed at an FDR<0.01 and $\log_2\text{FC}>1.5$, of those being 149 genes upregulated and 102 genes downregulated in SF compared to PB (Figure 1.15 a). Differential analysis in then CC-IL7R revealed a total of 37 modulated genes, with the majority (35 out of 37) upregulated in SF compared to PB (Figure 1.15 a). Due to the small size of the CC-IL7R and the small cohort size (only three samples), the differential analysis only identified as significantly differentially expressed (FDR<0.01) genes presenting large $\log_2\text{FC}$ (1.5 or greater). Of the 37 genes differentially expressed between the two tissues in the CC-IL7R cluster, 30 were shared with the CC-mixed cluster. Amongst the 7 distinctly modulated genes in the CC-IL7R cluster were *CD44*, *MT-CO2* or S-ribosomal protein (RPS) genes (*RPS29* and *RPS27*). *CD44* is a receptor hialuronic acid and the protein product from *SPP1* osteopontin an immune modulator increasing chemotaxis, cell activation and cytokine production. Interestingly, *SPP1* was also upregulated in SF when compared to PB in the two CD14^+ monocytes subpopulations (Figure 1.15 a and b).

Cross-tissue comparison analysis in PsA

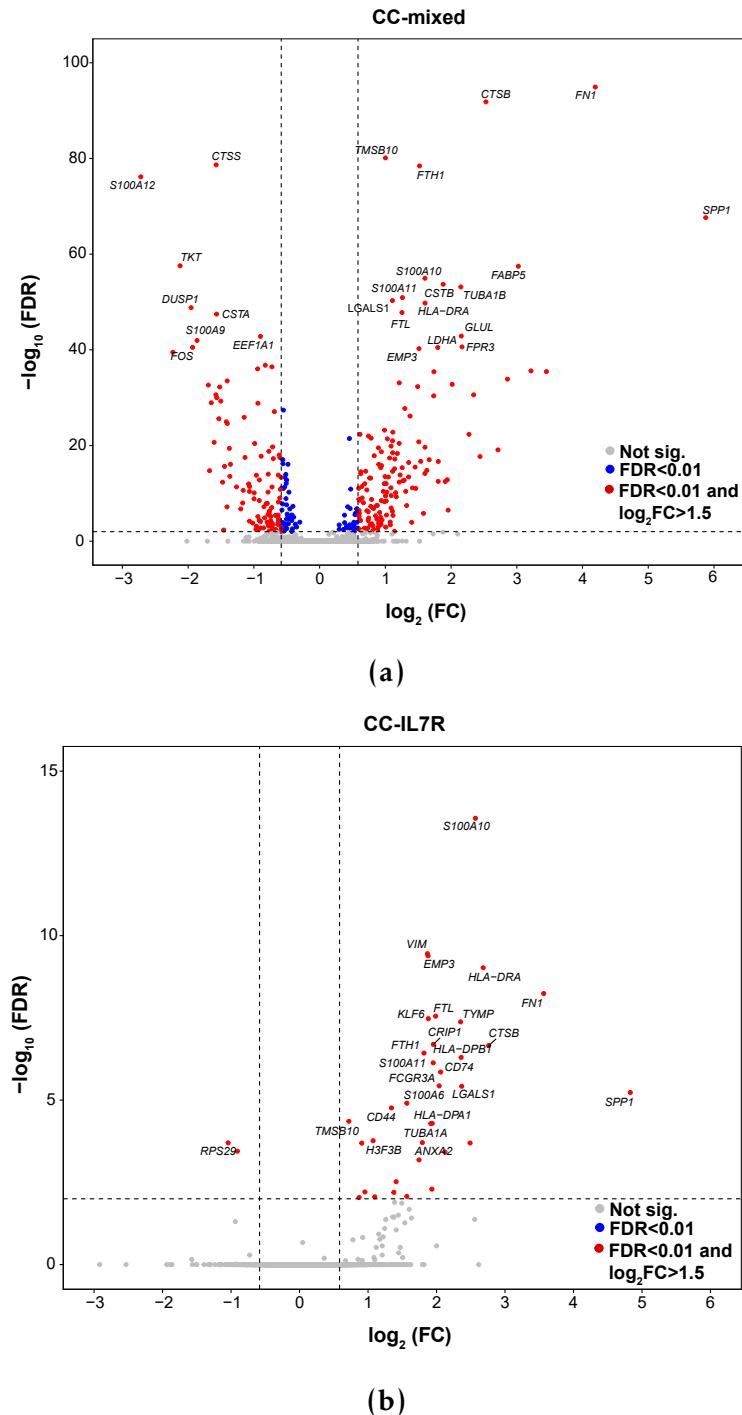


Figure 1.15: Sc-RNA-seq differential gene expression results between SF and PB in the CC-mixed and CC-IL7R CD14⁺ monocytes subpopulations. xxxx

