



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 in PsA and their limitations	2
1.1.3 Transcriptomics and proteomics at the single cell resolution	4
1.1.4 The challenges of using a multi-omics approach in the study of complex diseases	5
1.1.5 Integration of fine-mapping GWAS SNPs and functional data in PsA	6
1.1.6 Aims	8
1.2 Results	8
1.2.1 PsA patients cohort description and datasets	8
1.2.2 Immune cellular composition of blood and synovial fluid in the PsA cohort	11
1.2.3 Differential chromatin accessibility analysis in immune cells reveals differences between SF and PB	12
1.2.4 Pathway enrichment analysis highlights tissue functional differences in chromatin accessibility	22
1.2.5 Differential gene expression analysis in paired circulating and synovial immune cells	24
1.2.6 Characterisation of the CD14 ⁺ monocyte heterogeneity in PsA using scRNA-seq	40

CONTENTS

1.2.7 Mass cytometry reveals active cytokine production in CD14 ⁺ monocytes.	51
1.2.8 Prioritisation and interpretation of PsA GWAS SNPs	53
1.3 Discussion	66
Bibliography	70

List of Figures

1.1	Comparative percentages of PB and SF immune cellular composition from the PsA cohort.	12
1.2	Quality control assessment of ATAC data generated in four immune cell types isolated from PB and SF of PsA patients samples.	14
1.3	PCA analysis based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and SF.	15
1.4	Enrichment of eQTLs publicly available data in the combined cell type and tissue chromatin accessibility master list for the PsA cohort.	16
1.5	Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states.	19
1.6	Enrichment of PsA DARs for the FANTOM5 eRNA dataset.	20
1.7	Differentially accessible regions located within gene bodies in CD14 ⁺ monocytes and NK cells from PsA patients.	21
1.8	Distinct functional pathways enriched for DARs open in SF or open in PB in CD14 ⁺ monocytes, CD4m ⁺ ,CD8m ⁺ and NK.	24
1.9	Heatmap of gene expression FCs between SF and PB for those gene significantly modulated ($pval < 0.05$) in at least one cell type.	26
1.10	Gene expression changes in immune-relevant genes between SF and PB in CD14 ⁺ monocytes, mCD4 ⁺ and mCD8 ⁺ cells.	28
1.11	Chromatin accessibility landscape at the qPCR differentially expressed <i>FN1</i> gene in CD14 ⁺ monocytes.	30
1.12	Protein network analysis based on the immune qPCR array expression data.	36
1.13	Comparison of immune-relevant gene expression modulation across PsA tissues (SF vs PB) and in PsA patients verus healthy controls.	38
1.14	Identification of two main CD14 ⁺ monocytes subpopulations in the SF and PB combined analysis	41
1.15	Sc-RNA-seq differential gene expression results between SF and PB in the CC-mixed and CC-IL7R CD14 ⁺ monocytes subpopulations.	44
1.16	Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14 ⁺ monocytes.	47
1.17	Chromatin landscape in CD14 ⁺ monocytes upstream the differentially expressed gene <i>CCL2</i>	50
1.18	Mean expression of IL-8 and TNF- α markers in SF and PB from three PsA patients quantified by mass cytometry.	51
1.19	Comparison of TNF- α expression in SF and PB CD14 ⁺ monocytes before and after protein transport blockade with BFA using mass cytometry.	53
1.20	Experimental Factor Ontology terms enriched in GWAS Catalogue SNPs overlapping ATAC regions in four cell types.	59

LIST OF FIGURES

1.21 Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal.	63
1.22 Epigenetic landscape at the genomic location of three fine-mapped SNPs from the <i>STAT2/IL23A</i> PsA GWAS signal.	65

List of Tables

1.1	Description and metadata of the PsA patients cohort.	10
1.2	Datasets generated for each sample in the PsA cohort.	11
1.3	Summary results of the differential chromatin accessibility analysis between SF and PB in PsA samples.	17
1.4	Characterisation of the DARs located within genes in each of the four cell types from PsA samples.	20
1.5	Immune genes with significant modulated expression in SF and proximal to a DAR in Fast-ATAC.	29
1.6	Pathway enrichment analysis for the modulated genes between SF and PB in CD14 ⁺ and mCD4 ⁺	33
1.7	Most relevant enriched pathways for the DEGs between SF and PBCD14 ⁺ monocytes in CC-mixed and CC-IL7R.	46
1.8	scRNA-seq DEGs in SF versus PB CD14 ⁺ monocytes proximal to a DAR in Fast-ATAC.	48
1.9	Summary table of the PsA GWAS loci presenting $-\log_{10}ABF > 3$ for the fine-mapping lead SNP.	55
1.10	PsA fine-mapped SNPs from the 90% credible sets overlapping accessible chromatin identified by ATAC in four cell types.	58
1.11	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.	61

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.

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 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 (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).

1.1.2 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; Dolcino2015**; Stoeckman et al. 2006; Gu et al. 2002). Amongst the most comprehensives of these 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

Cross-tissue comparison analysis in PsA

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 (Ritchlin 1999; 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 (Kasela 2017; Fairfax et al. 2012; Fairfax et al. 2014; Raj et al. 2014; Peters et al. 2016). 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 (Antoniv 2006; Leipe 2010). The importance of investigating the transcriptional profile of patients' isolated discrete cell populations have yielded interesting findings in monocytes and Th-17 cells in AS, intestinal epithelial cells in CD and fibroblast-like synoviocytes in RA (Al-Mossawi et al. 2017; Smith et al. 2008; Howell et al. 2018; Ai et al. 2016). Overall, achieving a detailed and precise understanding of complex diseases requires the study of sorted cell populations and when possible isolated from the affected tissue.

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 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 (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-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.

1.1.4 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 enabled to characterise 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 (Villani2017; Jaitin et al. 2014; Bengsch et al. 2018). In complex diseases such as RA, scRNA-seq has revealed heterogeneity in the synovial fibroblast population and identified a potentially pathogenic cluster highly proliferative and active in pro-inflammatory cytokine secretion (Mizoguchi et al. 2018). Similarly, mass cytometry analysis performed in RA

identified an expanded CD4⁺ 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.

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

As already explained in Chapters ?? 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.

In this lines, the PsA GWAS study conducted by Bowes and colleagues successfully performed fine-mapping for seventeen of the associated regions (**Bowes2016**). This study provided a summary table for the overlap of SNPs from the 90% credible set (list of SNPs explainin 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.

Traditional Bayesian fine-mapping models assume in the model only one causal SNP driving the association at each locus. As a way to partially address this limitation, genotype level fine-mapping 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). A disadvantage of the methods performing fine-mapping from summary statistics data is the impossibility to perform this conditional analysis.

1.1.6 Aims

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

1.2 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

Cross-tissue comparison analysis in PsA

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 TNAi biologic therapy xxxx.

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

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 (Ross et al. 2000; Jongbloed et al. 2006). 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 (Spadaro et al. 2004). Similarly, a tendency towards reduced proportions of B cells in SF compared to PB reinforced the

Cross-tissue comparison analysis in PsA

lack of 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.

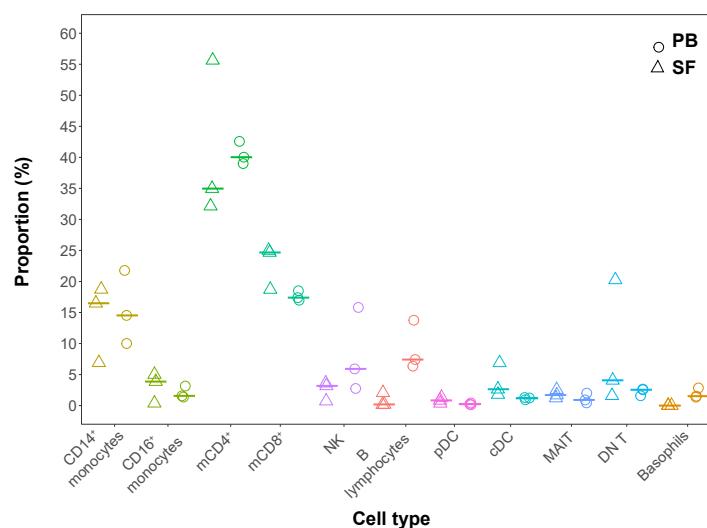


Figure 1.1: Comparative percentages of PB and SF immune cellular composition from the PsA cohort. Percentages of each of the twelve cell types identified by mass cytometry are shown by individual and tissue for PsA1718, PsA1719 and PsA1607. Horizontal line represents the median percentage for a particular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

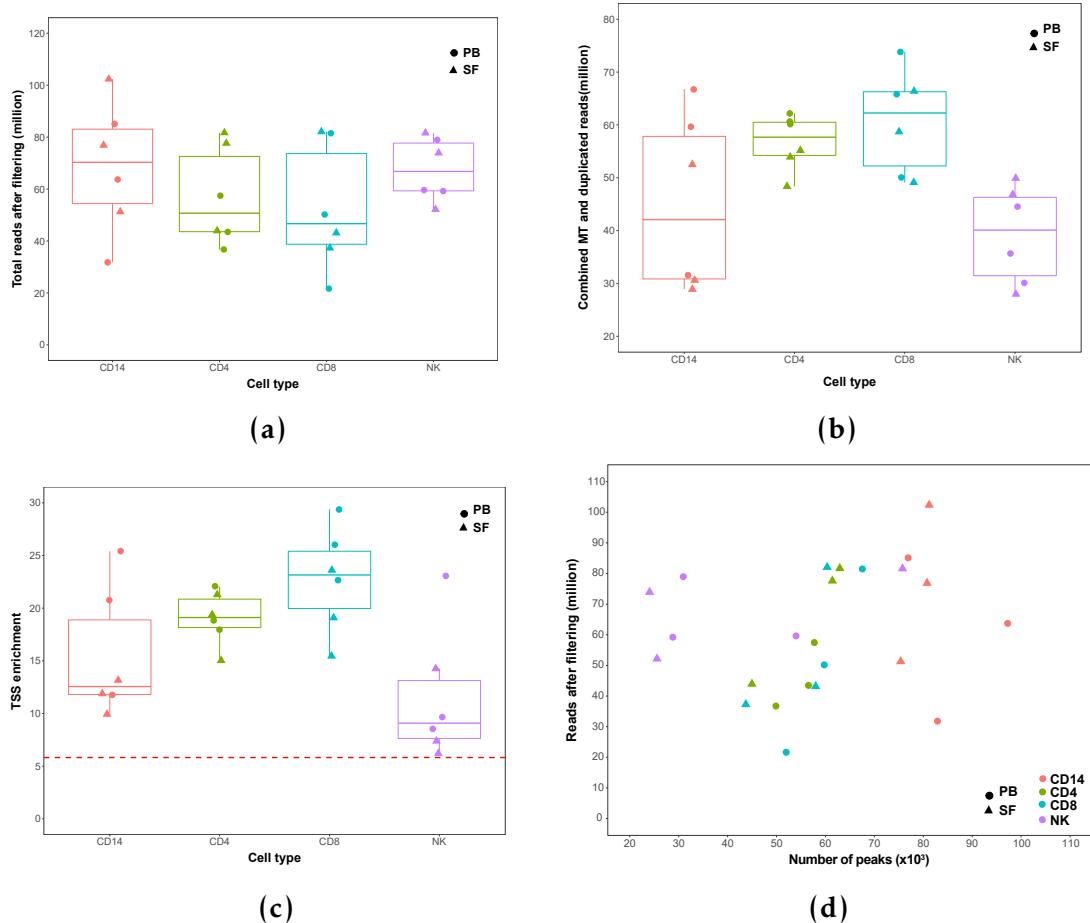


Figure 1.2: Quality control assessment of ATAC data generated in four immune cell types isolated from PB and SF of PsA patients samples. For each of the cell types boxplots representing a) million of reads after filtering, b) million of duplicated and MT reads combined and c) values for fold-enrichment 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 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 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).

