Chapter 5

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

# Introduction

## 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 in chapter 1, the dysregulated immune response in PsA is the results of the interaction between cellular components of the innate and adaptive immune response. Consequently, the molecular characterisation of the different immune cell types is pivotal not only for the understanding of the immune response but also to define disease state, comprehend the impact of genetic variants increasing disease risk and identify drugs with optimal efficacy and specificity.

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

circulating IL-17+ and IL-22+ CD4+ T cells have been reported in PsA patients compared to control individuals (Benham et al. 2013). Moreover, reduced percentage of pDCs and NK cells in peripheral blood have also been observed in PsA (Jongbloed et al. 2006; Spadaro et al. 2004). In terms of cytokine production, stimulated PBMCs from PsA patients released greater levels of IL-17 and IL-22 than the healthy control counterparts (Benham et al. 2013).

Nevertheless, PsA is characterised by involvement of the joints, where repeated local inflammatory response leads eventually to joint destruction. Oligoarticular PsA (involving four or fewer joints) is commonly managed by joint aspiration followed by intra-articular steroid injection to relieve pain, facilitating sample collection for research purposes (Kavanaugh and Ritchlin 2006). The importance of studying the synovium in PsA have been highlighted by differences in cell composition and cytokine production, amongst others, between peripheral blood and synovial fluid in PsA patients. For example, expansion mCD8+ but not mCD4+ T cells (Ross et al. 2000) and an elevated proportion of T cells expressing the cytokine receptors CCR6+ and IL-23R+ (Benham et al. 2013) was observed in synovial fluid when compared to peripheral blood in PsA paired samples. Moreover, the elevated TNF-*α*, IL-1, IL-6 and IL-18 production in PsA synovial fluid was comparable to RA (Kujik2006).

## Bulk transcriptomic studies in PsA and their limitations

Genome-wide transcriptomic studies in PsA have been mainly focused in characterising gene expression in bulk PBMCs and synovial fluid mononuclear cells (SFMCs). Several studies have been conducted to better understand gene expression differences in blood between PsA and controls and also specific differences between peripheral blood and synovial fluid from the same PsA patients or differences with other arthritic diseases (Batiwalla2005; Stoeckman et al. 2006; Gu et al. 2002; Dolcino et al. 2015). 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 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. Cytokine production assays have also been conducted in serum and synovial fluid, revealing increased levels of TNF-*α*, for example, in both tissues when compared to controls and osteoarthritis patients, respectively (Ritchlin et al. 1998; Li et al. 2017).

Studies using mixed cell populations can be influenced by the relative proportion of the different cell populations within the sample (Whitney et al. 2003). 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 na¨ıve CD4+ isolated from peripheral blood and synovial fluid

of PsA patients (Antoniv2006; Leipe2010). The importance of investigating the transcriptional profile of patients’ isolated discrete cell populations have yielded interesting findings in monocytes and Th-17 cells in AS, intestinal epithelial cells in CD and fibroblast-like synoviocytes in RA (Al-Mossawi et al. 2017; Smith et al. 2008; Howell et al. 2018; Ai et al. 2016). Overall, achieving a detailed and precise understanding of complex diseases requires the study of sorted cell populations and when possible isolated from the affected tissue.

## Transcriptomics and proteomics at the single cell resolution

In addition to the study of specific cell types, evidence for heterogeneity in the transcriptome within cells from the same population has accelerated the development of strategies providing single cell resolution. The establishment of single cell RNA sequencing (scRNA-seq) and mass-cytometry techniques represent an unbiased way to characterise and identify cell subpopulations within the samples, avoiding the pre-selection of particular cell types and thus providing a global overview of cell composition and interactions in the tissue of interest.

A wide range of approaches to study single-cell transcriptomics have been developed in the last few years, with Drop-seq, SmartSeq2 and 10X Chromium amongst the most widely used (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 2). As a result, 10X Chromium technology does not require pre-sorting of single-cells into plates and enables higher throughput than other with less manipulation and variability than other scRNA-seq methods such as SmartSeq2 (Baran-Gale and functional 2017).

Mass cytometry represents the next generation of fluorescence based flow- cytometry analysis to interrogate expression of cell surface and intracellular molecules. Mass cytometry is a hybrid technique between mass spectrometry and flow cytometry, where the Abs recognising the molecular markers have been labelled with stable isotopes instead of fluorophores (Bandura et al. 2009). The use of isotopes enables incorporating up to fourty-five Abs to profile cellular populations and assess molecular functions.

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

Incorporation of scRNA-seq and mass cytometry in addition to bulk RNA-seq and flow cytometry have led to a more detailed understanding of the immune system, accounting for the variability at the single-cell level in gene expression and protein translation (Jaitin et al. 2014; Villani et al. 2017; Bengsch et al. 2018). In complex diseases such as RA, scRNA-seq has revealed heterogeneity in the synovial fibroblast population and identified a potentially pathogenic cluster highly proliferative and active in pro-inflammatory cytokine secretion (Mizoguchi et al. 2018). Similarly, mass cytometry analysis performed

in RA identified an expanded CD4+ 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.

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

Fine-mapping of GWAS signals is required in order to reduce the putative number of causal SNPs accounting for a particular association in complex diseases. Fine-mapping using genotype level data incorporates a locus step-wise 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

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

The PsA GWAS study conducted by Bowes and colleagues successfully performed fine-mapping for seventeen of the associated regions (Bowes2016). This study provided a summary table for the overlap of SNPs from the 90% credible set (list of SNPs explaining 90% of the PsA GWAS association) with genomic annotations and ENCODE features (cell lines and healthy donors primary cells), to further narrow down the set of putative causal SNPs for each associations as well as the more relevant cell type where they may have an effect. Bowes *et al.* further investigated the PsA-specific association identified in this study at the 5q31 region and a pilot eQTL study confirmed the most significant correlation with *SLC22A5* expression for a SNP in high LD with the GWAS lead SNP.

Leveraging epigenetic data to further refine the candidate causal SNPs from fine-mapping studies would benefit from the generation of disease- specific chromatin regulatory maps in PsA affected tissue and, possibly, further integration of scRNA-seq and mass cytometry from the same individuals. Altogether, this data could represent an additional layer of information in the attempt to identify the causal variant driving GWAS associations with PsA and provide further insight into the disease pathophysiology.

# Aims

This chapter aims to establish and integrate multi- omic datasets in PsA blood and synovial immune cells including chromatin accessibility and transcriptomic profiling in peripheral blood, and application of single-cell transcriptomics in synovial fluid. The long term goal is to identify cell subsets contributing to pathophysiological relevant pathways in PsA and chronic inflammation, and facilitate interpretation of PsA GWAS. Specific aims of this chapter are:

1. To identify differences in the chromatin accessibility landscape and transcriptomic profiles between synovial fluid and peripheral blood in CD14+ monocytes, mCD4+, mCD8+ and NK cells isolated from PsA patients.
2. To explore transcriptional differences at the single-cell level in cell types of interest and perform a basic integration with mass cytometry data.
3. To conduct fine-mapping for a number of PsA GWAS loci using genotype data and integrate with cell and tissue-specific PsA chromatin accessibility maps and publicly available epigenetic and functional data to further narrow down the putative causal SNPs driving such associations.

# Results

## PsA patients cohort description and datasets

In this study peripheral blood (PB) and synovial fluid were collected from a cohort of six PsA patients, with equal numbers of males and females (Table 5.1). All the patients presented with oligoarticular joint involvement and had a diagnosis of 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, on a scale of 1 to 5. These four measurements include joint and global assessment PsARC disease activity scores), used by clinicians as the main indicator of response to treatment by recommendation of the National Institute for Health and Care Excellence (NICE) (Chapter 1).

The mean age of the cohort at the time of diagnosis was 44.3 years old and the mean disease duration 8.8 years. PsA1728 was diagnosed at a later age compared to the other patients in the cohort. The average C-reactive protein (CRP) level, a marker of inflammation, was 17.45 mg/L and was higher in PsA1719 and PsA1728 compared to the other patients. At the time of sample recruitment no patients were on active immunosupressive therapy. Following the recruitment visit, most of the patients qualified for TNFi biologic therapy xxxx.

Cross-tissue comparison analysis in PsA

217

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample ID | Sex | Age at diagnosis | Disease duration (months) | Type | TJC66/SJC66 | Physician assessment | Patient assessment | CRP  (mg/L) |
|  |  |  |  |  |  |  |  |  |
| PsA1718 | Female | 17 | 180 | Oligo | 2/2 | 3 | 3 | 6 |
| PsA1719 | Male | 33 | 24 | Oligo | 1/1 | 3 | 4 | 36.6 |
| PsA1607 | Male | 42 | 108 | Oligo | 1/1 | 4 | 3 | 8 |
| PsA1728 | Female | 72 | 48 | Oligo | 2/2 | 3 | 4 | 43.2 |
| PsA1801 | Female | 53 | 168 | Oligo | 2/2 | 3 | 3 | 9.9 |
| PsA1505 | Male | 35 | 108 | Oligo | 1/1 | 2 | 2 | 1 |
| Average | − | 44.3 | 106 | − | 1.5/1.5 | 3 | 3.2 | 17.4 |

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

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

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

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

The immune cellular composition of three PsA samples (PsA1718, PsA1719 and PsA1607) was characterised in synovial fluid and peripheral blood using the ICS mass cytometry panel (detailed in Chapter 2). For both tissues, mCD4+ (32.1 -55.6%) constituted the most abundant cell type followed by mCD8+ (16.9 -24.9%) and CD14+ ”non-classical” monocytes (6.9 - 21.7%). Consistent with previous studies, a trend of increased percentage of mCD8+ pDCs and cDCs was observed in synovial fluid compared to peripheral blood (Ross et al. 2000; Jongbloed et al. 2006). This data also showed reduced percentage of synovial fluid NK cells compared to peripheral blood, in line with previous studies suggesting the role of impaired non-MHC-restricted cytotoxicity in PsA (Spadaro et al. 2004). Similarly, a tendency towards reduced proportions of B cells in synovial fluid compared to peripheral blood reinforced the lack of contribution of

the humoral immune response in PsA pathophysiology. The observed differences in cell composition between synovial fluid and peripheral blood were not statistically significant for any of the 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.

60

**PB**

**SF**

55

50

45

40

**Proportion (%)**

35

30

25

20

15

10

5

0

CD14

+

monocytes

CD16

+

mCD4

mCD8

+

NK

B

pDC

cDC

MAIT

DN T

Basophils

+

**Cell type**

monocytes

lymphocytes

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

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

Quality control of Fast-ATAC data

Twenty four Fast-ATAC PsA samples from four different cell types and two tissues (synovial fluid and peripheral blood) were sequenced and processed using the in-house pipeline as previously detailed (Chapter 2 and 3). After filtering for low quality mapping, duplicates and MT reads, the median of total number of reads was 46.6 - 70.2 million (Figure 5.2 a) with MT and duplicated reads accounted for a median of 40 to 62.2% of the total number of unfiltered reads depending on cell type (Figure 5.2 b).

Regarding sample quality, TSS enrichment analysis showed differences in the levels of background noise across cell types and highlighted the variability of ATAC performance (Figure 5.2 c). A general trend towards greater TSS enrichment in peripheral blood samples compared to synovial fluid was observed. mCD4+ and mCD8+ showed 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. The fold enrichment for PsA1719 and PsA1607 in NK cells were close to the 6 fold enrichment threshold 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.

80

**PB SF**

**PB SF**

120

70

**Combined MT and duplicated reads(million)**

100

**Total reads after filtering (million)**

60

80

50

60

40

40

30

20

0

CD14 CD4 CD8 NK

**Cell type**

(a)

20

CD14 CD4 CD8 NK

**Cell type**

(b)

110

**PB SF**

**CD14 CD4 CD8 NK**

**PB SF**

30

100

90

25

**Reads after filtering (million)**

80

20 70

**TSS enrichment**

60

15 50

40

10

30

5

0

CD14 CD4 CD8 NK

**Cell type**

(c)

20

10

0

20 30 40 50 60 70 80 90 100 110

**Number of peaks (x103)**

(d)

Figure 5.2: Quality control assessment of ATAC data generated in four immune cell types isolated from peripheral blood and synovial fluid of PsA patients samples. For each of the cell types and samples, boxplots representing a) million of reads after filtering, 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 24x103 and 97x103 (Figure 5.2 d). The total number of called peaks passing filtering varied across cell types and was influenced by the quality sample however no significant 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 Chapters 2 and 3. PCA based on the normalised counts for each region showed that 65.5% of the variability (PC1) in the chromatin landscape correlated with cell type, leading to sample separation in four cluster (Figure 5.3). The myeloid (CD14+ monocytes) and lymphoid (mCD4+ and mCD8+) clusters were most distinct based on the Fast-ATAC profile consistent with biological expectations. Modest separation between synovial fluid and peripheral blood samples were found in the mCD4+, mCD8+ and NK clusters (Figure 5.3).

0

50

−200 −100 0 100

**CD14 CD4 CD8**

**NK**

**PC2 (11.2%)**

−50

−100

**SF**

**PB**

**PC1 (65.6%)**

Figure 5.3: PCA based on the ATAC chromatin accessibility landscape in four immune cell types isolated from blood and synovial fluid. PCA 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 of interest. The first two PCs (x-axis and y-axis, respectively) for the ATAC peaks included in the ML all are plotted. Each point represents a sample, where colour indicates cell type and shape tissue (SF and PB). The proportion of variation explained by each principal component is indicated.