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 enriched for Ag processing and presentation pathway contributed by up-regulation of expression in the SF of *CD74* and genes from the

Cross-tissue comparison analysis in PsA

HLA-D family (**Lamb1992**) (Table 1.7 and A.5). Enrichment for IFN signalling was driven by the *IFI6*, *IFITM3*, *LY6E*, (**ISG15**) and 0F *STAT1*, all of them upregulated in SF when compared to BD CD14⁺ monocytes. Interestingly, *IFI6*, *IFITM3*, *LY6E* have been identified as markers of a subpopulation of CD14⁺ monocytes for RA (**Zhang2018**). Another interesting enriched pathway was the extracellular matrix and extracellular matrix-associated proteins which involves proteins of the S100 family that interact with the receptor for advance glycosylation end products (RAGE) and induce production of matrix-degrading enzymes involved in joint erosion and development of arthritis (**Raghunatha2012**). Phagosome and lysosome formation also appeared to be more active in SF CD14⁺ monocytes through overexpression of genes *CTSL*, involved in protein degradation in lisosomes and an overall role in phagocytosis of apoptotic cells. Lastly, enrichment for cytokine signalling involved for example the insulin receptor substrate*IRS2* and up-regulation of the kinase *MAP3K8* which activates MAPK and JNK kinase pathways (ref). Pathway enrichment analysis for the DEGs in the CC-ILR subpopultaion identified common relevant pathways to the CC-mixed analysis (Table 1.7).

Table 1.7

Interestingly, latest studies have involved RPS proteins in immune signalling and modulation of the NF-B response (**Zhou2015**)

-Correlation with ATAC, overall picture -The biological relevance of the IL7R does not show distinct differences in gene expression between tissues and its relevance may be related to the distinctive biological function of this subset in both tissues. Seems more abundant in SF and that would make sens ewith the differences in the PCR array in bulk expression.

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

Table 1.7: 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.

1.2.7 Integration of PsA GWAS and epigenetics

ATAC peaks overlap with GWAS Check gene expression Fine-mapping of those loci

1.2.8 Mass cytometry reveals CD14⁺ monocytes as the most active cytokine producing cells

Measurement of cytokine production using mass cytometry was conducted in collaboration with Dr Hussein Al-Mossawi and Dr Nicole Yager. Cytokine production by CD14⁺ monocytes, mCD4⁺ T cells, mCD8⁺ T cells and NK was quantified in SF and PB after incubation with BFA for 6 h. BFA blocks cytokine release and enables to measure the cell cytokine production rate in absence of any inflammatory stimuli, as previously explained in Chapter ??(Figure 1.16). Following data QC, IL-2, IL-8, IFN- γ and TNF- α were the cytokines consistently released with the most reliable quantification results within each of the manual

Cross-tissue comparison analysis in PsA

gated cell type clusters. CD14⁺ monocytes isolated from both, SF and PB, was the only active cell type in producing IFN- γ , IL-2 and IL-8. Nevertheless, no significant differences in the production (mean) of those three cytokines was observed in CD14⁺ monocytes between SF and PB. On the other hand, only CD14⁺ monocytes isolated from SF showed positive detectable mean levels of TNF- α , which was undetectable in the PB counterpart cells.

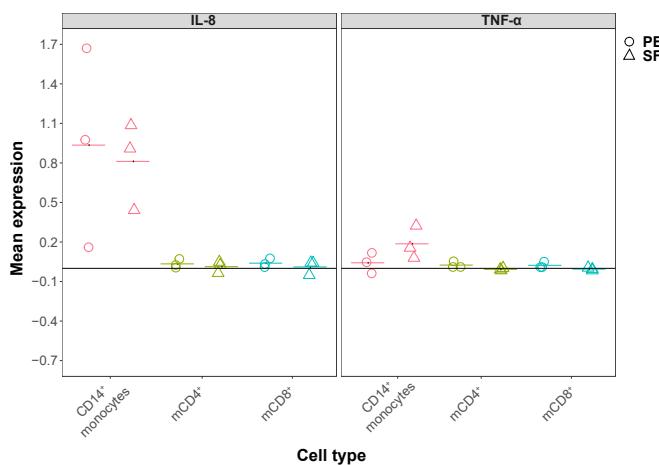


Figure 1.16: Mean expression of IL-8 and TNF- α in SF and PB from PsA patients. Expression levels quantified by mass cytometry. xxxx