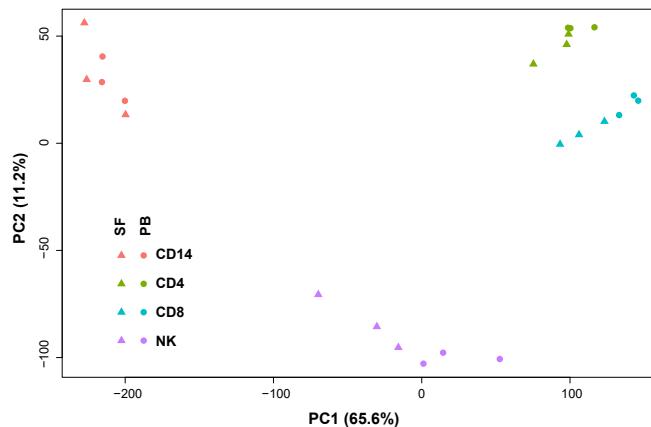


Figure 1.3: PCA analysis based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and SF. PCA analysis was performed using the normalised counts from the combined consensus master list (ML_all) across the four cell types (CD14⁺ monocytes, mCD4⁺, mCD8⁺ and NK cells) and two tissues (SF and PB) of interest. The first two PCs (x-axis and y-axis, respectively) of all 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 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 unstimulated, LPS 2h and IFN- γ 24h) followed by mCD8⁺ T cells (Figure 1.4 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.

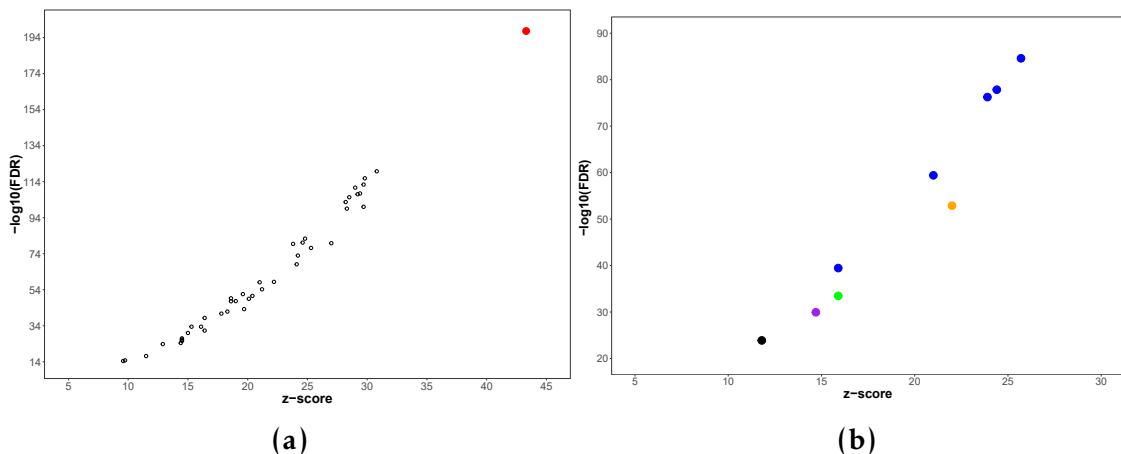


Figure 1.4: Enrichment of eQTLs publicly available data in the combined cell type and tissue chromatin accessibility master list for the PsA cohort. The dot plots showed the z-score values of the enrichment analysis in the x-axis and the significance ($-\log_{10}\text{FDR}$) in the y-axis for a) GTEx eQTL datasets and b) non-GTEx immune-related cell types including CD14⁺ monocytes (unstimulated, LPS 2h, LPS 24h and 24h IFN γ stimulated) 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

Cross-tissue comparison analysis in PsA

DESeq2 and a paired design (Table 1.3). 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 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. 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 ??).

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

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 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 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).

Cross-tissue comparison analysis in PsA

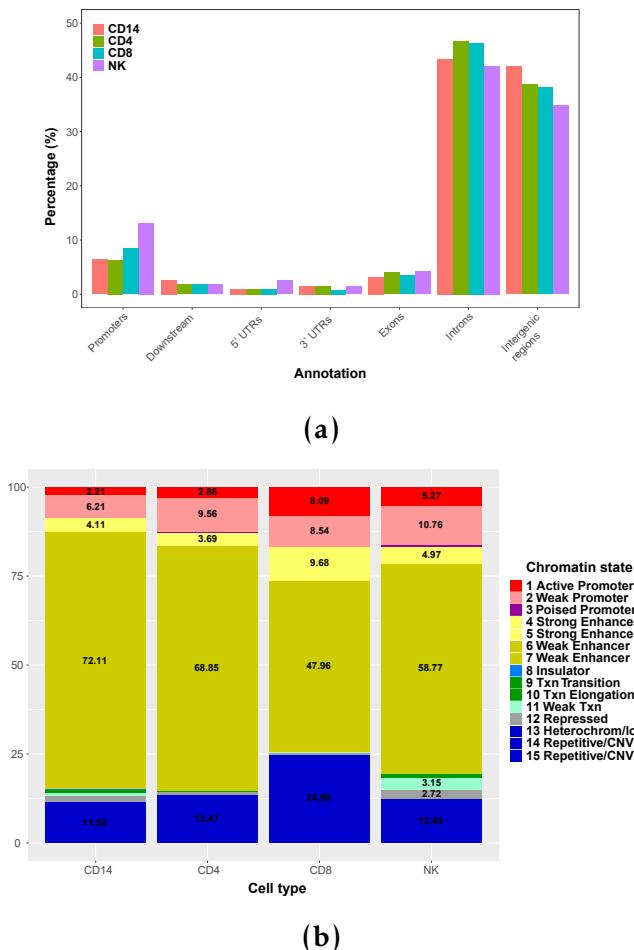


Figure 1.5: Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states. a) Barplot illustrating the percentage of nucleotides within DARs for each cell type that are annotated as promoters, downstream (regions at $\leq 1,000$ bp to a promoter), exons, introns, 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.

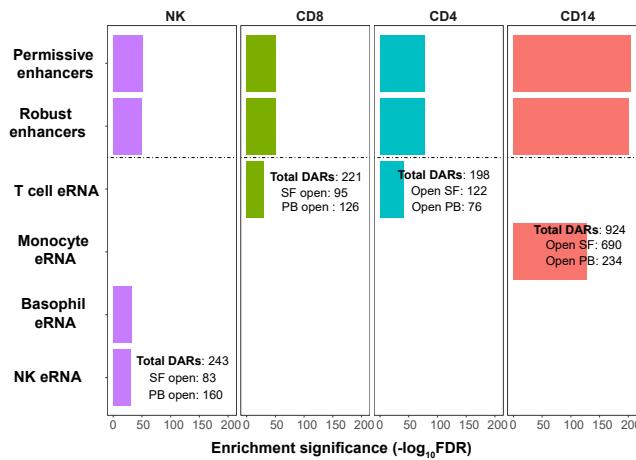


Figure 1.6: Enrichment of PsA DOCs for the FANTOM5 eRNA dataset. Robust enhancers have been defined as those detected at the genome-wide significant level in at least one primary cell type or tissue. Permissive enhancers are all detected eRNAs but not passing genome-wide filtering criteria (Andersson 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 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 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 1.4: 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 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).

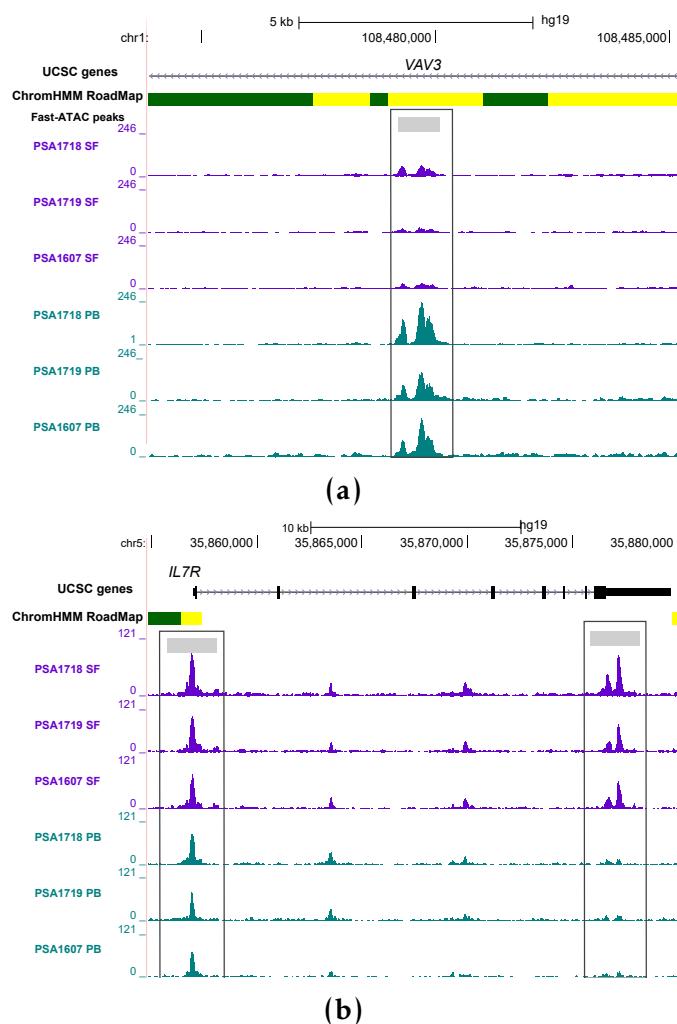


Figure 1.7: Differentially accessible regions located within gene bodies in CD14⁺ monocytes and NK cells from PsA patients. UCSC Genome Browser view illustrating the normalised ATAC read density (y-axis) in a) DAR located at an intron of *VAV3* gene (x-axis) in NK (less accessible in SF compared to PB) 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 abs(FC)>1.5.