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

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

|  |  |  |
| --- | --- | --- |
| 194 |  | 90 |
| 174 |  | 80 |
| 154 |  |  |
|  |  | 70 |
| 134  114  94  74  54  34  14 | **Whole blood Other tissues** | 60  50  40  30  20 |

**−log10(FDR)**

**−log10(FDR)**

5 10 15 20 25 30 35 40 45

**z−score**

(a)

5 10 15 20 25 30

**z−score**

**Monocytes tCD8+ tCD4+**

**Neutrophils B cells**

LPS 24h

LPS 2h

IFN

Unstimulated

(b)

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

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

A consensus master list of chromatin accessible regions was built for each of the four cell types of interest (ML CD14, ML CD4, ML CD8 and ML NK). Differential chromatin accessibility analysis between synovial fluid and peripheral blood was performed on the normalised counts retrieved for each of the cell type master lists using

DESeq2 and a paired design (Table 5.3). A 80% cut-off for background noise was used to filter the count matrix (Chapter 3). 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). Considering the origin of cells, in CD14+ monocytes the number of open DARS in synovial fluid were notably larger than in peripheral blood (3,779 and 1,506DARs, respectively) (Table 5.3). The number of synovial fluid and peripheral blood open DARs were similar for the other three cell types.

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

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

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

Genomic annotation of the DARs identified in each the cell types revealed that 80% or more of all regions with differential accessibility were located at intronic and intergenic regions (Figure 5.5 a). Universal promoter regions was the third most represented genomic feature, accounting for the annotation of

approximately between 5 to 15% of the DARs in each cell type. Chromatin states from the Roadmap Epigenomics maps were also used for annotation (Figure 5.5 b). For all four cell types, between 44.96 and 72.11% of the DARs were annotated as weak enhancers, which represented the most prominent category and the most significantly enriched (p-value and fold change and also plot in appendix data not shown). The over-representation of enhancers was consistent with large percentage of introns and intergenic regions found for the genomic features annotation, as those are the preferred location for enhancer elements. Modest percentages of DARs were annotated as heterochromatin and repetitive regions but not significant enrichment for these two chromatin states was found for any of the four cell types (Figure 5.5 b).

50



**CD14 CD4 CD8 NK**

40

30

**Percentage (%)**

20

10

0

Promoters

5’ UTRs

3’ UTRs

Exons

Introns

Intergenic

**Annotation**

Downstream

regions

(a)

100

**12.40**

**13.47**

**11.58**

**3.15**

**2.72**

**24.90**

**58.77**

**47.96**

**68.85**

**72.11**

**9.68**

**3.69**

**8.54**

**10.76**

**9.56**

**5.27**

**8.09**

**2.86**

**2.21**

**6.21**

**4.11**

**4.97**

**Chromatin state**

75

1. **Active Promoter**
2. **Weak Promoter**
3. **Poised Promoter**
4. **Strong Enhancer**
5. **Strong Enhancer**

**Percentage (%)**

1. **Weak Enhancer**
2. **Weak Enhancer**

50 **8 Insulator**

1. **Txn Transition**
2. **Txn Elongation**
3. **Weak Txn**
4. **Repressed**
5. **Heterochrom/lo**
6. **Repetitive/CNV**

25 **15 Repetitive/CNV**

0

CD14 CD4 CD8 NK

**Cell type**

(b)

Figure 5.5: Annotation of the PsA DARs identified in the four cell types with genomic annotations and chromatin states. a) Barplot illustrating the percentage of nucleotides within DARs for each cell type that are annotated as promoters, downstream (regions at 1,000bp to a promoter), exons, introns, 5’ or 3’UTR and intergenic regions. b) Stacked barplot representing the percentage of DARs annotated for each of the fifteen chromatin states defined in each of the four relevant cell types by Epigenome Roadmap chromatin

≤

segmentation maps (CD14+ peripheral blood 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 showed significant enrichment for the corresponding cell type eRNA set. The proportion of DARs overlapping the appropriate cell type set of expressed eRNAs ranged between

19.8% (83 open in synovial fluid and 160 open in peripheral blood) in NK and 31.8% (83 open in synovial fluid and 160 open in peripheral blood) in CD4+ cells.

**NK CD8 CD4 CD14**

**Permissive enhancers**

**Total DARs**: 221 **otal DARs**: 198

SF open: 95 Open SF: 122

PB open : 126 Open PB: 76

**Total DARs**: 243

SF open: 83

PB open: 160

**Total DARs**: 924

Open SF: 690

Open PB: 234

**T**

**Robust enhancers**

**T cell eRNA**

**Monocyte eRNA**

**Basophil eRNA**

**NK eRNA**

0 50 100 150 200 0 50 100 150 200 0 50 100 150 200 0 50 100 150 200

**Enrichment significance (-log10FDR)**

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

From the differential analysis between synovial fluid and peripheral blood, a number of DARs were overlapping a gene body (Table 5.4). Interestingly, the majority were located within introns instead of untranslated regions (UTRs) annotated as weak or strong enhancers according to the cell type specific chromatin segmentation map previously illustrated.

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

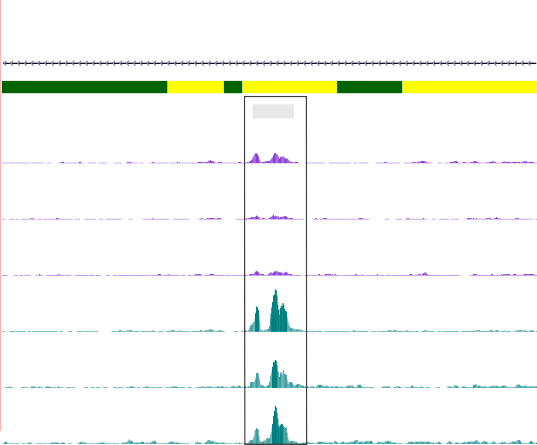
Table 5.4: Characterisation of the DARs located within genes in each of the four cell types from PsA samples. The number of DARs that overlapping a gene body for each of the cell types are indicated together with those genes harbouring more than one DARs. Further details about those regions includes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

For example, peripheral blood NK analysis identified an open DAR located in an intron of the *VAV3* gene that was more accessible in peripheral blood and also significantly expressed as an eRNA (Figure 5.7 a). By contrast in CD14+ monocytes two DARs located at the 5’ and 3’ UTRs of *IL7R* gene were more accessible in synovial fluid compared to peripheral blood (Figure 5.7 b).

chr1:

5 kb hg19

108,480,000 108,485,000



**UCSC genes** *VAV3*

**ChromHMM RoadMap**

**Fast-ATAC peaks**

246 \_

**PSA1718 SF**

**PSA1719 SF**

**PSA1607 SF**

**PSA1718 PB**

**PSA1719 PB**

**PSA1607 PB**

0 \_

246 \_

0 \_

246 \_

0 \_

246 \_

1 \_

246\_

0 \_

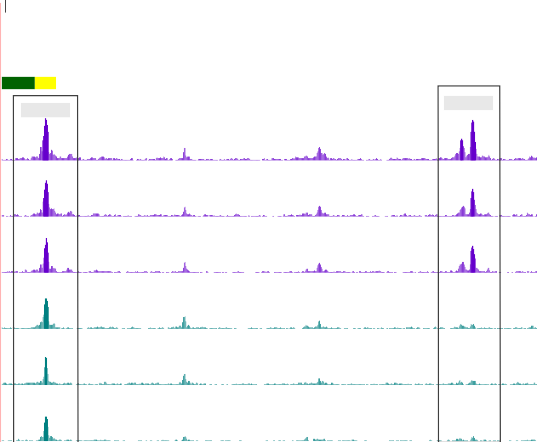
246 \_

0 \_

chr5:

(a)

10 kb hg19



35,860,000 35,865,000 35,870,000 35,875,000 35,880,000

*IL7R*

**UCSC genes**

**ChromHMM RoadMap**

121 \_

**PSA1718 SF**

**PSA1719 SF**

**PSA1607 SF**

**PSA1718 PB**

**PSA1719 PB**

**PSA1607 PB**

0 \_

121 \_

0 \_

121

0 \_

121 \_

0 \_

121 \_

0 \_

121 \_

0 \_

(b)

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

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

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

Pathway enrichment analysis for differentially open DARs in synovial fluid compared to peripheral blood for each cell type was conducted by defining DARs for each of the two contexts and defining those DARs open to a greater or lesser extent in each tissue. . Gene annotation of the DARs was performed based on location (physical proximity) as detailed in Chapter 2. Despite commonalities, differences in significant enriched pathways (FDR*<*0.01 or 0.05) were identified for the same cell type between synovial fluid and peripheral blood open DARs (Figure 5.8). In CD14+ monocytes synovial fluid open DARs showed enrichment for pathways involved in hemostasis, integrin interactions, calcium signaling and regulation of immunity, inflammation and cell survival such as the NF-*κ*B pathway and cytokine related pathways, including IL-2 and IL-3, 5 and granulocyte-macrophage colonystimulating factor (GM-CSF) signalling (Figure 5.8 a). By contrast, peripheral blood open DARs in CD14+ monocytes were not enriched for any of these pathways, showing enrichment rather for DAP12 interactions and regulation of PI3K/AKT network

mCD4+ synovial fluid open DARs, in contrast to peripheral blood, showed enrichment for TCR signalling as well as chemokine signalling, which included, amongst others, DARs in proximity to IFN-*γ* or *CXCL13* and *CXCR6*, respectively (Figure 5.8 b). Peripheral blood open DARs in this cell type were enriched for signaling by receptor tyrosine kinases and focal adhesion members, also involved in the T cell activation (Dustin 2001). Enriched pathways for synovial fluid or peripheral blood open DARs in mCD8+ were only significant when using an FDR*<*0.05 threshold. (Figure 5.8 c). The G protein coupled receptor (GPCR) signalling,

with varied roles in regulation of inflammation and mediation of the chemotactic recruitment of T cells to the inflamed tissue, was enriched for mCD8+ peripheral blood open DARs. mCD8+ synovial fluid open DARs showed enrichment for chemokine signaling and regulation of stem cell pluripotency.

NK synovial fluid open DARs showed enrichment for extracellular matrix organization and Rap1 signaling as well as Fc-gamma receptor (FC*γ*R)- mediated phagocytosis (Figure 5.8 d). Members of the HIF-1 pathway involved in oxygen homeostasis were also enriched in NK synovial fluid open DARs, in line with the hypoxic environment found in joint inflammation. Interestingly, enrichment of open peripheral blood DARs in the proximity of genes involved in NK-mediated toxicity.

**CD14+ monocytes**

**mCD4+**

DAP12 interactions Negative regulation of the PI3K/AKT network

Signalling by Receptor Tyrosine Kinases

Metabolism of vitamins and co−factors

Metabolism of lipids

Hemostasis

Integrin cell surface interactions

Calcium signalling IL−2 signalling GPCR ligand binding

Focal adhesion

TCR signalling

**PB** Phospholipase D signalling **PB**

**SF**

**SF**

Chemokine signalling

NF−kB signalling

Platelet activation

PI3K−Akt signalling pathway

IL−3, 5 and GM−CSF signalling

Signalling by ILs

0 1 2 3 4 5 6

**Enrichment significance**

**(−log10FDR)**

(a)

Signalling by ILs

0 1 2 3 4 5 6

**Enrichment significance**

**(−log10FDR)**

(b)

**mCD8+**

Signalling by GPCR

Signal transduction

**NK**

NK cell mediated cytotoxicity

Wnt signalling

T cell receptor signalling

Rho GTPase cycle

PI3K−Akt signalling

Signalling by ILs

Chemokine signalling

Signalling regulating pluripotency of stem cells

Regulation of actin cytoskeleton

**PB SF**

0 1 2 3 4 5 6

**Enrichment significance**

**(−log10FDR)**

(c)

Th-1 and Th-2 cell differentiation

Signal transduction MAPK family signalling cascades

Rho GTPase cycle Extracellular matrix organization

Rap1 signalling PI3K−Akt signalling Calcium signalling

Fc gamma R−mediated phagocytosis

HIF−1 signalling

**PB SF**

0 1 2 3 4 5 6

**Enrichment significance**

**(−log10FDR)**

(d)

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

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

Immune-relevant gene expression by qPCR

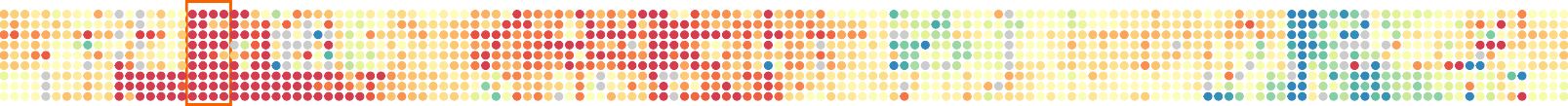
In order to contextualise the ATAC-seq data, qPCR gene expression analysis for 370 key genes in the inflammatory and autoimmune response was conducted in CD14+ monocytes, mCD4+ and mCD8+ cells isolated from synovial fluid and peripheral blood of three PsA patients (Table 5.2). Those appeared as the most abundant cell types in peripheral blood and synovial fluid from patients, and particularly mCD4+ and mCD8+ cells have been shown to expand in PsA inflammed synovium, as previously mentioned. The PCR array represented a cost-effective approach to study gene expression between peripheral blood and synovial fluid focusing on a relevant subset of genes given the inflammatory component in PsA. For each cell types, FC in expression was calculated pair-wise for synovial fluid respect to peripheral blood within each sample for each individual genes (detailed in Chapter 2). Likely due to the small sample size,the majority of

the modulated genes between synovial fluid and peripheral blood lacked of significance (FDR*<*0.05) after multiple testing correction. Therefore, to explore the potential biological relevance of this data, a less stringent pval*<*0.05 was used as the filtering threshold.