The increased rate of TNF- α production by SF CD14⁺ monocytes was reinforced when comparing the proportion of TNF- α positive cells after 6h of blocked cytokine secretion in each of the two tissues. The mass cytometry analysis of CD14⁺ versus TNF- α signal showed a greater number of SF CD14⁺ monocytes producing TNF α after 6 h of BFA treatment compared to PB for each of the three analysed patients (Figure 1.17 a). The basal (0h) percentage of CD14⁺ monocytes positive for TNF- α was negligible for all three patients in both, SF and PB (Figure 1.17 b). Conversely, upon inhibition of protein transport (6h), the percentage of CD14⁺ monocytes producing TNF- α experienced an increase, ranging between 2 and 11.8%, in SF whereas the increment in PB did not reach 1%. This trend of increased percentage of SF CD14⁺ monocytes TNF- α did not reach significance likely due to the small sample size.

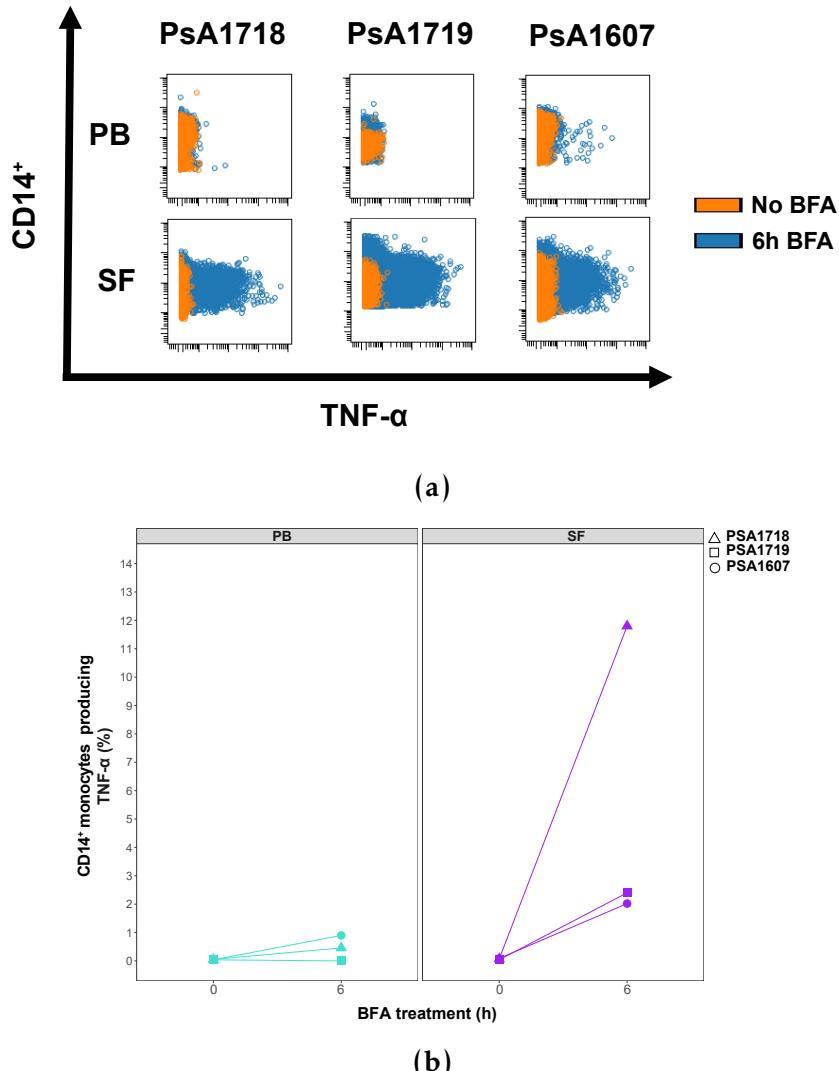


Figure 1.17: Differentially accessible regions located within gene bodies in CD14⁺ monocytes and NK cells from PsA patients. xxxx

1.3 Discussion

In Dolcino PSA vs HV SF comparison also detected more up genes than down in the array analysis Say which cell types drive more the top changed genes fGWAS analysis as Matthias did would be of interest but needs appropriate GWAS data I am going to try using XGR to do some of this

