



# Functional genomics of psoriasis

Alicia Lledo Lara  
Hertford College  
University of Oxford

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

# **Abstract**

**Functional genomics of psoriasis**

Alicia Lledo Lara, Hertford College, Trinity Term 2018

A thesis submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy of the University of Oxford

This is my abstract...

# Acknowledgements

Thank you, thank you, thank you.

# **Declarations**

I declare that unless otherwise stated, all work presented in this thesis is my own. Several aspects of each project relied upon collaboration where part of the work was conducted by others.

# **Submitted Abstracts**

<b>Title</b>	<b>Year</b>
Authors	

# **Associated Publications**

## **Title**

Journal

Authors

# **Other Publications**

## **Title**

Journal

Authors

# Contents

<b>Abstract</b>	i
<b>Acknowledgements</b>	ii
<b>Declarations</b>	iii
<b>Submitted Abstracts</b>	iv
<b>Associated Publications</b>	v
<b>Contents</b>	vi
<b>List of Figures</b>	viii
<b>List of Tables</b>	ix
<b>Abbreviations</b>	x
<b>1 Introduction</b>	1
1.1 Psoriasis and psoriatic arthritis . . . . .	1
1.1.1 Epidemiology and global impact . . . . .	1
1.1.2 Psoriasis and inflammatory dermatoses . . . . .	2
1.1.3 PsA and spondyloarthropathies . . . . .	3
1.2 Pathophysiology of psoriasis and psoriatic arthritis . . . . .	4
1.2.1 Clinical presentation and diagnosis . . . . .	4
1.2.2 Aetiology of psoriasis and PsA . . . . .	8
1.2.3 Cell types involved in psoriasis and PsA pathogenesis . . . . .	11
1.2.4 Therapeutic intervention . . . . .	17
1.3 Genetics of psoriasis and psoriatic arthritis . . . . .	18
1.3.1 Heritability . . . . .	18
1.3.2 Non-GWAS and linkage studies . . . . .	19
1.3.3 Genome-wide association studies . . . . .	20
1.3.4 Relevance of non-coding versus coding variants in disease susceptibility . . . . .	25
1.3.5 The role of GWAS studies in highlighting immune-relevant cell types and pathways . . . . .	26
1.3.6 Limitations and future of GWAS studies . . . . .	33
1.4 Functional interpretation of genome-wide association studies in complex diseases . . . . .	34
1.4.1 Overcoming the limitations of GWAS: post-GWAS studies . .	34
1.4.2 The use of fine-mapping to prioritise causal variants . . . .	35
1.4.3 Understanding the epigenetic landscape in complex diseases	37
1.4.4 The chromatin landscape . . . . .	40

## CONTENTS

---

1.4.5	Transcriptional profiles in disease . . . . .	47
1.4.6	Transcriptional regulation in complex diseases . . . . .	51
1.4.7	Integration and interpretation of genomic data . . . . .	52
1.4.8	Immunophenotyping . . . . .	54
1.4.9	Approaches to establish disease mechanisms and causality of genetic variant . . . . .	54
<b>2</b>	<b>Establishment of laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples</b>	<b>57</b>
2.1	Introduction . . . . .	57
2.2	Results . . . . .	58
2.2.1	Establishment of an ATAC-seq data analysis pipeline based on current knowledge . . . . .	58
2.2.2	Assessment of ATAC-seq transposition times and comparison with FAST-ATAC protocol in relevant cell types	70
2.2.3	Limitations of ATAC-seq and FAST-ATAC to assess chromatin accessibility in KC . . . . .	70
2.2.4	Discussion . . . . .	74
<b>3</b>	<b>Cross-tissue comparison of chromatin accessibility, gene expression signature and immunophenotypes in PsA</b>	<b>77</b>
3.1	Introduction . . . . .	77
3.2	Results . . . . .	77
3.2.1	PsA patients cohort description and datasets . . . . .	77
3.2.2	Differences in the chromatin accessibility landscape between circulating and SF immune cells . . . . .	80
3.2.3	Differential gene expression analysis in paired circulating and synovial immune cells . . . . .	88
3.3	Discussion . . . . .	88

# List of Figures

1.1	Crosstalk between innate and adaptive immunity in psoriasis . . . . .	16
2.1	Measurements for quality control assessment in ATAC-seq samples	63
2.2	Peak calling and sequencing depth in ATAC-seq samples . . . . .	66
2.3	Peak calling filtering using IDR analysis in ATAC-seq samples . .	67
2.4	Differential chromatin accessibility analysis for different background reads cut-offs. . . . .	69
2.5	Exploration of the differential chromatin accessibility analysis using 80% as the empirical cut-off. . . . .	70
2.6	Assessment of the effect of transposition times on the ATAC-seq QC parameters . . . . .	71
2.7	Differences in MT DNA abundance and signal specificity between ATAC-seq and FAST-ATAC protocols . . . . .	72
2.8	QC assessment of ATAC-seq in KC enriched cell suspension derived from a psoriatic lesional skin biopsy . . . . .	73
2.9	QC assessment of FAST-ATAC and Omni-ATAC in cultured NHEK	75
2.10	QC assessment of Omni-ATAC in NHEK and chromatin accessibility signal for the samples generated with the different ATAC-seq protocols . . . . .	76
3.1	QC of FAST-ATAC PsA samples in four cell types . . . . .	82
3.2	Combined PCA analysis of all four cell types isolated from blood and SF. . . . .	83
3.3	Enrichment of eQTLs in the combined cell types PsA accessible chromatin master list. . . . .	84
3.4	Annotation with genomic regions and chromatin states of the PsA DOCs from the four cell types differential analysis. . . . .	86

# List of Tables

1.1	Variables and scoring used in the Psoriasis Area and Severity Index (PASI) . . . . .	6
1.2	CASPARI criteria for diagnosis of PsA . . . . .	7
1.3	Main GWAS studies in psoriasis and PsA . . . . .	23
2.1	Summary table of ATAC-seq methodology analysis for peak calling, filtering and differential analysis. . . . .	59
2.2	ATAC-seq percentage of MT reads and fraction of reads in called peaks . . . . .	64
2.3	Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies. . .	74
3.1	Description of PsA patients cohort recruitment and metadata. . . . .	79
3.2	Datasets generated for the PsA cohort samples . . . . .	80
3.3	Summary results of the chromatin accessibility analysis between SF and PB in PsA samples . . . . .	85
3.4	Summary results of the chromatin accessibility analysis between SF and PB in PsA samples. . . . .	87

# Abbreviations

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

# **Chapter 1**

## **Introduction**

### **1.1 Psoriasis and psoriatic arthritis**

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

#### **1.1.1 Epidemiology and global impact**

Psoriasis represents a serious global health problem that currently affects about 100 million people worldwide, including children and adults with no sex bias (Organization 2016). Alike the minor correlation with geographic latitude, the development of psoriasis presents a strong ethnicity component (Jacobson et al. 2011). In fact, the prevalence of psoriasis in adults is lower among African, African American and Asian (between 0.4 and 0.7%) compared to American

## **Introduction**

---

and Canadian populations (4.6 and 4.7%, respectively). In the UK, psoriasis prevalence ranges between 2 and 3%, affecting approximately 1.8 million of people (Perera et al. 2012). On the other hand, the cases of PsA in the general population varies between 0.04 and 1.2% (Perera et al. 2012) but dramatically increase up to 10 to 30% within psoriasis patients (**Reich2008**; Gelfand et al. 2005), evidencing the strong association between the two diseases. Particularly, in the UK, 14% of the psoriasis patients develop chronic inflammatory arthritis in the form of PsA during the course of the disease (Ibrahim et al. 2009). Overall, data suggest an steady increase in both, psoriasis and PsA, prevalence over time (**Springate2007**; Organization 2016).

Onset of disease seems to have a bimodal distribution with psoriasis patients being classified as early-onset or type I (around 16-22 and 30-39 years) or late-onset or type II (between 50-60 years). This classification based on the age of onset has also correlates with distinctive clinical features including severity, relapse frequency and family history.

Accordingly, psoriasis and PsA represent a burden for the countries' economies due to treatments costs and associated morbidity. In the UK treatment and management-associated costs for psoriasis in 2015 accounted for 4,000 to 14,000, before and after requirements of biological therapy, respectively (Burgos-Pol and Dermo 2016; Poole et al. 2010) and the costs are further enhanced in the case of PsA.

### **1.1.2 Psoriasis and inflammatory dermatoses**

The group of inflammatory dermatoses affects up to 70% of the population and it represents the 4<sup>th</sup> leading cause of nonfatal burden (**ICD-10**; **Roderick2014**). The skin is the biggest organ in the human body constituting an effective barrier between the environment and the internal organs. The most external layer, the epidermis, plays a relevant role in the innate and adaptive

## **Introduction**

---

immunity and its alterations, due to exogenous or endogenous factors, can lead to development of inflammatory dermatose conditions such as psoriasis or atopic dermatitis (AD) (Johnson-Huang et al. 2009; Proksch et al. 2008). Lesions in psoriasis are very heterogeneous in type (pustular and non-pustular), location and severity, which difficult its clinical classification (Perera et al. 2012). As a result, several phenotypes including psoriasis vulgaris, guttate, pustular, erythroderma and nail pitting have been defined and it is under debate whether some of those should be considered independent disease entities (Marrakchi et al. 2011).

### **1.1.3 PsA and spondyloarthropathies**

PsA belongs to the family known as spondylarthropathies (SpA) which includes phenotypes such as ankylosing spondylitis (AS), reactive arthritis (ReA), idiopathic inflammatory bowel disease (IBD) and undifferentiated SpA (Baeten et al. 2013). All these SpA subtypes are characterised by structural damage (bone formation and erosion) as well as inflammation of joints and extra-articular sites such as eyes, gut and skin. Broadly, SpA has been classified into axial and peripheral base on the affected joints (spine/sacroiliac or peripheral) and the presence of extra-articular features (**Runwaleit2001**; **Runwaleit2001**). Studies in human families and rat models with HLA-B27 positive status have shown manifestation of different SpA forms, such as psoriasis and inflammatory bowel disease (IBD), within a single family or individual (**Said-Nahal2000**; Hammer et al. 1990). These observations support the hypothesis that SpA subtypes may be a single multifaceted condition with shared genetic, immunopathological and structural features and dynamic phenotypes (Baeten et al. 2013). Conversely, some studies suggest that the immunopathological differences between axial and peripheral arthritis could be partially explained by genetic factors (Porcher et al. 2005; Appel et al. 2011; Noordenbos et al. 2012).

## **Introduction**

---

As a phenotype, PsA can be further subdivided in five clinical groups as per Moll and Wright criteria: distal, destructive, symmetric, asymmetric and spinal (Moll et al. 1973). These subclasses mainly differ in the location, number and distribution of the affected joints and have been later modified to also include dactylitis (diffuse swelling of a digit), a distinctive feature of PsA (Reich et al. 2009). Altogether, this phenotypic heterogeneity of PsA increases the difficulty in the design and achievement of meaningful outcomes from clinical studies.

## **1.2 Pathophysiology of psoriasis and psoriatic arthritis**

### **1.2.1 Clinical presentation and diagnosis**

Amongst the aforementioned phenotypes, approximately 90% of all psoriasis cases are psoriasis vulgaris, which manifests with well demarcated plaques, erythema and scaling. The plaques formation is the result of the thickening (acanthosis) and vascularisation of the epidermis and can vary in size and distribution, being the most common locations elbows, knees and scalp (Perera et al. 2012; Griffiths and Lancet 2007). The second most common clinical presentation is psoriasis guttate (10% of all cases) characterised by acute onset of small droplike papules usually in the trunk and proximal extremities (Vence et al. 2015). Despite psoriasis vulgaris and guttate represents an important burden for patients wellbeing, they are not life threatening forms of disease. Conversely, the least prevalent phenotype, pustular psoriasis, presents higher lethality rate (Moura et al. 2015).

In addition to the impact in onset, type I and type II psoriasis also relate to the clinical presentations. Type I psoriasis patients, which have greater prevalence for HLA-C\*06:02 (85.4% of the cases) and stronger family history, commonly present guttate lesions followed very often by bacterial infection,

## **Introduction**

---

importantly *Streptococcus* throat infection (). In contrast, type II psoriasis where individuals are positive for HLA-C\*06:02 only in 14.6% of the cases, involves spontaneous chronic plaques such the psoriasis vulgaris (Perera et al. 2012).

In the case of PsA, symmetric/polyarticular constitutes the most common manifestation (more than 50% of the cases) followed by asymmetric/oligoarticular PsA (around 30%), that exclusively affects single or few distal interphalangeal or phalangeal joints (Reich et al. 2009; McGonagle et al. 2011). Skin psoriatic lesions precede joint inflammation in approximately 60 to 70% of the cases (Gladman et al. 2005; McGonagle et al. 2011). Particularly, nail pitting and scalp and intergluteal skin lesions constitute a predictive biomarker for development of joint inflammation (**Moll1976; McGonagle; 2011; Griffiths and Lancet 2007**). This observation reinforces the need of appropriate coordination between dermatologists and rheumatologists for an early diagnostic and treatment that could prevent functional joint disability.

Several comorbidities have been associated with psoriasis and PsA, with comparatively greater prevalence in PsA. For example, intraocular inflammation known as uveitis affects 8% of PsA patients compared to only a 2% of the psoriasis ones (Husted et al. 2011; Oliveira et al. 2015). Other comorbidities include IBD, cardiovascular disease (CVD), type II diabetes (T2D) and metabolic syndrome (**Saphiro2007; Cohrn20017; Gelfand et al. 2006**). Psoriasis and PsA have also been associated with an increased prevalence of depression and suicidal ideation (Sampogna et al. 2012).

The diagnosis of psoriasis and PsA is primarily based in the clinical assessment of patients symptoms due to the lack of appropriate molecular biomarkers at early stages of the disease (Villanova et al. 2013). The evaluation of skin lesions severity poses an additional challenge, and different measures have been implemented for criteria unification. The Psoriasis Area and Severity Index (PASI) is the most widely quantitative rating score of skin lesion severity

## Introduction

---

in research and clinical trials (Fredriksson and Dermatology 1978; Finlay 2005). PASI quantifies the lesional burden weighted by body part based on area of affected surface and the degree of erythemas severity, induration and scale at each location (Table 1.1). Disease is considered mild for PASI scores below 7 and is classified as moderate-to-severe for PASI scores between 7 to 12, depending on the study (Schmitt2005; Finlay 2005; Langewouters et al. 2008).

To diagnose PsA, a modified Moll and Wright criteria known as Classification Criteria for Psoriatic Arthritis (CASPAR) is the most widely used in a clinical setting (Taylor2006). A positive diagnostic based on CASPAR requires displaying of inflammatory arthritis, enthesitis, and/or spondylitis and three points from a list of associates elements (Table 1.2). In terms of disease activity and treatment efficacy, the PsA Response Criteria (PsARC) is the preferred measure (Philipp2011; Clegg et al. 1996). PsARC considers the number of tender joints (TJC) and swollen joints (SJC) over 68 and 66, respectively, as well as patients and physician global assessment of the patientss general health based on a short questionnaire.

PASI	description
Body location	Head and neck, upper limbs, trunk and lower limbs
Feature	Redness, thickness and scaling
Severity scale	Absent, mild, moderate, severe or very severe
Affected area (%)	0, 1-9, 10-29, 30-49, 50-69, 70-89 or 90-100

**Table 1.1:** For each of the four body locations the test quantifies the percentage of affected area and the severity of three intensity features: redness, thickness and scaling.