When considering the significantly modulated genes (pval*<*0.05) in at least one cell type, differences in magnitude and reproducibility in FCs were observed across samples and cell types (Figure 5.9). Some of the modulated genes showed up-regulation (FC*>*1.5) in synovial fluid compared to peripheral blood across the three cell types, for example *FN1*, *SPP1* or *CCL2*, amongst others (Figure 5.9 orange box). On the other hand, a number of genes showed reduced expression in synovial fluid (FC*<*1.5) in at least one of the three cell types, including *FOS*, *IL16*, *PPBP* and *TPST1* (Figure

5.9 purple box). Also, a number of genes were only consistently modulated in the three CD14+ monocyte samples but not in T cells (Figure 5.9 dark blue box).

**1**



***CMTM1***

***PREX1***

***IL10RA***

***CKLF***

***IL26***

***NFKB1***

***NFX1***

***CXCL5***

***GFRA2***

***IL18***

***STAT3***

***CCL17***

***TLR3***

***IL1RL2***

***PDGFB***

***CCL7***

***IL7R***

***CD28***

***FN1***

***SPP1***

***CCL18***

***CCL2***

***OLR1***

***CXCL9***

***CSF1***

***PDGFA***

***TNFSF18***

***C3***

***CXCL10***

***INHBA***

***SPRED1***

***CXCL12***

***VEGFB***

***CCR7***

***IL3RA***

***PROCR***

***CCL28***

***GPI***

***CXCL16***

***IL2RG***

***NFE2L1***

***APOL2***

***MYD88***

***ADGRE5***

***IRF7***

***CCR2***

***SCUBE1***

***FASLG***

***IL2RB***

***CCL5***

***TYMP***

***CCRL2***

***CCR1***

***CD70***

***FPR1***

***SRGAP1***

***CXCR6***

***IFNG***

***BCL6***

***CCR5***

***IL1R2***

***TNFSF9***

***CCL23***

***CXCL13***

***IL21***

***VEGFA***

***CCL3***

***C3AR1***

***CD180***

***XCR1***

***CCL4***

***TNF***

***MIF***

***BLNK***

***PRG2***

***CD74***

***GPR68***

***IL12RB2***

***IL1R1***

***NAMPT***

***YARS***

***NR3C1***

***CEBPB***

***IL12RB1***

***LY75***

***IFNK***

***S100B***

***IFNGR2***

***LY86***

***AIF1***

***FGF7***

***IL6R***

***TLR2***

***RIPK2***

***ERBB2***

***SLCO1A2***

***IL31RA***

***TLR1***

***IFNAR1***

***TRAP1***

***IL4R***

***EPOR***

***IRF4***

***HDAC4***

***IFNAR2***

***TLR6***

***IL10RB***

***TOLLIP***

***SDCBP***

***TNFSF14***

***IFNGR1***

***PRDX5***

***SYK***

***LY96***

***NMI***

***CCR4***

***IL17RA***

***IL32***

***IL18RAP***

***CSF2RA***

***HDAC9***

***LTB4R***

***MMP25***

***CLC***

***PPBP***

***TPST1***

***EPHX2***

***AREG***

***PTGS2***

***S100A12***

***PF4V1***

***PROK2***

***MEFV***

***S100A8***

***SIGIRR***

***FOS***

***IL16***

***TNFSF10***

***AZU1***

***NFAM1***

***IL2RA***

***TLR8***

***CSF3R***

***IL13RA1***

***CSF3***

***IFNL1***

***CAST***

***DOCK2***

***ITGB2***

***PARP4***

**mCD4+ 2**

**3**

**1**

**mCD8+ 2**

**3**

**1**

**CD14+ 2**

**3**

**Patients key:**

**1=PsA1607**

**2=PsA1719**

**3=PsA1728**

**log (FC)**

**4**



Cross-tissue comparison analysis in PsA

**2**

**2**

**0**

**−2**

**−4**

233

Figure 5.9: Heatmap of gene expression FCs between synovial fluid and peripheral blood for those gene significantly modulated (pval*<*0.05) in at least one cell type. Amongst the 370 genes measured by qPCR, the FC in gene expression between the synovial fluid and peripheral blood 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 log2*FC* in gene expression between synovial fluid and peripheral blood is colour-coded. Overall, the heatmap allows to observe the change in gene expression as well as the magnitude between synovial fluid and peripheral blood for each gene in each of the the three pairs of PsA samples in CD14+ monocytes, mCD4+ and mCD8+ cells.

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

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

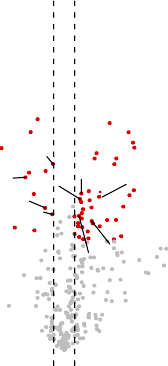
Correlation between gene expression and chromatin accessibility

Overlap between differentially modulated genes (in synovial fluid vs peripheral blood) and proximal DARs was only found to be significant in CD14+ monocytes (Fisher exact test pval=0.028) and not 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

synovial fluid also showed open DARs. For example, the increased expression of *IL7R* in synovial fluid correlated with increased chromatin accessibility at the 5’ and 3’ UTR of this gene (Figure 5.7 b). Another relevant example was the *FN1* gene, for which up-regulated expression in synovial biopsies compared to peripheral blood has been reported by others (Dolcino et al. 2015). In this

**CD14+ monocytes**

4.0



**Not sig.**

**pval<0.05 & abs(FC)>1.5**

**Up 56**

**Down 16**

*CXCL5*

*CCR1*

*TNFSF18*

*TNF*

*CCL5*

*CXCL9*

*IFNK CD180 IL1RL2*

*NFAM1*

*CSF3R IL18 TNFSF9 CXCL12*

*CCL2*

*GPI SPRED1PDGFB*

*CD74*

*IL16*

*SPP1*

*CXCL16*

*FN1*

*IL7R*

*CCL7*

*CCR5*

*IFNAR1*

*LTB4R*

*GFRA2 PRG2 CCL17*

*OLR1*

*CCL18*

*PROCR*

*VEGFB BLNK CXCL10 C3AR1*

*MYD88*

*TYMP*

*PROK2 PTGS2*

*FOS TLR8 S100A8 AZU1*

*SIGIRR IL2RG SRGAP1*

*C3*

*S100A12*

*MIF*

*INHBA*

3.5

3.0

2.5

2.0

**-log10(pval)**

1.5

1.0

0.5

0.0

−6 −4 −2 0 2 4 6 8 10 12 14 16 18

**log2(mean FC)**

(a)

4.0

3.5

3.0

2.5

**-log10(pval)**

2.0

1.5

1.0

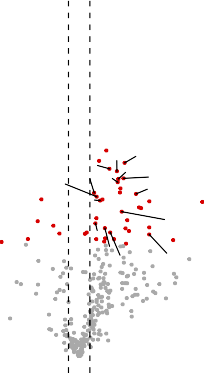
0.5

0.0

**mCD4+**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *CLC* | *NFX1 FGF7*  *EPHX2 S100A8*  *IFNGR2* | *IL10RA*  *NMI IL18RAP CD70 IFNGR1 GPR68*  *CSF3 MIF CCL4 IFNL1 CD74 TYMP*  *GPI CCL5 CCL3*  *LTB4R OLR1*  *IL17RA BCL6*  *CCR2 CCR1 CCRL2*  *SYK IL1R1 CXCR6* | *CXCL13*  *FN1* | *SPP1* |  |
|  |  | *IL2RA HDAC9 IL1R2* |  |  |  |
|  |  |  |  |  | **Not sig.** |
|  |  |  |  |  | **pval<0.05** |
|  |  |  |  |  | **Up 38** |
|  |  |  |  |  | **Down 8** |

−6 −4 −2 0 2 4 6 8 10 12 14 16 18



**log2(mean FC)**

(b)

4.0

3.5

3.0

2.5



**-log10(pval)**

2.0

1.5

1.0

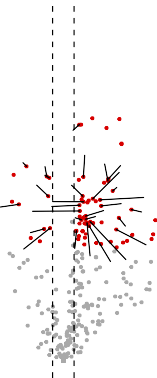
0.5

0.0

**mCD8+**

|  |  |  |
| --- | --- | --- |
| *CCR5 ADGRE5 CCRL2*  *IL2RG CD74*  *HDAC4 CXCR6 PDGFA*  *NR3C1 TYMP*  *IL6R IL3RA CCL5 TNF*  *EPHX2 ERBB2 CMTM1 SLCO1A2 NFE2L1 GPR68 CCR2*  *S100B DOCK2 CCL4 IL2RB TNFSF10 TNFSF14*  *IFNGR2 TOLLIP NMI IRF7 SDCBP CSF1*  *MYD88 IL10RA*  *CCR4 FASLG CCL18*  *AIF1*  *CCR1* | *CCL23*  *FN1 SPP1* |  |
|  |  |  |
|  |  |  |
| *IL12RB2* |  |  |
|  |  | **Not sig.** |
|  |  | **pval<0.05 & abs(FC)>1.5** |
|  |  | **Up 65** |
|  |  | **Down 12** |

−6 −4 −2 0 2 4 6 8 10 12 14 16 18



**log2(mean FC)**

(c)

Figure 5.10: Gene expression changes in immune-relevant genes between synovial fluid and peripheral blood in CD14+ monocytes, mCD4+ and mCD4+ cells. Volcano plots showing differences in gene expression measured by qPCR array for a) CD14+ monocytes,

b) mCD4+ and c) mCD8+ cells. The significance (log10pval) of the modulation in gene expression between the two tissues (y-axis) is plotted against the log2 of the mean FC across the three PsA patients. Positive FC indicates higher expression in SF. Genes showing pval*<*0.05 and mean FC*>*1.5 are coloured in red, with the most significant genes labelled.

*CXCL10*, *FN1*, *IL18*, *IL31RA*∗,

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

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

cohort, *FN1* expression was up-regulated in synovial fluid for all three cell types with the highest fold change found in CD14+ monocytes (Figure 5.10 a), concomitantly with more accessible chromatin at the promoter and downstream the 3’ UTR of the gene (Figure 5.11).

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 synovial fluid and peripheral blood, enrichment analysis was performed for each individual cell type. Up-

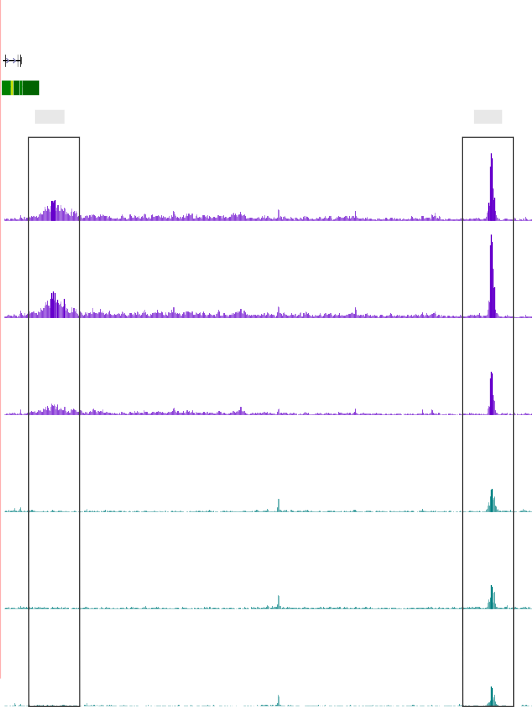
chr2:

**UCSC genes**

*FN1*

20 kb hg19

216,250,000 216,300,000



**ChromHMM RoadMap**

**Fast-ATAC peaks**

276 \_

**PSA1718 SF**

0 \_

276 \_

**PSA1719 SF**

0 \_

276 \_

**PSA1607 SF**

0 \_

276 \_

**PSA1718 PB**

0 \_

276 \_

**PSA1719 PB**

0 \_

276 \_

**PSA1607 PB**

0 \_

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

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 synovial fluid and peripheral blood in CD14+ monocytes were enriched for chemokine, NOD-like signalling and TLR signalling pathways (Table 5.6). All three pathways are involved in the activation of cytokines and chemokines gene expression, leading to T cell recruitment and inflammatory response.

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

The enrichment for the chemokine pathway in CD14+ monocytes (Table 5.6)

included genes highly up-regulated (mean FC*>*16) in synovial fluid compared to peripheral blood (e.g *CCL18* and *CCL2*) for all three cell types (Figure 5.10) as well as genes only consistently modulated between synovial fluid and peripheral blood in CD14+ monocytes (e.g *CCL28*, Figure 5.9 green box).