The integration of the varied types of datasets that can be generated from clinical samples of a wide range of complex diseases remains challenging. It

Cross-tissue comparison analysis in PsA

requires the implementation or development of new algorithms in order to integrate them into a systematic way Machine learning has been used for RA
<https://onlinelibrary.wiley.com/doi/epdf/10.1002/art.40428> CCA analysis also for RA Zhang2018

Appendices

A Appendices	48
A.1 Additional tables	48
A.1.1 Chapter 5 Tables	48
A.1.2 Chapter 4 Tables	54
A.1.3 Chapter 5 Tables	54
A.2 Additional figures	54
A.2.1 Chapter 3 Figures	54
A.2.2 Chapter 4 Figures	54
A.2.3 Chapter 5 Figures	56

List of Figures

A.1 FAST-ATAC and Omni-ATAC NHEK tapestation profiles.	50
A.2 Assessment of TSS enrichment from ATAC-seq and FAST-ATAC in healthy and psoriasis skin biopsies samples.	51
A.3 Percentage of MT reads in the ATAC-seq libraries generated in CD14 ⁺ monocytes, CD4 ⁺ , CD8 ⁺ and CD19 ⁺ isolated from psoriasis patients and healthy controls.	54
A.4 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 ⁺	55
A.5 PCA analysis illustrating batch effect in ATAC-seq and RNA-seq samples.	55
A.6 Permutation analysis SF vs PB in CD14 ⁺ ,CD4m ⁺ ,CD8m ⁺ and NK.	56
A.7 Heatmap for the top 20 marker genes of the CC-mixed and CC-IL7R CD14 ⁺ monocytes subpopulations.	57

A.8 Identification of the CD14 ⁺ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes.	57
---	----

List of Tables

A.1 Enrichment of ATAC-seq reads across the TSS for the CD14 ⁺ monocytes and CD4 ⁺ samples fresh, frozen and fixed.	48
A.2 Evaluation of ChIPm library complexity for the psoriasis and control chort 1B ChIPm assay.	49
A.3 Additional enriched pathways for DEGs between psoriasis and healthy controls in CD14 ⁺ monocytes and CD8 ⁺ cells.	51
A.4 Additional enriched pathways for DEGs between lesional and uninvolved epidermis isolated from psoriasis patients skin biopsies.	52
A.5 Additional enriched pathways for the DEGs between SF and PB CD14 ⁺ monocytes from the CC-mixed and CC-IL7R subpopulations.	53

Appendix A

Appendices

A.1 Additional tables

A.1.1 Chapter 5 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.

Appendices

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 chort 1B ChiPm assay. NRF, PBC1 and PBC2 are the three measures used according to the ENCODE standards as referred in Chapter ???. $0.5 \leq \text{NRF} < 0.8$ 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.

