Chapter 2

Material and Methods

# Ethical approval and recruitment of study participants

## Psoriasis patient recruitment

Patient blood samples and skin biopsies were collected in collaboration with Professor Graham Ogg at the Weatherall Institute of Molecular Medicine, University of Oxford, and the Dermatology Department research nurses at the Churchill Hospital, Oxford University Hospitals NHS Trust. This was under approval from the Oxfordshire Research Ethics Committee (REC 14/SC/0106 and REC 14/NW/1153). After written informed consent, up to 60mL of blood from eligible psoriasis patients were collected into 10mL anticoagulant ethylenediaminetetraacetic acid (EDTA)-containing blood tubes (Vacutainer System, Becton Dickson). Detailed clinical information describing the recruited psoriasis cohort is included in Chapter ?? Table 4.2.

*Eligibility criteria*. Psoriasis patients were eligible for recruitment when aged 18 years or older, previously or newly diagnosed as having psoriasis (fulfilling the Psoriasis Area and Severity Index (PASI) classification) and in a flare (active disease state). Recruited patients were required to have moderate to severe disease (PASI*>*5), to not have taken antibiotics in the two weeks before sampling and to be na¨ıve for biological therapy.

## PsA patient recruitment

Sample recruitment was performed as part of the Immune Function in Inflammatory Arthritis (IFIA) study established in 2006 (REC/06/Q1606/139) in collaboration with Dr Hussein Al-Mossawi at the Botnar Research Centre and research nurses at the Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Trust. Following informed written consent, approximately 30mL of both blood and synovial fluid aspirate (variable upon disease severity) were collected into 10mL anticoagulant sodium heparin coated tubes (Vacutainer System, Becton Dickson). Further details about the cohort and collected clinical information can be found in Chapter 5 Table 5.1.

*Eligibility criteria.* Eligible patients were aged 18 years or older, previously or newly diagnosed with PsA according to the PsA Response Criteria (PsARC) (Table 1.2) and a physician global assessment questionnaire, with concomitant psoriasis and in a flare. Patients had to present an oligoarticular phenotype, not having taken antibiotics in the two weeks before sampling and be na¨ıve for biological therapy and preferably for any other treatment.

## Healthy volunteer recruitment

Recruitment of healthy volunteers was conducted as part of the Genetic Diversity and Gene Expression in White Blood Cells study with approval from the Oxford Research Ethics Committee (REC 06/Q1605/55). Up to 80mL of blood were collected into 10mL anticoagulant EDTA-containing blood tubes. Healthy individuals recruited in the study were required to be 18 years old or older, preferably British or European, without family history of psoriasis, PsA, RA or SpA and not having suffered from an infection in the two weeks prior to sample recruitment. Written consent was required. Detailed information describing the recruited volunteers is included [where?].

# Sample processing

Blood, synovial fluid and skin biopsies were processed straight after recruitment, following the appropriate protocols.

## PBMCs and synovial fluid cells isolation

PBMCs were isolated from blood samples through density gradient separation using Ficoll-Paque (GE Healthcare) with centrifugation at 500g for 30 minutes at room temperature with minimum acceleration and no braking. Total synovial fluid (SF) mononuclear cells (SFMCs) were isolated by centrifugation at 500g for 5 min at room temperature in absence of density gradient. Samples were placed on ice, washed twice in ice cold Hanks balanced salt solution (HBSS) without calcium or magnesium (Thermo Fisher Scientific) and resuspended in phosphate saline buffer (PBS, Gibco) supplemented with 0.5% fetal calf serum (FCS, Invitrogen) and 2mM EDTA (Sigma), prior to separation of the different cell types. Cell numbers and viability were determined by manual counting using a haemocytometer with trypan blue (Sigma) for viability assessment.

## Skin biopsies processing and adherent assays

Keratinocytes enrichment from skin biopsies was performed as described in Gutowska-Owsiak and colleagues (GutowskaOwsiak and Schaupp 2012). Skin biopsies (approximately 4mm) were washed with PBS, cut in 1mm width strips and incubated in 2U/mL of dispase II (Sigma) overnight at 4◦C. Following incubation, the epidermis was separated from the dermis. For RNA extraction, the epidermis was snap-frozen in liquid nitrogen. For chromatin accessibility assay, the epidermis was further digested in trypsin (Invitrogen) at 37◦C for 5 min to obtain a cell suspension that was filtered through a 70µm nylon strainer (BD) and washed with PBS. In some instances cells were manually counted and

aliquoted in PBS prior to chromatin accessibility assay. In others, cells from each of the biopsies were resuspended in KGM-2 BulletKit (Lonza) supplemented with 0.06mM Ca2+ and cultured in a collagen IV coated 96-well plate at 37◦C 5% CO2 for 3 hours. After culture, cells in 96-well plates were washed twice with 200µL of PBS and kept at 37◦C for downstream chromatin accessibility processing.