Cross-tissue comparison analysis in PsA

239

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

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

In addition to pathway enrichment, network analysis was performed to understand the interaction and relationship of the genes differentially expressed between synovial fluid and PB. A gene subnetwork was identified from the STRING functional interaction database using as input all the genes from the qPCR array (regardless pval significance and FC) and ranking them based on the best pval across the three analysed cell types. The identified subnetwork predominantly included significantly modulated genes between synovial fluid and peripheral blood in at least one of the cell types. Amongst the most interesting nodes was the single Ig and Toll-interleukin domain containing gene (*SIGIRR*), which is a negative regulator of the TLR signalling pathway (Figure 5.12). *SIGIRR* was significantly down-regulated in synovial fluid CD14+ monocytes only consistent with the significant up-regulation (pval¡0.05) of the *TLR1*, *TLR2*, *MYD88* genes in synovial fluid as well as the enrichment for the TLR pathway in this cell type (Figure 5.9 and 5.6). Moreover, the significant up-regulation of *NFKB* and *TNF* in the synovial fluid 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 synovial fluid and PB.

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

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

**log2 mean FC**

**4**



**2**

**0**

**−2**

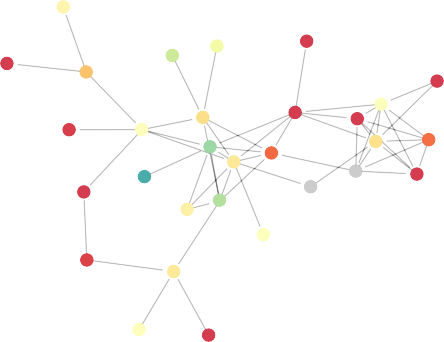
**−4**

**-log10pval**

 **Not sig. ** **pval<0.05**

**pval<0.01**

**CD14+**



TLR8

OLR1

IL7R

SIGIRR

IL2RG

CCL18

MYD88

CCL2

BLNK

SYK

FOS

NFKB1

CXCR6 CXCL10

CCR5 C3AR1

TNF

PROK2

CXCL13

C3

FN1

IL18RAP

PTGS2 NR3C1

HDAC4

VEGFB

ERBB2

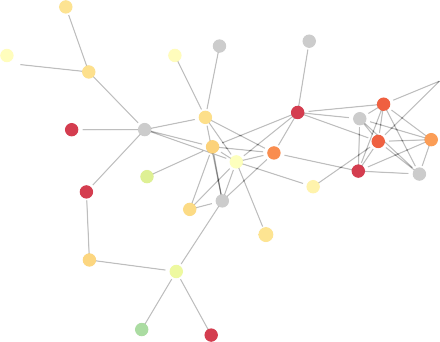
IL6R

PDGFA

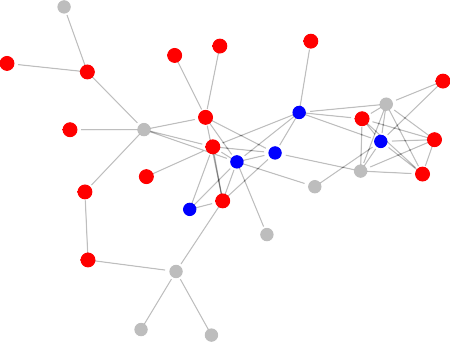
**monocytes**

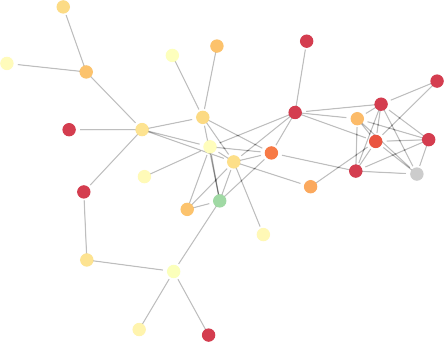
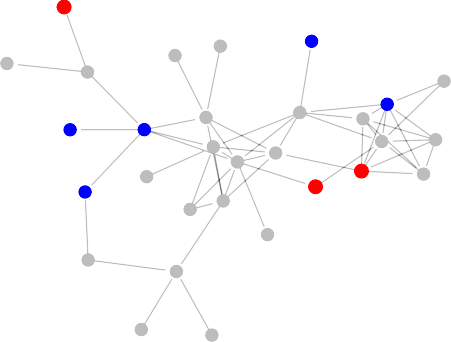
**mCD4+**

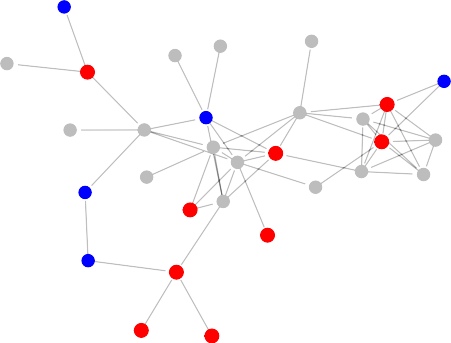
**mCD8+**



NMI



Figure 5.12: Protein network analysis based on the immune qPCR array expression data. The list of all the genes quantified in the qPCR array genes together with the best pval for significance of the mean FC across the three cell types was used to perform gene network analysis. STRING interaction network (including known and predicted proteinprotein interactions) was used to superpose the aforementioned list of genes and obtain a 30 gene size subnetwork common for all three cell types. This included maximal number of significant genes (pval*<*0.05) in at least one cell type and minimal presence of non-significant genes as linkers in the network. In the left hand panel, for each cell type each of the nodes (proteins) of the identified subnetwork (the same for each cell type, as previously explained) are colour-coded by the change of expression (log2 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 (log2 mean FC) in the qPCR array.

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 peripheral blood in three healthy controls using the same qPCR array. In each of the cell types, the FC in was calculated for the mean peripheral blood expression across the three PsA patients compared to the mean expression of the three healthy controls (as detailed in Chapter 2). Similar to the previous analysis, pvals for the FC significance were calculated for each particular genes. Integration of the previous results of modulated gene expression between synovial fluid and peripheral blood in PsA (see Immune-relevant gene expression by qPCR) with this analysis allowed the identification of three group of genes (Figure 5.13). First, the genes only significantly modulated (based on pval and FC threshold criteria) in

PB between controls and PsA were designated as systemic genes (Figure 5.13 green dots). Those genes were not significantly modulated in the prior analysis comparing synovial fluid versus peripheral blood within PsA patients and could then be considered as the circulating disease ”footprint”. In this respect, CD14+ monocytes was the cell type with lower number of systemic modulated genes (14), compared to mCD8+

(23) and mCD4+ (42) (Figure 5.13a, b and c). Amongst these genes were *CCL24*

and *CCL27* in CD14+ monocytes, *CCR7* and *TLR4* in mCD4+ cells and *CCR10* in mCD8+ cells.

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

**CD14+ monocytes**

|  |  |
| --- | --- |
| ***CCL2*** | **Putative disease-specific: 10**  **Systemic: 14**  **Tissue-specific: 62 None**  ***PRG2 GPI*** |
|  |  |

18

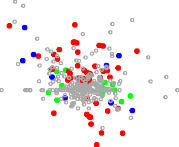
16

14

12

10

8



**PsA (SF vs PB)**

**(log2mean FC)**

6

4

2

0

−2

−4

−6

**mCD4+**

18

|  |  |
| --- | --- |
|  | **Putative disease-specific: 9**  **Systemic: 42**  **Tissue-specific: 37 None**  *GPR68* |
|  |  |

16

14

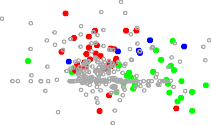
12

10

**PsA synovial fluid vs peripheral blood (log2mean FC)**

8

6



4

2

0

−2

−4

−6

−4 −2 0 2 4 6 8 10

**CTL (PB) vs PsA (PB)**

**(log2mean FC)**

(a)

18

|  |  |
| --- | --- |
|  | **Putative disease-specific: 8**  **Systemic: 23**  **Tissue-specific: 68 None** |
|  |  |

16

14

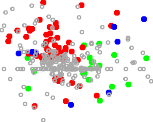
12

10

**PsA synovial fluid vs peripheral blood (log2mean FC)**

8

6



4

2

0

−2

−4

−6

**mCD8+**

−4 −2 0 2 4 6 8 10

**CTL (PB) vs PsA (PB)**

**(log2mean FC)**

(b)

−4 −2 0 2 4 6 8 10

**CTL (PB) vs PsA (PB)**

**(log2mean FC)**

(c)

Figure 5.13: Comparison of immune-relevant gene expression modulation across PsA tissues (SF vs PB) and in PsA patients verus healthy controls. The qPCR log2 mean FC for each of the genes in the PsA synovial fluid vs peripheral blood contrast are plotted against the log2 mean FC for the same genes in the PsA peripheral blood vs healthy control peripheral blood 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 peripheral blood between controls and PsA (systemic genes), only significantly modulated between synovial fluid and peripheral blood in PsA patients (tissue-specific) and significantly modulated between controls and PsA patients in peripheral blood as well as between synovial fluid and peripheral blood in PsA patients (putative disease-specific).

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 peripheral blood (Figure 5.13 a red dots). For example, the aforementioned *NFKB* and *MYD88*, *TLR2* genes were only up- regulated in PsA synovial fluid CD14+ monocytes and their expression was not significantly modulated between healthy controls and PsA circulating CD14+ monocytes. Similarly to CD14+ monocytes, mCD8+ cells also showed greater disease tissue-specific modulation than genes differentially expressed when compared to controls in peripheral blood (Figure 5.13 c red dots). For all three cell types, the two genes presenting the greatest FC between synovial fluid and peripheral blood *SPP1* and *FN1* appeared to be tissue-specific genes and no significant changes in peripheral blood between PsA and healthy controls were identified.

The third category comprised genes significantly modulated for each cell type between controls and PsA patients in peripheral blood as well as between synovial fluid and peripheral blood in PsA patients. These genes defined as putative disease-specific genes showed similar numbers across CD14+ monocytes, mCD4+ and mCD8+ (10, 9 and 8, respectively) (Figure 5.13 blue dots in a, b and c). In CD14+ monocytes two of those genes, *GPI* and *PRG2*, were up-regulated in both comparisons, with further exacerbation in synovial fluid (Figure 5.13 a). Evidence of the glucose-6- phosphate isomerase *GPI* up-regulation in disease has been found in RA synovial fibroblasts and linked to increased levels of TNF-*α* and IL-1*β* in the synovium (Zhong2015). Another example of exacerbated up-regulation in synovial fluid was the expression of *GPR68* in mCD4+. This gene was up-regulated in PsA peripheral blood mCD4+ when compared to the control counterparts and further up-regulated in synovial fluid when compared to peripheral blood in PsA individuals (Figure 5.13 b). *GPR68* was also up-regulated in synovial fluid compared to peripheral blood 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 peripheral blood compared to the controls but is down-regulated in PsA individuals when comparing synovial fluid versus peripheral blood (Figure

5.13 c). Despite the interesting aforementioned findings, the identification of disease-specific and disease tissue-specific genes is clearly limited by the impossibility of obtaining healthy controls synovial fluid to include in the experimental design.

When performing pathway enrichment analysis using the significantly modulated genes between healthy controls and PsA patients peripheral blood 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 synovial fluid and peripheral blood in CD14+ monocytes PsA patients and clearly suggest a more pronounced inflammatory phenotype of the pathological CD14+ monocytes in synovial fluid compared to PB.

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

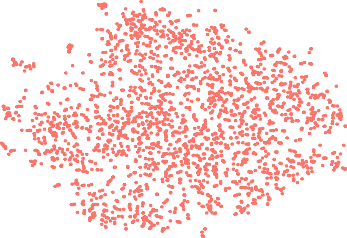
According to the analysis of chromatin accessibility and immune-related gene expression in this pilot cohort, CD14+ monocytes showed the greatest changes in chromatin accessibility and the most reliable modulation of expression for pro-inflammatory chemokines and cytokines between peripheral blood and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

scRNA-seq was performed in paired PBMCs and SFMCs isolated from three PsA patients (Table 5.2). scRNA-seq data from each of the PBMCs and SFMCs samples were filtered (as explained in chapter 2) and CD14+ monocytes were distinguished from other cell populations by expression of *CD14* and *LYZ*, two of the most accurate expression markers defining this cell population (Figure B.9 a and b). Across all six samples (three SFMCs and three PBMCs), 2,459 cells were CD14+ monocytes cells, representing approximately 17% of the bulk SFMCs and PBMCs cells included in the analysis and in line with the proportion of CD14+ monocytes previously reported using cell surface markers by mass cytometry (Figure 5.1). The CD14+ monocytes identified in each of the three paired PBMCs- SFMCs PsA samples were combined using CCA to correct for intrinsic batch effect, unavoidable due to patient samples recruitment on different days and generation of SFMCs and PBMCs 10X libraries separately. CCA alignment of the six CD14+ monocytes populations was followed by conservative unsupervised clustering (using resolution 0.1) and t-SNE visualisation.