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 (Dustin 2001). 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 (Boudousquié et al. 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

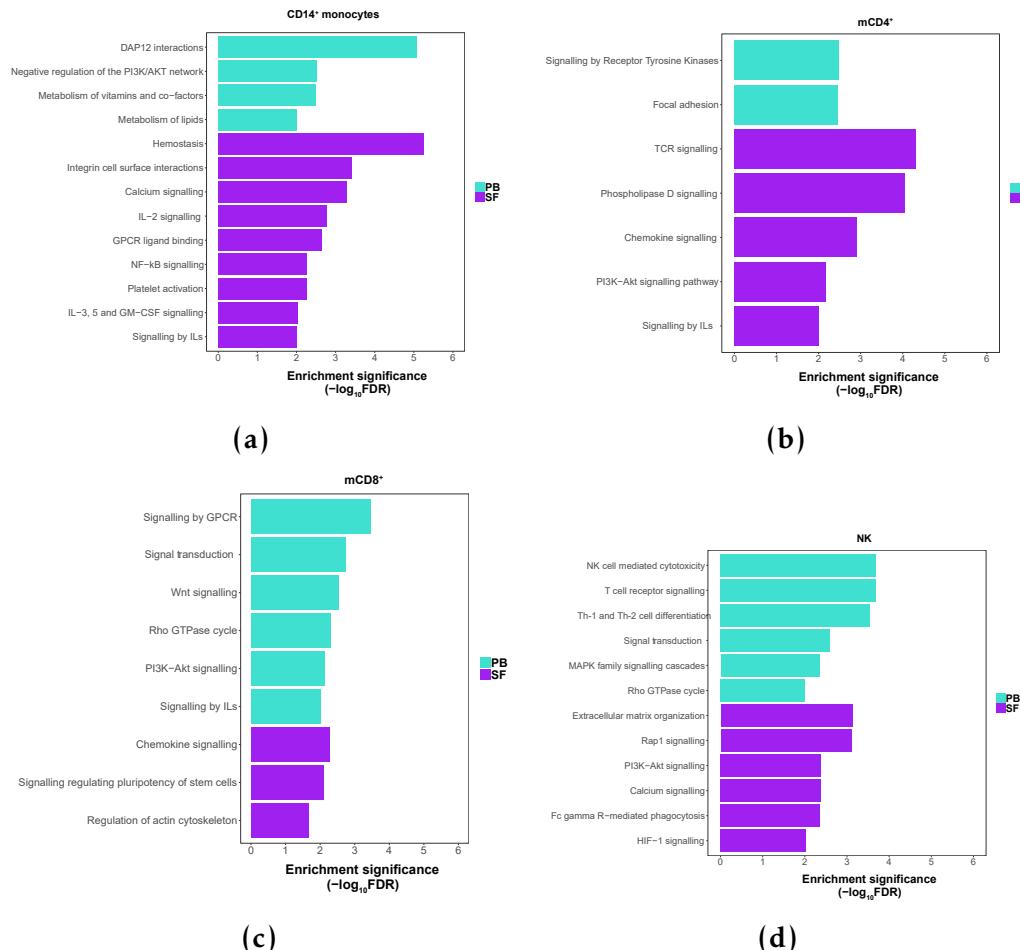


Figure 1.8: Distinct functional pathways enriched for DARs open in SF or open in PB in CD14⁺ monocytes, CD4m⁺, CD8m⁺ and NK. Enrichment analysis was performed separately for the DARs open in SF and DARs open in PB separately in a) CD14⁺ monocytes, b) mCD4⁺, c) mCD8⁺ and d) NK. The pathways presenting significant enrichment (FDR <0.01) only in open SF or open PB DARs have been shown here.

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

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 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<1.5) 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).

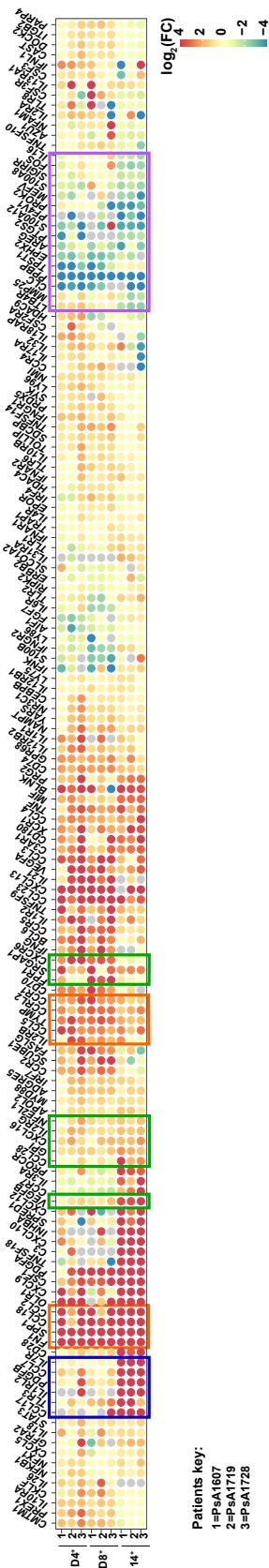


Figure 1.9: Heatmap of gene expression FCs between SF and PB for those gene significantly modulated ($pval<0.05$) in at least one cell type. Amongst the 381 genes measured by qPCR, the FC in gene expression between the SF and PB for each pair of samples has been represented only those genes which were 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 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).

Filtering of all the genes tested for expression in the qPCR array based statistical significance ($pval<0.05$) and $\text{abs}(\text{mean FC})>1.5$ revealed CD14⁺ monocytes and mCD8⁺ presenting greater number of significantly modulated genes (72 and 77, respectively) compared to mCD4⁺ cells (46 genes) (Figure 1.10 a, b and c). For the three analysed cell types, the majority of modulated immune genes showed 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 $\text{abs}(\text{mean FC})>1.5$ versus the 14 genes with mean FC<1.5 (Figure 1.10 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 1.5). 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 1.7 b). Another relevant example was the *FN1* gene, involved in cell adhesion, migration and osteoblast biology. Up-regulated expression in synovial biopsies compared to PB has already been reported by

Cross-tissue comparison analysis in PsA

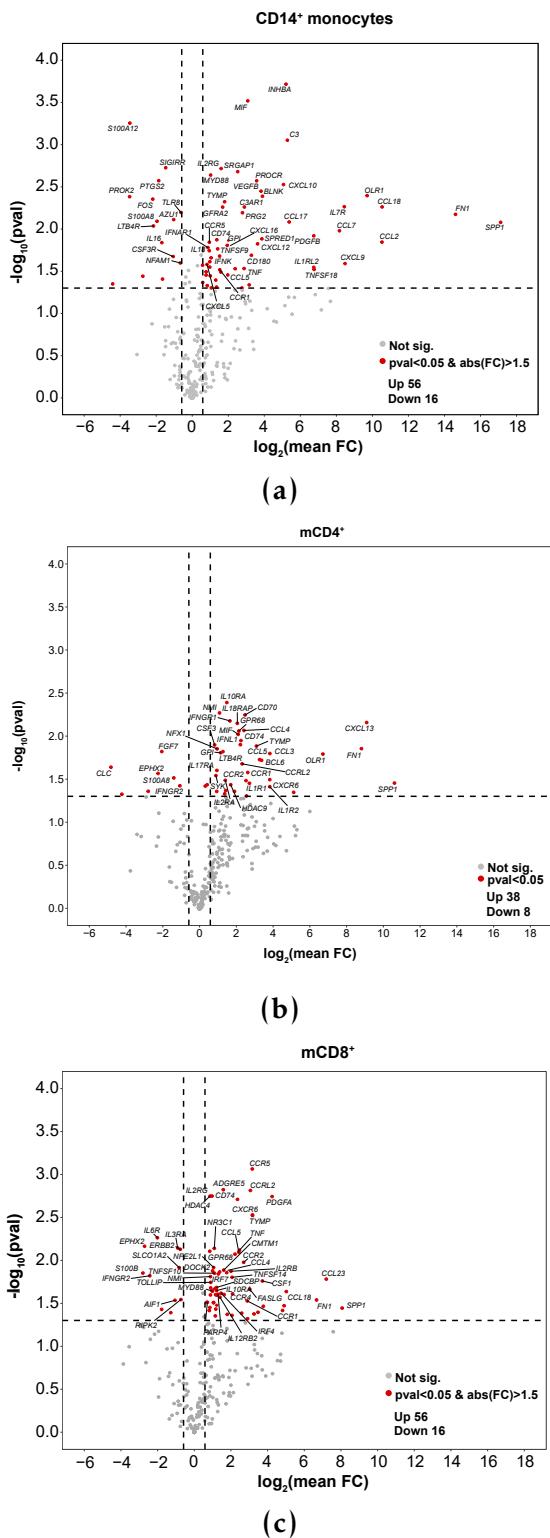


Figure 1.10: Gene expression changes in immune-relevant genes between SF and PB in CD14⁺ monocytes, mCD4⁺ and mCD8⁺ cells. Vulcano plots showing differences in gene expression measured by qPCR array between SF and PB for a) CD14⁺ monocytes, b) mCD4⁺ and c) mCD8⁺ cells. The significance ($\log_{10}pval$) 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 $pval < 0.05$ and $\text{abs}(\text{mean FC}) > 1.5$ 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*</i> , <i>CCR1*</i> , <i>CD180</i> , <i>CXCL10</i> , <i>FN1</i> , <i>IL18</i> , <i>IL31RA*</i> , <i>IL7R*</i> , <i>NFKB1*</i> , <i>PRG2</i> , <i>SRGAP1</i> , <i>STAT3</i>)	2 (<i>FOS</i> , <i>PROK2*</i>)
mCD4 ⁺	3 (<i>CXCL13</i> , <i>CXCR6*</i> , <i>IL2RA</i>)	0
mCD8 ⁺	6 (<i>CCL3</i> , <i>CCR2</i> , <i>CCR5</i> , <i>IRF4</i> , <i>TNFSF10</i> , <i>YARS</i>)	1 (<i>EPHX2</i>)

Table 1.5: 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).

others (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 downstream the 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). 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-

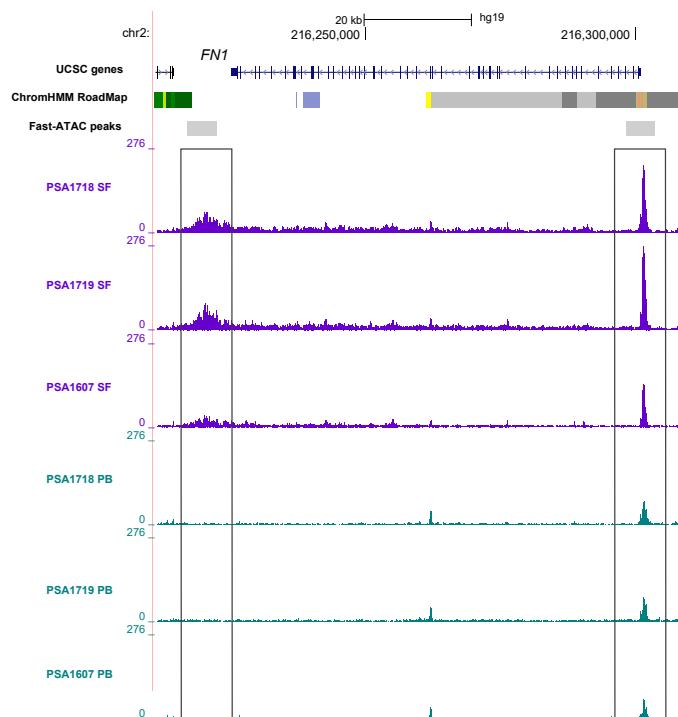


Figure 1.11: Chromatin accessibility landscape at the qPCR differentially expressed *FN1* gene in CD14⁺ monocytes. UCSC Genome Browser view illustrating the ATAC normalised read density (y-axis) in two DARs located at the promoter and downstream the 3' UTR of the *FN1* gene (x-axis) in CD14⁺ monocytes 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 abs(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 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 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

Cross-tissue comparison analysis in PsA

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 1.9 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).

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 abs(mean FC)>1.5. Reported enriched pathways were significant at an FDR <0.05.

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

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-Mossawi et al. 2018).

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.