## Fixation, cryopreservation and cell culture

Cells (50,000) were fixed using dithio-bis(succinimidyl propionate) (DSP) as described in Attar and colleagues and stored at 4◦C for 24h (Attar et al. 2018). Liquid nitrogen storage of 40-50x106 PBMCs was carried out using a modified version of the cryopreservation protocol described by Kent and colleagues (Kent et al 2009), where cells were pre-conditioned in RPMI 1640 complete medium (Lonza) supplemented with 2 mM L-glutamine (Sigma), 100U penicillin and streptomycin 100µg/mL (Sigma) and 50% FCS for 30 minutes and, afterwards, diluted 1 in 2 in complete RPMI 1640 (supplemented as previously described) with 20% dimethyl sulfoxide (DMSO, Sigma). PBMCs followed slow cryopreservation at -80◦C at -1◦C per minute and then transferred

and stored for a minimum of two weeks in liquid nitrogen.

PBMCs were thawed quickly in a 37◦C water bath, resuspended in supplemented complete RPMI 1640 with 10% FCS at a density of 106 cells/mL and rested for 30 min at 37◦C, 5% CO2 in 25mL non-adherent polypropylene cell culture flasks (Greiner) followed by filtering through a 40µm cell strainer to obtain a homogenous cell suspension for FACS separation. Cryopreserved normal human epidermal keratinocytes (NHEK, Lonza) in passage three were recovered and cultured at a cell density of 5x106 cells/mL in a 75 mL adherent cell culture flask (Greiner) in EpiLife basal medium (Gibco), following manufacturer’s instructions. After recovery, NHEKs were trypsinised at room temperature for 8 minutes and trypsin was inactivated with EpiLife 10% FCS. Cells were centrifuged at 180g for 10 min at room temperature and then

manually counted with trypan blue for viability staining. NHEKs (16,000 cells) were seeded in a 96-well plate in 100uL of medium and cultured for 2 days at 37C, 5% CO2 to reach 90-100% confluence (approximately 50,000 cells) before performing any ATAC protocol on the plate (further detailed in Chapter 3). When used for Omni-ATAC, NHEKs after trypsinisation were processed through Ficoll density gradient (as previously explained for PBMCs isolation) to remove dead cells as recommended by (Corces et al. 2017).

## Primary cell isolation using magnetic-activated cell sorting

Primary cell subpopulations were separated using magnetic-activated cell sorting (MACS, Miltenyi) following the manufacturer’s instructions. Consecutive positive selection was performed using Miltenyi beads for CD19+, CD8+, CD14+ monocytes and CD4+ cells (catalogue numbers 130-050-201, 130-045-101, 130- 045-201 and 130-050-301, respectively) and AutoMACS Pro (Miltenyi) followed by a manual cell count with trypan blue. MACS separation was chosen over fluorescence-associated cell sorting (FACS) due to time and logistic constraints during sample processing.

## Primary cell isolation using fluorescence-activated cell sorting

Isolation of cell subpopulations from PBMCs and SFMCs to study the effect of cryopreservation in the chromatin landscape (Chapter 3) (Chapter 5 was performed by FACS. PBMCs and SFMCs were resuspended in 1mM EDTA PBS (FACS buffer) at 10x106 cells/mL, stained with the appropriate antibody

cocktail (Table 2.1) for 30 min at 4◦C, washed with FACS buffer and centrifuged

at 500g for 5 min at 4◦C. For the cell separation in the Chapter 3 samples, a modified FACS buffer supplemented with 3mM EDTA , 2% FCS and 25mM 4-

(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Invitrogen) was used to avoid cell clumping after cryopreservation and short recovery in culture (as detailed previously). After removing the supernatant, cells were resuspended in FACS buffer prior to separation.

From the control samples of Chapter 3, CD14+ monocytes and CD3+ CD14−

CD4+ T cells were isolated using the SONY SH800 cell sorter. For the PsA samples, separation of CD19+ B cells, memory T cells (CD3+ CD14− CD4+ CD45RA− and CD3+ CD14− CD8+ CD45RA−), CD14+ monocytes and CD56+

NK from PBMCs and SFMCs was performed using FACS Aria (BD) cell sorter. Sorted cells were collected in 1.5mL tubes containing PBS 1% FCS when used for ATAC-seq or only PBS when processed for scRNA-seq to avoid potential RNAse contamination. OneComp eBeads (eBioscience) were used for compensation of fluorescence spill over.