## Introduction

---

CASPAR:	a patient with	must have three points	inflammatory articular points	disease from	(joint, spine, or five	or enthesial categories
					a. Current skin or scalp disease b. History of psoriasis c. Family history of psoriasis	
					Typical psoriatic nail dystrophy	
					a. Using preferrably by enzyme-linked immunosorbent assay (EMSA) b. Swelling of an entire finger c. History of dactylitis	
					Ossification near joint margins	
						Table 1.2: xxxx

## **Introduction**

---

### **1.2.2 Aetiology of psoriasis and PsA**

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

#### **Histopathological alterations in skin and joints**

The epidermis is the most external compartment of the skin, formed by approximately 90% keratinocytes (KCs) and organised in a layer-like structure that self-renews in an spatial and time-dependent manner (Wikramanayake et al. 2014). KC differentiation is associated with changes in morphology, replication ability and keratin composition of the intracellular matrix. In the context of psoriasis, impaired epidermis cell renewal leads to histological alterations and lesion development. Importantly, KCs undergo upregulation in the proliferation rate (hyperplasia) that causes aberrant cell differentiation (parakeratosis) thickening of the epidermis and the subsequent scale formation (Ruchusatsawat2011). Concomitantly, inflammation causes immune cell infiltration and hypervascularisation of the lesion driven by upregulation in the expression of angiogenic factors and activation of the endothelium (Perera et al. 2012).

In PsA, the joint affection, arising after skin lesions in the majority of the cases, involves a wide range of histological changes (Haddad and Chandran 2013). One of the most common structural changes is the arthritis caused by the swelling and inflammation of the joints (**Schett2011**). As a result of this inflammation, alterations in bone remodeling lead to osteolysis with the

## **Introduction**

---

subsequent bone resorption and erosion at the affected joints (**Mensah2017**). Bone erosion is also the main histopathological process driving dactylitis, where bone lysis resolves in shortening of the digits (Gladman et al. 2005). Moreover, 35% of the PsA patients also undergo inflammation of the connective tissue at the insertion of tendons or ligaments, phenomenon known as enthesitis (McGonagle et al. 2011; Polachek et al. 2017). As a result, the formation of bony spurs along the insertion sites causes structural debilitation of the joints (**Schett2011**).

### **Dysregulation of the innate and adaptive immune response**

The dysregulated immune response in psoriasis and PsA is the result of the interaction between innate and adaptive immune cells through feedback loops and a complex cytokine milieu.

Interferon (IFN)- $\alpha$  and  $\gamma$  are amongst the most relevant innate immune cytokines involved in disease initiation (**Leanne2009**). Both cytokines are mainly produced by circulating plasmacytoid dendritic cells (pDCs) and myeloid DC (mDCs), respectively, upon activation by KCs pro-inflammatory cytokines (Perera et al. 2012). Increased mRNA levels for both INFs have been detected in lesional skin and demonstrated to contribute to lymphocyte recruitment and maintenance of DCs activation (**Schmid1994**). TNF- $\alpha$  is another pivotal cytokine involved in the psoriasis and PsA dysregulated innate immune. TNF- $\alpha$  is produced by activated KCs, mast cells and also adaptive immune cell types, including T helper (Th)- 1 and Th-17 lymphocytes infiltrated in the skin lesions and inflamed joints (Perera et al. 2012; Lizzul et al. 2005). This cytokine causes activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-B), a master transcriptional regulator of both, the innate and adaptive immune system that induces expression of pro-inflammatory cytokines, antiapoptotic genes and genes involved in chronic inflammation maintenance (**Johansen2010**; Lizzul et al. 2005). Moreover, TNF- $\alpha$  has a prominent role in

## **Introduction**

---

bone turnover and bone remodeling, key features of the histopathological PsA joint alterations(Mensah et al. 2008).

Interleukin-23 (IL-23) and interleukine-17 (IL-17) constitute a link between the innate and adaptive immunity as well as a key loop for the perpetuation of the psoriasis and PsA inflammatory response. IL-23 is an innate immune cytokine mainly produced by the mDCs and macrophages homing the inflamed skin (ref). IL-23 exerts its function through binding to the IL-23 receptor (IL-23R), highly expressed by the lesion resident DCs and T cells and the circulating CD4 lymphocytes (ref). In psoriasis, IL-23 mediates the pathogenic loop between activated KCs and T cells (ref). The activation of the IL-23 pathway importantly leads to increased IL-17 cytokine levels as a result of NF-B activation More recently, interleukin 22 (IL-22) has gained relevance as mediator of the dysregulated crosstalk between the innate and adaptive immune response. IL-22 levels are increased in the skin lesions and the serum of psoriatic patients and is mainly produced by a subset of CD4<sup>+</sup> cells known as Th22 (ref). IL-22 contributes to some of the histological changes in skin as well as to AMP production by KCs (ref).

## **Environmental factors and disease**

Several environmental factors are known to be associated with increased risk and worsening of psoriasis and PsA development. A wide range of drugs including antidepressant, antihypertensive and anticytokine therapies have been clinically associated with initiation, exacerbation and worsening of psoriasis (Kim et al. 2010). Infectious agents such streptococcal throat infection have also been associated with development of type I psoriasis (**Valdimarsson2009**; Gudjonsson and of 2003; Diluvio et al. 2006). Consistently with other chronic inflammatory disease such as IBD and AS, recent studies have also observed perturbation in the composition of the gut and skin microbiota in psoriasis and

## **Introduction**

---

PsA patients (**add reference**). Furthermore, physical trauma, including tattoos, surgical incisions and mechanical stress can trigger the appearance of skin lesions and digits joint inflammation (**Weiss2002**; Nestle et al. 2009). Lastly, behavioral factors including smoking, alcohol and stress have been linked to psoriasis and PsA but no clear association with disease development has been established yet(Meglio et al. 2014).

### **1.2.3 Cell types involved in psoriasis and PsA pathogenesis**

Psoriasis and PsA are complex dynamic pathophysiological processes, and the understanding of the relative importance of different cell types at different disease stages still remains challenging.

Several studies have shown the role of KCs as immune sentinels through MHC-II antigen presentation and production of antimicrobial peptides (AMP), cytokines and chemokines (Black and of 2007). Indeed, complex formation between the cationic AMP LL-37 and self-DNA/RNA released by KCs has been observed upon damage triggered by environmental factors (Lande et al. 2007). This complex acts as an antigen for activation of the skin-resident DCs that initiate and perpetuate the skin inflammatory response through secretion of pro-inflammatory cytokines, including IL-1, IL-6 and TNF- $\alpha$  (**Feldmeyer2007**; Arend et al. 2008; Nestle et al. 2009; Nestle et al. 2005). Furthermore, *in vivo* studies have described the development of psoriatic lesions in immunodeficient mice upon human xenotransplant of psoriatic skin(Boymann et al. 2004). Altogether, these findings support the role of epidermis dysfunction in the initiation of the psoriatic chronic inflammatory response (**Proskch2008**). The relevance of KCs at early stages of psoriasis pathogenesis is reinforced by the genetic association between KC-specific genes from the late cornified envelope (LCE) family and increased psoriasis risk (Tsoi et al. 2012)

## **Introduction**

---

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

Neutrophils are also thought to be closely involved in disease initiation through their ability to form neutrophil extracellular traps (NET)that contain host DNA and LL-37 (Hu et al. 2016). There is evidence of increased NET formation in peripheral blood and lesional skin of psoriasis patients and they seem to be contributing to pDC and CD4 $^{+}$  T activation (Hu et al. 2016). Neutrophils have also been identified in recent studies as one of the main sources of IL-17 production in the skin lesions (Lin et al. 2011) and they also release a wide range of proteases which some induce KC proliferation (**Mahil2006**).

In the context of the innate immunity, the involvement of monocytes and macrophages in psoriasis and PsA has not been extensively studied. Resident macrophages in the healthy dermis undergo a 3-fold increase upon skin lesion and they are involved in disease development through TNF $\alpha$  production (Perera et al. 2012; Mahil et al. 2016). Similarly, mice models for chronic psoriasiform skin inflammation have shown macrophage migration into the affected skin and TNF- $\alpha$  production for maintenance of the skin lesions (**Stratis2006; Wang2006**).

## **Introduction**

---

Some studies using isolated monocytes from psoriasis patients PBMC have shown greater phagocytic and bactericidal activity compare to those from healthy individuals (**Bar-Eli1979**). Later studies have also shown increased circulating intermediate monocytes (CD14<sup>+</sup> high CD16<sup>+</sup> high) and monocyte aggregation in psoriasis patients causing enhanced platelet activation and angiogenesis (**Golden2015**). In PsA, synovial membranes levels of monocytes/macrophage metalloproteinases which mediate bone erosion through differentiation into osteoclasts are comparable to those found in RA joints (Hitchon et al. 2002). Overall, these observations highlight the systemic aspects of both pathologies.

In the context of the innate immunity, the involvement of monocytes and macrophages in psoriasis and PsA has not been extensively explored. Resident macrophages in the healthy dermis undergo a 3-fold increase upon skin lesion and contribute to disease development through TNF $\alpha$  production (Perera et al. 2012; Mahil et al. 2016). Similarly, mouse models for chronic psoriasiform skin inflammation have shown the role of macrophage migration into the affected skin and production of TNF- $\alpha$  in maintenance of the skin lesions (**Stratis2006**; **Wang2006**). Initial studies showed greater phagocytic and bactericidal activity of PBMCs isolated monocytes from psoriasis patients compared to those from healthy individuals (**Bar-Eli1979**). Additionally, increased circulating intermediate monocytes (CD14<sup>+</sup> high CD16<sup>+</sup> high) and monocyte aggregation was also observed in psoriasis patients, resulting in enhanced platelet activation and angiogenesis (**Golden2015**). In PsA synovial membranes, the levels of monocytes/macrophage metalloproteinases responsible for bone erosion through differentiation into osteoclasts have been found to be similar to those found in rheumatoid arthritis (RA) joints (Hitchon et al. 2002).

Regarding the adaptive immunity, T lymphocytes have been considered the most relevant cell types in the initiation and maintenance of psoriasis and PsA. Report cases in humans have demonstrated that bone marrow

## **Introduction**

---

transplantation can initiate or terminate psoriasis (**Gardembas1990**; Eedy et al. 1990). Reduced numbers of circulating T cells but increased percentages of the memory populations CD4<sup>+</sup>CD45RO<sup>+</sup> and CD8<sup>+</sup>CD45RO<sup>+</sup> have been observed in moderate-to-severe and severe psoriasis patients when compared to milder phenotypes and healthy controls (**Lecewicz-Toru2001**; Langewouters et al. 2008). Different studies have reported controversial results regarding the total abundance and ratios of CD4<sup>+</sup> and CD8<sup>+</sup> in PBMC, likely due to the phenotype heterogeneity of the psoriasis cohorts between studies (**Lecewicz-Toru2001**; Cameron and of 2003; Langewouters et al. 2008). In PsA, no differences in abundance of circulating T cells have been identified when compared to healthy individuals (Costello et al. 1999).

In homeostasis, CD8<sup>+</sup> and CD4<sup>+</sup> lymphocytes are found in the epidermis and dermis, respectively (Clark et al. 2006). An increase in activated memory CD4<sup>+</sup>CD45RO<sup>+</sup>and CD8<sup>+</sup>CD45RO<sup>+</sup> cells can be detected by the third day from the lesion appearance (**Perera2012** ; Clark et al. 2006). *In vivo* studies showed that development of psoriasis following engrafted human pre-lesional skin was only dependent on local T cell proliferation, highlighting the importance of circulating T cells recruitment during the priming event rather than at later stages of the disease (**Boyle2013**; Perera et al. 2012). The relative importance of CD4<sup>+</sup>versus CD8<sup>+</sup> cells in psoriasis initiation has been explored in pre-lesional skin mouse xenografts where CD4<sup>+</sup> but not CD8<sup>+</sup> T cells were required in the transition from uninvolved to lesional skin (Nickoloff and Wrone-Smith 1999). Interestingly, the injection of activated CD4<sup>+</sup> cells in mice was followed by an acute increase in activated resident CD8<sup>+</sup> T cells. Overall, these results supported the hypothesis of skin CD4<sup>+</sup> cells being drivers of resident T-cell activation and the population of resident activated CD8<sup>+</sup> the main effector of the immune response. In synovial tissues of PsA patients, CD4<sup>+</sup> are significantly more abundant than CD8<sup>+</sup> (Diani et al. 2015). However, amongst the CD8<sup>+</sup>

## **Introduction**

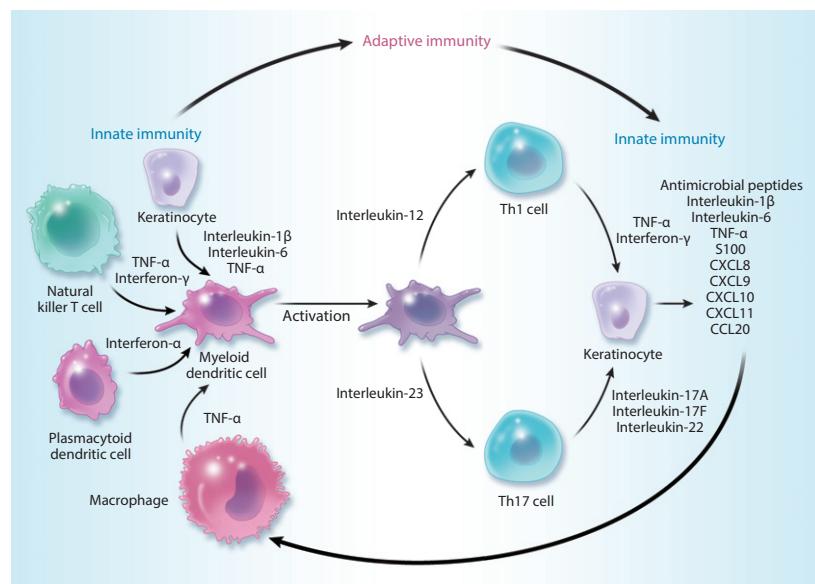
---

populations, the memory cells are prevalent in the patients synovial fluid (SF) with a significant enrichment compared to controls (Costello et al. 1999). The contribution of regulatory T (Treg) remains controversial in both, psoriasis and PsA (Perera et al. 2012).

Based on the cytokine profile, psoriasis and PsA have been classified as a type 1 Th/Tc disease, where activation of naive CD4<sup>+</sup> and CD8<sup>+</sup> cells is driven by IL-12 and IFN- $\gamma$  (Austin et al. 1999; Perera et al. 2012). In addition, T-cell subsets including Th-17/Tc-17 and Th-22/Tc-22, producing high levels of IL-17 and IL-22, respectively, have been identified to be relevant for the perpetuation of the inflammatory response (Mahil et al. 2016). The importance of Th-17 cells and IL-17 production has been evaluated in skin, joints and blood, with elevated mRNA and protein levels of IL-17 and also IL-23 reported in psoriasis and PsA patients compared to controls (**Cai2012; Dolcino2015**). The relevance of IL-17 has been further highlighted by the presence of CD8<sup>+</sup> populations in patients SF that are predominantly IL-17 producers and whose abundance correlates with markers of inflammation and structural changes in the joint (Menon et al. 2014). This finding is in line with observations in skin and suggests a prominent role for CD8<sup>+</sup> IL-17-producing cells in the different stages of both pathologies. Studies directed to understand the importance of IL-17 have led to the discovery of other immune cells producing this pivotal cytokine, including innate immune lymphoid (ILC) cells and  $\gamma\delta$  T cells, opening new research avenues in the context of psoriasis and PsA pathophysiology and treatment (Meglio et al. 2014; Leijten et al. 2015). IL-17-producing cells have also been hypothesised to be at the link between skin and joint lesions. Although the precise mechanisms for transition between psoriasis and PsA is still poorly understood, the study of psoriasis and RA in mouse models revealed that skin lesions facilitate arthritis and joint inflammation ( ).