Appendices

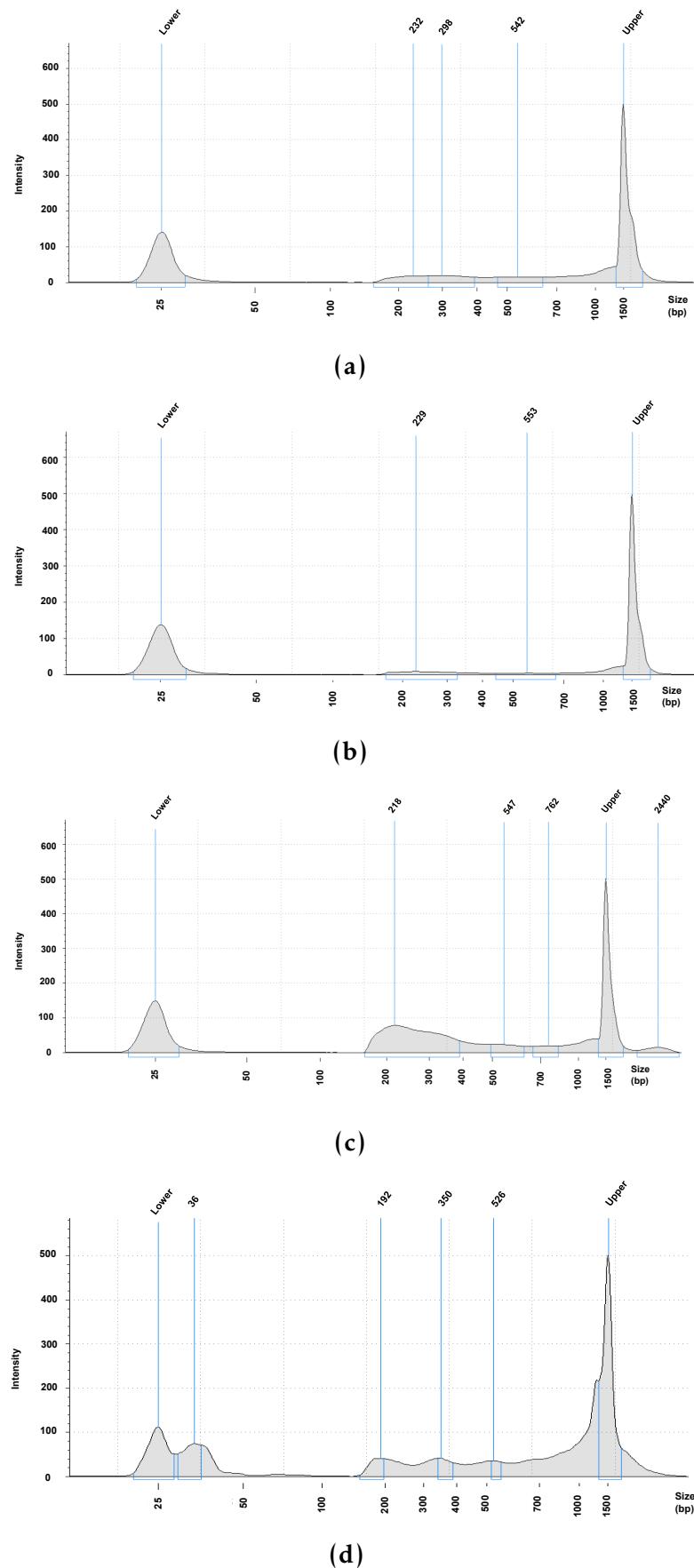


Figure A.1: FAST-ATAC and Omni-ATAC NHEK tapestation profiles.

Appendices

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.

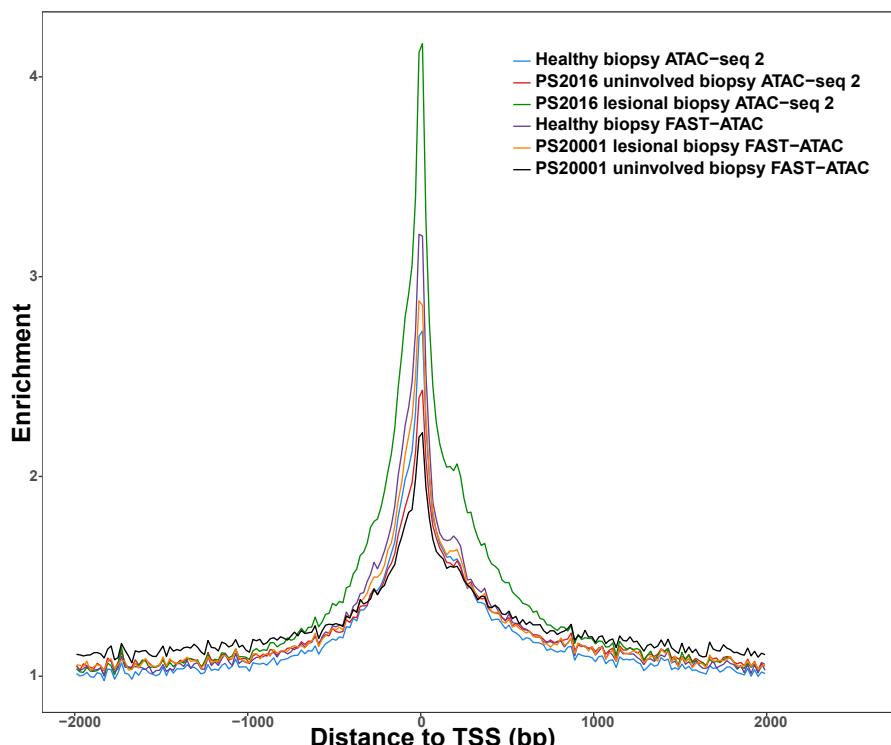


Figure A.2: Assessment of TSS enrichment from ATAC-seq and FAST-ATAC in healthy and psoriasis skin biopsy samples.

Appendices

Lesional versus unininvolved 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 unininvolved 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.

Appendices

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.5: 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.