40



20

**t-SNE2**

 **CC-mixed**

0 **CC-IL7R**

−20

−20 0 20

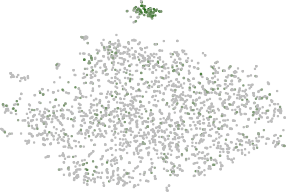
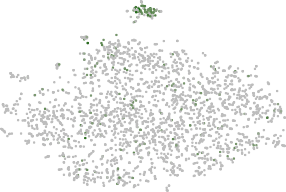
**t-SNE1**

(a)

***IL7R***

***IL32***

40



40

20

0

−20

20

**t-SNE2**

**t-SNE2**

0

−20

−20 0 20

**t-SNE1**

−20 0 20

**t-SNE1**

(b)

Figure 5.14: Identification of two main CD14+ monocytes subpopulations in the synovial fluid and peripheral blood combined analysis. a) Visualisation using t-SNE dimensional reduction of the two cluster (CC-mixed and CC-IL7R) identified in the combined synovial fluid and peripheral blood 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 synovial fluid and peripheral blood CD14+ monocytes.*IL32* and *IL7R* gene expression appeared as markers for the CD14+ monocytes from the CC-IL7R cluster.

Using this conservative approach for cluster definition, two robust clusters were identified (Figure B.9 a). The smallest cluster, named 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 synovial fluid versus 29 from PB) (Figure ?? b and B.8) ((Al-Mossawi et al. 2018), in revision). The proportion of IL7R+CD14+ monocytes when compared to the totalCD14+ monocyte population was very similar in synovial fluid and peripheral blood (3 and 2.7%, respectively) in this data. The largest cluster, named as CC-mixed, consisted of 2,387 (1,356 synovial fluid and 1,031 PB). CC-mixed

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

Differential gene expression between synovial fluid and peripheral blood CD14+ monocytes in CC- mixed and CC-IL7R

DGE analysis was performed in order to explore differences between synovial fluid and peripheral blood within each of these two main CD14+ monocyte subpopulations. For the CC- mixed cluster, a total of 251 genes were differentially expressed at an FDR*<*0.01 and FC*>*1.5 between synovial fluid and peripheral blood, of which 149 and 102 showed up- and down- regulation, respectively (Figure 5.15 a). Differential analysis within the CC-IL7R

cluster revealed a total of 37 modulated genes, with the majority (35 out of 37) up-regulated in synovial fluid compared to peripheral blood (Figure 5.15 a). Due to the low number of cells in the CC-IL7R cluster and the limited sample size (n=3), the analysis only identified as significantly differentially expressed (FDR*<*0.01) genes presenting FC*>*1.5. Out of the 37 DEGs in the CC-IL7R cluster between the two tissues, 30

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

**CC-mixed**

100

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *CTSB* | *FN1* |  |
| *CTSS* |  | *MSB10* |  |  |
| *S100A12* |  | *FTH1* |  |  |
|  |  |  |  | *SPP1* |
| *TKT* |  |  |  |  |
|  |  | *S100A10 FABP5* |  |  |
| *DUSP1*  *CSTA S100A9*  *EEF1A1* | L | *S100A11 CSTB TUBA1B*  ALS1 *HLA−DRA*  *FTL GLUL*  *LDHA FPR3* |
| *FOS* |  | *EMP3* |
|  |  |  | *CCL2* |  |
|  |  |  |  | **Not sig. FDR<0.01 FDR<0.01 &**  **log2FC>1.5** |
|  |  |  | | |

80

**−log10 (FDR)**

60

40



20

0

−3 −2 −1 0 1 2 3 4 5 6

**log2 (FC)**

(a)

**CC-IL7R**

15

|  |  |  |
| --- | --- | --- |
| *RPS29* | *TMS* | *S100A10*  *VIM*  *EMP3*  *HLA−DRA*  *KLF6 FTL TYMP FN1*  *CRIP1 CTSB FTH1 HLA−DPB1*  *S100A11 CD74 FCGR3A*  *S100A6 LGALS1 SPP1*  *CD44 HLA−DPA1 B10 TUBA1A*  *H3F3B ANXA2* |
|  |  | **Not sig. FDR<0.01 FDR<0.01 &**  **log2FC>1.5** |

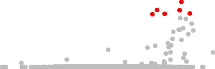
10

**−log10 (FDR)**

5

0

−3 −2 −1 0 1 2 3 4 5 6



**log2 (FC)**

(b)

Figure 5.15: Sc-RNA-seq differential gene expression results between synovial fluid and peripheral blood in the CC-mixed and CC-IL7R CD14+ monocytes subpopulations. Vulcano plots showing differences in gene expression between synovial fluid and peripheral blood a) CC-mixed and b) CC-IL7R CD14+ monocyte cluster. In a and b, the significance (log10FDR) of the differential expression (y-axis) is plotted against the log2FC. A positive FC indicates higher expression in CD14+ monocytes from synovial fluid compared to PB. Genes showing FDR*<*0.01 are coloured in blue and genes presenting and FDR*<*0.01 and FC*>*1.5 are coloured in red. The most significant genes are labelled.

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 (pval*<*0.05) detected by qPCR between synovial fluid and peripheral blood in CD14+ monocytes, only 12 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 synovial fluid and peripheral blood by the two approaches included genes with the largest FCs in both assays, such as *SSP1*, *FN1*, *OLR1* and *S100A12*, being the direction of change also consistent for all of them.

6

**qPCR**

**CC−mixed CC−IL7R**

5

4

3

2

**log2(FC)**

1

0

−1

−2

−3

*CCL2*

*CCR1*

*CD74*

*FN1*

*FOS*

*MIF*

*OLR1*

*SPP1*

*TYMP*

**Shared DEGs**

*CSF3R*

*S100A12*

*S100A8*

(a)

7

6

5

4

3

2



1

0

−1

−2

−2 −1 0 1 2 3 4 5

**ATAC (log2FC)**

(b)

Figure 5.16: Correlation between scRNA-seq, qPCR and chromatin accessibility in PsA CD14+ monocytes. a) Overlap between the scRNA-seq DEGs (significant based only on FDR*<*0.01) from the CC-mixed and CC-IL7R CD14+ monocyte cluster in synovial fluid versus peripheral blood and the corresponding DEGs (pval¡0.05) detected by qPCR in the bulk CD14+ monocytes population. For each of the shared DEGs, the log2FCs from the qPCR and the 10X scRNA-seq analysis are represented. b) Correlation plot comparing synovial fluid and peripheral blood differences in scRNA-seq expression of the CC-mixed CD14+ monocytes and ATAC chromatin accessibility in total CD14+ monocytes. The log2FCs for scRNA-seq differential expression of all transcripts in the CC-mixed CD14+ monocytes are plotted against the log2FC for total CD14+ monocytes ATAC differential chromatin accessibility analysis in regions proximal ( 5Kb) to the same genes. Blue colouring indicates significant differential expression in scRNA-seq only; green represents ATAC significant DAR only; red indicates significant differential expression and chromatin accessibility; grey indicates no significant differential expression or chromatin accessibility in CD14+ monocytes. Significance is based on FDR and FC thresholds (FDR*<*0.01 and FC*>*1.5) in both assays.

Pathway enrichment analysis was performed for the significantly DEGs between synovial fluid and peripheral blood in the CC-mixed and CC-IL7R subpopulations. The DEGs in the CC-mixed cluster were significantly enriched (FDR*<*0.01) a number of pathways (Table 5.7 and A.5), some of them of particular pathological relevance (Table 5.7).

Cluster Pathways

Ag processing and presentation via MHC-II Extracellular matrix and extracellular matrix-associated proteins

CC-mixed Phagosome and lysosome formation

IFN signalling Cytokine signalling ∗ Apoptosis

Innate immunity

Adaptive immunity

Ag processing and presentation

CC-IL7R Phagosome Extracellular matrix and extracellular

matrix-associated proteins

Table 5.7: Most relevant enriched pathways for the DEGs between synovial fluid and PBCD14+ monocytes in CC-mixed and CC-IL7R. Significantly DEGs based on the FDR and FC threhold were used for the analysis. Most relevant enriched pathways based on FDR*<*0.01. (∗) Enrichment for FDR*<*0.05. Additional pathways included in Table A.5.

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

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

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

In order to determine the overall correlation between scRNA-seq expression and chromatin accessibility, comparison between the log2FCs for all the expressed genes in the CC-mixed CD14+ monocytes cluster and all the accessible chromatin regions in bulk CD14+ monocytes between synovial fluid and peripheral blood was conducted

. Changes in expression and chromatin accessibility only showed a moderate correlation in this data (R=0.214, pval=2x10−16)(Figure 5.16 b). In the CC-mixed

cluster, 64 genes out of the 251 DEGs (FDR*<*0.01 and FC*>*1.5) were proximal (≤ 5Kb) to one or more ATAC DARs (Table 5.8). This overlap was significant and

highlighted the enrichment (Fisher exact test pval=1.5x10−3) of DEGs in the CC-

mixed cluster for proximal DARs identified by ATAC in bulk CD14+ monocytes. The majority of the overlaps corresponded to matched increase or decrease (40 and 12 genes, respectively) of gene expression and chromatin accessibility in synovial fluid vs PB. However, 14 DEGs in the CC-mixed cluster showed opposite direction of change between expression and chromatin accessibility (Table 5.8).

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

Cluster Up-regulated genes Up-regulated genes Opposite direction

|  |  |  |  |
| --- | --- | --- | --- |
|  | with proximal | with proximal | in expression |
| SF open DAR | PB open DAR | and DAR |
|  |  |  |  |
| CC-mixed | 40 | 10 | 14 |
| CC-IL7R | 9 | 0 | 4 |

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

up-regulated in the same direction by the qPCR expression analysis in synovial fluid bulk CD14+ monocytes compared to peripheral blood (Figure 5.16 a).

Similarly to the CC-mixed results, the DEGs in the CC-IL7R cluster

between synovial fluid and peripheral blood were also enriched for those genes with at least one DARs nearby (Fisher exact test pval=1.85x10−9). Amongst the 22 DEGs overlapping a proximal DAR, 13 of them had correlated dysregulation of expression and chromatin accessibility only 4 showed opposite directionality in the variation of the two features (Table 5.8). The CC-IL7R cluster and CD44 data not shown.

Overall, these 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.

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

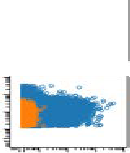
To determine the effect of chromatin accessibility and genes expression differences in CD14+ monocytes between synovial fluid and peripheral blood at the protein level mass cytometry was conducted in collaboration with Dr Hussein Al-Mossawi and Dr Nicole Yager. For those samples with paired ATAC and/or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

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

Mass cytometry expression of CD14+ versus TNF-*α* demonstrated greater

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

**No BFA 6h BFA**



**TNF-α**

(a)

13

12

11

10

9

8

7

6

5

4

3

2

1

0

PB SF

**Tissue**

(b)

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

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

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

1.8

1.6

1.4

1.2

1.0

0.8

0.6

0.4

0.2

0.0

PB SF

**Tissue**

(a)

44

42

**pval=0.048**

**pval=0.001**

40

38

36

34

**CD14+ monocytes producing osteopontin (%)**

32

30

28

26

24

22

20

18

16

14

12

10

8

6

4

2

0

PB SF

**Tissue**

(b)

24

**pval=0.003**

3

2

1

0

PB

SF

20

16

**CD14+ monocytes producing MCP-1 (%)**

12

8

4

0

PB SF

**Tissue**

(c)

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

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

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

of *CCL2* was shown to be significantly modulated (pval*<*0.05 and mean FC*>*1.5) between synovial fluid and peripheral blood by qPCR in CD14+ monocytes only, whereas no significant changes were observed for mCD4+ and mCD8+ in this data from the same patients (Figure 5.10 a, b, c and Table 5.5). Up-regulation of *CCL2* was not found in peripheral blood CD14+ monocytes compared to PsA patients and healthy controls, being defined as one of the tissue-specific genes in the previous analysis (Figure

?? a). Furthermore, *CCL2* was also identified by scRNA-seq as one of the up- regulated genes in the CC-mixed cluster (Figure 5.15 a and 5.16 a and b). Expression of *CCR2*, the receptor for the chemokine MCP-1 (protein product of *CCL2*), appeared up-regulated by qPCR in synovial fluid mCD4+ and mCD8+ cells in the same individuals, which could suggest increased chemotaxis driven by CD14+ monocytes and leading to T cell infiltration in the synovium. Interestingly, in this data no significant up-regulation of *CCR2* was observed in PsA peripheral blood when compared to healthy controls in any of the three cell types.

**UCSC genes**

Scale chr17:

2 kb hg19

32,579,000 32,580,000 32,581,000 32,582,000 32,583,000 32,584,000



*CCL2*

**ChromHMM RoadMap**

13.03 \_

**Roadmap H3K27ac**

0.22 \_

23.6162 \_

**Roadmap H3K4me1** 0 \_

**ATAC peaks**

72 \_

**PSA1718 SF**

0 \_

72 \_

**PSA1719 SF**

0 \_

39 \_

**PSA1607 SF**

1 \_

72 \_

**PSA1718 PB**

0 \_

72 \_

**PSA1719 PB**

0 \_

39 \_

**PSA1607 PB**

0 \_

(a)

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

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

## Prioritisation and interpretation of PsA GWAS SNPs

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

Bayesian fine-mapping using genotyping data

In order to further refine the PsA Immunochip GWAS signals identified by Bowes and colleagues, Bayesian fine-mapping was conducted using genotype data from 1,103 patients and 8,900 controls (PsA Immunochip UK cohort from 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 se were provided in the supplementary material (Bowes et al. 2015). Nevertheless *de novo* fine-mapping was performed here to obtain the identity of the SNPs in the credible set and integrate with chromatin accessibility data. Compared to the fine-mapping performed in Bowes *et al.*, 2015, the power of the analysis presented here was limited by a smaller sample size (only UK cohort as previously mentioned).