## Introduction

---



**Figure 1.1: Figure adapted from (Nestle et al. 2009)**

## **Introduction**

---

### **1.2.4 Therapeutic intervention**

Nowadays, psoriasis or PsA are still incurable diseases and the different treatments available are solely focused in alleviating the symptoms. For instance, topical therapies are the choice in cases of mild-to-moderate psoriasis, represented by the extended emollients and short-term corticosteroids, due to associated side-effects (Menter et al. 2009). In psoriasis, other topical treatments are used in combination with corticosteroids such ultraviolet (UV) light therapy and vitamin D analogues, directed to inhibit T-cell and KC proliferation, and stimulate KC differentiation (**Rizova2001**). In the case of PsA patients presenting swelling of two or less joints, intra-articular injection of glucocorticosteroids together with joint aspiration is used as a short-term solution to reduce pain and inflammation (Coates et al. 2016). Nonetheless, treatment of most forms of PsA and moderate-to-severe psoriasis require the use of systemic therapies. Patients presenting mild cases of PsA commonly receive nonsteroidal anti-inflammatory drug (NSAID) to control the inflammatory symptoms (Coates et al. 2016). More severe forms of PsA require the use of disease-modifying antirheumatic drugs (DMARDs) including the antagonist of folic acid methotrexate (MTX) and the phosphodiesterase 4 inhibitor Apremilast that act as immunosuppressors of activated T cells and cytokine production, respectively (**Schmitt2014; Keating2017; Gossec et al. 2016; Polachek et al. 2017**).

Remarkably, biologic systemic agents represent the most specific treatment option for severe psoriasis and PsA. This category encompasses an array of cell-based molecular species that modulate the immune response in a physiological manner (Perera et al. 2012). Specifically, the relevance of TNF-alpha in psoriasis and PsA has led to the extensive therapeutic use of TNF-alpha inhibitors during the past five decades, making them the drugs of choice amongst all the biologic agents targeting cytokine pathways. Three TNFi have been approved for the

## **Introduction**

---

treatment of psoriasis: etanercept, infliximab and adalimumab (**Ahil2016**). In addition to those, certolizumab pegol and golimumab are often applied in the management of PsA and other rheumatoid diseases (**Coates2016b**). Despite of TNF- $\alpha$  blockade being one of the most effective treatments, side effects such as increased risk of infection or reactivation of latent infections have been identified (Nickoloff and Nestle 2004). Moreover, between 20 to 50% of the patients fail to respond to the first TNFi administrated, requiring switching to alternative TNFi (Abramson and and 2016). New biologic therapies have been developed to target other key cytokines in the pathogenesis of PsA and psoriasis, such as IL-12, IL-23 (ustekinumab) or IL-17 (secukinumab and ixekizumab) and represent a substantial benefit for treating patients failing to respond to TNFi (**Coates2016b**; Mahil et al. 2016).

## **1.3 Genetics of psoriasis and psoriatic arthritis**

As complex diseases, the risk to develop psoriasis and PsA is not only influenced by the surrounding environmental conditions but also by the genetic background of each individual. Determining the magnitude of contribution of the genetic factors in the development of these diseases and identifying the exact genes or genomic regions involved in the predisposition to psoriasis and PsA remains challenging.

### **1.3.1 Heritability**

Several studies have shown a trend towards the increase of psoriasis and PsA prevalence over the last 30 years in different countries (Organization 2016). This importantly reflects changes in life style habits and it highlights the need to better understand the genetic factors that predispose to disease upon interaction with environmental stresses.

## **Introduction**

---

The contribution of genetics in the development of psoriasis has also been demonstrated in several twins studies. The concordance of psoriasis has been shown to be greater in monozygotic (33-55%) compared to dizygotic (13-21%), estimating an 80% of heritability in this condition (**Faber1974; Penderson2008**; Duffy et al. 1993). Conversely, similar concordance between mono- and dizygotic twins has been reported in the case of PsA, probably due to lack of statistical power and appropriate diagnosis (**Penderson2008**). In the general population, approximately 40% of the patients with psoriasis or PsA have family history in first degree relatives (Gladman et al. 1986). Interestingly, the recurrence in first-degree relatives has been shown to be greater in PsA (40) compared to psoriasis (8) in a study in Icelandic population (Chandran et al. 2009). This suggests differences in the heritability between the two phenotypes and a stronger genetic contribution in PsA.

### **1.3.2 Non-GWAS and linkage studies**

Different approaches have been undertaken to uncover the genetic variability predisposing to psoriasis and PsA. The appearance of next generation sequencing (NGS) techniques and the progressive reduction of cost has allowed to move from single locus gene candidate studies to a genome-wide approach.

The study of psoriasis and PsA genetic architecture started with linkage analyses in family pedigrees presenting an autosomal dominant condition. This approach yielded nine psoriasis susceptibility loci (PSORS1-9) with PSORS1 showing the strongest genetic association (**International2003; Capon 2017**). The PSORS1 locus lies within the MHC I region in chromosome 6p21, previously associated with psoriasis susceptibility in serological studies (**Rusell1972; Tiilikainen1980**). Importantly, Mendelian forms of disease with rare highly penetrant mutations have also been identified in family studies for two genes within PSORS2 (17q25): zinc finger protein 750 (ZNF750) and caspase domain

## **Introduction**

---

family member 14 (*CARD14*) (**Tomfohrde1994**; Jordan et al. 2012). Rare gain of function and *de novo* mutations and also common variants in *CARD14* have been identified in psoriasis and PsA patients, suggesting an important role of the genetic variation in this gene for Mendelian and multi-factorial forms of disease (Jordan et al. 2012; Tsoi et al. 2012). Nevertheless, the inability of independent studies to reproduce the results for regions other than PSOR1, 2 and 4, highlighted the limitations of the linkage studies to understand the genetics of complex diseases (Capon 2017). Additionally, gene based studies in psoriasis and PsA disclosed the importance of genetic variability in the activating killer immunoglobulin receptors 2DS1 (*KIR2DS1*) gene, also reported for AS and RA, which interestingly is mainly triggered by interaction with HLA-Cw<sup>\*</sup>06:02 (**uszczek2004**; **Williams2005**; **Yen2001**; Carter et al. 2007).

### **1.3.3 Genome-wide association studies**

The dramatic advances in sequencing and genotyping technologies have allowed the implementation of association studies at a genome-wide scale. The genome-wide association studies (GWAS) have benefitted from the understanding of common single base-pair changes known as single nucleotide polymorphisms (SNPs) in different populations through whole genome sequencing (WGS) projects such as HapMap (**The international HapMaP Consortium**) and the 1000 Genomes (**The 1000 Genomes**). GWAS generally focus in identifying disease-associated common SNPs (with minor allele frequency (MAF) >5% ) showing differences in allele frequency between patients and controls (Ku et al. 2010). GWAS design are thus based on the hypothesis that complex diseases are more likely to be caused by the interaction of multiple common variants, showing greater power than the previous linkage studies to identify multiple loci with low penetrance and moderate to small effects (Schork et al. 2009; Cui et al. 2010).

## **Introduction**

---

Due to the organisation of the genome into segments of strong linkage disequilibrium (LD) where genetic variants are strongly correlated with each other, the genotyped SNPs in GWAS are solely used a proxy for the disease causative variant. Therefore, disease causal variants can be non-genotyped SNPs or other type of genetic variability such as copy number variants (CNVs), also highly frequent in the genome but less widely studied by GWAS (Hirschhorn 2005; Ku et al. 2010).

Since 2007, when the first psoriasis and PsA GWAS were published, a total of sixty-three genetic associations have been identified at a genome-wide significance ( $p\text{val}>5\times 10^{-8}$ ) which explain 28% of the psoriasis and PsA heritability (Table ??) (Tsoi2017). The majority of the studies have been performed in Caucasian European or North American cohorts but increasing numbers of GWAS in large Chinese cohorts are also being published (Zhang et al. 2009; Sun et al. 2010; Yin et al. 2015). The early GWAS performed in discrete size cohorts with moderate power confirmed association with loci overlapping the PSOR1 and PSOR2 genomic regions from the linkage studies (Cargill et al. 2007; Strange et al. 2010). Specifically, HLA-C has been consistently identified as the most significant locus with the greatest effect size. Additional MHC-I and MHC-II associations with disease risk have been identified for HLA-A, HLA-B and HLA-DQA1 through step-wise conditional analysis(Okada et al. 2014). The information extracted from GWAS studies was significantly enhanced with the use of the Immunochip genotyping platform, which covers 186 immune relevant loci identified in previous GWAS studies across different inflammatory diseases at a greater genotyping density (Tsoi et al. 2012). The psoriasis Immunochip study uncovered fifteen new associations, including the PSOR4 *CARD14* association and also included meta-analysis with the largest available psoriasis cohorts at the time(Tsoi et al. 2012). This meta-analysis has been further expanded yielding sixteen additional associations in the latest

## **Introduction**

---

study and reinforcing the importance of NF $\kappa$ B and cytotoxicity pathways in disease pathophysiology (Tsoi2015; Tsoi2017). Meta-analysis of GWAS across Caucasian and Chinese populations have showed the value of this trans-ethnic approach to identify new associations and understand the differences in the genetic associations contributing to disease risk in different populations (Yin et al. 2015). The importance of conducting psoriasis and PsA independent GWAS studies has been supported not only by their discrepancy in first degree relatives heritability but also by differences in HLA-C and HLA-B alleles frequencies within each phenotype population (Winchester2012; Okada et al. 2014). In fact, cohort stratification confirmed specific GWAS association with PsA for previously identified psoriasis loci such as *TRAF3IP*, *IFNLR1*, *IFIH1* and *NFKBIA* as well as PsA-specific independent signals for *IL23R* and *TNFAIP3* (Stuart2015; Ellinghaus et al. 2010). Interestingly, the association for *LCE3C/B*, identified in combined phenotypic studies, showed greater strength in those patients presenting psoriasis for over ten years without developing joint affection (Stuart2015). Lately, PsA GWAS using the Immunochip platform revealed a PsA-specific association in chromosome 5q31, an independent secondary signal in the *IL23R* region and failed to show genome-wide significance for the *LCE3* locus (Bowes et al. 2015).

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

## Introduction

---

**Table 1.3: Main GWAS studies in psoriasis and PsA.** Summary table describing the most relevant psoriasis and PsA GWAS studies. Information regarding sample size, patients phenotypes and the main reported associations in each study is included. The Ellinghaus *et al.*, 2010 and the Stuart *et al.*, 2015 studies included stratified association analysis of psoriasis and PsA independently. WA=white American; Eur=European; \* Meta-analysis performed.

Study	Etnicity	Sample size (Cases/Controls)	Phenotype	Main associations (putative genes)
(Cargill et al. 2007)	WA	1,446/1,432	Psoriasis, PsA	HLA-C (PSOR1) and <i>IL12B</i>
(Nair et al. 2009)	Eur	1,409/1,436	Psoriasis, PsA	<i>IL23A</i> , <i>IL23R</i> , <i>IL12B</i> , <i>TNIP1</i> , <i>TNFIP3</i> , <i>IL4</i> and <i>IL13</i>
(Stuart et al. 2010)	WA, Eur	1,831/2,546	Psoriasis, PsA	<i>NOS2</i> , <i>FBXL19</i> , <i>PSMA6</i> - <i>NFKBIA</i>
(Ellinghaus et al. 2010)	German	472/1,146	Psoriasis	<i>TRAF3IP2</i>
(Strange et al. 2010)	Eur	2,622/5,667	Psoriasis, PsA	<i>LCE3D</i> (PSOR2), <i>IL28RA</i> , <i>REL</i> , <i>IFIH1</i> , <i>ERAP1</i> , <i>TYK2</i> and <i>HLA-C/ERAP1</i>
(Zhang et al. 2008)	Chinese	1,139/1,132	Psoriasis	(type <i>LCE</i> gene family and <i>IL12B</i> ) I)

## Introduction

---

(Sun et al. 2010)	Chinese	8,312/12,919	Psoriasis, PsA	<i>ERAP1</i> , <i>PTTG1</i> , <i>CSMD1</i> , <i>GJB2</i> ,
(Tsoi et al. 2012)★	WA, Eur	10,588/22,806	Psoriasis, PsA	<i>SERPINB8</i> , <i>ZNF816A</i>
(Tsoi2015)★	WA, Eur	15,000/27,000	Psoriasis, PsA	<i>CARDI4</i> ( <i>PSOR4</i> ), <i>RUNX3</i> , <i>B3GNT2</i> , <i>ELMO1</i> , <i>STAT3</i>
(Bowes et al. 2015)	British, Australians	Irish, 1,962/8,923	PsA	<i>1q31.1</i> , <i>5p13.1</i> , <i>PLCL2</i> , <i>NFKBIZ</i> , <i>CAMK2G</i>
(Stuart2015)	WA and Eur	1,430/1,417	Psoriasis, PsA	5q31 PsA-specific PsA-specific secondary signals (main text), 1p36.23 psoriasis-specific, stronger psoriasis LCE association
(Yin et al. 2015)	WA, Eur, Asian	15,369/19,517	Psoriasis, PsA	<i>LOC144817</i> , <i>COG6</i> , <i>RUNX1</i> and <i>TP63</i> ; signals with ethnic heterogeneity
(Tsoi2017)★	WA, Eur	19,032/39,498	Psoriasis, PsA	<i>CHUK</i> , <i>IKBKE</i> , <i>FASLG</i> , <i>KLRK1</i> , <i>PTEN</i>

## **Introduction**

---

### **1.3.4 Relevance of non-coding versus coding variants in disease susceptibility**

Approximately 88% of all GWAS associations map within non-coding regions and only the remaining 12% account for coding variants likely to cause non-synonymous mutations impacting in the protein function (Welter2013).

Exome psoriasis association studies in Chinese and Caucasian populations have increased the number of coding variants with putative effect in the protein structure (Tang2014; Zuo2015; Dand et al. 2017). These studies have confirmed some of the previously identified missense associations in *CARD14* or *ERAP1*, revealed new common coding variants at previously associated loci and identified protective rare missense changes, for example in the *TYK2* gene(Tang2014; Dand et al. 2017). Nevertheless, results from extensive exome studies suggest that non-synonymous SNPs have a limited contribution to the overall genetic risk of psoriasis compared to non-coding variants (Tang2014).

The association of non-coding variants with disease can be explained by their ability to regulate gene expression in a cell and context specific manner (Fairfax et al. 2012). These variants can be located at different regulatory elements, including enhancer, silencers, promoters and the 5' and 3' untranslated region (UTR) of genes (Ward2012). Non-coding GWAS variants can alter the expression of target genes through different mechanisms including changes in chromatin accessibility, histone modifications, protein binding such as transcription factors (TFs), DNA methylation and binding of non-coding RNA molecules(Knight and C 2014) (Section 1.4.X).

Identification of the target gene regulated by non-coding variants represents a challenge in the field of functional genetics. This can be partially addressed by conducting expression quantitative trait loci (eQTL) analysis, which identifies genome-wide statistical associations between gene transcript levels and SNPs in *cis* (<1Mb) or *trans* to the gene. For instance, in T2D such approach

## **Introduction**

---

revealed a *cis*-eQTL for expression of the TF *KLF4* and a haplotype of non-coding GWAS SNPs located 14Kb up-stream (**Small2011**). Moreover, this haplotype also showed association with genes in *trans*, highlighting downstream targets regulated by KLF4. Nevertheless, eQTL mapping alone only provides instances for transcriptomic regulation and additional functional assays, such as chromatin conformation, are required to demonstrate causality (Edwards et al. 2013).