Appendices

A.1.2 Chapter 4 Tables

A.1.3 Chapter 5 Tables

A.2 Additional figures

A.2.1 Chapter 3 Figures

A.2.2 Chapter 4 Figures

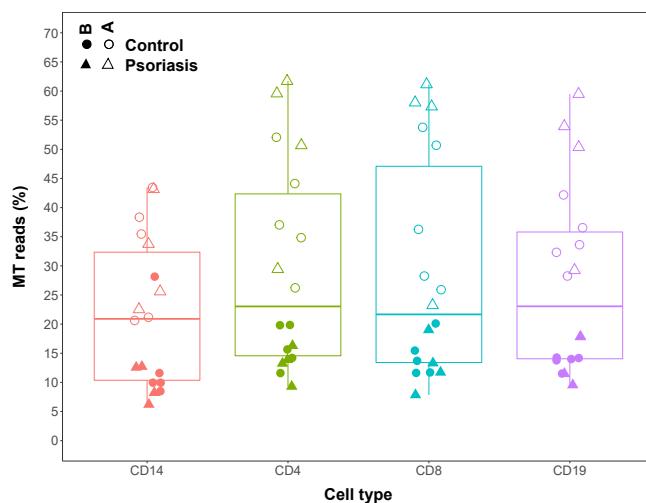


Figure A.3: Percentage of MT reads in the ATAC-seq samples generated in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺ isolated from psoriasis patients and healthy controls. Samples from cohort 1A (open circles and triangles) were generated with the standard ATAC-seq protocol from Buenrostro *et al.*, 2013 whereas samples from cohort 1B (filled circles and triangles) were processed using FAST-ATAC (**Corces2016**).

Appendices

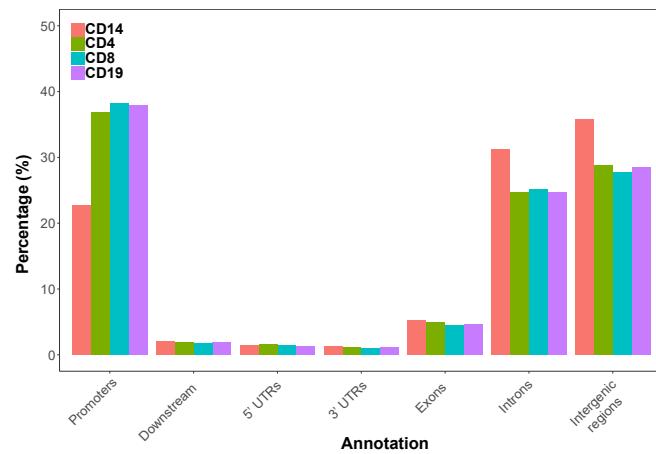
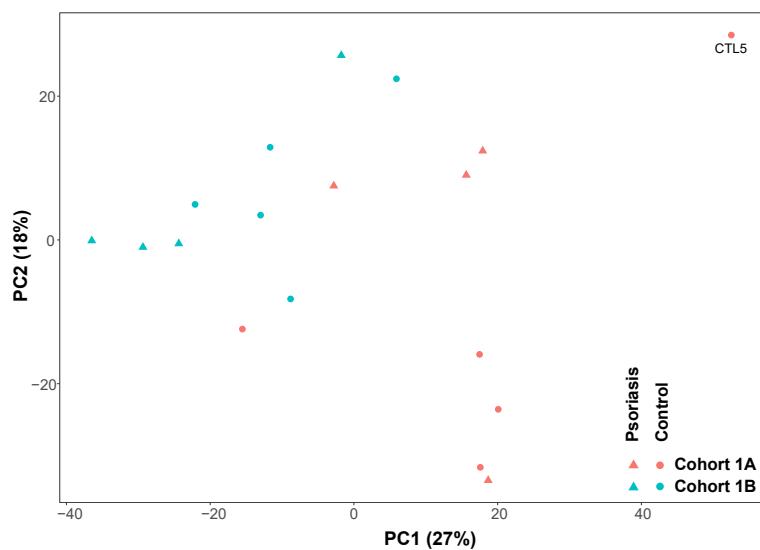
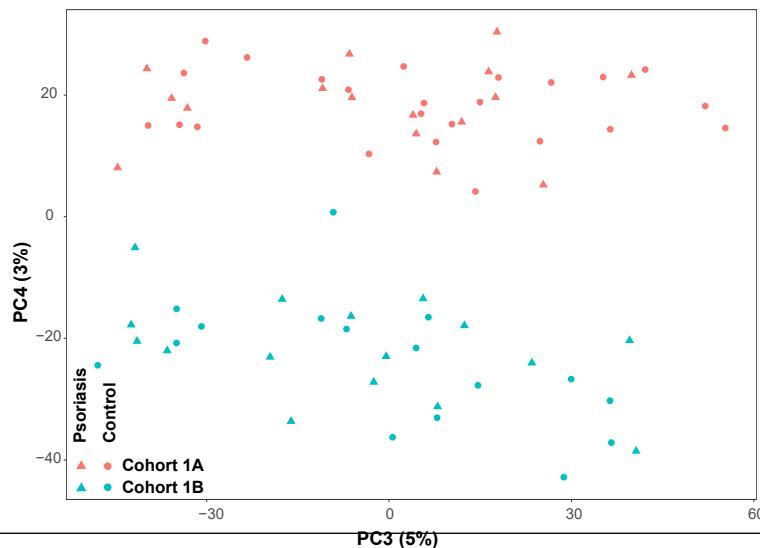