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

Cross-tissue comparison analysis in PsA

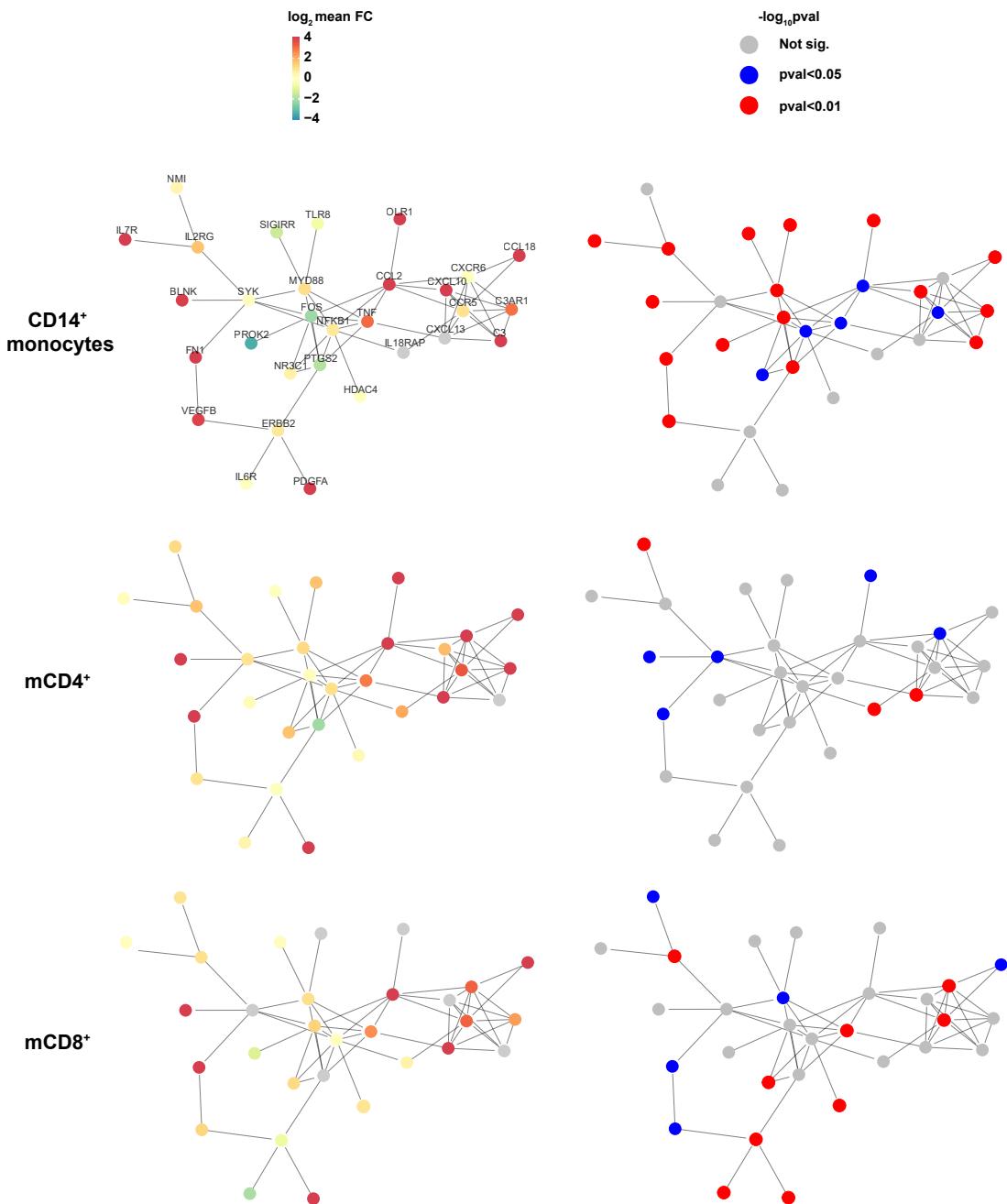


Figure 1.12: Protein network analysis based on the immune qPCR array expression data. The list of all the genes quantified in the qPCR array genes together with the best pval for significance of the mean FC across the three cell types was used to perform gene network analysis. STRING interaction network (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 ($p\text{val}<0.05$) 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

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).

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

Cross-tissue comparison analysis in PsA

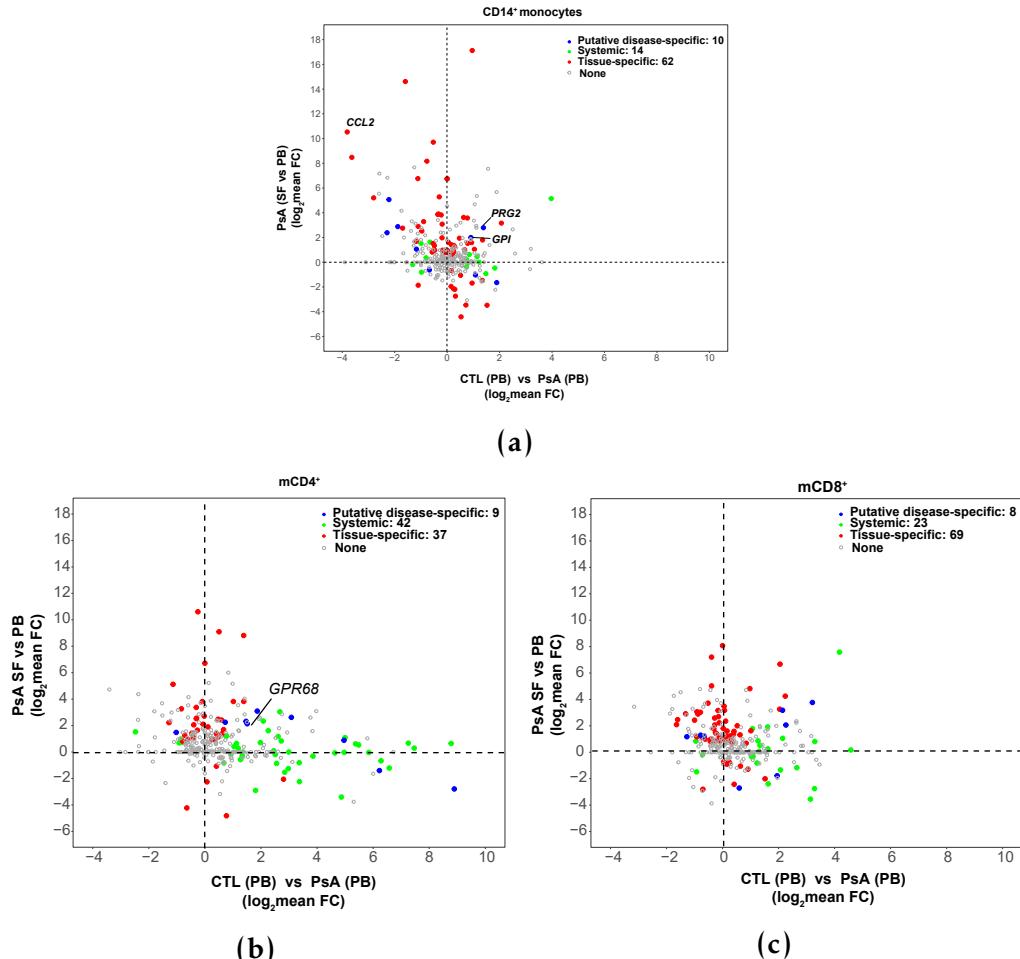


Figure 1.13: Comparison of immune-relevant gene expression modulation across PsA tissues (SF vs PB) and in PsA patients verus healthy controls. The qPCR \log_2 mean FC for each of the genes in the PsA SF vs PB contrast are plotted against the \log_2 mean FC for the same genes in the PsA PB vs healthy 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

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 (Biniecka et al. 2016). *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 behavior 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 (Zaiss et al. 2013), 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 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.

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 cell populations by expression of *CD14* and *LYZ*, two of the most accurate expression markers defining this cell population (Figure 1.14 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 1.1). 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

Cross-tissue comparison analysis in PsA

the six CD14⁺ monocytes populations was followed by conservative unsupervised clustering (using resolution 0.1) and t-SNE visualisation.

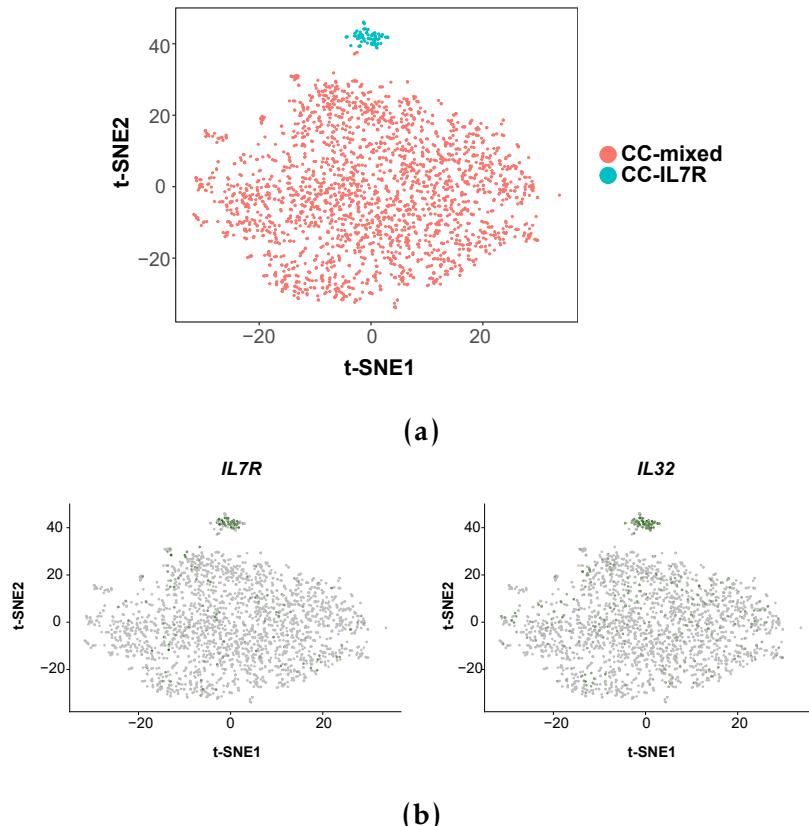


Figure 1.14: Identification of two main CD14⁺ monocytes subpopulations in the SF and PB combined analysis. a) Visualisation using t-SNE dimensional reduction of the two cluster (CC-mixed and CC-IL7R) identified in the combined SF and PB CD14⁺ monocyte cells using a very conservative resolution (res=0.1) 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 1.14 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 ??) ((Al-Mossawi et al. 2018), in revision). The proportion of IL7R⁺CD14⁺ monocytes when compared to the totalCD14⁺ monocyte population was very

Cross-tissue comparison analysis in PsA

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 was an heterogeneous cluster, without consistent expression pattern for those genes identified as cluster markers (Figure ??). 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 and the complexity in the definition and identification of stable clusters, increasing cohort size and a more exhaustive analysis could be used to identify additional subpopulations within the combined SF and PB CD14⁺ monocytes in the CC-mixed cluster. 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 abs(FC)>1.5 between SF and PB, of which 149 and 102 presented up- and down-regulation, respectively (Figure 1.15 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 1.15 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 abs(FC)>1.5. Out of the 37 DEGs in the CC-IL7R cluster between the

Cross-tissue comparison analysis in PsA

two tissues, 30 were also shared by the CC-mixed cluster. The seven distinctly modulated genes in the CC-IL7R cluster included *CD44*, *MT-CO2* or S-ribosomal protein (RPS) genes (*RPS29* and *RPS27*). *CD44* is a receptor of the hialuronic acid and osteopontin, the the protein product from the *SPP1* gene, which acts as an immune modulator increasing chemotaxis, cell activation and cytokine production. Although *SPP1* presents the greatest up-regulation in SF when compared to PB in the CC-mixed and CC-IL7R clusters (Figure 1.15 a and b), SF monocytes from the CC-IL7R subpopulation may be more responsive to this chemokine.