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

GWAS represent a biologically unbiased approach to shed some light into pathophysiological relevant cell types and molecular pathways associated with disease. In the field of common immune-mediated diseases, GWAS have underlined some of the most important cell types for which genetic variation is functionally relevant. For example, the strongest GWAS associations in T2D are enriched at pancreatic  $\beta$  cells and affect genes involved in insulin secretion, consistently with the characteristic insulin resistance of this pathology (**Visscher2017**). The better understanding of immune-related diseases has likewise led to identification of shared susceptibility loci and the use of therapeutic interventions across diseases, such as the anti- IL-23 and anti- IL-17 antibodies to treat psoriasis, PsA, AS and IBD (**Visscher2017**). Immune diseases have also benefited from the use of the immune targeted genotyping array Immunochip to perform a systematic comparison of the genetic architecture across the different conditions. For example psoriasis and PsA share risk loci, in the same or opposite directions, with AS, CD, MS, RA and T1D. (**ImmunoBase**). Interestingly, cross-disease association studies performed for AS, UC, primary sclerosing cholangitis (PSO), CD and psoriasis have revealed significant enrichment of the 206 multi-trait associated loci for regulatory elements in bone marrow, NK and T cells as well as immune response pathways

## **Introduction**

---

(Ellinghaus et al. 2016). This study has also demonstrated genetic pleiotropy of psoriasis with AS and CD, reflecting how the same alleles predispose to different diseases and supporting the contribution of GWAS to the biological understanding of disease. In the case of psoriasis and PsA, the majority of the GWAS risk loci have been linked to genes that belong to a limited number of pathways and show enrichment for regulatory elements in several cell types (Capon 2017). Nevertheless, non-coding variants from GWAS studies often lack of functional characterisation and tend to be associated arbitrarily to the nearest gene in line with current knowledge about pathophysiology. This is being progressively addressed by the incorporation of tissue and context specific eQTL studies, chromatin conformation assays as well as by the therapeutic efficacy shown by drugs targeting some of those genes.

## **Antigen presentation**

In psoriasis *HLA-Cw<sup>\*</sup>0602* represents the strongest GWAS, also shared with other diseases such as Hepatitis C, PSO and Graves disease (Blais et al. 2011). No differences at the transcript level have been identified for *HLA-Cw<sup>\*</sup>0602* when comparing psoriasis patients versus controls, suggesting alterations in antigen presentation as the mechanism explaining disease association (Hundhausen et al. 2012). The relevance of antigen presentation in psoriasis and PsA has been reinforced by the GWAS association of the endoplasmic reticulum aminopeptidase 1 *ERAP-1* gene, involved in the trimming of peptide antigens. Moreover, GWAS studies identified that *ERAP-1* was associated with psoriasis and PsA only in individuals carrying one copy of the rs10484554 *HLA-C* risk allele (Strange et al. 2010). Similarly, the same study identified a dependent association between *HLA-Cw<sup>\*</sup>0602* and SNPs in the vicinity of the zeta chain of T cell receptor associated protein kinase 70 (*ZAP70*) gene, involved in the regulation of CD8<sup>+</sup> cells auto-reactivity (Picard et al. 2009). These epistatic

## **Introduction**

---

phenomena, whereby association of one gene is dependent on the presence of another, have also been reported between other HLA class I molecules, including *HLA-B\*27* and *HLA-B40*, and *ERAP1* in AS (**Cortes2015b**; Evans et al. 2011). In fact, the AS *ERAP1* GWAS association, which shares signal and direction with psoriasis, increases *ERAP-1* and *ERAP-2* expression and also alters splicing, resulting in an ERAP-1 protein isoforms with increased activity (**Constatino2015**; Hanson et al. 2018).

## **Skin barrier**

GWAS have highlighted KC specific genes such the previously mentioned *LCE* gene cluster and genes with a key role in skin biology such as *CARD14*. Further studies in the *PSORS4* region have revealed that association with diseases is driven by a deletion in two of the genes within this family, *LCE3B* and *LCE3C* (*LCE3C\_LCE3B\_del*)(Cid et al. 2009). The expression of *LCE3B* and *LCE3C* is induced upon barrier disruption, where these proteins participate in the formation of the cornified envelope at the most external layer of the epidermis and are likely involved in the KCs terminal differentiation (**Bergboer2011**). Overall, the lack of *LCE3B* and *LCE3C* expression in psoriasis patients could lead to an impaired repair following skin disruption and potentially facilitate microorganisms infection and trigger a disregulated immune response. In fact, the use of UVB radiation have proved upregulation of *LCE3E* expression 48 hours after treatment, contributing to amelioration of the skin lesions(Jackson et al. 2005). Similarly to the *LCE* gene cluster, *CARD14* is primarily expressed in epithelial tissues mediating the recruitment and activation of the NF- $\kappa$ B pathway in this tissue (Blonska and research 2011). Common and rare pathogenic mutations of *CARD14* in KC cell lines lead to increased activation of NF- $\kappa$ B as well as overexpression of psoriasis-associated genes including IL-6, *TNFA* and *TNFAIP2*, among others (**Jordan2012b**).

## **Introduction**

---

### **NF-κB and TNF pathways**

The NF-κB pathway is involved in the regulation of the innate and adaptive immune response. Several psoriasis and PsA GWAS loci have been mapped to gene members of the NF-κB and TNF signaling pathways including *TNIP1*, *TNFAIP3*, *NFKBIA*, *REL*, *TRAF3IP2* (Table ??). Dysregulation of the feedback loop between TNF- $\alpha$  and NF-κB contributes to the development of many chronic inflammatory diseases and neutralisation of TNF- $\alpha$  is widely used for treatment of immune-mediated diseases, as previously detailed (Liu2017). In psoriasis, elevated levels of NF-κB are present in lesional compared to uninvolved and normal skin (Lizzul et al. 2005). The described psoriasis and PsA GWAS association with the NF-κB inhibitor *NFKBIA* and the NF-κB subunit *REL* are solely driven by nearby intergenic SNPs, lacking of direct experimental evidence for their role of these genetic variants in regulating the expression of both genes (**GWAS studies**). *REL* has been associated with other inflammatory diseases, including CD and RA (**ImmunoBase**) and, interestingly, the RA risk allele has a protective effect in PsA showing opposite direction effects (Bowes et al. 2012). The relevance of members downstream TNF- $\alpha$  signaling is highlighted by GWAS associations with *TNIP1* and *TNFAIP3*, which participate in the regulation of NF-κB activation. SNP variants in these regions have also been identified in CD, UC and SLE, amongst other immune diseases, reinforcing the relationship between these two pathways and chronic inflammation (**ImmunoBase**). For example, a haplotype including missense mutations and intronic variants in *TRAF3IP2* has been reported to drive psoriasis and PsA association by reducing the affinity for TRAF interacting proteins and concomitantly altering NF-κB activation and the IL-17/IL-23 axis(H“”–u“ffmeier2010; Ellinghaus et al. 2010). Moreover, the latest psoriasis and PsA meta-analysis study has also revealed three additional associations with genes belonging to the NF-κB pathway,

## **Introduction**

---

reinforcing the implication of NF- $\kappa$ B activation in psoriasis and PsA development (**Tsoi2017**).

### **Type I IFN and innate host defense**

Psoriasis and PsA GWAS associations have highlighted genes involved in innate immunity including the hosts response to virus and bacteria, importantly represented by members of the type I IFN signaling pathway. Mapping of several GWAS loci to genes from the type I IFN signaling pathway together with clinical and experimental data has reinforced the role of pathogen response in psoriasis and PsA (**Nextle2005**). GWAS associations involved in this IFN response include *IL28RA*, *IFIH1*, *TYK2*, *RNF114*, *ELMO1* and *DDX58*, some of which have been previously reported as susceptibility loci for other immune-mediated diseases. For instance, GWAS-lead SNPs causing missense mutations in *TYK2* have been identified in CD, IBD, T1D, RA and MS, in addition to psoriasis and PsA (**ImmunoBase**). *TYK2* is one of the Janus kinases (JAK) protein family members that initiates the IFN type I downstream response (**Calamonti1994**). Exome-sequencing and GWAS studies have identified two independent protective missense mutations predicted to impair its catalytic activity and thus the initiation of the downstream inflammatory cascade (Strange et al. 2010; Tsoi et al. 2012; Dand et al. 2017). Tofacitinib, the only available JAK inhibitor approved solely for use in RA treatment, is currently undergoing clinical trials in other immune-related diseases alongside with the development of more specific JAK inhibitors (Baker and diseases 2017). Moreover, drugs targeting additional type I IFN pathway members are also under development. For example, monoclonal antibody (Ab) against IFN- $\alpha$  subtypes have failed to suppress the IFN gene signature in psoriasis patients and new approaches towards blocking the IFN- $\alpha$  receptor have shown greater efficacy in SLE (Furie et al. 2017). Psoriasis and PsA GWAS associations with upstream elements of the

## **Introduction**

---

IFN I pathway such intronic variants at *ELMO1* gene may distort the activation of the pathogen-sensing receptors *TLR7* and (**TLR9**) and the subsequent IFN- $\alpha$  production in pDC (Tsoi et al. 2012). Currently, clinical trials testing inhibitors of these TLR receptors are being conducted in SLE (Baker and diseases 2017).

### **IL-17/IL-23 axis**

Together with the TNF pathway, the IL-17/IL23 axis is the most widely targeted by biological therapeutics. GWAS studies have suggested the relevance of this pathway in psoriasis and PsA by several associations including *IL23A*, *IL23R* and *IL12B* (Table ??). The cytokine IL-23, involved on a wide range of pro-inflammatory processed as previously explained, is formed by two subunits: IL-23A/p19 and IL12-B/p40, which is also a component of IL-12 protein. GWAS association has been established by proximity of non-coding lead SNPs to *IL23A* and *IL12* genes but direct functional evidence in regulation of their expression has not yet been established (Cargill et al. 2007; Strange et al. 2010; Tsoi et al. 2012). Nevertheless, transcriptional studies have shown increased levels of p40 and p19 in psoriasis lesional skin and a role for both subunits in abnormal KC differentiation (Lee2004; Zhu2011). Similarly, psoriasis and PsA GWAS associations with *IL23R* has been reported in several studies, including a protective two SNPs haplotype shared with CD in a German and American Caucasian cohort study (Nair et al. 2008; Strange et al. 2010; Tsoi et al. 2012).

Interestingly, a secondary *IL23* signal to the reported by Tsoi *et al.*, 2012 has been specifically associated with PsA and the independency from AS secondary signals for the same locus has also been demonstrated *Tsoi2012,Bowes2015*. The genetic relevance of the Th-17 pathway is partly explained through the cross-talk with the IL-23 response and its role in Th-17 cell differentiation and activation. Nevertheless, psoriasis and PsA GWAS associations with intronic variants at TFs regulating Th-17 differentiation, such as (*IRF4*) and *STAT3*, have

## **Introduction**

---

also been identified (**Huber2008; Harris2007**; Tsoi et al. 2012). These GWAS associations been also reported for other immune-mediated disease including CD and MS (**Immunobase**). The relevance of this axis in the aetiology of psoriasis and PsA is reinforced by the greater efficacy of individual IL-17A or IL-23 blockade compared to TNF inhibition reported by some studies (**Griffiths2015; Blauvelt2017**). Interestingly, the inhibition of IL-17A using secukinumab is effective in the treatment of psoriasis, PsA and AS, whereas it worsens CD (**Patel2012; Hueber2012**). Overall, this stressed the importance of the Th17/IL-23 axis in inflammation and demonstrates that blocking the pathway at different levels translates into different effects within and across inflammatory diseases.

### **Intergenic regions and genome-wide pathway enrichment analysis**

As previously mentioned, the majority of the GWAS associations are located at intergenic regions or gene deserts, difficulting the understanding of their functional and biological relevance. Some examples in psoriasis and PsA include chr1p36.23, chr2p15, chr6q25.3 and chr9q31.2. One of the most interesting regions is chr2p15, which lead SNP and direction of association is shared with AS (**Immunobase**). Within this locus, genes with a role in the immune response involve *B3GNT2* and *CMMID1* (**Maine2007; Tsoi et al. 2012**). The association at chr1p36.23 is shared with UC and proximal to a number of gene candidates including *RERE*, *SLC45A1*, *ERRFI1* and *TNFRSF9* (Tsoi et al. 2012). Unpublished capture-HiC data using the immortal KC cell line HaCaT has revealed interaction of SNPs in this locus with the promoter of the *ERRFI1* gene, inhibitor of the epidermal growth factor receptor signaling required for normal KC proliferation (**Ray-Jones2017**).

New approaches using genetic association data have allowed to perform genome-wide pathway analysis to disclose relevant biological processes in disease. In psoriasis, genome-wide pathway analysis has revealed association

## **Introduction**

---

of novel processes, such as retinol metabolism, transport of inorganic ions and aminoacids and post-translational protein modifications (PTMs) not previously related with the disease aetiology (**Aterido2015**). Interestingly, the *B3GNT2* gene, proximal gene to the chr2p15 association, belongs to the post-translational protein modifications pathway presenting significant enrichment in the study.

### **1.3.6 Limitations and future of GWAS studies**

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

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

Another concern is the heritability missed to be explained by the GWAS associations in relation to the estimated heritability in family studies (Ku et al. 2010; Yang et al. 2010). Since complex traits are influenced by polygenic effects, where the genetic contribution is driven by multiple variants with small effect size, larger experimental cohorts have led to the discovery of new genome-wide significant associations (**Visscher2017**). For example, in human height, most of

## **Introduction**

---

the missing heritability could be explained by GWAS associated variants with nominal significance that failed to pass the stringent threshold due to their small effect size (Yang et al. 2010).

Another source of unexplained heritability may be rare putative causal variants poorly tagged by common SNPs in the genotyping platforms due to differences in the allele frequencies (Wray2005). Such limitations have partly been overcome by improved genotyping arrays like the Immunochip, which incorporates SNPs with MAF<1% (Cortes2011). Moreover, exome studies have also demonstrated the contribution of coding and intronic rare variants (MAF<5%) in the genetic architecture of complex traits such as height or psoriasis (Marouli et al. 2017; Dand et al. 2017). In addition to rare variants, other sources of common variation such as CNV, small (<1Kb) insertions/deletions (indels) and inversions could contribute to the missing heritability. Incorporation of new genotyping platforms has allowed the genome-wide identification of CNV with autism and schizophrenia, among others (Glessner et al. 2009; Marshall et al. 2017). The accurate detection of translocations and inversions relies on the implementation of long reads WGS technologies (Visscher2017). Lastly, the missing heritability may also be the consequence of the overestimated heritability in complex traits as the result of assuming additive genetic effect instead of epistatic interaction between the different associated loci (Zuk2012).

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

### **1.4.1 Overcoming the limitations of GWAS: post-GWAS studies**

GWAS studies shed limited light on the link between genetic variants and disease mechanisms. As previously mentioned, GWAS report associations with disease for a particular locus tagged by a lead SNP in close LD with a number of

## **Introduction**

---

variants, but they fail to identify the true causal variant(s) within the haplotype block Edwards2013. Statistical fine-mapping approaches have been designed to partially overcome those limitations and further refine the association of each GWAS locus towards the most likely causal variant driving disease association within each LD block. In addition to this, the overlap between the fine-mapped variants and functional genomic data will further help to narrow down the candidate set of causal SNPs. The integration of cell type and context specific epigenetic data, including chromatin accessibility, histone modifications and DNA methylation can help to determine the chromatin state where the variant is located and its potential in regulating gene expression Petronis2010. Additionally, the incorporation of gene expression, eQTL analysis and chromatin interaction data will leverage the establishment of a relationship between non-coding variants and putative gene targets. Finally, validation of the functional relationship between the genetic variant and the disease phenotype will involve conducting appropriate cellular assays and *in vivo* animal models.

### **1.4.2 The use of fine-mapping to prioritise causal variants**

Fine-mapping strategies can partially overcome two of the main limitations of the GWAS studies: the association of hundreds of SNP per locus due to LD and the incomplete coverage of the human genetic variation. The aim of fine-mapping analysis is reducing the size of the GWAS genomic intervals and yielding a minimal set of SNPs which will contain the causal variant and explain most of the association for that particular locus (**Spain2015**).