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

Table 2.1: Antibody panel used for FACS separation of primary cell populations in Chapter 3 controls and Chapter 5 PsA samples. Details regarding target molecule, fluorochrome, clone, supplier and dilution used for PBMCs and SFMCs staining are provided for each surface marker in the panel. For cell separation from Chapter 3, control PBMCs staining was only performed for CD3, CD4, CD14 and viability markers.

# Experimental protocols

## ATAC - Chromatin Accessibility

Three different versions of the ATAC-seq protocol were progressively used in this thesis for assessment of chromatin accessibility in different primary cells, including CD14+ monocytes, CD4+ and CD8+ T cells, CD19+ B cells and CD56+ NK cells, as well as in epidermal keratinocytes isolated from skin biopsies. Fast-ATAC and Omni-ATAC were two subsequent protocols published following the standard ATAC-seq protocol from (Buenrostro2013), aiming to reduce the amount of mitochondrial DNA in the sequencing libraries and improve the signal-to-noise ratio of the original protocol. When using MACS separation, primary cells were manually counted, as specified above, and resuspended in PBS 1% FCS.

ATAC-seq

ATAC-seq was used to generate data from NHEKs, skin biopsies (Chapter

??) and healthy volunteers to test the effect of cryopreservation in the chromatin landscape (Chapter ??) as well as cohort 1A primary immune cells isolated from blood of psoriasis and control samples (Chapter ??). ATAC-seq was performed using an estimated 50,000 cells as described in Buenrostro *et al.*, 2013 with minor modifications. Cells lysis was carried out for 10 min, the isolated nuclei were transposed for 40 min at 37◦C using the Nextera Tn5 transposase (Illumina) and DNA was purified with the PCR MinElute kit (Qiagen), following the manufacturer’s instructions. When using DSP fixed cells, two washes with 50µL of PBS were performed to remove any remaining fixative prior to the ATAC- seq protocol. After the transposition reaction, the Tn5 enzyme was inactivated with 500 mM EDTA for 30 min at 70◦C followed by de-crosslinking using 50 mM dithiothreitol (DTT) for 30 min at 37◦C and DNA column purification, as previously detailed. All transposed samples were simultaneously amplified

and singled indexed for 11 PCR cycles using modified Nextera primers from Buenrostro *et al.*,2013, after appropriate assessment of the approximate required number of qPCR cycles. The resulting DNA libraries were purified using the MinElute PCR purification kit (Qiagen) and a 1.8X (v/v) of Agencourt AMPure XP Magnetic Beads (Beckman Coulter) to remove adapters excess and primer dimers.

Additional modifications of the protocol were implemented when processing keratinocytes isolated from skin biopsies and NHEKs in 96-well plates (Bao et al. 2015) (as later detailed in Chapter 3).

Fast-ATAC

An improved ATAC-seq protocol was published in Nature Methods by Corces and colleagues (Corces et al. 2016), called Fast-ATAC. Optimised for hematopoietic cells, the protocol combined cell lysis and transposition into a single step. In this thesis, Fast-ATAC was performed in skin biopsies (Chapter ??), cohort 1A primary immune cells isolated from blood of psoriasis and control samples (Chapter ??) and primary immune cells isolated from blood and SF of PsA patients (Chapter ??). Fast-ATAC was conducted as described by Corces et al., 2016 with minor modifications. Following the advice from the authors, approximately 20,000 cells (MACS or FACS sorted) were washed with 200µL of PBS, centrifuged at 500g for 5 min at 4◦C and incubated in the lysis/transposition buffer containing digitonin (Roche) for 30 min at 37◦C and agitation at 400rpm. Following transposition, DNA was prepared and purified as per the protocol except 13 cycles of PCR amplification were used after appropriate cell cycle determination in a pilot set of samples.

Omni-ATAC

Omni-ATAC was performed in 50,000 viable NHEKs in suspension as described by (Corces et al. 2017). Transposed DNA was simultaneously amplified and indexed, as detailed in the ATAC-seq standard protocol, for 8 PCR cycles and purified using MinElute PCR purification columns (Qiagen) only.

Quality control and sequencing

Indexed and amplified ATAC samples were assessed for tagmentation profile on an Agilent 2200 or 4200 Tapestation with the D1000 high sensitivity DNA tape (Agilent) as part of the quality control. Quantification of the library concentration was performed by qPCR using the Kapa assay from Roche, following the manufacturer’s instructions. Pools of 12 to 16 libraries were sequenced on up to 3 lanes of the HiSeq4000 Illumina platform by the Oxford Genomics Centre, at the Wellcome Centre for Human Genetics (WCHG), aiming for 30 million paired-end reads.