Cross-tissue comparison analysis in PsA

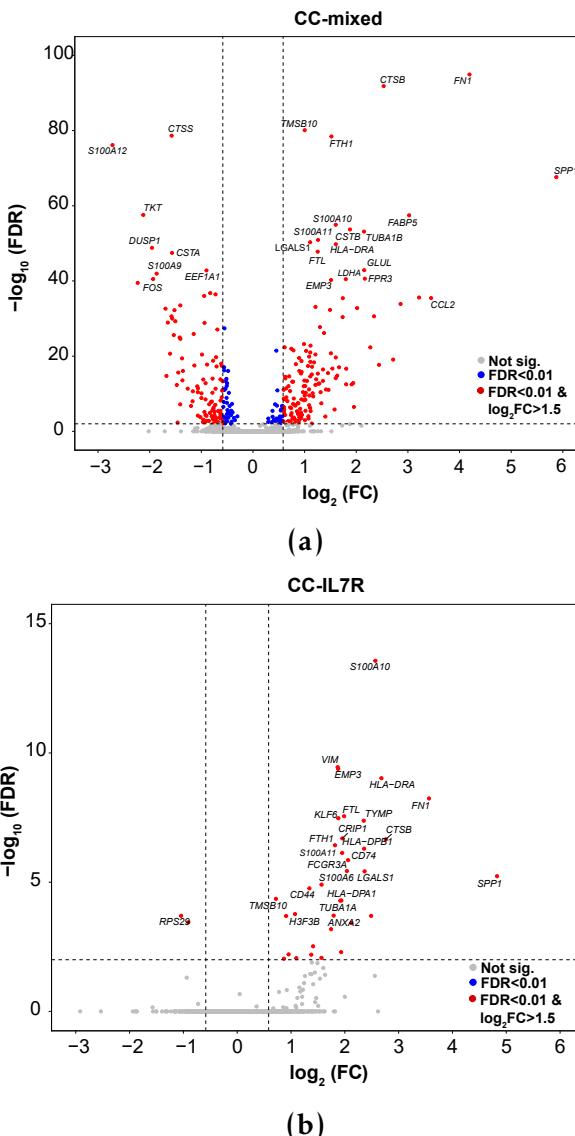


Figure 1.15: Sc-RNA-seq differential gene expression results between SF and PB in the CC-mixed and CC-IL7R CD14⁺ monocytes subpopulations. Vulcano plots showing differences in gene expression between SF and PB a) CC-mixed and b) CC-IL7R CD14⁺ monocyte cluster. In a and b, the significance ($\log_{10}\text{FDR}$) 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 and $\text{FDR}<0.01$ and $\text{abs}(\text{FC})>1.5$ are coloured in red. The most significant genes are labelled.

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 the Ag processing and presentation pathway, contributed by up-regulated expression in SF CD14⁺ monocytes of CD74 and genes from the HLA-D family (Lamb and Immunology 1992) (Table

Cross-tissue comparison analysis in PsA

1.7 and ??). Enrichment for IFN signalling was driven differential expression of genes such as *IFI6*, *IFITM3*, *LY6E*, *ISG15* and the TF *STAT1*, all of them up-regulated in SF when compared to PB CD14⁺ monocytes. Interestingly, *IFI6*, *IFITM3*, *LY6E* have also been identified as markers of a subpopulation of IFN- γ activated monocytes in from RA synovial tissue (Zhang et al. 2018). 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 involved in joint erosion and development of arthritis (Raghunatha 2012). Two genes of this family, *S100A8* and *S100A9* also appeared to be dysregulated in lesional skin compared to uninvolved in Chapter ?? and contributed to significant enrichment for the IL-17 signalling pathway. Interestingly, *S100A8* and *S100A9* were up-regulated in lesional skin whereas they have shown down-regulation in SF in comparison to PB. The phagosome 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. 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 1.7).

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 13 and 4 genes were also differentially expressed ($FDR<0.01$ 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*,

Cross-tissue comparison analysis in PsA

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.

FN1, *OLR1* and *S100A12*, being the direction of change also consistent for all of them. The limited overlap between qPCR and scRNA-seq DEGs could also partly explain the absence of overlap between enriched pathways across the two gene expression analysis.

Cross-tissue comparison analysis in PsA

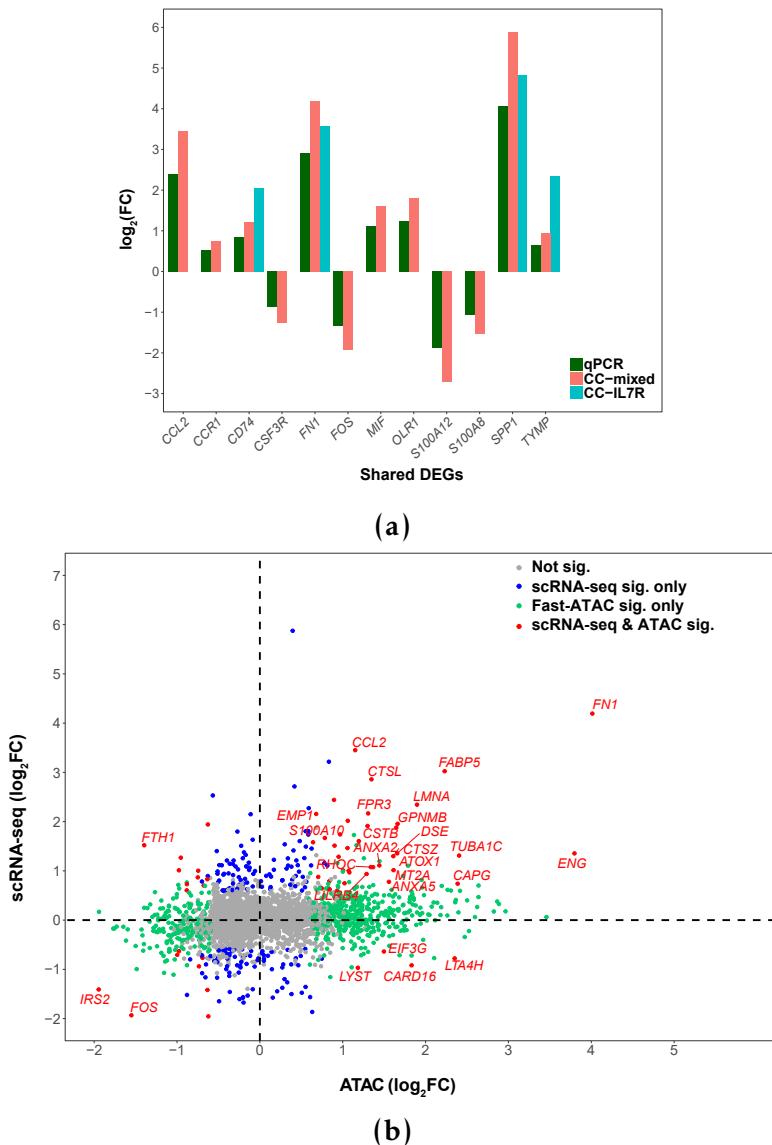


Figure 1.16: Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14⁺ monocytes. a) Overlap between the scRNA-seq DEGs (significant based only on FDR<0.01) from the CC-mixed and CC-IL7R CD14⁺ monocyte cluster in SF versus PB and the corresponding DEGs ($pval<0.05$) detected by qPCR in the bulk CD14⁺ monocytes population. For each of the shared DEGs, the \log_2FCs 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_2FCs for scRNA-seq differential expression of all transcripts in the CC-mixed CD14⁺ monocytes are plotted against the \log_2FC 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 (FDR<0.01 and $\text{abs}(FC)>1.5$) in both assays.

Moderate genome-wide correlation 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 \log_2 FCs for all the expressed genes in the CC-mixed CD14⁺ monocytes cluster and all the accessible chromatin regions in total CD14⁺ monocytes between SF and PB was conducted (Figure 1.16 b). Changes in expression and chromatin accessibility only presented a moderate correlation in this data ($R=0.214$, $pval=2\times 10^{-16}$). In the CC-mixed cluster, 64 genes out of the 251 DEGs ($FDR<0.01$ and $\text{abs}(FC)>1.5$) were proximal ($\leq 5\text{Kb}$) to one or more ATAC DARs (Table 1.8). This overlap represented an enrichment (Fisher exact test $pval=1.5\times 10^{-3}$) of CC-mixed cluster DEGs for genes in the proximity of at least one DARs in total CD14⁺ monocytes. The majority of the overlaps corresponded to matched increase or decrease (40 and 12 genes, respectively) of gene expression and chromatin accessibility when comparing SF vs PB. However, 12 DEGs in the CC-mixed cluster showed opposite dysregulation of expression and chromatin accessibility (Table 1.8).

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.

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

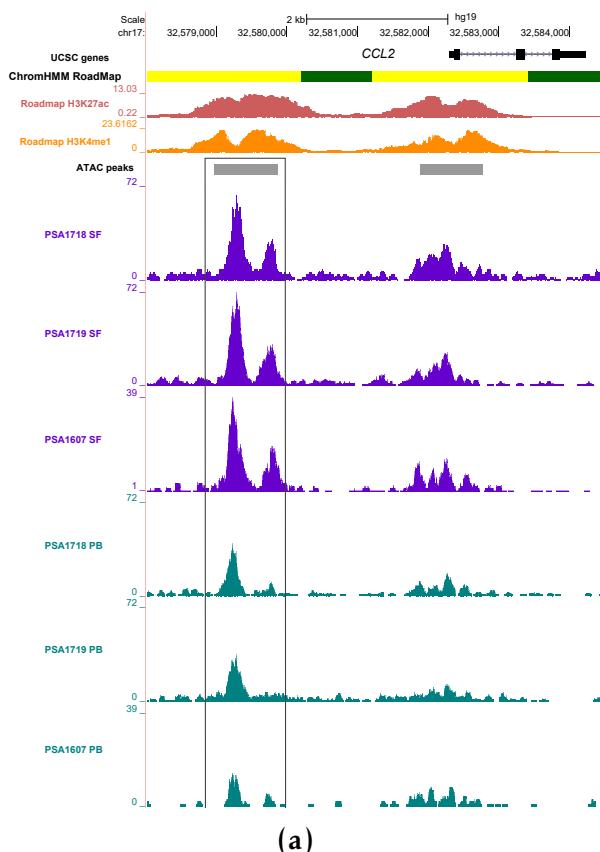
Table 1.8: scRNA-seq DEGs in SF versus PB CD14⁺ monocytes proximal to an ATAC DAR.. For each of the two CD14⁺ monocytes cluster identified by scRNA-seq analysis, an overlap is defined when a gene is differentially expressed ($FDR<0.01$ and $\text{abs}(FC)>1.5$) between SF and PB and a proximal significant DAR ($\leq 5\text{Kb}$) showing same or opposite direction of change is also found.

Cross-tissue comparison analysis in PsA

An example of a relevant DEGs in the CC-mixed cluster overlapping a proximal DAR was *CCL2* (Figure 1.16 b). The expression 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 the same data (Figure 1.10 a, b, c and Table 1.5). In contrast, *CCR2* which is the receptor for the chemokine MCP-1 (protein product of *CCL2*) appeared up-regulated in mCD4 $^{+}$ and mCD8 $^{+}$ cells in the same individuals, likely highlighting increased chemotaxis driven by CD14 $^{+}$ monocytes and leading to T cell infiltration in the synovium. Consistently with the up-regulated expression of *CCL2* in SF versus PB, a cell type-specific SF open DAR upstream *CCL2* has also been identified by ATAC in this data (Figure 1.17). 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. However, no eQTL or Hi-C interaction in CD14 $^{+}$ monocytes have been reported for SNPs overlapping this DAR, lacking of additional data to determine the relationship between differential expression of *CCL2* and changes in chromatin accessibility in this nearby locus.