Figure A.4: Genomic annotation of the consensus master list of ATAC-seq enriched sites built for downstream differential chromatin accessibility analysis in CD14⁺ monocytes, CD4⁺, CD8⁺ and CD19⁺. Annotation is expressed in percentage over the total number of ATAC-seq sites included in each particular cell type master list.



(a)



(b)

Appendices

A.2.3 Chapter 5 Figures

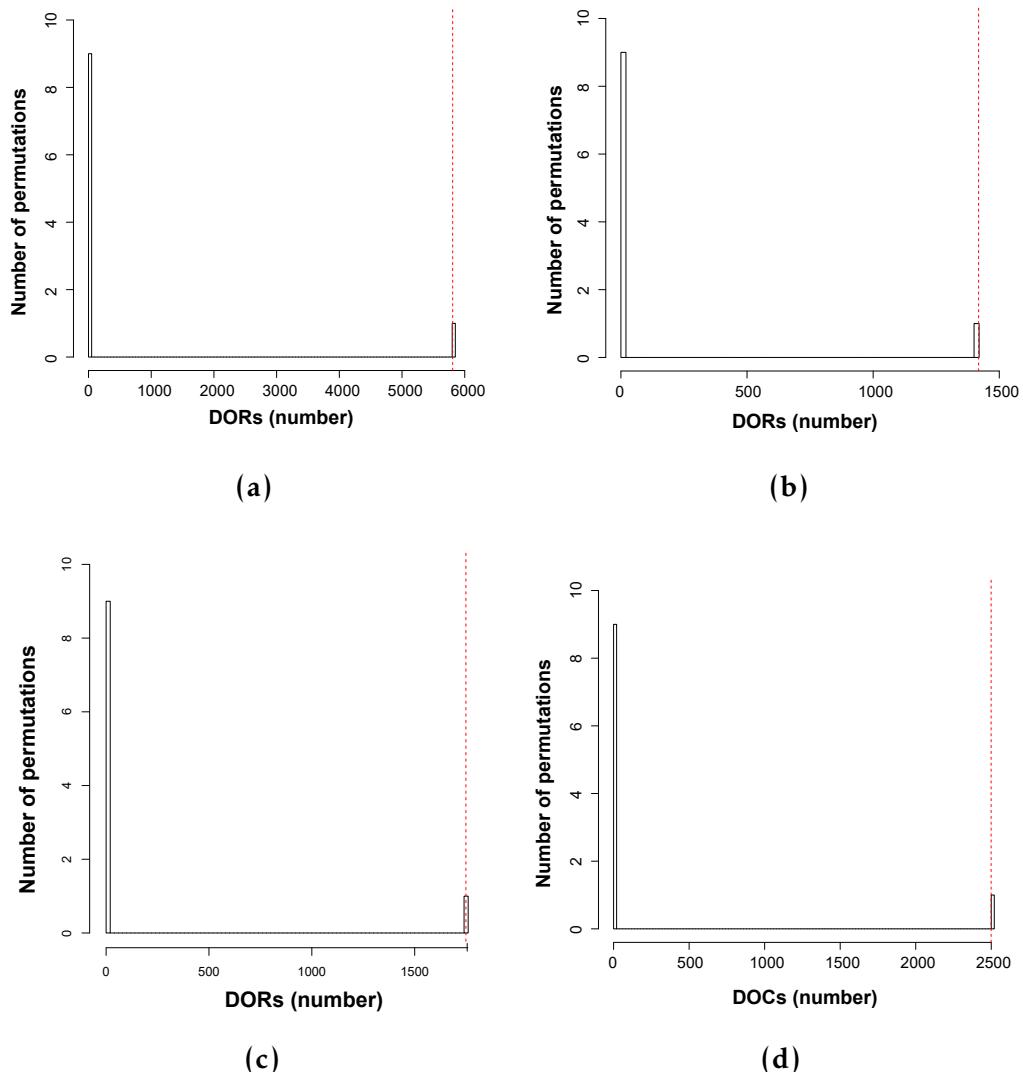


Figure A.6: Permutation analysis SF vs PB in CD14⁺,CD4m⁺,CD8m⁺ and NK.