## Chromatin immunoprecipitation with sequencing library preparation by Tn5 transposase

For chromatin immunoprecipitation (ChIP) a low cell input protocol known as ChIPmentation (ChIPm) was used (Schmidl2015). The H3K27ac histone mark (active enhancer and promoter marker) was assayed in four cells types (CD14+ monocytes, total CD4+, total CD8+ and CD19+). For ChIPm, 600,000 MACS sorted cells, as described in 2.2, were fixed with 1% formaldehyde (Sigma) and

snap frozen in dry ice and ethanol prior to storage at -80◦C. Fixed cells were

thawed, resuspended in SDS lysis buffer, sonicated for 8 min using Covaris M220(Covaris) with a duty factor of 5%. After sonication, chromatin was split into 6 aliquots (100,000 cells per aliquot), snap frozen and stored at -80◦C.

Aliquots as needed were thawed on ice and then processed downstream for ChIPm as described in Schmidl *et al.*, 2015. Immunoprecipitation was carried out with 1µg of the Diagenode Ab (C15410196). For each sample, an aliquot of chromatin was processed in parallel without incubation with the anti-H3K27ac Ab (control input). Tagmentation of the control input was performed using 1ng of DNA.

Amplification by qPCR was carried out in each of the samples and control inputs to determine the appropriate number of full cycles required to reach one-third of the final fluorescence to minimise the presence of PCR replicates upon NGS. Libraries were then amplified for the number of determined cycles minus one and simultaneously dual indexed using the primers optimised by (Buenrostro et al. 2015). A pool of 64 libraries (including control input samples) were sequenced over a number of lanes in the HiSeq4000 Illumina platform by the Oxford Genomics Centre, WCHG, aiming for 25 million paired-end reads

## RNA extraction and gene expression quantification

RNA extraction

Following MACS isolation of the different cell types, 2-3x106 cells were resuspended in 350µL of RNAProtect (Qiagen) or RLT buffer (Qiagen) supplemented with 0.1% of beta-mercaptoethanol (BM, Sigma) and snap frozen in dry ice before storage at -80◦C. Cells isolated from Cohort 1A psoriasis and control samples (Chapter ?? cohorts overview table, Table 4.2 and 4.3) were preserved in RNAProtect, which stops any biochemical reaction and transcriptional activity whilst maintaining cell integrity. At early stages of the project, the time frame to process the acquired samples was uncertain and RNAProtect was chosen as the most appropriate strategy to preserve cells for future RNA extraction to guarantee high quality in case storage exceeded 6 months. In the psoriasis and control samples from Cohort 1B (Chapter ?? cohorts overview table, Table 4.2 and 4.3) and PsA samples (Chapter 5), cells were resuspended in 0.1% BM supplemented RLT

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

Total RNA was extracted using the AllPrep DNA/mRNA/microRNA Universal kit (Qiagen) following the manufacturer’s instructions. RNA extractions were performed in batches of 12 samples, including all cell types from each individual processed and balanced numbers of psoriasis and control samples, to minimise batch effect correlating with phenotype. Basic

quantification was performed with NanoDrop (Thermo Scientific) before storage at -80◦C.

RNA-seq

RNA-seq quality control (QC), quantification, library preparation and sequencing were carried out by Oxford Genomics Centre at the WCHG in two independent batches of samples, each including Cohort 1A or Cohort 1B, respectively. Processing of samples in two batches was due to logistics of patient recruitment in the project. RNA quality control and quantification were assessed with the Bioanalyzer (Agilent). Samples were depleted from ribosomal RNA using Ribo-Zero rRNA Removal kit (Illumina) prior to cDNA synthesis and library preparation using TruSeq Stranded Total RNA (Illumina). This method preserves non-polyadenylated transcripts including nascent pre-mRNA (unspliced) and functionally relevant lncRNAs. For each of the cohorts, all libraries were pooled together and sequenced over several lanes of HiSeq4000 aiming a depth of approximately 50 million total reads per sample to maintain an appropriate level of sensitivity for subsequent expression analysis.

Gene expression quantification by qPCR array

Expression of immune-relevant genes was profiled by qPCR using the RT2 Profiler PCR Array (PAHS-3803Z, Qiagen) in collaboration with UCB. This platform included primers to test expression for 370 key genes involved in immune response during autoimmunity and inflammation, as well as appropriate house-keeping genes for normalisation. In brief, RNA was extracted, as detailed previously, from CD14+ monocytes, mCD4+ and mCD8+ cells. Reverse-transcription for cDNA synthesis and qPCR gene expression quantification was performed by UCB following the PCR array’s manufacturers instructions.