In terms of the pathological relevance, *CCL2* did not appear up-regulated in PB CD14 $^{+}$ monocytes when comparing PsA patients and healthy controls, being defined as one of the putative disease-specific genes in the previous analysis (Figure ?? a). Similarly, Dolcino and colleagues did not report *CCL2* dysregulation neither in PBMCs when comparing PsA patients and healthy controls nor when comparing synovial membrane biopsies from PsA patients versus controls (**Dolcino2015**). Moreover, collaborators in University of XX have performed measurements of cytokine levels in PB and SF from ten PsA patients. High levels of MCP-1 (approximately 1,000 pg/mL) were reported in SF whereas this cytokine failed to be detected in plasma, supporting the up-regulation of *CCL2* expression also at the protein level (Figure ??). Overall, these observations

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.



(a)

Figure 1.17: 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 abs(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.

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 $pval=1.85\times10^{-9}$). 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 1.8). The CC-IL7R cluster and CD44 data not shown.

1.2.7 Mass cytometry reveals active cytokine production in CD14⁺ monocytes.

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⁺ and mCD8⁺ cells was quantified in SF and PB following incubation with BFA for 6h. BFA blocks cytokine release and enables to measure the cell cytokine production rate in absence of any inflammatory stimuli, as previously explained in Chapter ??.

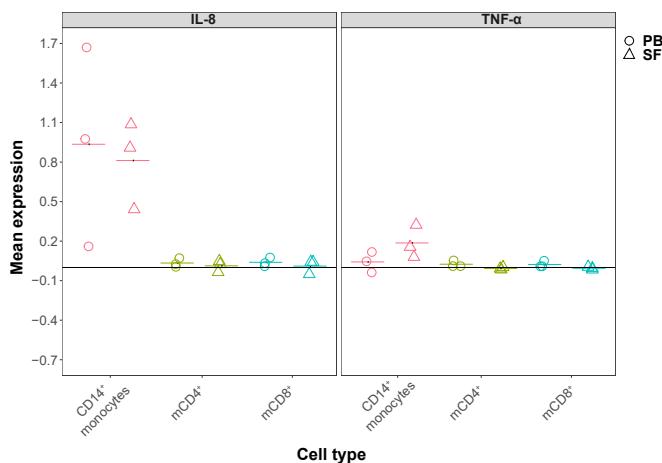


Figure 1.18: Mean expression of IL-8 and TNF- α in SF and PB from three PsA patients quantified by mass cytometry. For each of three PsA patients, mean levels of IL-8 and TNF- α expression detected by mass cytometry were calculated for the three cell populations of interest from matched SF and PB. SFMCs and PBMCs were incubated with BFA for 6h prior to identification of cell populations and quantification of cytokine expression by mass cytometry. CD14⁺ monocytes, mCD4⁺ and mCD8⁺ populations were defined by manual gating based expression of appropriate cell surface markers.

Mass cytometry staining for IL-8 and TNF- α cytokines were the two showing the best QC and therefore the most reliable quantification within each of the manually gated cell populations. TNF- α is released by all four cell types and IL-8 is secreted by monocytes, CD4⁺ T cells and some subsets of CD8⁺ T cells,

Cross-tissue comparison analysis in PsA

making these two cytokines a representative panel for the three cell types with ATAC and qPCR array data four cell types. CD14⁺ monocytes isolated from SF and/or PB were the only cell type with active production of IL-8 and also TNF- α , to a lesser extent (Figure ??). Measured IL-8 mean expression did not appear to be different between SF and PB CD14⁺ monocytes. In contrast, CD14⁺ monocytes isolated from SF presented detectable mean levels of TNF- α , which not detected for the matched cells in PB.

The increased rate of TNF- α production by SF CD14⁺ monocytes further investigated by comparing the proportion of TNF- α positive cells after 6h of blocked cytokine secretion in each of the two tissues. Mass cytometry expression of CD14⁺ versus TNF- α demonstrated greater number of SF CD14⁺ monocytes producing TNF α after 6h of BFA treatment compared to PB in all three PsA patients (Figure 1.19 a). The basal (0h) percentage of CD14⁺ monocytes positive for TNF- α was negligible for all three patients in both, SF and PB (Figure 1.19 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.

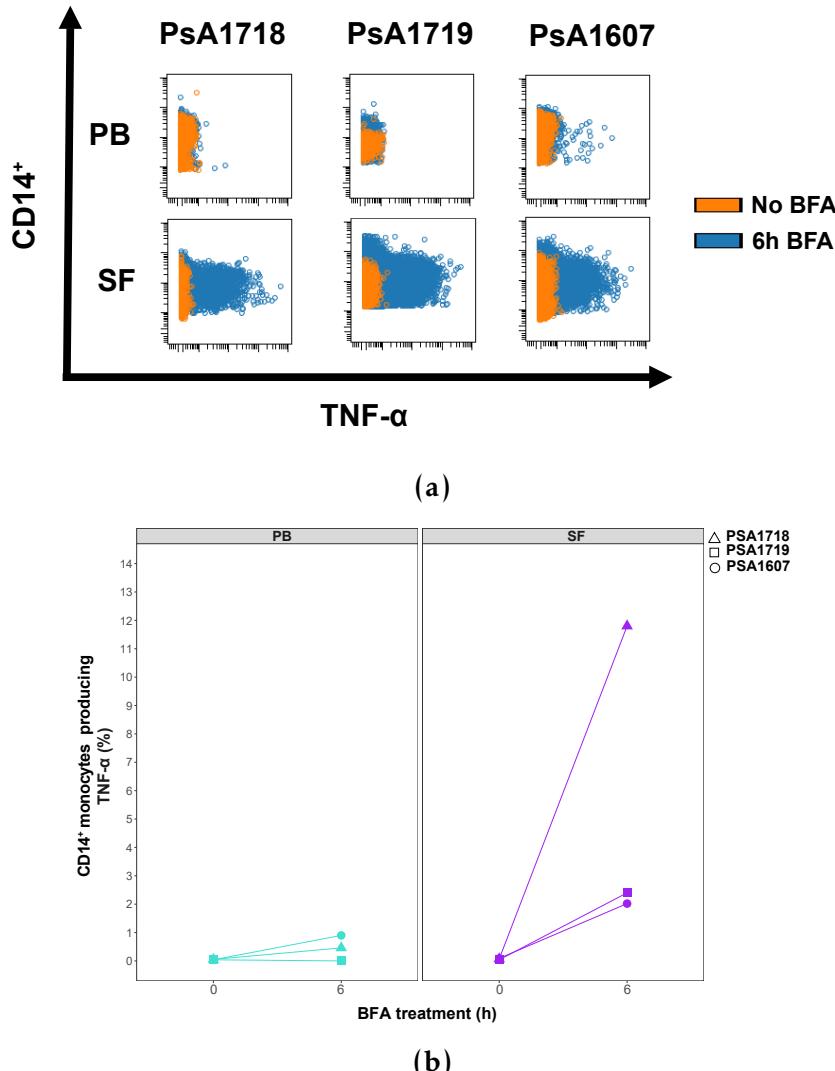


Figure 1.19: Comparison of $TNF-\alpha$ expression by SF and PB $CD14^+$ monocytes before and after protein transport blockade with BFA using mass cytometry. a) For each of the three PsA patients, representation of $CD14^+$ (y-axis) versus $TNF-\alpha$ (x-axis) intensity of expression in matched SFMCs and PBMCs without protein transport blockade (blue dots) or after 6h treatment with BFA (orange dots). b) Quantification of the proportion of $CD14^+$ monocytes (percentage) from SFMCs or PBMCs in each PsA patient with detectable $TNF-\alpha$ production levels in absence or presence of BFA treatment in a.

1.2.8 Prioritisation and interpretation of PsA GWAS SNPs

The generation of epigenetics 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, mapping of the epigenetic landscape in different cell types

Cross-tissue comparison analysis in PsA

isolated from clinical samples and overlaying it with GWAS associations fine-mapped credible set of SNPs may add power to refine the number of putative functional relevant that could be causal of particular associations, when compare to integration of epigenetic data from cell lines or healthy controls.

Bayesian fine-mapping using genotype 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 (PsA Immunochip UK cohort from the Bowes *et al.*, 2015). Bowes and colleagues performed fine-mapping for some of the loci and the number of independent signals for each locus as well as the number of SNPs in each 90% credible were provided in the supplementary material (Bowes et al. 2015). Nevertheless *de novo* fine-mapping was performed here to obtain the individual identity of each of the SNPs in the credible set and conduct further integration with the in-house chromatin accessibility data generated in SF and PB. Compared to the fine-mapping performed in Bowes *et al.*, 2015, the power of the analysis here presented will be limited by a smaller sample size (only UK cohort as previously mentioned) and could be improved by incorporation of additional samples or extended collaboration with Dr Anne Barton group.

Initially, fine-mapping was performed in thirty-six of reported loci in Bowes *et al.*, all of them showing at least nominal significance in their GWAS study. As expected, Bowes study's GWAS signals with the lowest significance for the lead SNP ($p\text{val} < 10^{-4}$) also showed $-\log_{10}\text{ABF}$ under 3 (cut off used in the analysis following (Bunt et al. 2015)) for the fine-mapping lead SNP in the association analysis and were not taken forward for the calculation of PP and credible set of SNPs.

Table 1.9: Summary table of the PsA GWAS loci presenting $-\log_{10}ABF > 3$ for the fine-mapping lead SNP. For twelve PsA loci $-\log_{10}ABF$ of the fine-mapping lead SNP was 3 or greater. In four of those loci (*) the fine-mapping lead SNP was in low LD ($r^2 < 0.5$) with the PsA GWAS SNP, indicating spurious signals identified by the association analysis. MAD= minor allele frequency; OR=odds ratio; ABF=approximate Bayes factor; PP=posterior probability; FM=fine-mapping

chr	Gene	GWAS lead SNP	MAF	OR	$-\log_{10}ABF$	PP	90% credible set	lead SNP	Bowes FM	Bowes 90% credible set
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</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

In this analysis, fine-mapping was also unsuccessful ($-\log_{10}ABF < 3$) in four loci previously fine-mapped by Bowes *et al.* (*B3GNT2*, *NOS2A*, *REL* and *TNIP1*) highlighting the already acknowledged reduced power of the analysis performed here. Amongst the twelve loci with fine-mapping lead SNP in the association analysis passing the $-\log_{10}ABF \geq 3$ cut-off, eight have also been successfully fine-mapped by Bowes and colleagues (Table 1.9). For two of the loci, *IL23R* and *TRAF3IP2*, the same fine-mapping lead SNPs were reported by the two analysis. Regarding *IL12B* locus, the second signal reported by Bowes and colleagues was also identified in this analysis by the step-wise conditional analysis performed within the fine-mapping. For the four additional fine-mapped loci reported here but not by Bowes *et al.*, the fine-mapping lead SNP were in very low LD ($r^2 < 0.5$) with the PsA GWAS lead SNP, suggesting identification of spurious signals or signals from other locus nearby in the association analysis (Table 1.9 labelled with *). For example, the association analysis performed for the fine-mapping of *IL13* was confounded by the *TYK2* signal, with no LD found between the fine-mapping and the GWAS index SNPs in the *IL13* locus. Therefore, these four signals were also removed from downstream integration analysis.