Fine-mapping studies require extensive genotyping to meet the assumption that the putative causal variant will be likely interrogated in the analysis. This can be achieved by WGS, dense genotyping arrays and *in silico* imputation using publicly available data. The use of the Immunochip array across most of the immune-mediated inflammatory diseases has increased the genotyping

## Introduction

---

density at previously associated immune-relevant loci in a cost-effective manner (**Trynka2011**). Similarly, imputation methods using WGS reference panels, such as the aforementioned HapMap and 1000 Genomes Project, have offered genome-wide coverage for SNPs and CNVs with MAF >1% across different ancestry groups (Abecasis GR 2012). More recently, the UK10K project has improved the quality of imputation specifically for rare variants with MAF between 0.01% and 0.5% (**Chou2016**).

Bayesian statistical analysis has been chosen over the frequentist approach (based on p-value calculations) to increase the resolution of the GWAS associations and facilitate the identification of relevant genes and disease mechanism. Bayesian fine-mapping quantifies the evidence of association for each of the genotyped or imputed SNPs as Bayes Factor(BF) later used to calculate posterior probabilities (PP) which represent the probability of each SNP to drive a particular association (**Wakefield2007**). Since including only the most significant fine-mapped SNP would miss the causal variant in approximately 97.6% of loci, the Bayesian strategies reported a credible set of SNPs that account for 95 or 99% of the overall PP in each loci(Bunt et al. 2015). Fine-mapping has been systematically applied to a number of immune-mediated GWAS, including T2D, IBD, AS and SLE (**Gaulton2015; Sun2016; Huang2017**; Maller et al. 2012; Bunt et al. 2015). In contrast, systematic fine-mapping studies for all the sixty-three psoriasis GWAS loci have not been performed yet. In PsA, Bayesian fine-mapping has been conducted for some of the Immunochip GWAS associations, including the 5q31 PsA-specific locus (Bowes et al. 2015).

Regardless the contribution towards dissecting causality, traditional Bayesian fine-mapping models assume only one causal SNP driving the association at each locus. As a way to partially address this limitation, these methods include a locus step-wise conditional analysis to identify independent secondary signals, prior to calculate PP and credible sets for each of them

## **Introduction**

---

Maller2012,Bunts2015. Improved methods such as stochastic Bayesian fine-mapping outperforms the step-wide Bayesian approached by avoiding the biases of the conditional analysis and considering all the possible models regarding number of putative causal SNPs to explain association of each locus (**Wallace2016**)

The resolution of fine mapping studies could be enhanced by the integration of trans-ethnic fine-mapping meta-analysis, particularly by the inclusion of Yoruba (YRI) and Chinese Han (CHB) descendants 1000 Genomes samples with reduced LD blocks, that can shed light on the true independence of secondary signals (**Kichaev2015**; Bunt et al. 2015). Additionally, inclusion of functional data from publicly available sources as priors of the approximate Bayesian model demonstrated reduction of the number of SNPs in the credible set and also increased the proportion of successfully fine-mapped loci (**Kichaev2015**; Bunt et al. 2015). These observations were further reinforced by the integration of fine-mapping data generated by the Bayesian probabilistic identification of causal SNPs (PICS) methods and a map of genomic regulatory elements, showing approximately 60% of the top fine-mapped SNPs overlapping enhancer elements, particularly stimulus-dependent T cells enhancers, as well as significant overlap with accessible chromatin and TF binding sites (TFBS) (**Farh2015**). Recently, publicly available tools such as fGWAS and PAINTOR have leveraged cell type-specific annotation to inform the Bayesian analysis and output a further refine credible set of SNPs with functional relevance (**Pickrell2014; Kichaev2015**).

### **1.4.3 Understanding the epigenetic landscape in complex diseases**

Epigenetic modifications, previously mentioned, consist of heritable changes in gene function independent of mutations in the DNA sequence(**Feil2012**). Environmental and intrinsic factors can trigger changes in the epigenome that

## **Introduction**

---

will result in alteration of gene function through regulation of expression. For example, dietary components such as vitamin B12 intake can results in changes in methylation with locus specific effect (**Wolff1998**). In addition, the genetic background can increase the predisposition to epigenetic changes caused by extrinsic factors. In fact, studies have demonstrated differences in response to environmental factors by different mice breeds as well as greater differences in the epigenetic landscape in human dizygotic twins when compared to those monozygotic (**Pogribny2009; Kaminsky2009**).

The plasticity of the epigenetic landscape is determinant for cell differentiation and identity and particularly important in the immune system to ensure adaptation and response to different pathogen infections (**Yosef2016**). The epigenetic landscape is responsible for the regulation of gene expression and its cell type specific effect has been demonstrated in eQTL studies, proving that between 50 to 90% of eQTLs are cell type and stimulus dependent(**Nica2011; Naranbhai2015; Kasela2017; Dimas et al. 2009; Fairfax et al. 2012; Fairfax et al. 2014; Raj et al. 2014**). The disease-associated GWAS variants have consistently shown enrichment for DNA regulatory elements, characterised by the combination of a number of epigenetic marks including accessible chromatin, histone modifications and DNA methylation **Trynka2013, Trynka2013b, Gusev2014**. For example, 76.6% of all non-coding GWAS SNPs together with those in complete LD have been located within broadly accessible chromatin tagged by DNaseI hypersensitive sites (DHSs) (**Maurano et al. 2012**). Consequently, to leverage the functional interpretation of GWAS hits, consortiums such as ENCODE, The Roadmap Epigenomics and Blueprint have increased the efforts to extensively characterise differences in epigenetic marks including DHS, histone modifications, TF binding and DNA methylation, in a wide range of cell types and conditions (**ENCODE2012; Adams2012; Bernstein et al. 2010**). The integration of those epigenetic datasets has provided a more

## **Introduction**

---

precise insight into the functionality of the genome, allowing the elaboration of cell type specific chromatin states maps. This has been achieved through development of algorithms such as ChromHMM that uses Hidden Markov Model to segment the genome and label it with a chromatin state based on concurrence of several epigenetic marks (**Ernst2010; Hoffman2012**). Amongst the most comprehensive chromatin segmentation maps, The Roadmap Epigenome Project has released chromatin state maps defining eight active and seven repressed regulatory states for a total of 111 elements, including primary cells and tissues (**Kundaje2015; Ernst et al. 2011**).

The latest methodological advances in the field are enabling the personalised study and understanding of the epigenome by the implementation of low cell input and high throughput techniques coupled to NGS (**Buenrostro2013; Schmidl2015; Oudelaar2017**). One step further, the understanding of cell-to-cell epigenomic heterogeneity is also being addressed and may help to elucidate the impact of genetic variability in regulation of gene expression and disease mechanisms (**Cusanovich2015; Rotem2015; Nagano2013; Smallwood2014; Buenrostro et al. 2015**).

Since this epigenetic technical revolution started, systematic studies have been conducted to identify inter- and intra- individual differences and pathological changes in chromatin accessibility and DNA methylation (**Qu2015; Corces2016; Liu2013. Add other ATAC**). In addition to this, personalised epigenomes can also provide insight into disease activity and drug response. For example, differences in DNA methylation of genes responsible for CD4<sup>+</sup> T cell activation correlated with clinical activity in juvenile idiopathic arthritis and different methylation patterns in RA also explained the failure to respond to DMARDs therapy in some patients (**Spreafico2016; Glossop2017**). Overall, the technical feasibility of refining the specificity of the epigenetic maps in a cost

## **Introduction**

---

effective manner will allow to expand the number of epigenomes available to inform the functional follow up and characterisation of GWAS variants.

### **1.4.4 The chromatin landscape**

In the nucleus of the cells DNA is compacted into a highly organised structure known as chromatin. The nucleosome is the basic repeating unit of chromatin and is formed by a 147bp segment of DNA wrapped around an octamere core of histone proteins regularly spaced by 10bp of linker DNA (**Luger1997**). In general, highly compacted DNA will remain more inaccessible for the assembly of the transcriptional machinery preventing gene expression. Chromatin accessibility can be altered by PTM of the histone proteins that will affect their affinity with the DNA within the nucleosome as well as the interaction between nucleosomes in the vicinity (**Polach2000; Pepenella2014**). Additionally, chromatin structure is also influenced by adenosine triphosphate (ATP)-remodelling complexes that facilitate sliding of individual nucleosomes to neighboring DNA segments, increasing temporary chromatin accessibility at particular sites (**Cosma1999**). From the biochemical point of view, the signature of chromatin accessibility, histone modifications, transcription factor occupancy and DNA methylation has been used to define *cis*-regulatory elements such as promoters, enhancers, silencers, insulators and locus control regions, amongst others, as previously mentioned in the chromatin segmentation maps (**Kundaje2015; Boyle et al. 2012**).

#### **Chromatin accessibility**

Accessible chromatin constitutes about 1% of the human genome and represents a very robust marker for histone modifications, early replication regions, TSS and TFBS (**ENCODE2007**). The informativeness chromatin accessibility for understanding gene regulation has driven the development

## **Introduction**

---

of several high-throughput techniques towards accurately tagging these parts of the genome. The "golden standard" technique is DNase I hypersensitive sites sequencing (DNase-seq), which uses the non-specific doublestrand endonuclease DNase I to preferentially cut on nucleosome-free regions known as DHSs. In this approach, isolation of the chromatin-free DNA is followed by further enzymatic digestion and DNA library preparation prior to NGS (**John2013**). DNase-seq also provides high quality information regarding TFBS, generating footprints that allow to identify TF binding in relation to chromatin structure (**Hesselberth2009; Boyle2010**). Another method to interrogate the accessible genome is formaldehyde-assisted isolation of regulatory elements (FAIR-seq), which uses formaldehyde cross-linking, sonication and phenol-chloroform extraction to remove the DNA-protein complexes and retain only the nucleosome-depleted regions that undergo NGS (**Giresi2006**). Both methods have enabled ENCODE to map regulatory elements in several cell lines, primary cells and tissues when abundant collection was possible (**ENCODE2007; Buck2014; Gaulton2010**). Indirect measurement of the chromatin accessibility has also been carried out using micrococcal nuclease-sequencing (MNase-seq), which relies on the endo-exonuclease activity of the micrococcal nuclease to degrade chromatin-free DNA on cross-linked nuclei and retain only the nucleosome-bound material for downstream sequencing (**Axel1975; Ponts2010**). MNase-seq provides a qualitative and quantitative comprehensive map for nucleosome positioning and also TF occupancy. The high number of cells (ideally 5 to 10 M) required by these assays to provide good quality data limits their application to particular biological and clinical samples.

Recently, a new methodology known as assay for transposase-accessible chromatin using sequencing (ATAC-seq) has represented a groundbreaking step to characterise the genomic regulatory landscape (**Buenrostro2013**). ATAC-seq is based on an engineered hyperactive transposase enzyme, known as Tn5,

## **Introduction**

---

that preferentially access nucleosome-free and inter-nucleosomes DNA inserting sequencing adapters at both end of those fragments (Gradman and Reznikoff 2008; Adey et al. 2010). The main advantage of ATAC-seq over DNase-seq is the lower number of cells and the simplicity of the protocol. ATAC-seq requires between 500 to 50,000 cells and is a fast two-steps protocol that yields information about open chromatin and nucleosome positioning simultaneously. These two aspects make ATAC-seq a very versatile technique to interrogate the chromatin landscape in a clinical set-up, where sample availability and time-efficiency are key factors (Scharer2016; Qu2015; Qu2017). Regardless the strengths of this new technique, ATAC-seq sensitivity is not comparable to DNase-seq for some cell types and tissues, and further optimisations of the first released protocol by Buenrostro and colleagues have been implemented (Corces2016; Sos2016; Corces2017).

## **Histone modifications and TF occupancy**

The combination of histone modifications and TF binding to the DNA are essential mechanisms to fully understand transcriptional regulation. Characterising the regulatory elements based on the co-localisation of histone marks in defined combinations is known as the "histone" code (Jenuwein2001). Histone modifications take place in the NH<sub>2</sub>terminal tail that protrudes from the nucleosome and modulate their affinity for DNA-binding proteins and the interaction with neighboring nucleosomes (Bannister2011). Amongst the most common modification are acetylation, phosphorylation and methylation, (Bayarsaihan2011). These PTM are reversely catalysed by histone acetylases and deacetylases (HATs and HDACs), histone kinases and histone methyl transferases and histone demethylases (HMTs and HDMs), which activity and recruitment gets affected by the surrounding histone modifications (Bannister2011; Shi2006; Nelson2006).

## **Introduction**

---

Acetylation of histones increases chromatin accessibility due to the negative charges reducing the affinity for the DNA and is associated with active transcription (). Conversely, histones methylation can exert an activating or repressing effect, depending on other marks. Overall, the combination of histone modifications can be used to broadly divide chromatin into condensed non-transcribed heterochromatin and accessible transcriptionally active euchromatin. Further studies have identified facultative and constitutive heterochromatin, which distinguishes spatially and temporally regulated genes from those permanent silenced, respectively. Facultative heterochromatin is enriched for H3K27me3 and the polycomb repressor complexes (PRCs), whilst constitutive heterochromatin is marked by H3K9me3 (**Hansen2008; Bannister2001**). Several types of chromatin corresponding to different regulatory elements have also been defined. Enhancers and promoters, regardless the activation state, are tagged by high levels of H3K4me1 or H3K4me3, respectively, and both features co-localise with H3K4me2 modifications (**Heintzman2007; Hon2009**). H3K9ac is specifically enriched at active promoter whereas H3K27ac designated activation at both, promoters and enhancers (**Hon2009; Creyghton2010**). Conversely, H3K27me3 together with the heterochromatin mark H3K9me3 informs of gene repression at promoter elements (**Hansen2008; Bannister2001; Pan2007**). Interestingly, GWAS variants for different complex diseases have demonstrated to be relatively enriched for some of those modifications, importantly H3K4me3, H3K9ac, H3K79me2, H3K4me1 and H3K36me3 (Ernst et al. 2011; Trynka and Raychaudhuri 2013). Altogether, the functional understanding and interpretation of thistone marks co-localisation at a particular locus still remains challenging and incorporation of additional epigenetic information is usually required.

TF also play a role, together with histone modifications, in nucleosome positioning as well as in acting as boundary elements to separate chromatin

## **Introduction**

---

states (**Vierstra2014; Bell2000**; Zhang et al. 2009). TF occupancy is indirectly tagged by chromatin accessibility assays such as DHS mapping through reduced cutting sensitivity of DNase I due to protein binding and steric hindrance. The enrichment of GWAS variants within DHS regions highlights the potential role of many disease-associated SNPs to become pathological by altering TF binding and consequently gene expression.

Chromatin immunoprecipitation sequencing (ChIP-seq) has been widely used in the last few years to precisely locate histone modification and TF binding into the genome. This technique allows assaying protein-DNA binding *in vivo* using Abs that specifically recognise histone modifications or TF after DNA-protein cross-link and sonication. Following immunoprecipitation of the desired DNA-protein complexes with the appropriate Ab, the cross-linking is reversed and the proteins digested prior to DNA library preparation and sequencing (**Solomon1988; Barski2007; Johnson2007**). ChIP-seq has been used to analyse a wide range of histone modifications and TF binding in different cell lines, primary cells and tissues (**ENCODE2012; Adams2012; Bernstein et al. 2010**). Similarly to the first generation of chromatin accessibility techniques, ChIP-seq requires at least between 5 to 10 M cells per experiment, restraining its application to the availability of biological material. In order to overcome this limitation, a wide range of protocols have been developed, of which ChIPmentation (ChIPm) stands as the simplest and most cost-effective method, only requiring 10,000 and 100,000 cells to assay histone modifications or TF, respectively (**Schmidl2015**). ChIPm has incorporated the use of the Tn5 transposase to simultaneously fragment and add adapters to the immunoprecipitated DNA, accelerating library preparation and increasing the sensitivity of the results. ChIPmentation has been successfully used to identify, for example, subtype-specific epigenome signatures in chronic lymphocytic leukaemia (**Rendeiro2015**).