Single-cell RNA-seq

scRNA-seq data was generated using 10X Genomics technology Chromium single cell 3’ expression library preparation kit (PN-120267) by the Oxford Genomics Centre at the WCHG. Briefly, PBMCs and SFMCs were made into a cell suspension. Approximately 3,000 cells from the PBMCs or SFMCs suspensions were partitioned into single-cell gel beads in emulsion (GEMs) using the 10X Chromium controller system. Reverse-transcription for cDNA synthesis was performed within the GEMs, which included a 16bp 10x barcode, a 10bp unique molecular identifier (UMI) and poly-dT primers. The cDNA was released from the GEMs, followed by PCR amplification, enzymatic fragmentation and size selection. Afterward, appropriate sequencing Illumina indexes were incorporated into the samples through library preparation. Sequencing was performed using PE HiSeq4000 with 26bp for read 1 and 98bp for read 2 at a depth of approximately 50,000 reads per cell, following standard 10X library sequencing requirements.

## DNA extraction and rs4672405 genotyping

DNA isolation was performed using the AllPrep DNA/mRNA/microRNA Universal kit (Qiagen) following the manufacturer’s instructions. Quantification was performed using NanoDrop (Thermo Scientific) and samples were kept at -80◦C for long term storage. The extracted DNA was amplified by PCR using forward (5’-CACTGTGGAGGGAGGAACAA-3’) and reverse (5’- CGTGTTGGCCAGGATAGTCT-3’) primers annealing up and down stream the SNP rs4672505, respectively. An aliquot of the sample was run on a 1

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

Mass cytometry assay was performed by Dr Nicole Yager in collaboration with UCB following their in-house protocol for the CyTOF instrument. Briefly, an aliquot of whole blood and SF were fixed for 5 min with 1.6% paraformaldehyde (PFA) within 30 min of venipuncture/aspiration, respectively. These samples were defined as time 0h. In addition, another aliquot of whole blood and SF

were incubated at 37◦C for 6h in the presence of the protein transport inhibitors

1X BD GolgiStop (BD) and 1X BD GolgiPlug (BD), containing monesin and brefeldin A, respectively. Treatment with monensin and brefeldin A prevents the extracellular transport of cytokines from the cells and allowed measuring the intrinsic cytokine production rate in basal conditions. After 5h 45 mins the samples were treated with cisplatin to facilitate discrimination of dead cells, and then fixed 5 min with 1.6% PFA. These samples were defined as time 6h. After fixation of time 0h or 6h samples, red blood cells were lysed and cell suspensions were washed with PBS and stained with Abs against the cell surface markers of the intra-cellular staining (ICS) panel (Table 2.2). The samples were further

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

Markers from the ICS mass cytometry panel

CD248, CD19, GP38, FAP, CD8a, IL8, CD16, CD25, CD123, IL-17F, IL-17A, IL-10, CD11c, CD14, IL6, IFN-*γ*, CD-11b

CD45RO, CD56, HLA-DR, IL-13, CD117, CD4, IL4, IL-2, TNF*α*, IL-21, FceR, CD3, CD161, MCP-1(∗),GM-CSF, CD45,

MCP-1(∗), osteopontin(∗)

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

# Computational and statistical analysis

## ATAC data analysis

ATAC-seq, Fast-ATAC and Omni-ATAC data were analysed using an in house pipeline, towards which I made an important contribution. The pipeline performs single sample data processing and also builds a combined master list for each of the conditions of interest for chromatin accessibility characterisation and differential analysis.

Next generation sequencing data processing

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

following parameters were used -k 4 -X 2000 -I 38 −−mm -1, consistent with other publications (Buenrostro2013; Corces et al. 2016). Samtools (Li et al. 2009) was used to remove PCR duplicate reads marked with Picard Tools ( <http://broadinstitute.github.io/picard/)> as well as low MAPQ (*<*30), non- uniquely mapping and non-properly paired reads. The resulting bam file was additionally filtered to remove mitochondrial DNA and reads were adjusted by +4bp on the plus strand and by −5bp on the minus strand to represent the center of the transposition binding event. Pileup tracks (bigWig files) representing the number of reads per bp position were generated using bedtools genomeCoverageBed (Quinlan and Bioinformatics 2010) and the UCSC genome browser bedGraphToBigWig tool (Kent et al. 2010). For visualisation purposes, normalised bigWig files were generated from normalised bedgraph files with bedtools genomecov and the library size factor estimated by the differential analysis algorithm.