Fine-mapping was performed in thirty-six loci reported in Bowes *et al.*, all showing at least nominal significance in their GWAS study. As expected, the

GWAS signals with the lowest significance for the lead SNP (pval*<* 10−4) also

showed -log10*ABF* under 3 (cut-off used in (Bunt et al. 2015)) for the lead SNP

in the fine-mapping association analysis and were not taken forward for the calculation of PP and credible set of SNPs.

Cross-tissue comparison analysis in PsA

263

Table 5.9: Summary table of the PsA GWAS loci presenting -log10ABF*>*3 for the fine-mapping lead SNP. For twelve PsA loci -log10*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 (r2 *<* 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 | -log10*ABF*  FM lead SNP | PP | 90% credible  set | Bowes FM  lead SNP | Bowes 90%  credible set |
|  |  |  |  |  |  |  |  |  |  |
| 2 | *IFIH1* | rs13406089 | 0.33 | 0.78 | 4.58 | 0.48 | 2 | rs35667974 | 4 |
| 5 | *IL12B/ADRA1B* | rs2546890 | 0.48 | 0.76 | 6.53 | 0.6 | 23 | rs4921482 | 3 |
| 5 | *IL13* | rs2069616∗ | 0.44 | 1.25 | 5.16 | 0.05 | 55 | NA | NA |
| 1 | *IL23R* | rs12044149 | 0.25 | 1.41 | 9.83 | 0.14 | 29 | rs12044149 | 34 |
| 1 | *IL28RA/GRHL3* | rs2135755∗ | 0.50 | 1.20 | 3.06 | 0.03 | 49 | NA | NA |
| 19 | *ILF3* | rs11085727∗ | 0.30 | 0.79 | 3.83 | 0.22 | 35 | NA | NA |
| 17 | *PTRF* | rs730086∗ | 0.34 | 0.81 | 3.39 | 0.39 | 400 | NA | NA |
| 1 | *RUNX3* | rs6600250 | 0.50 | 1.20 | 3.07 | 0.03 | 48 | rs7523412 | 52 |
| 12 | *STAT2/IL23A* | rs12368739 | 0.06 | 1.70 | 4.04 | 0.02 | 110 | rs2020854 | 121 |
| 6 | *TRAF3IP2* | rs33980500 | 0.07 | 1.71 | 8.26 | 0.87 | 2 | rs33980500 | 7 |
| 19 | *TYK2* | rs11085727 | 0.30 | 0.79 | 3.83 | 0.21 | 32 | rs34725611 | 5 |
| 5 | *CSF2/P4HA2* | rs11242104 | 0.48 | 1.24 | 5.31 | 0.07 | 58 | rs715285 | 35 |

In this analysis, fine-mapping was unsuccessful (-log10*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. Of the twelve loci passing the -log10*ABF* ≥ 3 cut-off, eight had been successfully fine-mapped by Bowes and colleagues (Table 5.9). For two loci, *IL23R* and *TRAF3IP2*, the same fine-mapping lead SNPs were reported by the two analyses. For the *IL12B* locus, the second signal reported by Bowes and colleagues was also identified here by step-wise conditional analysis. For the four fine-mapped loci reported here but not by Bowes *et al.*, the fine-mapping lead SNPs were in very low LD (r2 *<* 0*.*5) with the PsA GWAS lead SNP, suggesting identification of spurious signals or signals from other loci nearby in the association analysis (Table 5.9 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 comprised a total of 294 unique SNPs. These SNPs were used to perform overlap with the significant (FDR*<*0.01 and FC*>*1.5) DARs identified between synovial fluid and peripheral blood in CD14+ monocytes, mCD4+, mCD8+ and NK cells. Unfortunately, none of the 294 SNPs were spanned by a DAR in any of these cell types. Additional overlap was performed between these SNPs and the accessible chromatin regions (consensus peaks without filtering based on the differential analysis) in each of the four cell types assayed by ATAC. The largest number of SNPs (17) was found to overlap accessible chromatin in CD14+ monocytes, followed by mCD8+, mCD4+ and NK cells (Table 5.10). The 43 unique SNPs from the 90% credible

set overlapping ATAC accessible chromatin were located across the *CSF2* (8),

*IL12B* (3), *IL23R* (4), *RUNX3* (6), *STAT2* (14), *TRAF3IP2* (1) and *TYK2* (7) loci.

A number of these SNPs were found to only overlap accessible chromatin in one particular cell type.

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

Table 5.10: PsA fine-mapped SNPs from the 90% credible sets overlapping accessible chromatin identified by ATAC in four cell types. The number of SNPs in the 90% credible set union from the eight fine-mapped loci overlapping each cell type ATAC master list are reported. Furthermore, the number of SNPs only found to overlap open chromatin in one cell type are indicated together with the locus in which the SNP was fine-mapped.

SNPs from the fine-mapping credible set were significantly enriched in ATAC peaks when compared to all other GWAS Catalog SNPs and SNPs in LD (r2=8) in all four cell types (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 Catalog SNPs overlapping ATAC accessible regions were significantly enriched (FDR*<*0.001) for particular terms from the Experimental Factor Ontology (EFO) (Figure 5.20). The EFO is a hierarchical tree-like ontology where each term represents a (disease) trait or group of related (disease) traits with which disease-risk SNPs may be annotated (Figure 5.20). Enrichment for general terms (towards the root of the tree) including autoimmune diseases, rheumatic diseases and skin diseases were found across all four cell types. Disease-specific terms (amongst the branches of the tree) related to these general terms, such as CD and IBD, were also found to be enriched for SNPs

overlapping ATAC in all four cell types. Conversely, other ”branches” from more general terms, including psoriasis and MS, showed significant enrichment (FDR*<*0.001) only in CD14+ monocytes and mCD4+ cells, respectively. Overall, this reinforced the specificity of the overlap between GWAS Catalog genetic variants not included in the fine-mapping credible set with accessible chromatin across the immune cell types investigated in this chapter.

**autoimmune disease digestive system disease inflammatory bowel disease**

**skin disease Crohn's disease skeletal system disease rheumatic disease**

**neoplasm lymphoid neoplasm hematological system disease**

**ulcerative colitis**

**cancer protein measurement**

**lipid or lipoprotein measurement cardiovascular measurement CVD biomarker measurement**

**lipid measurement**

**allergy triglyceride measurement lipoprotein measurement hematological measurement**

**platelet count inflammatory skin disease**

**infectious disease platelet measurement**

**bacterial disease high density lipoprotein cholesterol**

**glycoprotein measurement**

**atopic eczema eye measurement**

**psoriasis endocrine neoplasm**

**multiple sclerosis systemic lupus erythematosus**

**NK mCD8+ CD4+ CD14+**

|  |  |  |  |
| --- | --- | --- | --- |
| FDR=1.4e−10 FDR=9.1e−8 FDR=4.3e−6 FDR=1.3e−4 FDR=2.0e−4 FDR=3.8e−4  FDR=4.8e−4 | FDR=1.4e−10 FDR=9.1e−8 FDR=4.3e−6 FDR=1.3e−4 FDR=2.0e−4 FDR=3.8e−4  FDR=4.8e−4 | FDR=1.1e−20  FDR=6.3e−11  FDR=1.7e−11 FDR=6.0e−4 FDR=4.6e−8 FDR=2.5e−4 FDR=2.9e−5 | FDR=3.1e−13  FDR=1.1e−12  FDR=3.5e−12 FDR=2.2e−5 FDR=1.5e−9 FDR=6.6e−5 FDR=3.4e−4 |
| FDR=7.2e−5 FDR=8.2e−5 FDR=2.3e−4 FDR=2.8e−4  FDR=3.5e−4 | FDR=7.2e−5 FDR=8.2e−5 FDR=2.3e−4 FDR=2.8e−4  FDR=3.5e−4 |  | FDR=2.7e−6 FDR=2.8e−4 FDR=1.0e−4 FDR=3.7e−7  FDR=4.4e−6 |
|  |  | FDR=4.6e−8 FDR=6.0e−4 | FDR=2.9e−12  FDR=2.3e−9  FDR=2.3e−9 FDR=1.1e−8 FDR=2.5e−8 FDR=1.6e−7 FDR=3.2e−7 FDR=2.0e−6 FDR=3.0e−6 FDR=1.5e−5 FDR=4.5e−5 FDR=1.8e−4 FDR=2.2e−4 FDR=4.3e−4 FDR=4.3e−4 FDR=4.4e−4 FDR=5.2e−4 FDR=6.9e−4 FDR=7.9e−4 FDR=9.5e−4 |

0 5 10 15 20 0 5 10 15 20 0 5 10 15 20 0 5 10 15 20

**Enrichment significance: -log10FDR**

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

Further investigating the PsA-specific 5q31 locus

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

rs7721882) of the four variants highlighted by Bowes and colleagues as the most functionally relevant according to ENCODE epigenetics data.

|  |  |  |
| --- | --- | --- |
| SNP | Cell type ATAC overlap | Top eGene , cell type and condition |
|  |  |  |
| rs10065787 | CD14+, mCD4+ | *P4HA2* (monocytes LPS2, LPS24, IFN-*γ*) |
| rs11242104 | All | *SLC22A5* (monocytes UT)  NA |
| rs11242105 | All | NA |
| rs2069803 | All | *SLC22A5* (CD4+(∗), CD8+) |
| rs27437 | CD14+, mCD4+ | *SLC22A5* (CD4+, CD8+) |
| rs4705908 | All | *SLC22A5* (CD4+(∗), CD8+) |
| rs2089855 | All | *P4HA2* (monocytes LPS2, LPS24, IFN-*γ*,)  *SLC22A5* (monocytes UT, IFN-*γ*, |

CD4+(∗), CD8+(∗))

rs7721882 mCD4+ *SLC22A5* (CD4+(∗), CD8+)

Table 5.11: Publicly available and unpublished *cis*-eQTL datasets reporting an effect for the PsA 5q31 GWAS locus fine-mapped SNPs (90% credible set) overlapping ATAC accessible regions. For each of the SNPs, the cell type for the ATAC overlap, the gene which expression is reported to be regulated by the SNP (eGene) and the cell type where the eQTL study was conducted are specified.The eQTLs datasets included in the analysis were monocytes (UT, LPS 2h, LPS 24h, IFN-*γ*) (Fairfax et al. 2014), B cells (Fairfax et al. 2012), NK untreated (Naranbhai et al. 2015), neutrophils untreated (unpublished), tCD4+ and tCD8+ (Kasela et al. 2017) and whole blood (Jansen et al. 2017). (∗) for eQTLs extremely significant (FDR*<*2.2x10−308).

The SNP rs10065787 overlapped accessible chromatin in CD14+ monocytes and mCD4+ cells, showing also moderate enrichment for the enhancer histone mark H3K4me1 in mCD4+ (Figure ?? right hand side panel red line). Similarly, the nearby SNP rs27437 overlapped an ATAC peak in CD14+ monocytes and mCD4+ and the same TFBS site cluster as rs10065787 (Table 5.11 and Figure ?? right hand side panel green line). rs10065787 appeared to be part of an eQTL signal for *SLC22A5* and *P4HA2* expression in unstimulated and stimulated monocytes (LPS2, LPS24, IFN-*γ*), respectively (Table 5.11). However, no *cis* eQTL for tCD4+ was reported for this SNP in the Kasela and colleagues dataset (Table 5.11). In contrast,

rs27437 was part of a *cis*-eQTL in tCD4+ and tCD8+ for *SLC22A5*, the same eGene reported by Bowes and colleagues in their pilot eQTL study. Chromatin conformation data using promoter capture-HiC (Javierre2016) in unstimulated monocytes does not clearly reveal interaction for rs10065787 with the promoter of *P4HA2* or *SLC22A5*. Conversely, rs27437 is relatively close to the bait in the *IL3* promoter, which interacts with *SLC22A5*, potentially bringing this SNP in proximity with the promoter of *SLC22A5*.