## **Introduction**

---

### **DNA methylation**

Together with histone modifications, DNA methylation has a pivotal role in the immune system, importantly in the differentiation of haematopoietic stem cells and the maturation and activation of immune cells (**Sellars2015; Lai2013**). For example, regulation of TNF- $\alpha$  production upon inflammatory stimuli involves a complex network of DNMTs that alter the methylation signature at the locus (**Sullivan2007**). DNA methylation involves the transference of a methyl group to the 5' carbon of the a cytosine that precede a guanine nucleotide or CpG sites by a group of enzymes known as DNA methyl-transferase (DNMTs). CpG islands are found along the entire genome and their methylation generally associates with repression of gene expression (**Herman2003**). Interestingly, DNA methylation is tightly coordinated with histone methylation and the repressive mark H3K9me3 is involved in driving DNA methylation at the same site (**Rottach2009**). The pathogeneity of changes in the methylome has been studies in a wide range of complex diseases including RA, SLE, psoriasis and PsA, amongst others (**Lei2009; Liu2013; Zhang2010**). For example, an genome-wide study of the methylome in PsA PBMCs revealed a hypomethylated pattern in naïve patients compared to those under MTX treatment (**Kim1996**). Whole-genome bisulfite sequencing (WGBS) and bead array hybridisation are currently the most widely used methods to characterise DNA methylation. Both are based on the bisulfite treatment of DNA that converts cytosine into uracil previously to sequencing or probe hybridisation (**Frommer1992; Miura2014; Dedeurwaerder2013**). The use of methylome arrays such as the Illumina HumanMethylation450 Bead Chip is the most cost-effective strategy to detect functionally relevant differences in methylation focusing in the main location of regulatory CpG islands (**Tserel2015; Bonder2017**). This ability to ascertain the epigenome profile in clinical samples has also allowed to conduct epigenome-wide association studies (EWAS) that identify CpG methylation changes between

## **Introduction**

---

patients and controls in a cell type specific manner. As an example, a DNA methylation EWAS in psoriasis skin samples revealed nine disease-associated differentially methylated sites underlying disease status and environmental factors rather than genetic effects (**Zhou2016**).

### **Chromatin interactions and gene expression**

Functional understanding of non-coding variants have benefited from eQTL studies. Nevertheless, eQTLs only provide an indirect evidence of the effect of an SNP on regulating expression of a particular gene. Interrogation of chromatin interactions provides additional evidence for physical contact between enhancers and gene promoters coordinating assembly of the transcriptional machinery and consequently regulating expression. Chromatin is organised into topologically associating domains (TADs) of several hundreds Kb insulated from other TADs by the binding of CTCF protein, amongst others (**Nora2017**). Chromatin loops between promoters and the corresponding regulatory elements mostly take place within the same TAD and are highly cell- and context-specific (**Smith2016**). Since enhancers may not control expression of the closest gene, functional interpretation of GWAS variants requires genome-wide mapping of those chromatin interactions (Smemo et al. 2014). A wide range of genome-wide and high-throughput methods to investigate the 3D chromatin conformation have been developed, showing differences in performance and suitability depending on the application (**Davies2017**). Of particular interest, Capture-C has simultaneously scaled up the number of interactions investigated at high resolution and minimised the number of input cells required (**Davies2016; Oudelaar2017**). Other techniques such as promoter capture HiC have yielded comprehensive immune-specific maps of promoter-enhancer interactions in seventeen human primary hematopoietic cell types (**Javierre2016**). Lately, HiChIP has improved the integration of ChIP and chromatin interaction methods

## **Introduction**

---

to enhance the specificity of the assay while reducing sequencing depth and input material (**Mumbach2016**). Importantly, HiChIP based on the active histone mark H3K27ac has produced interesting results about the specificity of chromatin interactions in immune-relevant cells type in the context of chromatin accessibility and histone modifications (**Mumbach2017**). Cell-type specific HiChIP data has also demonstrated ability to further fine-map immune-mediated GWAS loci and nominate functional causal variants within haplotype blocks.

### **1.4.5 Transcriptional profiles in disease**

The role of environmental and genetic factors in altering gene expression regulation in complex diseases has been investigated through extensive comparison of case-control transcriptional profiles. The informativeness of this approach is conditioned by studying the relevant disease tissue, which sometimes remains challenging due to the lack of pathophysiological understanding of disease mechanisms or difficulties in accessing it. In immune-mediate diseases, PBMCs differential gene expression (DGE) analysis between patients and controls has enabled to identify relevant pathways and biochemical functions in RA, UC, SLE, AS, psoriasis and PsA, amongst others (**Miao2013; Junta2009; Baechler2003; Assassi2010; Batliwalla2005**). The growing evidence supporting cell type and context specificity relevance in the regulation of gene expression has driven more disease-specific targeted studies. Such studies include synovial-isolated macrophages in RA, B cells and monocytes in SLE or skin biopsies in psoriasis (**Katschke2001; Dozmorov2015; Jabbari et al. 2012**). The extensive overlap of GWAS variants with non-coding regions potentially dysregulating gene expression has highlighted the role of eQTL studies, previously explained, as a very informative tool to link GWAS variants with targets genes. In this lines, consortium such as the Genotype Tissue Expression (GTEx) have generated publicly accessible comprehensive tissue-specific eQTL studies that have greatly

## **Introduction**

---

contributed to the functional understanding of GWAS risk alleles in many complex diseases Lmondsdale2013,Fagny2017. Lately, eQTL studies have also been performed in the context of disease. For instance, an eQTL study in five immune relevant cell types isolated from IBD and anti-neutrophil cytoplasmic antibody-associated vasculitis patients have revealed disease specific eQTLs, some of which disappear following treatment (**Peters2016**).

### **micro-RNAs**

In the last few years, the understanding of transcription has experienced a profound revolution, revealing that most of the genome undergoes transcription (**ENCODE2007**). In addition to protein coding mRNAs, a number of non-coding RNAs have been characterised and demonstrated to have a role in regulation of transcription and gene expression. In fact, between 30 and 80% of human genes are predicted to be under transcriptional control of micro-RNAs (miRNAs) (**Lewis2005; Friedman2008**). miRNAs are generated as larger precursors through transcription of non-coding regions of the genome and undergo processing into 21 to 24 nucleotides long (**Lee2002**). Under particular conditions, expression of genes containing complementary sequences to miRNAs are commonly negatively regulated through assembly of the miRNA-induced silencing complex followed by mRNA degradation, mRNA destabilisation or translational repression (**Ameres2010; Braun2011; Petersen2006**). Several studies highlighting the role miRNAs transcriptional regulation in disease have been conducted, particularly in the context of immune-mediated complex diseases, including psoriasis (**Lerman2011**).

### **Long non-coding RNAs and enhancer RNAs**

Another category of non-coding RNAs are the long non-coding RNAs (lncRNAs), transcripts between 200 and 100Kb long that undergo splicing, 5'

## **Introduction**

---

capping and 3' poly-adenylation (**Derrien2012**). A large number of lncRNAs with different functions have been identified and different categories have been established based on location (nucleus or cytoplasm) and mechanisms of action (**Rinn2012; Fatica2014**). LncRNAs can positively and negatively regulate transcription through different mechanisms including guidance of chromatin modifiers such as DMTs and PRCs to specific loci, alteration of mRNA stability, translational control and decoy for miRNAs and other regulatory proteins(**Pandei2008; Faghihi2008; Gong2011; Carrieri2012; Kino2010**). According to the latest GENCODE annotation release, 15,778 lncRNAs have identified in humans (**Derrien2012**). Amongst the characterised lncRNAs, many have been demonstrated to play a role in the regulation of the innate and adaptive immune response, for example in T cells activation and host-pathogen interactions (**Pang2009; Rossetto2009**). Moreover, differential case-control gene expression analysis have underscored the contribution of lncRNAs in several chronic inflammatory conditions, including RA, SLE and psoriasis (**M””-u”ller2014; Shi2014; Ahn2016; Li et al. 2014**).

A particularly relevant type of lncRNAs are the enhancer RNAs (eRNAs), shorter molecules compared to the canonical lncRNAs (approximately 346 nucleotides) that do not undergo splicing or poly-adenylation (**FANTOM2014**). Although traditionally chromatin segmentation maps have defined enhancers as DNA regions with particular characteristics, later studies have shown their ability to be bi-directionally transcribed into eRNAs molecules (**De Santa2010; Kim et al. 2010**). Importantly, the transcriptional activity of enhancer has been demonstrated to be an excellent proxy to identify functionally active regulatory regions successfully validated by reporter assays (**FANTOM2014; Anderssen2014**).

Methods to assay gene expression The use of micro-arrays based methods to perform genome-wide expression studies have been increasingly replaced

## **Introduction**

---

by RNA sequencing (RNA-seq), as result of NGS technologies becoming more cost-effective. RNA-seq is a method that performs sequencing of all the RNA species contained in the sample, overcoming the bias introduced by the use of pre-designed complementary probes in the hybridisation techniques. RNA-seq involves retro-transcription of the extracted RNA into cDNA and PCR amplification preserving relative abundance of each transcript prior to library preparation and NGS (**Mortazavi2008**). Systematic comparison of gene expression has shown superior dynamic range of detection for RNA-seq compared to micro-arrays, particularly for low abundance transcripts (**Zhao2014**). Moreover, RNA-seq also allows capturing additional information such as identification of new exons, alternative splicing events and allele-specific expression (ASE). Regulation of gene expression through protein isoform abundance is very common at different tissues and during particular biological processes. For example, RNA-seq isoform quantification has highlighted that differentiation of CD4<sup>+</sup> T cells into the pro-inflammatory Th-17 is particularly driven by one of the nuclear receptor RORC $\gamma$  isoforms (**Zhao2014**). RNA-seq has also allowed to quantify ASE for transcripts in individuals heterozygous for exonic SNPs haplotypes in a particular gene, which previously required additional molecular assays (**Yan2002**). Importantly, ASE has provided direct evidence for local/*cis*-eQTLs to be explained by allele-specific mechanism, showing significant differences in haplotype transcript abundance for up to 88% of the genes with an associated cis-eQTLs (**Pickrell2010**). Furthermore, the development of single-cell RNA-seq (scRNA-seq) has enabled the identification of cell sub-populations within a tissue in an unbiased way (**Tang2009; Tang2010**). scRNA-seq does not require prior isolation of populations based on a panel of surface or intra-cellular molecular markers by FACS and identifies cell-to-cell variation and rare populations using the transcriptomic profiles of thousand of cells. scRNA-seq has importantly contributed to the field of

## **Introduction**

---

immunology identifying new subsets of DCs and monocyte populations involved in mounting the immune response and re-defining the atlas of human blood myeloid cells (**Jaitin2014; Villani2017**). Furthermore, precise identification of TSS and the associated promoters for each transcript has been enabled by cap analysis of gene expression (CAGE) and other 5-end RNA-sequencing methods, designed to clone short sequence tags from the 5 ends of cDNAs (**Yamashita2011; FANTOM2014**). The CAGE data generated by the functional annotation of the mammalian genome 5' (FANTOM5) Consortium has also sequenced thousands of eRNAs and contributed to better characterise the definition of enhancers and their spatial and temporal specificity in hundreds of human primary cells and tissues (**Andersson2014**). Importantly, this data has stressed the relevance of mapping not only histone marks and DHSs but also the eRNAs to confidently distinguish active enhancers from putative regulatory regions non-functional in a particular cell type, tissue or condition.

### **1.4.6 Transcriptional regulation in complex diseases**

Non-coding GWAS variants can exert the pathogenic effect by affecting one or many of the previously described mechanisms responsible for the tight regulation of gene expression in homeostatic conditions. For example, intronic SNPs can influence mRNA splicing through exon skipping and resulting truncated but functional proteins. This is the case for the intronic risk allele at the TNF Receptor Superfamily Member 1A (*TNFRSF1A*) associated with multiple sclerosis (MS), which results in a soluble isoform of the TNFRS1A protein with TNF antagonistic function (Gregory et al. 2012). Non-coding variants at enhancers, silencers and promoters can dysregulate gene expression by altering affinity at TFBS, histone modifications and chromatin accessibility. For example, in thyroid autoimmunity, the risk allele of an intronic SNP in the thyroid stimulating hormone receptor (TSHR) gene reduces *TSHR* protein

## **Introduction**

---

expression (which cell type) (Stefan et al. 2014). The risk variant increases the affinity of the repressor promyelocytic leukemia zinc finger protein (PLZF) that recruits HDACs to the locus, resulting in impaired tolerance to thyroid auto-antigens. Alterations in TF binding can also affect looping and long-range chromatin interactions between enhancers and promoters. For instance, in prostate cancer this phenomenon causes upregulated expression of the oncogene SOX9 due to increased enhancer activity and enhancer-promoter interaction (Zhang2012). Alternatively, non-coding SNPs regulate gene expression by creating a new promoter-like element, as in the  $\alpha$ - thalassemia disease, where this phenomenon leads to dysregulated downstream activation of all  $\alpha$ -like globin genes in (cell type)(Gobbi et al. 2006). Since few enhancers have shown to exert regulation of gene expression through the eRNA molecules, enhancer GWAS variants of immune-related disease could affect that function(Shechner2015; Fahr2014). lastly, Non-coding variants placed at UTRs and intergenic regions can affect binding of miRNAs and lncRNA to the target genes. For example there is a CD associated variant at the 3'UTR of the gene immunity related GTPase M *IRGM* that reduces binding of the miR-196 which increases mRNA stability and translation, ultimately resulting in disrupted autophagy (Brest et al. 2011). In psoriasis and PsA, some specific SNPs located at 3' UTR of genes such as *IL-23*, *TRAF3IP2* or *SOCS1* have been hypothesised to disrupt or create *de novo* miRNA binding sites, but no experimental evidence has been provided yet (Pivarcsi et al. 2014).

### **1.4.7 Integration and interpretation of genomic data**

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

## **Introduction**

---

effect that involved several genes and signaling pathways through alteration of epigenetic features and subsequent dysregulation of gene expression. Tools such as RegulomeDB allows querying a large number of human publicly available epigenetic and functional data, including DHSs, TFBS, histone modification and DNA-protein interactions, at the SNP level (Boyle et al. 2012). Other powerful tool to interpret and integrate different -omics data is the University of California Santa Cruz (UCSC) genome browser. This enables display of in-house and up to date publicly accessible annotation data, including epigenetic features, chromatin segmentation maps, expression data, eQTLs, TFBS or chromatin interactions, amongst others, at the single bp resolution (**Kent2002**). In addition to this, the international consortiums generating large-scale epigenetic and expression data such as ENCODE, Blueprint, The epigenome Roadmap, the International Human Epigenome Consortium (IHEC), GTEx or FANTOM have created comprehensive website resources for browsing and downloading data.

In addition to data integration, other bottleneck encountered by functional genomics is the clinical relevance of the GWAS SNPs, eQTLs, differentially expressed genes or differentially epigenetic modified regions. This can be addressed by performing enrichment analysis, which tests for statistically significant over-representation of particular annotation terms within the entities of interest. Examples of annotation sources include ontologies, signalling pathways or functional elements. Pathway enrichment analysis uses functional units containing related genes defined by prior knowledge. Amongst the most comprehensive and informative pathways sources are The Kyoto Encyclopedia of Genes and Genomes (KEGG) and the REACTOME, which also considers biochemical reactions such as binding, activation or protein translocation (**Kanehisa2000**). Such annotation sources may be used to interpret for example a set of differentially expressed genes; additionally, pathway enrichment analysis can take as input a list of genes obtained from annotation of non-coding

## **Introduction**

---

regions using chromatin interaction or eQTL studies. Similarly, enrichment analysis of functionally annotated regions can also be performed for a varied collection of genomic and epigenomic features such as DHSs or histone modification, particularly useful to help identifying regulatory elements in relevant cell types in complex diseases. For example, enrichment analysis using Minimum distance-based Enrichment Analysis for Genetic Association (MEAGA) showed enrichment of the latest meta-analysis psoriasis GWAS hits for enhancers in Th-1 and Th-17 CD4<sup>+</sup> and CD8<sup>+</sup> T cells ([Tsoi2017](#)).