Peak calling, filtering and sample quality assessment

Peak calling was performed using MACS2 callpeak (Zhang et al. 2008) and the parameters −−nomodel −−shift -100 −−extsize 200 −−p 0.1 −−keep-dup all −−call-summits. Peaks overlapping blacklisted features from the ENCODE project [(https://ww](http://www.encodeproject.org/annotations))w[.encodeproject.org/annota](http://www.encodeproject.org/annotations))tions) were removed. The

−−shift and −−extsize parameters were set according to the recommendations of MACS2 for DHS and following other ATAC-seq publications (Buenrostro et al. 2015; Corces et al. 2016). The pval cut off for filtering called peaks was determined for each cell type using Irreproducibility Discovery Rate (IDR)

analysis (as further detailed in Chapter ??ch:Results1). For this, the filtered bam file of each sample was partitioned into two equal size files (pseudoreplicates). Peak calling was performed in each pseudoreplicate, followed by filtering for a

range of pvals (from 0 to 10−45) and IDR analysis done for the resulting pairs

of filtered peak sets. For each of the filtering pvals, the percentage of peaks sharing IDR rank between the two pseudoreplicates was determined, and the optimal pval filter identified. When more than one summit was identified in a peak, the median of the summits was used. For all peaks, summits were extended

±250bp to create a non-overlapping homogenous 500bp peak list for each sample

(Buenrostro et al. 2015; Corces et al. 2016).

Sample quality was determined by the fold enrichment of ATAC-seq signal across all the hg19 TSSs annotated by Ensembl (175,114 features), computed as in (Buenrostro et al. 2015) using a script provided by Dr Silvia Salatino. In brief, transposition events were calculated in 1bp bins ±1,000bp surrounding all TSSs and normalised to the mean value of background reads (signal from

-1,000 to -800). For overall library quality assessment all ATAC fragments were considered. When assessing chromatin structure within or across the TSS, fragments of *<*150bp or 260340bp were used, respectively (Scharer et al. 2016). Fraction of reads in peaks (FRiP) was calculated for samples in Chapter 3 as the overlap between the peak list filtered for FDR*<*0.01 and all ATAC fragments using bedtools intersect with the parameter −−f 0.1.

Combined peak master list and differential analysis

To perform differential open chromatin analysis a non-overlapping 500bp peak master list including all samples for a particular contrast was built. Each master list was built by union of all the peaks present in at least 30% of the samples included, regardless of the subgroup to which they belonged (e.g patients or controls, SF or PB). Reads overlapping each of the peaks in the master list were retrieved for each sample using the HTSeq-count algorithm (Anders et al. 2015). Principal component analysis (PCA) was performed on all counts normalised with the vsd function from the DESeq2 v1.20 R package (Love et al. 2014).

For differential chromatin accessibility analysis, an empirical 80% confidence cut-off was calculated and used to pre-filter peaks with high noise that could be confounding the analysis (detailed in Chapter 3). Differential analysis was performed using DESeq2 with a paired design (in Chapter 3 and Chapter 5) or including batch as a covariate in the model (in Chapter ??). Peaks were annotated with proximal genes (≤5Kb) using xGR2nGenes function from the XGR R package (Fang et al. 2016). These gene lists were intersected with differentially expressed genes (DEGs) from RNA-seq/scRNA-seq analysis or genes reported by psoriasis and PsA GWAS.

## ChIPm data analysis

Next generation sequencing data processing

ChIPm NGS data from samples and inputs were processed similarly to ATAC-seq (see Section ATAC-seq, Fast-ATAC and Omni-ATAC data analysis) for trimming, mapping and filtering with minor modifications. Specifically, the MAPQ30 score for filtering reads was lowered to 10. For visualisation, bedgraph files with noise subtracted using the control input were generated using MACS2 bdgcmp -m subtract followed by conversion to bigWig with bedGraphToBigWig tools.

Peak calling, filtering and sample quality

Peak calling for each ChIPm sample was performed accounting for background signal using the control input samples with MACS2 callpeak −−bw 200 −−p 0.1 −−keep-dup all −−call-summits. In this case the average library fragment size (−−bw) was used by MACS2 to first empirically find the model that best represents the precise protein-DNA interactions and calculate the appropriate −−shift parameter. For ChIPm PCA, filtering and downstream analysis peak homogenisation was performed similarly to Section 2.4.1 to build

a combined master list for all samples and cell types from cohort 1B (Chapter ??

Table 4.2 and 4.3).

Sample quality was determined by a combination of measurements. For library complexity the non-redundant fraction and PCR bottleneck coefficients (PBC1 and PBC2) were calculated following ENCODE guidelines (<https://www.encodeproject.org/chip-seq/histone/>) from unfiltered bam files. Enrichment of the ChIPm signal was evaluated based on the normalised strand cross-correlation coefficient (NSC) and relative strand cross-correlation coefficient (RSC), calculated with SPP using bam files filtered for low MAPQ30, duplicated and non-properly paired reads.