Integrating fine-mapped SNPs and chromatin accessibility from PsA samples

The union of the 90% credible sets for the eight successfully fine-mapped loci represented a total of 294 unique SNPs. These SNPs were used to perform overlap with the significant ($FDR < 0.01$ and $\text{abs}(FC) > 1.5$) DARs identified between SF and PB in $CD14^+$ monocytes, $mCD4^+$, $mCD8^+$ and NK cells. Unfortunately, none of the 294 SNPs were found to be contained by a DAR in any of these cell types. Additional overlap was performed between the SNPs from the credible set and the accessible chromatin regions (consensus peaks in each cell type master list without the additional filtering performed for the differential analysis) in each of the four cell types assayed by ATAC. The

Cross-tissue comparison analysis in PsA

largest number of SNPs (17) was found to overlap accessible chromatin in CD14⁺ monocytes, followed by mCD8⁺, mCD4⁺ and NK cells (Table 1.10). Overall, the 46 unique SNPs from the 90% credible set overlapping ATAC accessible chromatin were distributed across *CSF2* (11), *IL12B*(3), *IL23R*(4), *RUNX3*(6), *STAT2*(14), *TRAF3IP2*(1) and *TYK2*(7) loci. A number of those SNPs were found to only overlap accessible chromatin at one particular cell type. However, none of the locus credible set of SNPs appeared to be completely cell type specific. Moreover, no overlap was found for *TRAF3IP2* credible set of SNPs with accessible chromatin for any of the four cell types here included.

ATAC cell type master list	90% credible set overlapping SNPs (number)	Cell type specific overlap
CD14 ⁺ monocytes	32	<i>STAT2</i> (5), <i>CSF2</i> (1) <i>TYK2</i> (2), <i>RUNX3</i> (1), <i>TRAF3IP2</i> (1) mCD4 ⁺
29 mCD8 ⁺	27 <i>CSF2</i> (2), <i>IL23R</i> (1)	<i>RUNX3</i> (1)
NK	18	<i>TYK2</i> (1)

Table 1.10: 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, from the overlapping SNPs those ones only found to overlap open chromatin in one cell type are indicated together with the locus where the SNP was fine-mapped.

The overlap found for SNPs in the 90% credible set of the eight fine-mapped loci with ATAC peaks across the four cell types was compared to the corresponding overlap found for GWAS Catalogue SNPs and SNPs in LD ($r^2 \geq 0.8$). SNPs from the fine-mapping credible set presented a significant enrichment for ATAC peaks in all four cell types when compared to all GWAS Catalogue SNPs (Fisher exact test: CD14⁺ monocytes pval=7.12x10⁻⁸, mCD4⁺ pval=1.69x10⁻¹⁰, mCD8⁺ pval=6.40x10⁻⁹ and NK pval=1.86x10⁻⁵). Notably, the GWAS Catalogue SNPs overlapping ATAC accessible regions were significantly

Cross-tissue comparison analysis in PsA

enriched ($FDR < 0.001$) for sets of SNPs annotated to terms from the Experimental Factor Ontology (EFO) (Figure 1.20). The EFO is a hierarchical tree-like ontology where each term represents a set of disease risk SNPs annotated to a disease trait or group of related disease traits a trait or group of related disease traits (Figure 1.20). 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 the previous general terms, such as CD and IBD, were also found to be enriched for all 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 also found for the overlap between GWAS Catalogue genetic variants not included in this analysis fine-mapping credible set with the ATAC accessible chromatin across these four immune cell types.

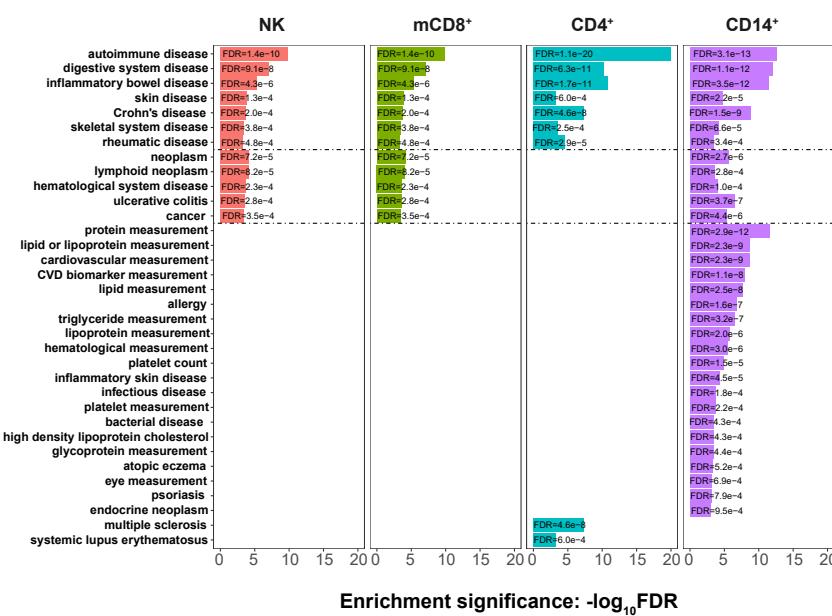


Figure 1.20: Experimental Ontology Factor terms enriched for GWAS Catalogue SNPs overlapping ATAC regions in four cell types. Each term represents a set of risk SNPs annotated to a disease trait or a group of related disease traits. Enrichment analysis was performed using as input data the GWAS Catalogue SNPs overlapping ATAC accessible chromatin regions. A minimum of ten SNPs overlap and $FDR < 0.001$ was required for enrichment to be considered significant.

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-specific compared to psoriasis GWAS. Amongst the 58 SNPs in the 90% credible set were included the four SNPs (rs10065787, rs3846728, rs27437 and rs7721882) highlighted by Bowes and colleagues as the most functionally relevant. Out of the 58 SNPs, only 11 overlapped ATAC accessible chromatin (ATAC peaks from the different cell types master lists) in at least one of the four cell types included in this study. Despite the overlap with ENCODE epigenetic features indicated by Bowes and colleagues, rs3846728 and rs7721882 did not overlap accessible chromatin in any of the four cell types here studied, either in SF or PB form PsA patients (Table 1.11).

Rs10065787, which overlaps ENCODE clusters of occupancy for TFs relevant in CD4⁺ and CD8⁺ biology, presented accessible chromatin in CD14⁺ monocytes and mCD4⁺ cells but no eQTLs for tCD4⁺ and tCD8⁺ cells were reported by the Kasela and colleagues dataset (Table 1.11 and Figure 1.22 right hand side panel red line).

In contrast, rs10065787 was found to be a *cis*-eQTL in stimulated monocytes (LPS 2h, LPS 24h and IFN- γ) regulating *P4HA2* expression (Table 1.11). The nearby SNP rs27437 overlapped an ATAC peak in CD14⁺ monocytes and mCD4⁺ and the same TFBS site cluster than rs10065787 (Table 1.11 and Figure 1.22 right hand side panel green line). The enhancer histone mark H3K4me1 was absent in CD14⁺ monocytes and only moderate in mCD4⁺ (Figure 1.22 right hand side panel). Interestingly, *cis*-eQTL were only reported for tCD4⁺ and tCD8⁺ associated to modulation of expression for *SLC22A5*, the same eGene reported by Bowes and colleagues in their pilot eQTL study. Nevertheless, chromatin conformation data has not revealed any interaction for rs10065787 with the

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 ⁺)
rs1469149	mCD4 ⁺	<i>P4HA2</i> (monocytes LPS2, LPS24) <i>SLC22A5</i> (monocytes UT, LPS2, IFN- γ , CD4 ^{+(*)})
rs2089855	CD14 ⁺	<i>P4HA2</i> (monocytes LPS2, LPS24, IFN- γ), <i>SLC22A5</i> (monocytes UT, IFN- γ , CD4 ^{+(*)} , CD8 ^{+(*)})
rs3846728	mCD4 ⁺	<i>SLC22A5</i> (CD4 ^{+(*)} , CD8 ⁺)
rs743564	mCD4 ⁺ , mCD8 ⁺	<i>SLC22A5</i> (CD4 ^{+(*)} , CD8 ^{+(*)})
rs7721882	mCD4 ⁺	<i>SLC22A5</i> (CD4 ^{+(*)} , CD8 ⁺)

Table 1.11: Publicly available *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 (Naranbhai2015), neutrophils untreated (unpublished), tCD4⁺ and tCD8⁺ (Kasela2017) and whole blood (Jansen et al. 2017).

Cross-tissue comparison analysis in PsA

promoter of *P4HA2*. Conversely, rs27437 is relatively close to the bait in *IL3* which interacts with *SLC22A5*, potentially bringing this SNP in proximity with the promoter of the eQTL eGene reported in tCD4⁺ and tCD8⁺ Kasela's dataset.

Another two relevant SNPs from the 90% credible set here reported were rs2069803 and rs4705908, both overlapping ATAC peaks in all four cell types (Figure 1.22 left hand side panel black and brown lines, respectively). Rs4705908 is located upstream the promoter of *ACSL6* gene and is overlapping a region enriched for H3K4me1, supporting a regulatory role for that region. Notably, rs4705908 is mapping to a CTCF binding site reported in GM12878 and LCLs cell lines. Likewise, rs2069803 is also overlapping open chromatin and moderate H3K4me1 signal in mCD4⁺, mCD8⁺ and NK cells (Figure 1.22 left hand side panel). The region has also been annotated as enhancer and weakly transcribed in mCD4⁺, mCD8⁺ and NK cells by the Epigenome Roadmap chromatin segmentation maps (yellow and light green in 1.22 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 1.11). 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 appeared also within the bait of the *ACSL6* promoter interacting with the *IL3* promoter bait, which also includes the rs2069803 as previously mentioned (Javierre2016). Overall, promoter-capture HiC data revealed potential physical interaction between rs27437, rs2069803 and rs4705908 in CD4⁺ and CD8⁸ cells with potential functional relevance in gene expression regulation of *SLC22A5*.