From the number of tools designed to perform this type of analysis, eXploring Genomic Relations (XGR) and functional mapping and annotation of genetic associations (FUMA) are particularly powerful ([Fang2016](#); [Watanabe2017](#)). For example, XGR is an open source R package and web-app that allows to handle different types of input data (SNPs, genes and regions). XGR integrates a wide range of ontologies and up to date publicly available functional data and perform different types of annotation and enrichment analysis, facilitating background customisation for reliable and meaningful output results. Moreover, XGR also performs gene network analysis from the same inputs as the pathway analysis. This leverages experimentally validated interaction information to identify gene networks modulated by putative pathogenic variants, improving interpretation through consideration of network connectivity.

### **1.4.8 Immunophenotyping**

Not sure if needed as such of merge with sth else CyTOF

### **1.4.9 Approaches to establish disease mechanisms and causality of genetic variant**

Prioritisation of non-coding variants by integrating fine-mapping, epigenetics and expression data, as previously described, still does not

## **Introduction**

---

unequivocally addresses the functional mechanisms conferring them pathogenic effect in a cell type and context specific manner. To overcome this, a wide range of experimental approaches can be applied to perform functional validation and test the predicted effect of the variant in regulating gene expression. *In vitro* assays to investigate the effect of genetic variants in regulating gene expression, include transfection of constructs containing the promoter or enhancer element to test followed by luciferase expression (**Niimi2002**). Other molecular assays to interrogate allelic differences in affinity for TF binding include electrophoretic mobility shift assay (EMSA) and ChIP using Ab for the particular TF of interest followed by qPCR quantification for the target region (**Vernes2007**). The need to perform these assays at a genome-wide scale has yielded to development of high-throughput technologies, such as massively parallel reporter assays (MPRAs), which test putative enhancers and the effect of genetic variability in their functionality (**Kheradpour2013**). In addition to this, mass spectrometry (MS) techniques has been used to perform allele-specific quantitative proteomics, which has revealed allele-dependent binding of TF and co-regulators at the T2D *PPARG* GWAS locus (**Lee2017**). *In vivo* validation has traditionally involved the use of mice models, including knock-outs for the potentially pathological genes or the regulatory elements containing GWAS prioritised variants. Nevertheless, the use of mice models to study human genotype-phenotype relationships has shown to have limitations that need to be taken into account when interpreting the results (**Ermann2012**).

Both, *in vitro* and *in vivo*, models for functional studies have benefited from the development a the genome-editing technology known as clustered regularly interspaced short palindromic repeats (CRISPR/Cas) (**Cong2013**). CRISPR/cas enables monoallelic and biallelic modifications of primary cells and embrionic stem cells (ESCs) for the particular SNP or region of interest. The limitations of CRISPR to edi certain primary cells is being overcome

## **Introduction**

---

by the use of human induced pluripotent stem cells (hiPSCs), which can undergo terminal differentiation into the cell type of interest after CRISPR modification (**Ding2013**). For example, neurons differentiated from CRISPR/cas edited iPSCs have been used in schizophrenia to functionally dissect a risk haplotype in the promoter of *MIR137* (**Forrest2017**). Moreover, CRISPR/cas has also been used for high-throughput interference screens (CRISPRi) to discover regulatory elements and identify their target genes by altering chromatin state at particular locus **Fulco2016**. Similarly, CRISPR activation (CRISPRa) assays have been used to identify stimulus-responsive enhancers independently of stimulus exposure, which represents allows simplification of the experimental designs (**Simeonov2017**). Both approaches have been used in combination with HiChIP for validating linkage of non-coding variants to putative gene targets (**Mumbach2017**).

# **Chapter 2**

## **Establishment of laboratory methods and analytical tools to assess genome- wide chromatin accessibility in clinical samples**

### **2.1 Introduction**

**Previous and current methods to identify the accessible genome  
in cells and tissues**

**Implementation of ATAC-seq to define the chromatin landscape**

**Technical limitations and recent advances in optimisation**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4473780/>

Talk about ATAC being more variable, a native chromatin accessibility assessment without cross-linking. Role of transposase ability in accessing the chromatin, debris and DNA from dead cells adding noise

Paper to justify peak calling: A comparison of peak callers used for DNase-Seq data.

New ATAC but also explanations of the limitations: Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay

## **Challenges of working with clinical samples**

### **2.2 Results**

#### **2.2.1 Establishment of an ATAC-seq data analysis pipeline based on current knowledge**

When the first ATAC-seq publication (**Buenrostro2013**) appeared, there were not well established protocols for the complete processing of the data. Since then, several publications have used ATAC-seq and modifications of this protocol together with a wide range of data analysis strategies to answer different biological questions (Table 2.1). There are several limiting aspects in the process of analysing ATAC-seq data, including QC assessment, peak calling/filtering and differential analysis of chromatin accessibility regions between groups. Using the current knowledge in the field as well as on my own analysis, I agreed on the most appropriate criteria and parameters to implement in our in-house pipeline. For this purpose, I used ATAC data generated with the first protocol (**Buenrostro2013**) in paired CD14<sup>+</sup> monocytes and CD4<sup>+</sup> total T cells from the same three healthy individuals, all of them downsampled to 30 million of reads, in order to facilitate the comparison across all of them.

Table 2.1: .

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

Alasoo <i>et al.</i> , 2018	MACS2 (-nomodel -shift -25 -extsize 50 -q 0.01	Merge of peaks from all conditions present in at least three samples of the same condition.	Peak based: TMM normalisation and lima voom (FDR<0.01).
Qu <i>et al.</i> , 2017	ZINBA PP>0.99.	Merging of filtered peaks from each individual sample.	Quantile normalisation and peak based in house Pearson correlation method.
Rendeiro <i>et al.</i> , 2016	MACS2 (-nomodel -extsize 147)	Merge of peaks from all samples in an iterative process including permutations	Peak based: quantile normalisation and Fisher exact test (FDR<0.05).
Schareret <i>et al.</i> , 2016	HOMER (-style dnase)	Merge of all overlapping peaks between all samples using HOMER mergePeaks	Peak based: TMM normalisation and edgeR package (FDR<0.05).

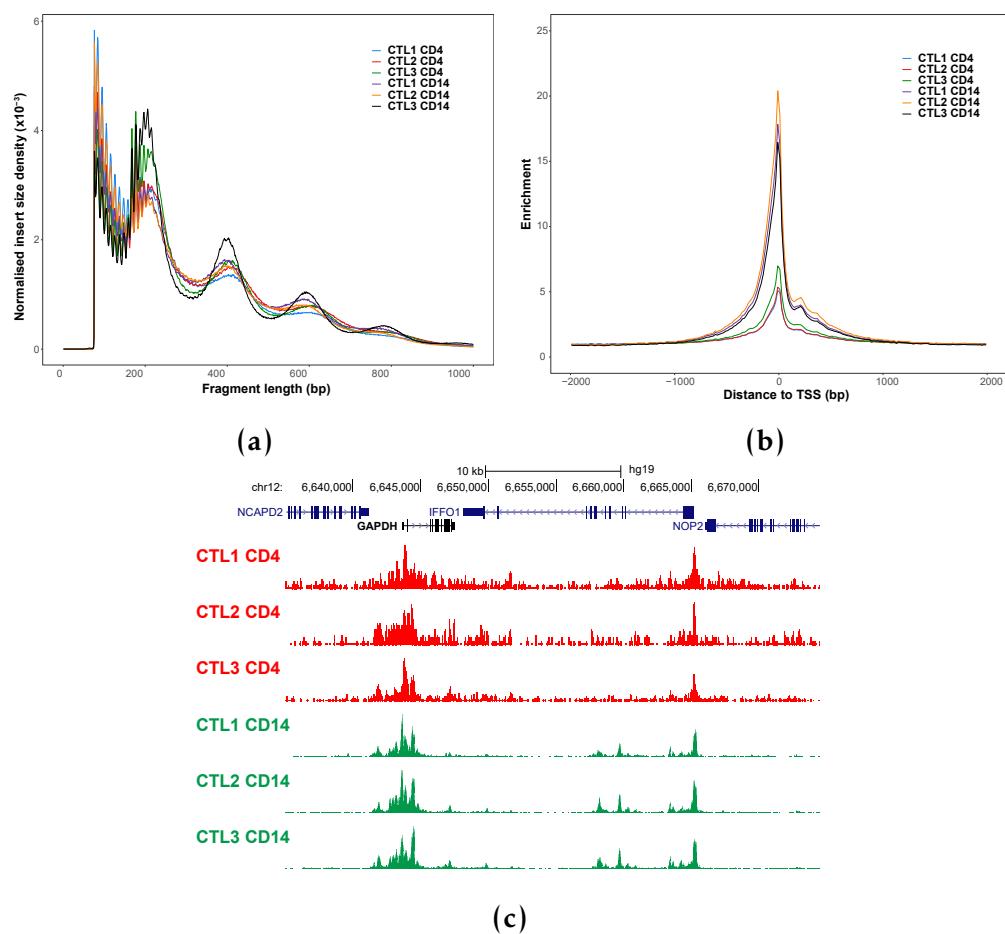


### **Sample quality control**

Regarding QC measurements, the variability in performance of the methodology, particularly ATAC-seq and Fast-ATAC, has required to agree on appropriate parameters to determine the quality of the samples before proceeding with downstream differential analysis. After reviewing the different read-outs implemented across different publications as well as the recently ENCODE update, I have identified the most informative ones showing supporting correlation between them.

Firstly, I analysed the fragment size distribution for each of the six samples in order to determine if they recapitulated the expected periodicity of nucleosomes protecting the DNA during the transposition event (Figure 2.1a). All the samples showed periodicity every ~200bp up to 600bp, clearly distinguishing chromatin organisation into mono-, di- and tri-nucleosomes. The relative intensity of nucleosome-free DNA fragments (<~147pb) compared to nucleosome-bound DNA was greater for some of the samples (e.g CTL1 CD4<sup>+</sup> and CD14<sup>+</sup>) and similar or lower for others (e.g CTL3 CD4<sup>+</sup> and CD14<sup>+</sup>). Nucleosome-free fragments(peak<~147bp) are also clearly distinguished in all of the samples, meeting the ENCODE QC recommendations (**ENCODE**).

Another QC measurement was the enrichment of ATAC-seq signal over a random background of reads across all the TSS identified for Ensemble genes (Figure 2.1b). It is well established that nucleosome repositioning and an increase in chromatin accessibility take place at TSS to allow formation of the transcriptional machinery and initiation of transcription. Fold-enrichment signals ranged between 5-7 for the CD4<sup>+</sup> samples and they were much higher(between 17-20) for the CD14<sup>+</sup> samples. The lower sample quality of the CD4<sup>+</sup> compared to CD14<sup>+</sup> shown by the TSS signal were recapitulated by the ATAC-seq signal at the promoter of the constitutively expressed gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) (Figure 2.1c).



**Figure 2.1: Measurements for quality control assessment in ATAC-seq samples**

As part of the QC assessment I looked at the percentage of mitochondrial reads and the fraction of reads in peaks (FRiP)(Table 2.2).

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

**Table 2.2:**

FRiP score is a way of assessing the background signal in different types of assays that are based on peak calling, including ChIP-seq. Positive correlation between the TSS fold-change enrichment and FRiP was observed (data not shown), being both appropriate inter-dependent QC measures to evaluate sample noise. Regarding TSS and FRiP cut-off values, Alsooo *et al.*, 2018 and, recently, ENCODE have recommended minimum FRiP between 10-20% and TSS between 6-10. ENCODE has prioritised the use of TSS over FRiP as the measurement to determine the noise in the sample (**ENCODE**). According to this recommendations all these samples passed QC; however clear differences were seen between CD4<sup>+</sup> and CD14<sup>+</sup> samples. The mitochondrial content ranged between 14.9-43.3% and, alike FRiP and TSS, it was higher in CD14<sup>+</sup> than in CD4<sup>+</sup> and not directly related with any of the other QC measurements.

### **Peak calling and filtering**

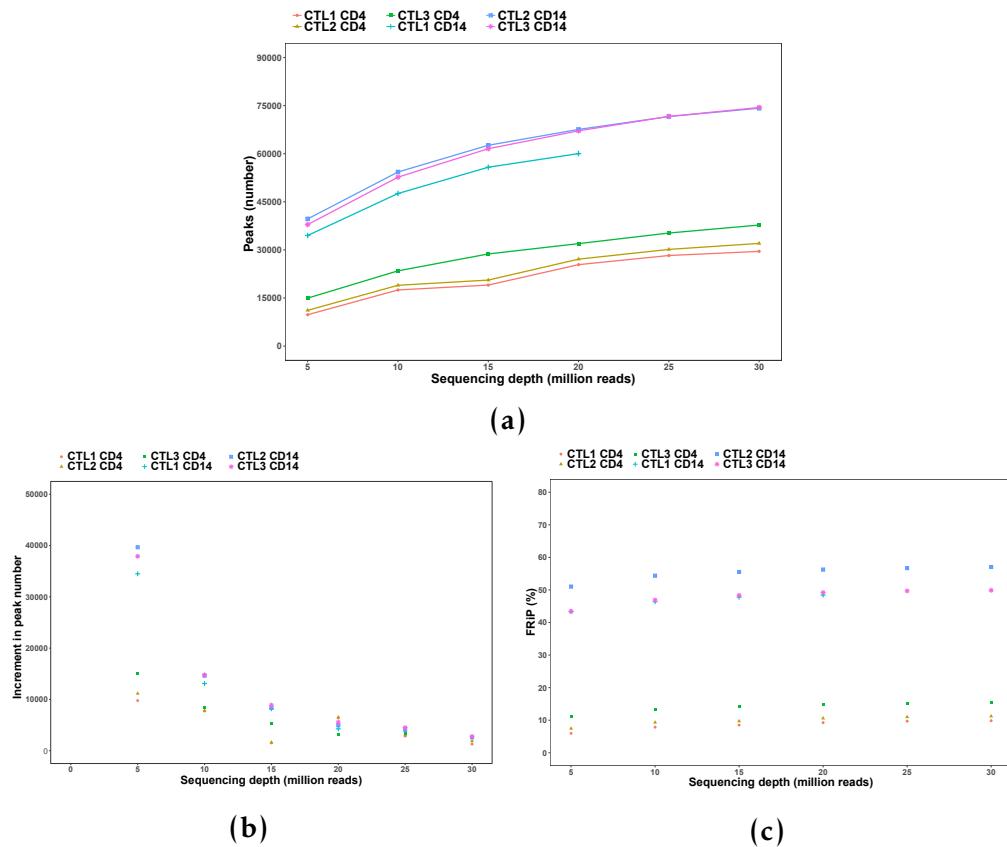
As part of the ATAC-seq pipeline implementation, peak calling and the criteria for filtering where another two aspects to determine. Although different peak callers have been used, most of the publications as well as ENCODE has been using MACS2 as the preferred methodology (Table 2.1). MACS2 has been initially developed for ChIP but it has also been used for DHS and ATAC-seq

with disabling the model and agreeing in an extension size (`-extsize`) and a shift (`-shift`), which indicate the direction and number of bp for reads to be shifted and the number of bp for them to be extended, respectively. The `-extsize` should correspond to the average fragment size, which in my libraries is  $\sim$ 200bp and the `-shift` is set to -100, as it is recommended to be set to  $-1/2$  of the fragment size for chromatin accessibility assays. This parameter could be further optimised but it escapes from the aim of this thesis.