Combined peak master list and differential analysis

DiffBind (default parameters unless specified) was used to build a peak master list and perform differential H3K27ac analysis between psoriasis patients and healthy controls for each cell type. DiffBind used the unfiltered sample peak files generated by MACS2 and the filtered bam files (from samples and control inputs) to generate a master list including high quality reproducible peaks present in at least 50% of the samples (modification from default parameters) and retrieve counts of the reads mapping at the location of each peak.

## Gene expression analysis

qPCR analysis

Pre-processing of the qPCR data up to calculation of fold changes (FCs) for each gene was conducted by UCB collaborators. When comparing SF and PB from patients, FC was calculated for each patient as the ratio between the 2−*dCt* in each of the tissues. Therefore, in a particular cell type three FCs (one per individual) were provided for each gene. When comparing expression in PB from PsA patients versus healthy controls, the FC for each gene and each individual

was calculated as the ratio between the average 2−*dCt* in the three controls and

the 2−*dCt* of a particular patient. In order to determine the significance of the FCs, one sample t-test was performed with the null hypothesis being *H*0 : *µ* = 1 (no change) and the alternative hypothesis *H*1 : *µ* -:t 芠 뤀怀뤀怀뤠怀 1.

Bulk RNA-seq analysis

NGS RNA-seq data processing was performed using an in-house pipeline developed by Dr Katie Burnham. Ribo-depleted RNA-seq data was mapped against the reference genome (build hg19) using the aligner STAR (Dobin et al. 2013). Mapping allowed multiple alignments and only retained those with the best score and a mis-match percentage lower than 0.04%. Duplicates were marked and removed using Picard Tools. Gene counts were retrieved using HTSeq-count and the gencode hg19 annotation file comprising 2,840,278 gene entities, including lnc-RNAs. Differential gene expression analysis was performed with DESeq2 on genes with five or more reads in at least eight samples (smallest group size corresponding to the psoriasis patients’ samples). Independent filtering of genes with low expression levels, outlier removal using Cook’s distance and moderation of log2FC were enabled when using DESeq2. Differentially expressed transcripts were identified based on False discovery rate (FDR)*<*0.05. Batch effect was included as a covariate in the contrast between psoriasis patients and healthy controls. This effect related to the RNA extraction, library preparation and sequencing of cohort 1A and cohort 1B samples from the psoriasis study (including healthy controls). Lnc-RNAs were annotated using the list provided by gencode.v19 [https://www.gencodegenes.org/releases/](https://www.gencodegenes.org/releases/19.html) [19.html](https://www.gencodegenes.org/releases/19.html). The paired design of the psoriasis skin DGE analysis was taken into account by the DESeq2 model.

Single-cell RNA-seq analysis

Raw Illumina sequencing data from the 10X Genomics technology Chromium single cell 3’ expression libraries generated in bulk PBMCs and SFMCs from three PsA patients (see 5) were first processed using Cell Ranger v2.2 software provided by 10X Genomics technology [https://support.](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/) [10xgenomics.com/single-cell-gene-expression/software/pipelines/](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/).

Illumina sequencing base call files (BCLs) were demultiplexed and converted into fastq files using cellranger mkfastq. For each of the samples, mapping of the fastq files to the compatible human transcriptome reference (GRCh38-1.2.0) and retrieval of counts for each transcript included in the reference genome were performed with cellranger count using default parameters.

The count matrix files were fully processed downstream using the R package Seurat 2.3.4 (Butler et al. 2018). Each of the PBMCs and SFMCs individual count matrices were downsampled to 3,500 cells after removing all the genes expressed in fewer than 30 cells (drop-out events). Additional filtering was conducted to remove cells presenting more than 7.5% of mitochondrial reads, a number of genes larger than 500±1SD (approximately 1,800 genes in all

the samples) and a maximum of 10,000 UMI. After filtering, individual PBMCs

and SFMCs count matrices were processed for data scaling and normalisation (regressing number of UMI and percentage of mitochondrial counts) and PCA analysis. The first eight PCs (capturing most of the variation in the data according to jackstraw analysis) were then used to identify clusters (groups of cells with similar expression profiles) built by a shared nearest neighbor (SNN) modularity optimisation algorithm (default resolution 0.6) followed by visualisation using t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensional reduction. CD14+ monocyte clusters were selected based on co- expression of appropriate cell specific markers (*CD14* and *LYZ*). For each of the CD14+ monocyte populations, the top variable genes were quantified by

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

## Genomic region annotation, enrichment analysis, gene- network analysis and pathway visualisation

Genomic region annotation with chromatin states was performed by

overlapping peaks with the appropriate cell type chromatin segmentation map from Epigenome Roadmap ([https://egg2.wustl.edu/roadmap/web\_portal/](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html) [chr\_state\_learning.html](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html)).