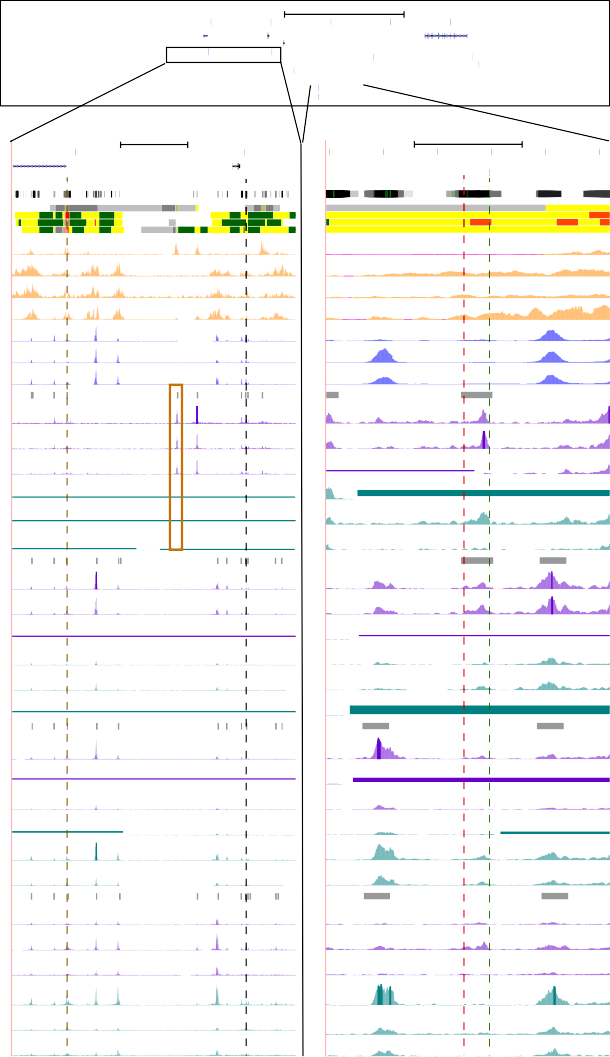
Other SNPs such as rs2089855 also appeared in an eQTL signal for *P4HA2* and *SLC22A5* in untreated and stimulated monocytes, and were associated with *SLC22A5* expression in tCD4+ cells (Table 5.11). This SNP is proximal to rs11955347, which in Bowes *et al.* showed the most significant correlation with expression of *SLC22A5* in tCD4+ and tCD8+, and has also shown to be within the *cis*-eQTL signal for *SLC22A5* expression in unstimulated and IFN-*γ* stimulated monocytes and for *P4HA2* in monocytes stimulated with LPS (2 and 24h) (Fairfax et al. 2014). The effect of rs11955347 genotype in modulation of *SLC22A5* expression was also confirmed in the larger data set of Kasela *et al.* in both tCD4+ and tCD8+ cells. These observations suggest a role for the 5q31 PsA- specific GWAS association in regulating expression of *P4HA2* and *SLC22A5* not only in T cells but also in monocytes under different conditions.

Another two relevant SNPs from the 90% credible set reported here were rs2069803 and rs4705908, both overlapping ATAC peaks in all four cell types (Figure ?? left hand side panel black and brown lines, respectively). rs4705908 is located upstream from the promoter of the *ACSL6* gene in a region showing H3K4me1 enrichment, supporting a regulatory role. Notably, rs4705908 maps to a CTCF binding site reported in GM12878 and LCLs cell lines. Likewise,

rs2069803 also overlaps moderate H3K4me1 signal in mCD4+, mCD8+ and NK

cells (Figure ?? left hand side panel). The region has also been annotated as an enhancer and weakly transcribed in mCD4+, mCD8+ and NK cells by the

100 kb hg19



131,400,000 131,450,000 131,500,000 131,550,000

ACSL IL3

CSF2

rs4705908 rs2069803 rs7721882

P4HA2

rs715285 rs11955347 rs2089855

rs3846728 rs27437

rs10065787

rs11242104 rs11242105

Scale chr5:

20 kb hg19

131,350,000 131,400,000

Scale chr5:

2 kb hg19

131,435,000 131,436,000 131,437,000 131,438,000 131,439,000

**UCSC genes** *ASCL5*

*IL3 CSF2*

rs3846728

rs27437

**Fine-mapped SNPs ENCODE TF**

rs4705908 rs2069803

rs10065787

**Roadmap chromatin segmentation**

**CD14 mCD4 mCD8**

**NK**

11.1699 \_

**CD14**

chromHMM tracks from Roadmap

E029 Primary monocytes from peripheral blood H3K4me1 foldChange signal

42 \_

0 \_ 2 \_

15.0838 \_

**mCD4**

E037 Primary T helper memory cells from peripheral blood 2 H3K4me1 foldChange signal

42 \_

**Roadmap H3K4me1**

**mCD8**

0 \_

12.86 \_ 0 \_

E048 Primary T CD8+ memory cells from peripheral blood H3K4me1 foldChange signal

2 \_

42 \_

2 \_

54.4461 \_

E046 Primary Natural Killer cells from peripheral blood H3K4me1 foldChange signal

42 \_

**Roadmap digital footprint**

**CD14+**

**monocytes**

**NK CD4**

**CD8 NK**

**Peaks ATAC CD14**

**PSA1718\_SF\_CD14**

**PSA1719\_SF\_CD14**

**PSA1607\_SF\_CD14**

**PSA1718\_PB\_CD14**

**PSA1719\_PB\_CD14**

**PSA1607\_PB\_CD14**

0 \_

586 \_

1 \_

1021 \_

1 \_

491 \_

1 \_

131 \_

1 \_

131 \_

0 \_

131 \_

0 \_

131 \_

0 \_

131 \_

0 \_

131 \_

0 \_

2 \_

306 \_

0 \_

306 \_

2 \_

306 \_

0 \_

56 \_

1 \_

57 \_

1 \_

56 \_

0 \_

56 \_

0 \_

56 \_

0 \_

56 \_

0 \_

**Peaks ATAC mCD4**

500 \_

96 \_

**mCD4+**

**PSA1718\_SF\_CD4**

**PSA1719\_SF\_CD4**

**PSA1607\_SF\_CD4**

**PSA1607\_PB\_CD4**

**PSA1718\_PB\_CD4**

**PSA1719\_PB\_CD4**

1 \_

500 \_

0 \_

500 \_

0 \_

500 \_

0 \_

500 \_

0 \_

500 \_

0 \_

1 \_

96 \_

1 \_

96 \_

0 \_

96 \_

0 \_

96 \_

0 \_

96 \_

0 \_



**Peaks ATAC mCD8**

457 \_

168 \_

**mCD8+**

**PSA1718\_SF\_CD8**

**PSA1719\_SF\_CD8**

**PSA1607\_SF\_CD8**

**PSA1607\_PB\_CD8**

**PSA1718\_PB\_CD8**

**PSA1719\_PB\_CD8**

0 \_

457 \_

0 \_

457 \_

0 \_

457 \_

0 \_

457 \_

1 \_

457 \_

1 \_

168 \_

0 \_

168 \_

0 \_

168 \_

0 \_

168 \_

0 \_

168 \_

**Peaks ATAC NK**

**PSA1719\_SF\_NK**

**PSA1718\_SF\_NK**

**PSA1607\_SF\_NK**

**NK PSA1718\_PB\_NK PSA1719\_PB\_NK**

**PSA1607\_PB\_NK**

0 \_

219 \_

0 \_

219 \_

1 \_

219 \_

0 \_

219 \_

0 \_

219 \_

0 \_

219 \_

0 \_

0 \_

123 \_

0 \_

123 \_

0 \_

123 \_

0 \_

123 \_

1 \_

123 \_

0 \_

123 \_

0 \_

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

Epigenome Roadmap chromatin segmentation maps (yellow and light green in

?? left hand side panel). Although accessible chromatin has been identified at rs2069803 and rs4705908 for all cell types, *cis*-eQTL for these two SNPs have only been reported in CD4+ or CD8+, with the genotypes of both SNPs correlating with regulation of *SLC22A5* expression with extremely high significance in tCD4+ cells (Table 5.11). Promoter capture Hi-C data in na´ıve and total CD4+ CD8+ cells revealed interaction of the *IL3* promoter bait containing rs2069803 with the promoter of *SLC22A5*. Interestingly, rs4705908 also appeared within the bait of the *ACSL6* promoter interacting with the *IL3* promoter bait, which also includes the previously mentioned rs2069803 variant. Overall, promoter-capture HiC data revealed potential physical interactions between rs27437, rs2069803 and rs4705908 in CD4+ and CD8+ cells with potential functional relevance in regulation of *SLC22A5* expression.

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

In addition to epigenetic evidence, further investigation of the suitability of *SLC22A5* and *P4HA2*, two of the genes showing eQTLs for some of the SNPs in the 5q31 credible set, as drug targets for PsA was investigated. Dr Hai Fang has developed an algorithm named Priority Index (Pi) based on random forest to levearage 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, amongs others) as well as prediction for high

druggability based on protein structure data ([http://galahad.well.ox.ac.uk:](http://galahad.well.ox.ac.uk/)

3010/pidb/discovery/PSO/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.

# Discussion

## Characterising chromatin accessibility in PsA samples

In this chapter the epigenetic landscape of immune cells isolated from peripheral blood and synovial fluid in PsA patients has been determined and compared in this chapter. The cell types included (CD14+ monocytes, mCD4+, mCD8+ and NK) are key players in the innate and adaptive immune response dysregulated in PsA pathophysiology. In particular, expansion of mCD4+ and mCD8+ cells in the synovium of PsA patients has been reported previosuly, and GWAS have highlighted significant association with MHC-I and other pivotal genes involved in T cell immune response. (Taams2018).

CD14+ monocytes were found to show the largest number of DARs between synovial fluid and peripheral blood (23.3% of the investigated regions). For all the cell types, the majority of the DARs were located at intergenic and intronic regions annotated as enhancers and were enriched for eRNAs identified in the FANTOM project (Fantom et al. 2014) consistent with differential regions having long-range effects on regulation of gene expression (Qu et al. 2015). Pathway enrichment analysis of the genes proximal to the DARs revealed both commonalities and differences across cell types. In synovial fluid, enrichment for genes in the NF-*κ*B pathway was found in CD14+ monocytes,

consistent with the TNF-*α* signature and the efficacy of anti-TNF therapies already reported for PsA (Ahil2016). Interestingly, in mCD8+ T cells enrichment

of DARs open in peripheral blood was found proximal to genes in the Wnt signaling pathway, such as *SMAD3*. Wnt signalling is involved in many biological processes and its dysregulation is associated with a number of autoimmune diseases, such as RA (Miao et al. 2013). In this data, the significantly increased chromatin accessibility near Wnt signalling genes in peripheral blood mCD8+ cells may be related to an increased recall proliferation capacity of the circulating cells compared to the tissue-resident (Boudousqui et al. 2014). In NK cells, increased accessibility proximal to gene members of the NK-mediated cytotoxicity pathway was found in peripheral blood compared to synovial fluid cells. NK CD56*bright* cells resident at sites of inflammation are more specialised for cytokine production than cytotoxicity (Michel et al. 2016). Notably, matched mass cytometry data in these patients have shown a reduced proportion of CD56*bright* cells in peripheral blood compared to synovial fluid which would be consistent with the NK-mediated cytotoxicity enrichment.

Overall, this approach has identified robust differences in chromatin accessibility between synovial fluid and peripheral blood in relevant immune cells isolated from PsA patients. This is in line with other studies that have revealed changes in chromatin accessibility between patients and healthy controls and also across different tissues involved in disease (Scharer et al. 2016; Wang et al. 2018; Corces et al. 2016). Although these findings suggest meaningful functional differences in chromatin accessibility between circulating and affected tissue cell populations, a limitation of pathway analysis in ATAC is linking putative regulatory regions to genes. Annotation of ATAC peaks with genes in proximity has been widely used in the literature (Scharer et al. 2016; Ackermann et al. 2016; Corces et al. 2016; Wang et al. 2018). However, this approach fails to evaluate long range interactions, assuming accessible chromatin regulates neighbouring regions. Moreover, accessible chromatin is not a definitive marker for regulatory regions and mapping of histone marks such as H3K4me1 and H3K27ac together

with eRNA quantification will help to refine the functional relevance of the identified DARs.

## Bulk gene expression profiling and integration with chromatin accessibility data

In contrast to chromatin accessibility profiling, recent research has investigated differences in the transcriptional profile between PBMCs, bulk T cells, SFMCs and synovial tissue from PsA patients (Dolcino et al. 2015; Fiocco et al. 2015). However, comparative analysis in matched discrete cell subpopulations from synovial fluid and peripheral blood from the same patient have yet to be reported.

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

The subsequent integration of transcriptome profiling with paired-ATAC data in CD14+ monocytes, mCD4+ and mCD8+ revealed that genes with differentially modulated expression in synovial fluid and peripheral blood corresponded with nearby DARs showing changes in the same direction. Some of those DARs had also been identified as eRNAs in CD14+ monocytes (e.g *NFKB1*) and mCD4+ cells (e.g *CCR6*) by the FANTOM project. Although the characterization of those DARs as

eRNAs evidences the downstream role of those regions in active transcriptional regulation, it does not unequivocally link this regulatory role to the proximal gene found to be differentially expressed between tissues by qPCR.

The integration of differences in peripheral blood gene expression between PsA patients and controls with the cross-tissue comparison within patients led to identification of systemic, tissue-specific and putative-disease specific modulated genes in this pilot data. According to my data, more profound transcriptional changes across PsA tissues (tissue-specific genes) were identified when compared to changes in expression between PsA patients and controls in peripheral blood for CD14+ monocytes and mCD8+ cells. Systemic genes for mCD8+ cells included for example *CCR10*, a chemokine receptor co-expressed by a subset of memory cells that preferentially migrate to skin, which has also been identified as an up- regulated gene in tCD8+ cells in the psoriasis cohort (Chapter ?? and in patients with atopic dermatitis (Hijnen et al. 2005). Interestingly, *SPP1* and *FN1* appeared in the tissue-specific category in the three cell types. Dolcino and colleagues had reported these genes as the two most dysregulated when comparing bulk synovial membrane transcriptome from PsA patients and healthy individuals (Dolcino et al. 2015). My data suggests that CD14+ monocytes, mCD4+ and mCD8+ cells may all contribute to the up-regulation of *SPP1* and *FN1* in the PsA synovium membranes. The *SPP1* protein product, osteopontin, is a cytokine and chemokine expressed by many cell types, including monocytes/macrophages and T cells. It is involved in cell migration, adhesion and cell-mediated immune response through regulation of T cells, importantly in Th-17 (Morimoto et al. 2010). *SPP1* also plays a role in other chronic inflammatory and autoimmune diseases (Rittling and research 2015). *FN1* encodes fibronectin-1, a main component of the cartilage matrix, involved in cell adhesion, migration, growth and differentiation and found to be highly expressed in RA inflamed synovium (Chang et al. 2005). Moreover, *FN1* has been shown to induce bone resorption

mediated by pro-inflammatory mediators, such as nitric oxyde and IL-1*β* (Gramoun et al. 2010). In Dolcinos data *FN1* up-regulation was only found when comparing synovial membranes of PsA versus controls, and no changes were reported in peripheral blood samples, altogether suggesting the tissue-specificity of this dysregulation in PsA. Furthermore, the identification of DARs open in synovial fluid at the promoter and 3 downstream of *FN1* in CD14+ monocytes may suggest a link between changes in modulation of gene expression and chromatin accessibility in this particular cell type.