Appendices

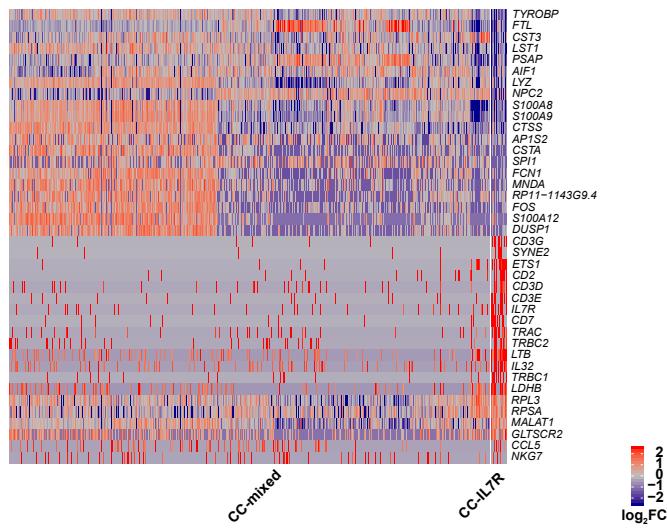


Figure A.7: Heatmap for the top 20 marker genes of the CC-mixed and CC-IL7R CD14⁺ monocytes subpopulations. Rows are the top 20 marker genes for each of the two subpopulations (total of 40 genes). The columns represent each of the cells members of the CC-mixed (left) or CC-IL7R (right) clusters. The 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.

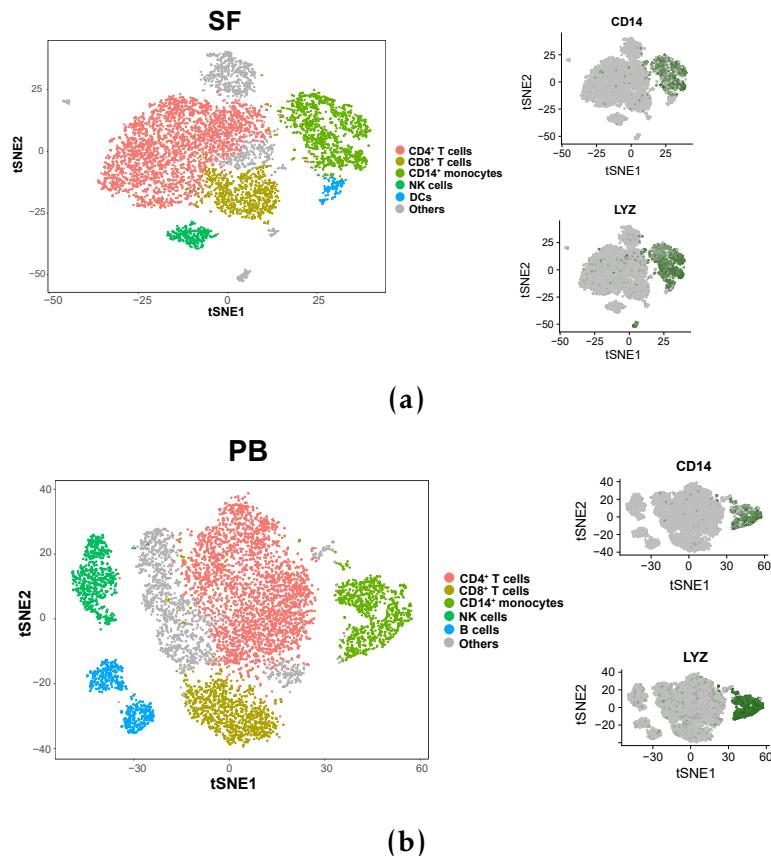


Figure A.8: Identification of the CD14⁺ monocytes populations from bulk SFMCs and PBMCs using scRNA-seq transcriptomes. XXXX