Genomic feature enrichment analysis was performed using XGR (reference) built-in data (including eQTLs, FANTOM enhancers, histone marks and TFBS, amongst others) and the function xGRviaGenomicAnno. When annotating regions (ATAC or ChIPm peaks) the summit of the peaks were used to increase the specificity of the analysis. Pathway enrichment analysis was conducted for the built-in KEGG, Reactome MsigdbC2BIOCARTA and MsigdbC2CPall pathways with the xEnricherGenesAdv XGR functionality. Input data used were genes annotating differentially accessible ATAC peaks or differentially expressed genes, and background was defined as all the annotated ATAC peaks from the differential

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

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

Visualisation of the signalling-pathway enrichment from the RNA-seq results was performed using the R package Atlas and Analysis of systems- biology-led pathways (A2), developed by Dr. Hai Fang and towards which I have contributed (manuscript in preparation). The manually curated KEGG pathways expanded with genes for each gene family were coloured based on the FC from the corresponding RNA-seq differential analysis and highlighted in bold when passing the FDR threshold for significance.

## Statistical fine-mapping

Fine mapping of the psoriasis and PsA GWAS signals was performed using a Bayesian approach, aiming to overcome the incomplete coverage of genotyping arrays and the hundreds of associations per locus due to the LD structure of the genome. Fine-mapping was conducted using two different strategies due to availability of summary statistics or genotyping data from the psoriasis and PsA Immunochip GWAS studies, respectively. Both strategies include the same main steps of imputation, association testing and calculation of PP and were implemented in collaboration with Dr Adria´n Corte´s.

PsA fine-mapping using Immunochip genotyping data

Fine-mapping was performed for a number of non-MHC PsA Immunochip GWAS susceptibility loci using 1,103 patients and 8,900 controls from the Bowes et al. PsA Immunochip UK cohort. Access to the data post-quality control was kindly provided by Dr Anne Barton (The University of Manchester) PCA analysis was performed using only pruned SNPs with flashpca (Abraham and Inouye 2014) and the calculated PCs were used as covariates in the association analysis to correct for population stratification.

For each of the fine-mapped loci a 2Mb window around the lead SNP was defined and SNPs extracted from the data using PLINK 1.9 (Chang et al. 2015). Phasing of the genotype data was performed with SHAPEIT (Delaneau et al. 2012) and used to impute missing genotypes with IMPUTE2 (Howie et al. 2009) and the 1000 Genomes Project Version 3 as the reference panel (October 2015 release). SNPs for which imputation was not successful in at least 70% of the samples (info-score*<*0.7) were filtered out using QCtool. The association and conditional analysis was conducted using a Bayesian additive model implemented with SNPTEST and including the previously calculated PCs as covariates (Burton 2007). Approximated Bayes Factors (ABF) were calculated for the lead signal and step-wise conditional analysis was performed if ABF*>*3. Credible sets of SNPs containing the variants likely to explain 50% and 90% of that association were identified for each of the signals in the locus, along with their corresponding posterior probabilities (PP) as further detailed in Bunt *et al.*, 2015.

Psoriasis fine-mapping using Immunochip summary statistics

Fine mapping was performed for X of the risk loci reported by the psoriasis Immunochip GWAS study from Tsoi *et al.*, 2012, for which only the summary statistics of the GAPC Immunochip cohort (2,997 cases and 9,183 controls) were

available. The summary statistics file included the pval and the OR calculated for each of the genotyped SNPs using a logarithmic regression model and correcting for ten principal components. The statistic z-scores from the genotyped SNPs were used in the direct imputation of summary statistics (DIST) method to impute the z-scores for allele 1 of the missing SNPs based on the correlation in linkage disequilibrium (r2) from the 1000 Genome Project Version 3 (Lee et al. 2013). Imputation was performed genome-wide for all autosomes and the results were filtered based on the quality of the imputation (*>*0.8). Association analysis and calculation of the ABF were performed using Wakefield approximation for a 2Mb window around the GWAS lead SNP of each locus of interest. This approximation was applied under the priors of (1) normally distributed OR with mean and variance (v2); (2) the greater the variance the bigger the size effects obtained will be; (3) mean is 0 and variance is fixed to 0.2 (accepted variance for GWAS studies). In this approach ABF was calculated using effect size (*β*) and standard error (SE) derived from the variance formula. *β* was calculated using the z-scores of each of the interrogated SNPs as: *β*=z-score∗SE. It is important to note that step-wise conditional analysis is not performed when using summary statistics imputation. Similarly to the genotyping fine-mapping approach, PP for the SNPs in a particular window (signal) were calculated and ranked to set the threshold of the 50 and 90 % credible set of SNPs.

## Mass cytometry data analysis

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

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