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

An example of an interesting putatively disease-specific gene identified by this study was *GPR68*, up-regulated in mCD4+ PsA peripheral blood compared to controls with expression further increased in SF. *GPR68* is a G protein-coupled receptor (GPCR), expressed in T cells and others, that undergoes activation through pH acidification, characteristic of synovial tissues under inflammation (Saxena2011; Biniecka et al. 2016). Interestingly, *GPR65*, another member of the acid-sensing GPCR family, has been associated with a number of immune mediated diseases, including AS, CD and MS (Cortes et al. 2013; Lassen et al. 2016; Wirasinha et al.

2018). GPR65 was found to be a marker of pathogenic Th17 cells in the murine and human systems (Al-Mossawi 2017; Gaublomme et al. 2015). Unfortunately, GPR65 was not included in the gene array used in this study. Indeed, the use of a gene array rather than a transcriptomics approach using RNA sequencing is a limitation of this study.

In terms of relevant biological processes, pathway analysis using consistently modulated genes between synovial fluid and peripheral blood in this data revealed enrichment for TLR and NOD-like signalling pathways in CD14+ monocytes. This was consistent with the relevance of TLR and NOD-like receptors for rheumatic diseases (WJ and Arthritis 2009). Up-regulated expression of *TLR1* and *TLR2* was significant in synovial fluid CD14+ monocytes compared to peripheral blood and a similar trend in *TLR4* expression was also observed but failed to reach significance in this pilot study. These finding were in line with some studies that have identified increased *TLR-2* and *TLR-4* expression in SFMCs compared to PBMCs in patients with juvenile idiopathic arthritis (Myles and Aggarwal 2011). The relevance of NOD-like signaling has also been highlighted in the genome-wide

trancriptomic analysis in lesional and uninvolved psoriatic skin presented in Chapter ??, reinforcing the role of NOD-like signalling in the inflammatory response at the site of inflammation. The cross-talk between the TLR and NOD- like signalling pathways was further evidenced by network-based analysis in this data highlighting increased activation of the NF*κ*B TF in SF, particularly in CD14+ monocytes. In my data, enrichment of DARs open in synovial fluid in the proximity

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

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

## The relevance of monocytes and investigation of other cell types

In this exploratory data, CD14+ monocytes showed the most DARs and confidently modulated genes between synovial fluid and peripheral blood as well as functionally relevant pathway enrichment. The work presented is part of a collaborative multi-omics PsA pilot study. In light of this thesis, and for the cohort size available at the time of analysis, CD14+ monocytes were chosen as the cell type to further explore by scRNA-seq and integrate with mass cytometry data. mCD4+ and mCD8+ also showed relevant changes in chromatin accessibility and gene expression, and have been identified as the two cell types undergoing greater expansion in the synovium of PsA patients. Colleagues involved in this project have since performed differential TCR clonality analysis between synovial fluid and peripheral blood using PsA samples included in this thesis. Interestingly, a significant number of mCD8+

TCR clones with potentially shared antigen recognition across patients were found to be enriched in synovial fluid compared to peripheral blood (manuscript in preparation).

## Characterisation of monocytes by scRNA-seq

Monocytes are very plastic cell populations that undergo cell differentiation at the site of inflammation, with differences that may be better captured at the single-cell level. In this pilot experiment, cluster identification in scRNA-seq from combined synovial fluid and peripheral blood monocytes using a conservative approach (see 5.4.6) revealed two main subpopulations (CC-mixed and CC-IL7R). CC- mixed appeared as a large heterogenous cluster in contrast to the CC-IL7R subpopulation, characterised by cells consistently expressing *IL7R*, *IL32* and textitCCL5, amongst other markers. According to the scRNA-seq data, the CC-IL7R cluster represented around 3% of the total monocytes and had approximately the same number of cells from synovial fluid and peripheral blood (3 and 2.7%, respectively). Conversely, flow cytometry data in PsA and AS patients have shown up to 35% of the total synovial fluid CD14+ monocyte to be IL-7R+ versus approximately 1% in peripheral blood (in revision (Al-Mossawi et al. 2018)). This could be due to discrepancies between gene expression and protein translation acknowledged in the literature but may also be a consequence of lower sensitivity of scRNA- seq in quantifying gene expression (Liu et al. 2016; Islam et al. 2014). This may be leading to underestimate the CC-IL7R cluster size and the predominance of synovial fluid monocytes IL-7R+ in the contribution towards this cluster. Differences in sensitivity between qPCR and scRNA-seq may also partly explain the limited overlap of differentially expressed genes between synovial fluid and peripheral blood when using qPCR and scRNA-seq. Although the same top dysregulated genes (including *SPP1*, *FN1* or *OLR1*) were identified by both techniques, a modest number of significantly modulated genes found by qPCR were reproduced by the scRNA-seq analysis in the CC-mixed cluster. In terms of chromatin accessibility, the comparison of

FCs from contrasting synovial fluid and peripheral blood chromatin accessibility in CD14+ monocytes and scRNA-seq expression in the CC-mixed cluster only showed moderate correlation. This limited correspondence between chromatin accessibility and gene expression has also been reported by other studies and may be also the result of aforementioned limitations in annotating accessible chromatin with a putative target gene (Ackermann et al. 2016; Wang et al. 2018).

The identification of sub-populations of monocytes expressing IL-7R+ is

of biological interest as *IL7R* polymorphisms are associated with a number of chronic inflammatory diseases, including AS and MS (Cortes2007; Gregory et al. 2007). Although the role of IL-7 and IL-7R in mediating the immune response was only characterised in T cells, the relevance of IL-7R in CD14+ monocytes under LPS stimulation has been demonstrated in eQTL studies and also at the protein level in a manuscript under review, to which I have contributed (Fairfax et al. 2014; Al-Mossawi et al. 2018). Al-Mossawi and colleagues identified a distinct transcriptional profile of the IL7R cluster in PsA synovial fluid monocytes very similar to the gene expression profile from IL7R+ *in vitro* stimulated monocytes. Interestingly the *IL-7R* locus showed differentially accessible chromatin in PsA synovial fluid and was one of the top DEGs in the CD14+ monocyte qPCR array. Moreover, my analysis *CD44* appeared to be up-regulated in synovial fluid compared to peripheral blood in the CC-IL7R cluster. Interestingly, *CD44* is the receptor for osteopontin and this observation may suggest that the synovial fluid CC-IL7R cells may be more responsive to this cytokine. Taken together, this data supports a role for IL-7 signalling in PsA circulating and tissue monocytes in chronic inflammation.

Pathway enrichment analysis using scRNA-seq DEGs between synovial fluid and peripheral blood in the CC-mixed cluster identified biologically relevant processes, including MHC- II Ag processing, IFN signaling and extracellular-matrix components, amongst others. Interestingly, up-regulation of *IFI6* and *IFITM3*, two of the genes contributing to the enrichment of this pathway, have also been identified as

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

## Challenges and future perspectives in multi-omic approaches

The work presented in this chapter is an exploratory study and a proof of principle for the implementation of a multi-omics approach, which represents a very powerful strategy to dissect disease pathophysiology in a cell type specific manner. Nevertheless, a number of limitations and challenges were encountered and need to be taken into account to contextualise these results. One limitation is the small sample size (n=3) and the lack of paired data across all the techniques presented. This results from difficulties of recruiting PsA patients na´ıve for any treatment, the logistical difficulties to coordinate all of the techniques from the same sample, and the high cost of this approach. A

further limitation in this study is the lack of peripheral blood from healthy controls or synovial fluid from another autoimmune or non-inflammatory joint disease, as included by other studies (Fumitaka2018; Dolcino et al. 2015; Zhang et al. 2018). The definition and categorisation of the qPCR significantly modulated genes into systemic, tissue-specific and putative disease-specific was particularly limited by a lack of control samples to compare to the synovial fluid cells and the use of a biased transcriptomic analysis using a qPCR array. Another challenge in this study relates to the analysis and integration of scRNA-seq and mass cytometry data. Both techniques still represent emerging fields were no consensus has been reached on the best strategy to combine samples across patients and experiments, accounting for batch effect. In this exploratory study, monocytes were identified from each SFMCs and PBMCs scRNA-seq sample and combined using CCA for further subpopulation identification. However standard resolution for cluster identification yielded potentially spurious subpopulations, leading to the adoption of a more conservative approach to define clusters in this particular analysis. This may be the consequence of remaining batch effects, and alternative methods of combining samples from the different experiments should be investigated. In this respect, the identification of robust and stable subpopulations through cluster analysis will benefit from the implementation of algorithms designed for cluster validation such as Silhouettes, which has recently been used successfully in the field of single-cell (Rousseeuw 1987; Zhang et al. 2018). In addition to this, incorporation of bulk RNA-seq data from CD14+ monocytes will help interpretation and validation of the scRNA-seq results. In mass cytometry, to reduce batch effects patient samples are undergoing *ex-vivo* fixation and cryopreservation followed by simultaneous staining and barcoding. Moreover, different methodologies for cluster identification and annotation are also being explored and so far no clear clusters have confidently been found.

In terms of data integration, this pilot study used relatively simplistic multi-omics data integration limited by sample size, technical aspects and time scale, and provides a platform for future validation studies. A more systematic integrative approach should be implemented for the expanded cohort to establish appropriate correlation across datasets. Currently Zhang and colleagues have presented the most exhaustive methodology to integrate bulk-RNA-seq, scRNA- seq, mass cytometry and flow cytometry into multi-modal transcriptomic and proteomics profiles, but their work is still under peer review (Zhang et al. 2018). This strategy has revealed disease-specific functional expanded subpopulations amongst the most relevant cell types in RA pathophysiology. Additionally, the correlation between bulk ATAC and scRNA-seq is clearly limited by the different scales of the two approaches. Therefore, generation of scATAC-seq data, identification of clusters based on chromatin profiling and appropriate methods for the overlap with scRNA-seq populations should be used to have a better understanding of the correlation between chromatin accessibility and gene expression at the single-cell level (Duren et al. 2018).

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

Integration of epigenetic data with fine-mapped SNPs from GWAS studies has been widely demonstrated to be a powerful tool to further narrow down candidate causal variants, particularly for intergenic or intronic signals not driven by LD with missense coding SNPs (Bunt et al. 2015; Farh et al. 2014). Although DARs between synovial fluid and peripheral blood in four cell types did not show any overlap with SNPs from the credible set, significant enrichment of fine-mapped SNPs for accessible chromatin was demonstrated in all four cell types. A number of SNPs from the 5q31 credible set in my analysis overlap accessible chromatin

and eQTL signals in the same cell type. Integration of tCD4+ and tCD8+ eQTL

from Kasela *et al.* confirmed the association between SNPs from the 5q31 GWAS signal and T cells *SLC22A5* expression previously reported in a smaller study by Bowes *et al.*, and also suggested a potential role for the 5q31 PsA-specific GWAS association in regulating *P4HA2* and *SLC22A5* expression in unstimulated and stimulated monocytes. *SLC22A5* is a cell membrane transporter of carnitine involved in fatty acids metabolism and has been prioritised by an in-house pipeline as the third most promising druggable candidate for the treatment of psoriasis. Supporting evidence highlights the implications of this gene in other inflammatory conditions such as CD and the potential as a druggable therapeutic target in PsA (Leung et al. 2006).

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

## Conclusions

The strategy and analysis presented in this chapter illustrates how a multi-omics approach can be applied to clinical samples in a cell type and tissue- specific manner. The study of chromatin accessibility in immune cells from synovial fluid and peripheral blood of PsA patients has demonstrated differences across the two tissues and shown specific enrichment for relevant pathophysiological functional pathways.

Transcriptional analysis conducted on the same samples for a number of genes involved in the immune response has also revealed differential expression between the two tissues in all the cell type and shown some of those genes to be proximal to regions presenting changes in chromatin accessibility in the same direction. In this study, both data types highlighted CD14+ monocytes as the cell type presenting the largest number of significant changes in chromatin accessibility and consistent modulation of gene expression between synovial fluid and peripheral blood, with enrichment for pathways leading to inflammation through NF-*κ*B activation and subsequent cytokine and chemokine production. Implementation of scRNA- seq has shown the ability of this approach to identify sub-populations within synovial fluid and peripheral blood CD14+ monocytes. Overall, this chapter has shown that in PsA the pro-inflammatory environment at the site of inflammation drives changes in chromatin accessibility and gene expression that distinguish circulating cells from those infiltrating the involved tissue.