Regarding the PsA GWAS lead SNP (rs715285) and the SNP reported to be associated with *SLC22A5* expression (rs11955347) by Bowes and colleagues, both

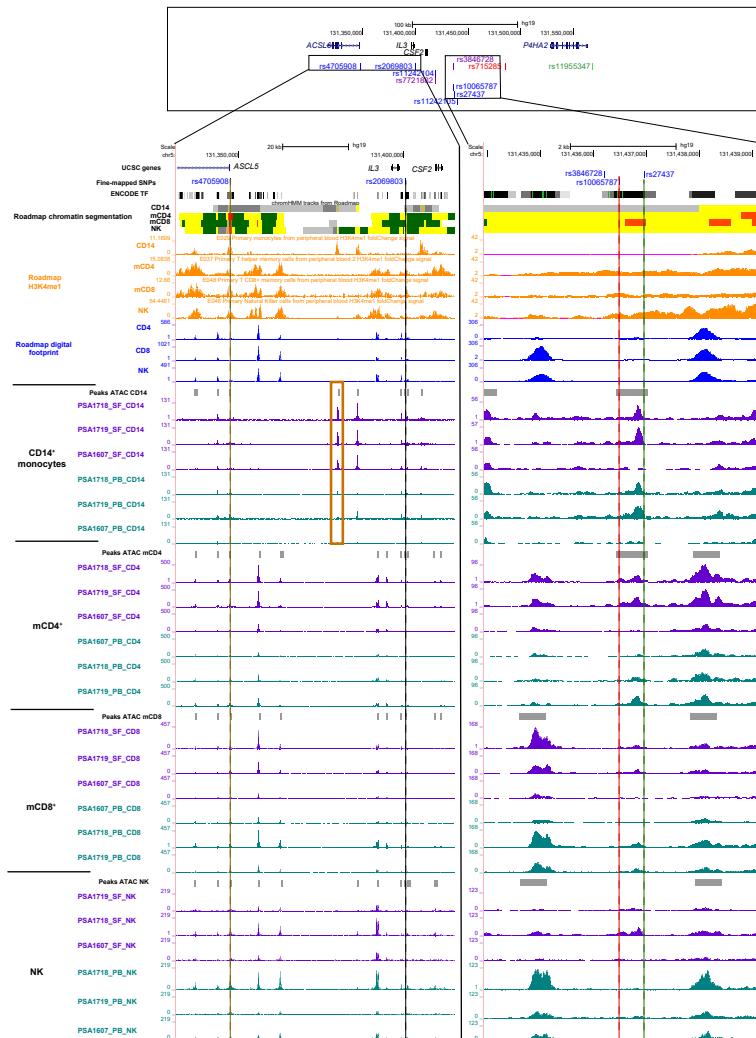


Figure 1.21: Epigenetic landscape at the genomic location of fine-mapped SNPs for the 5q31 PsA GWAS signal. The top panel is a schema of the genomic location for the six SNPs in the 5q31 fine-mapping 90% credible set overlapping PsA ATAC accessible regions in at least one of the four cell types (in blue). 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 two SNPs from the credible set overlapping several ENCODE annotation features with no overlap with the in-house PsA ATAC data (in purple). The left and right hand side panels are a zoom in onto the a 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 (H3Kme1, chromatin segmentation maps, digital footprint and ENCODE cluster for TF binding) generated in the same cell types as the in-house ATAC has also been included. The yellow box highlights a DARS found in CD14⁺ monocytes nearby rs2069803 and rs4705908.

Cross-tissue comparison analysis in PsA

are found to overlap closed chromatin and depleted signals for H3K4me1 and digital footprint in any of the four cell types. Rs11955347 is very significantly associated with expression of *SLC22A5* in CD4⁺ and CD8⁺ from Kasela *et al.* and also shows association with *P4HA2* expression in UT and stimulated monocytes in all three conditions to a lesser level of significance compare to the aforementioned signal. However, the lack of epigenetic marks mapping to rs11955347 location may suggest that one or more of the previously highlighted SNPs from the 90% credible set may be responsible for the effect on gene expression regulation of *P4HA2* or *SLC22A5*.

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.

Cross-tissue comparison analysis in PsA

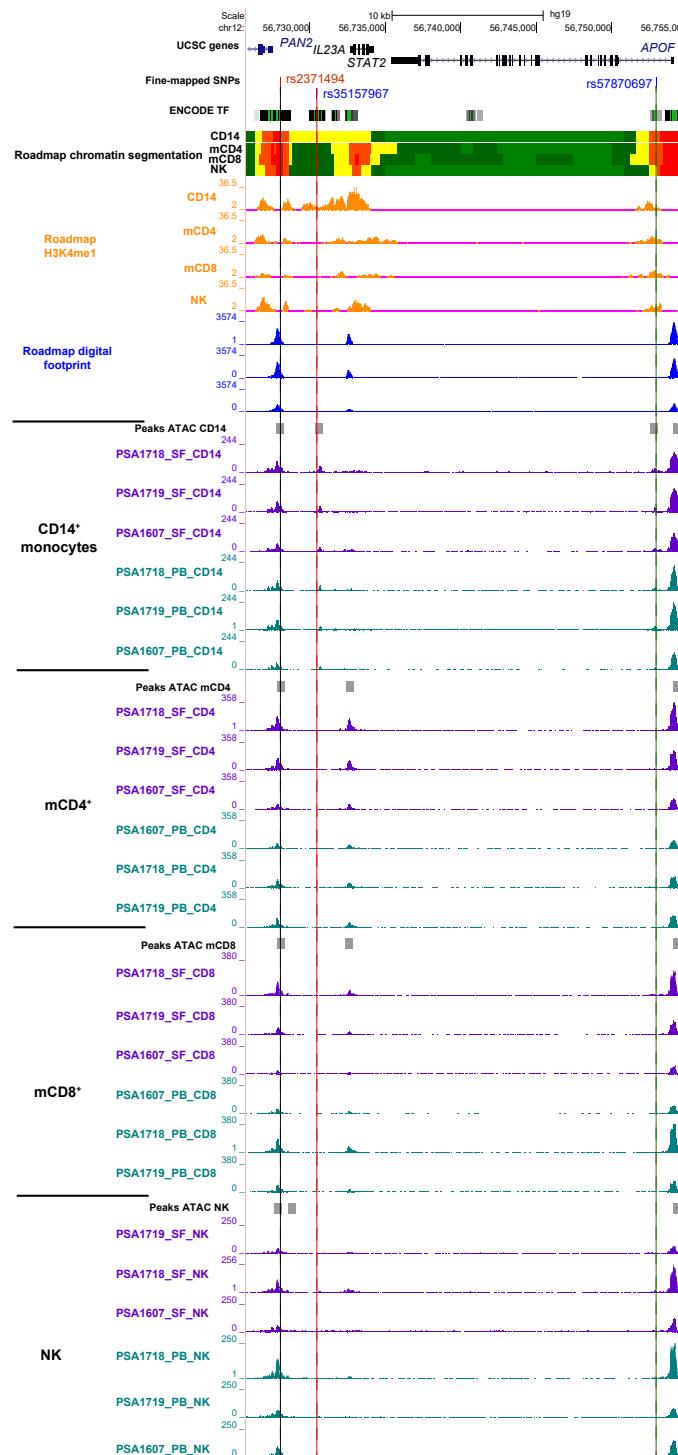


Figure 1.22: Epigenetic landscape at the genomic location of three fine-mapped SNPs from the STAT2/IL23A PsA GWAS signal. UCSC visualisation of the epigenetic landscape for rs2371494 (black line), rs35157967 (red line) and rs57870697 (green line) (x-axis), three relevant fine-mapped SNPs at the STAT2/IL23A locus. 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 (H3Kme1, chromatin segmentation maps, digital footprint and ENCODE cluster for TF binding) generated in the same cell types as the in-house ATAC has also been included.

Overview on the integration results for other loci

The *STAT2/IL23A* locus

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 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 Zhang 2018

Bibliography

- 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*.
- Andersson, Robin, Gebhard, Claudia, Miguel-Escalada, Irene, Hoof, Ilka, Bornholdt, Jette, Boyd, Mette, Chen, Yun, Zhao, Xiaobei, Schmidl, Christian, and Suzuki, Takahiro (2014). "An atlas of active enhancers across human cell types and tissues". *Nature* 507: p. 455.
- 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*.
- Baran-Gale, J. and functional, Chandra-T. in (2017). "Experimental design for single-cell RNA sequencing". *Briefings in functional*.
- 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*.
- Biniecka, M., Canavan, M., McGarry, T., and the, Gao-W. of (2016). "Dysregulated bioenergetics: a key regulator of joint inflammation". *Annals of the*.
- Boudousquié, Caroline, Danilo, Maxime, Pousse, Laurne, Jeevan-Raj, Beena, Angelov, Georgi S., Chennupati, Vijaykumar, Zehn, Dietmar, and Held, Werner (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, John, Budu-Aggrey, Ashley, Huffmeier, Ulrike, Uebe, Steffen, Steel, Kathryn, Hebert, Harry L., Wallace, Chris, Massey, Jonathan, Bruce, Ian N., Bluett, James, et al. (2015). "Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis". *Nature communications* 6: p. 6046.

BIBLIOGRAPHY

- 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.
- Dustin, Michael L. (2001). "Role of adhesion molecules in activation signaling in T lymphocytes". *Journal of clinical immunology* 21: pp. 258–263.
- Fairfax, Benjamin P., Makino, Seiko, Radhakrishnan, Jayachandran, Plant, Katharine, Leslie, Stephen, Dilthey, Alexander, Ellis, Peter, Langford, Cordelia, Vannberg, Fredrik O., and Knight, Julian C. (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, Lau, Evelyn, Jostins, Luke, Plant, Katharine, Andrews, Robert, and McGee, Chris (2014). "Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression". *Science* 343: p. 1246949.
- Gu, J., MrkerHermann, E., Baeten, D., and, Tsai-W. C. (2002). "A 588gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients" ..
- 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*.
- 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*.
- Jongbloed, Sarah L., Lebre, M. C., Fraser, Alasdair R., Gracie, J. A., Sturrock, Roger D., Tak, Paul P., and McInnes, Iain B. (2006). "Enumeration and phenotypical analysis of distinct dendritic cell subsets in psoriatic arthritis and rheumatoid arthritis". *Arthritis research & therapy* 8.
- 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.
- Lamb, C. A. and Immunology, Cresswell-P. of (1992). "Assembly and transport properties of invariant chain trimers and HLA-DR-invariant chain complexes". *The Journal of Immunology*.
- Li, Chun-xiao, Li, Hua-guo, Huang, Lin-ting, Kong, Yu-wei, Chen, Fu-ying, Liang, Jian-yin, Yu, Hong, and Yao, Zhi-rong (2017). "H19 lncRNA regulates keratinocyte differentiation by targeting miR-130b-3p". *Cell death & disease* 8.

BIBLIOGRAPHY

- Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J. M., Auton, A., Myers, S., Morris, A., et al. (2012). "Bayesian refinement of association signals for 14 loci in 3 common diseases". *Nat Genet* 44: pp. 1294–301.
- Mizoguchi, Fumitaka, Slowikowski, Kamil, Wei, Kevin, Marshall, Jennifer L., Rao, Deepak A., Chang, Sook, Nguyen, Hung N., Noss, Erika H., Turner, Jason D., and Earp, Brandon E. (2018). "Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis". *Nature communications* 9: p. 789.
- 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*.
- Picelli, S., Faridani, O. R., Björklund, K, and protocols, Winberg-G. (2014). "Full-length RNA-seq from single cells using Smart-seq2". *Nature protocols*.
- 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., Replogle, Joseph M., Feng, Ting, Lee, Michelle, Asinovski, Natasha, and Frohlich, Irene (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*.
- 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.
- Smith, Judith A., Barnes, Michael D., Hong, Dihua, DeLay, Monica L., Inman, Robert D., and Colbert, Robert A. (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.
- Spadaro, A., Scrivo, R., Moretti, T., Bernardini, G., Riccieri, V., Taccari, E., Strom, R., and Valesini, G. (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.
- 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*.
- Whitney, Adeline R., Diehn, Maximilian, Popper, Stephen J., Alizadeh, Ash A., Boldrick, Jennifer C., Relman, David A., and Brown, Patrick O. (2003). "Individuality and variation in gene expression patterns in human blood". *Proceedings of the National Academy of Sciences* 100: pp. 1896–1901.

BIBLIOGRAPHY

- Zaiss, Dietmar M. W., Loosdregt, Jorg, Gorlani, Andrea, Bekker, Cornelis P. J., Grne, Andrea, Sibilia, Maria, Henegouwen, Paul M. P., Roovers, Rob C., Coffer, Paul J., and Sijts, Alice (2013). "Amphiregulin enhances regulatory T cell-suppressive function via the epidermal growth factor receptor". *Immunity* 38: pp. 275–284.
- 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*.