I was interested in understanding the effect of sequencing depth and the sample quality on the peak calling to have a better control of both variables in the downstream analysis. I performed random read sub-sampling every 5M total reads (from 5M to 30M) followed by peak calling with arbitrary filtering for  $\text{FDR} < 0.01$  in each of the six aforementioned samples.

The number of called peaks passing filtering showed an steady increased over the read depth which seemed to reach a *plateau* around 25M reads (Figure 2.2a). This was consistent with the decay in the increments of called peaks over read depth, almost invariable, from 20M reads onwards (Figure 2.2b). Moreover, lower number of peaks were detected in CD4 $^{+}$  samples compared to CD14 $^{+}$  highlighting the influence of sample quality on the total number of called peaks. Interestingly, sample quality measured by FRiP reflected very low changes over read depth and was stable from 15M reads for all six samples (Figure 2.2c). Overall, this confirmed that measurement of sample quality by FRiP or TSS is not biased by sequencing depth.

Regarding peak calling filtering, most of the ATAC-seq publications using MACS2 have arbitrarily used an  $\text{FDR} < 0.01$  (Table 2.1). In collaboration with Dr. Gabriele Migliorini and following ENCODE pipeline, we explored the use of IDR to experimentally identify the most appropriate p-val for filtering each individual sample. Each sample was partitioned in two, peaks were called in each half and the percentage of peaks (over the total number shared peaks)



**Figure 2.2: Peak calling at different sequencing depth in ATAC-seq samples**

sharing IDR at a particular p-val was calculated (Figure 2.3 a and b). Both of the representative samples showed variation in the percentage of shared peaks upon sequencing depth under 10M reads, being the effect more pronounced and extended in the lower quality (CTL2 CD4<sup>+</sup> Figure 2.3 a) compared to the counterpart CD14<sup>+</sup> (Figure 2.3 b). The shape of the curves was also influenced by the sample quality, presenting a smoother profile reaching a single maximum percentage of shared IDR peaks for samples with TSS enrichment >~10 compared to samples with lower quality. All the CD14<sup>+</sup> samples reached the maximum percentage of IDR shared peaks at approximately -log10 pval 8 (data not shown). Filtering the CD4<sup>+</sup> peaks at the -log10 pval of the first maximum of IDR shared peaks reduced the percentage of peaks overlapping noise ( e.g heterochromatin, repetitive sequences and repressed regions) when compared to peaks filtered based on FDR<0.01 (Figure 2.3 b). In summary, this

## Establishment of methods to assess genome-wide chromatin accessibility

IDR analysis appeared as systematic method to identify an optimum p-val to perform individual filtering in a sample-specific manner and in a less arbitrary way than the extended 1% FDR.

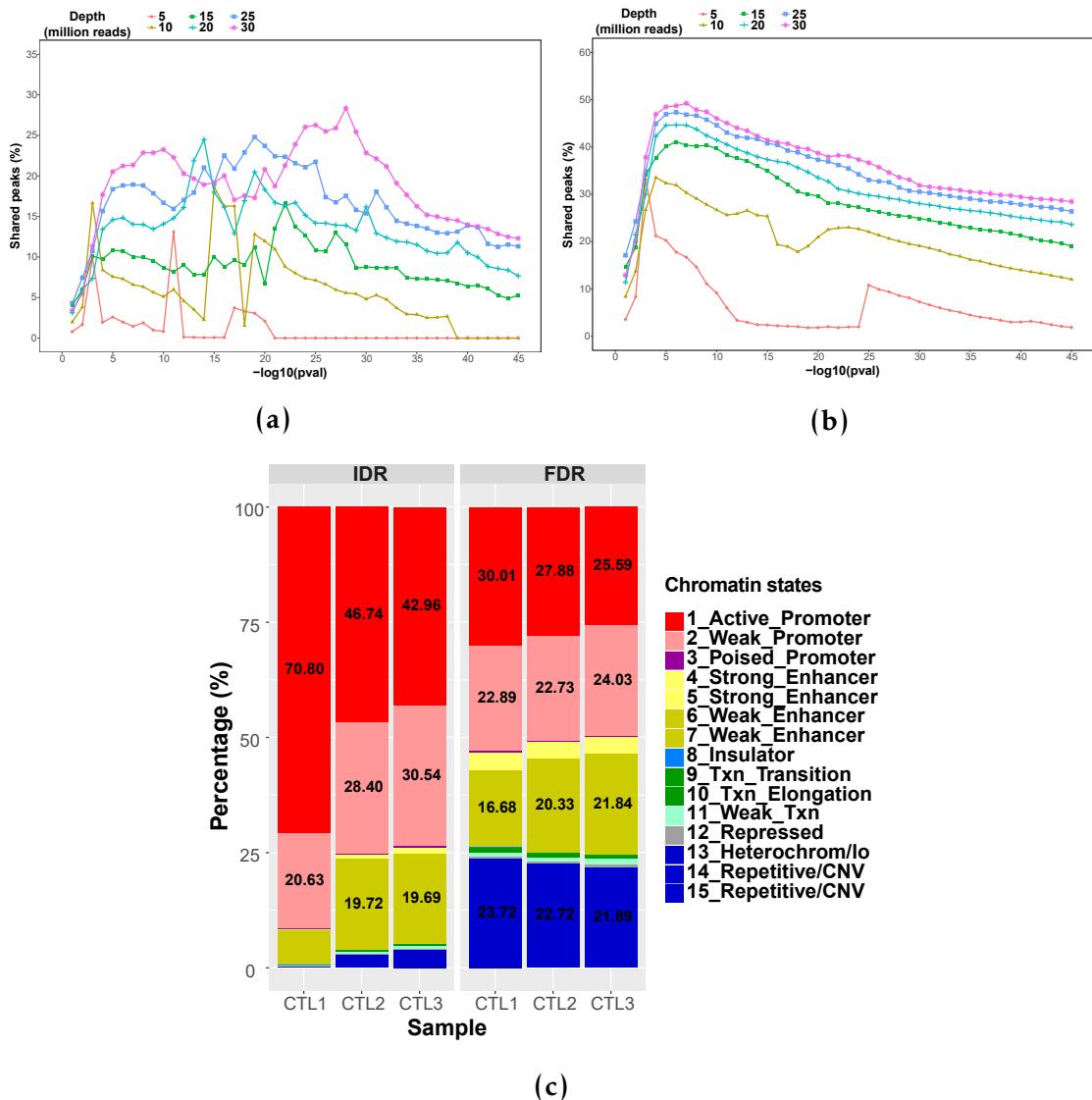


Figure 2.3: Peak calling filtering using IDR analysis in ATAC-seq samples

## Differential chromatin accessibility analysis

From the methods that can be used to perform differential chromatin accessibility analysis (Table 2.1), I chose a peak-based approach where a consensus master list between all samples was built and the number of reads overlapping the master list peaks were retrieved for each sample. As previously

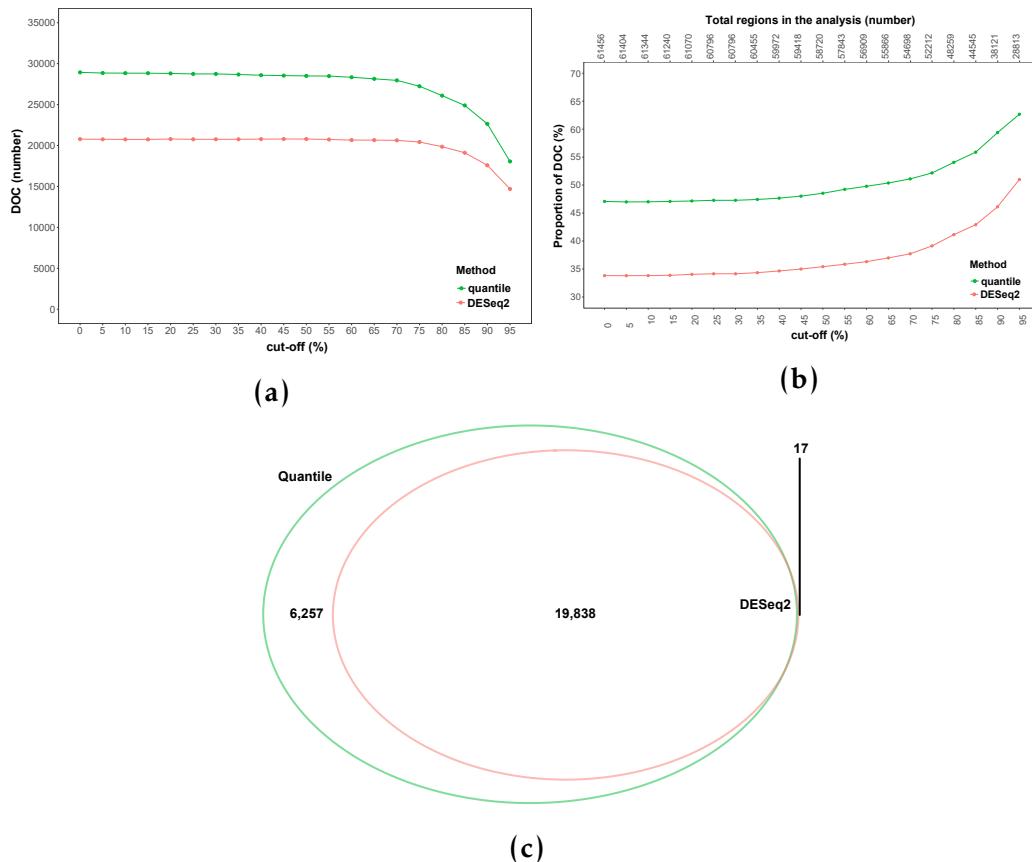
## **Establishment of methods to assess genome-wide chromatin accessibility**

---

mentioned in the Chapter the master list was composed of non-overlapping 500bp with peaks present in at least 30% of the samples, regardless the group they belonged to (e.g patients or controls). One of the main limitations of the ATAC-seq and FAST-ATAC protocols (discussed in the next section) is the background signal. Therefore, it was calculation of an empirical cut-off, similarly to the strategy use in micro-array technology, was performed to minimise the impact of background read counts on the differential analysis (Xinmin2005; Jonker et al. 2014). Moreover, due to the lack of consistency found across the ATAC-seq publications, two methods for normalisation/differential analysis were assayed.

From the count matrix of the same six samples as before, the combined distribution of read density from all the absent peaks in each sample was used to define a sequence of twenty cut-offs. These cut-offs corresponded to the number of counts showed by a particular percentage of absent peaks (supplementary info). Each cut-off was used to filter out from the raw count matrix those peaks from the master list for which the number of counts was  $\leq$  than that particular cut-off in more than three samples (being three the number of the smallest group of replicates in this particular experimental design). Quantile normalisation followed by differential analysis with limma voom showed greater number of differential open chromatin regions (DOCs) at an FDR $<0.01$  compared to DESeq2 across all the cut-offs (Figure 2.4 a). The two approaches presented progressive decrease in the number of DOC sites from the 75% cut-off. Conversely, the proportion of DOC calculated over the total number of regions considered in the differential analysis for each cut off significantly increased from the 50% cut-off onwards, indicating a progressive reduction in the false positive hits reported 2.4 b). From this analysis, 80% was chosen as a conservative filtering cut-off for which almost all the 19,855 DOCs identified by the most conservative method

(DESeq2) at an FDR<0.01 were recapitulated by limma voom at the same FDR (Figure 2.4 c).



**Figure 2.4: Differential chromatin accessibility analysis using limma voom and DESeq2.**

Both methods performed appropriate normalisation of the counts at each of the master list peaks across the six samples, being the median of the quantile normalisation slightly more consistent across the two cell types compared to DESeq2 (Figure 2.5 a). When looking at the first FDR ranked 19,855 limma voom DOCs, 18,768 of them were the same as the retrieved by DESeq2. Moreover, very significant positive correlation was found between the fold changed of those 18,768 significant DOCs in both differential analysis methods ( $r^2=0.999$ ,  $p\text{-val}=2.2^{-16}$ ) (Figure 2.5 b). These observations suggested that the differences in the number of FDR significant DOCs reported by each of the methods could be partly due to differences in the way of calculating the false-positive rate.

Clustering and heat map and pathway analysis-briefly

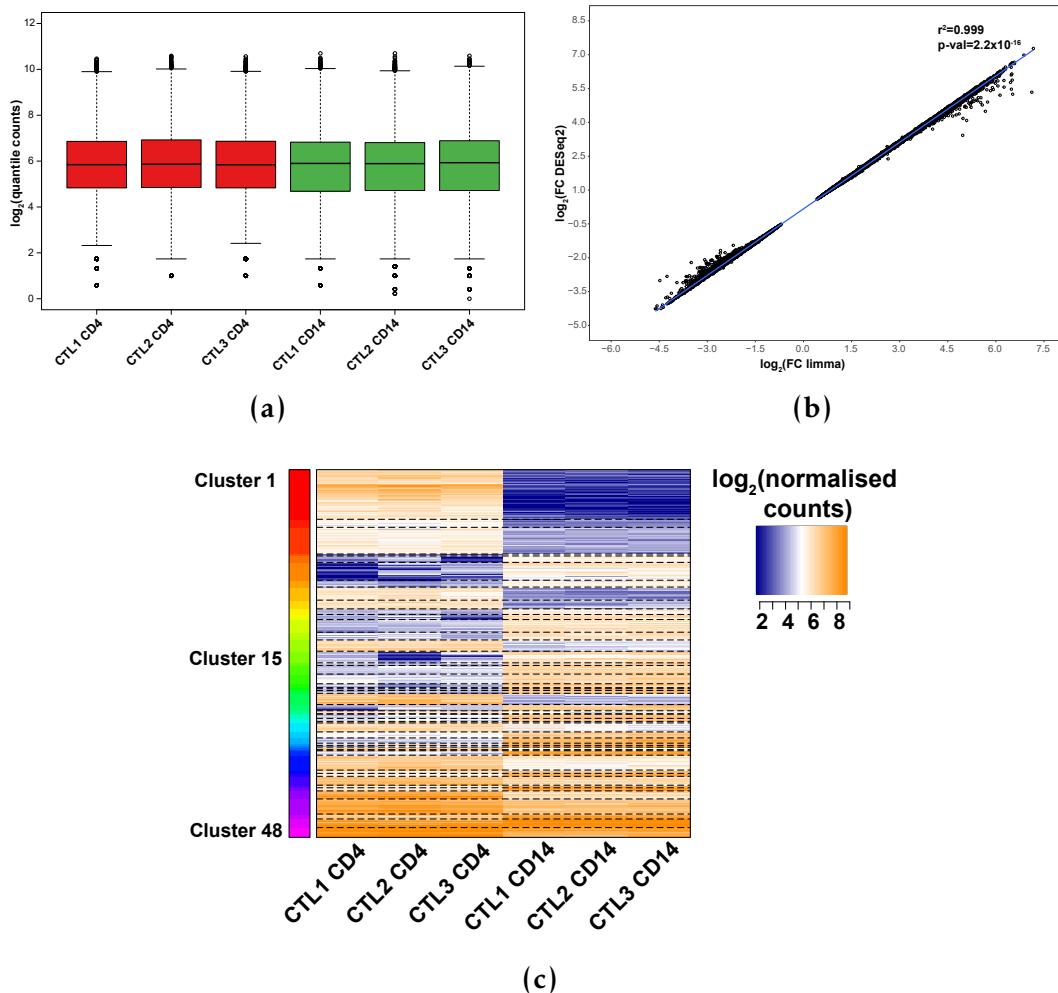


Figure 2.5: Exploration of the differential chromatin accessibility analysis using 80% as the empirical cut-off.

## 2.2.2 Assessment of ATAC-seq transposition times and comparison with FAST-ATAC protocol in relevant cell types

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

Due to the fact that KC is one of the most relevant cell types in psoriasis pathophysiology, ATAC-seq as described in Buenrostro *et al.*, 2013 (named as ATAC-seq 1 here) was performed in 50,000 cells of a suspensions isolated from a

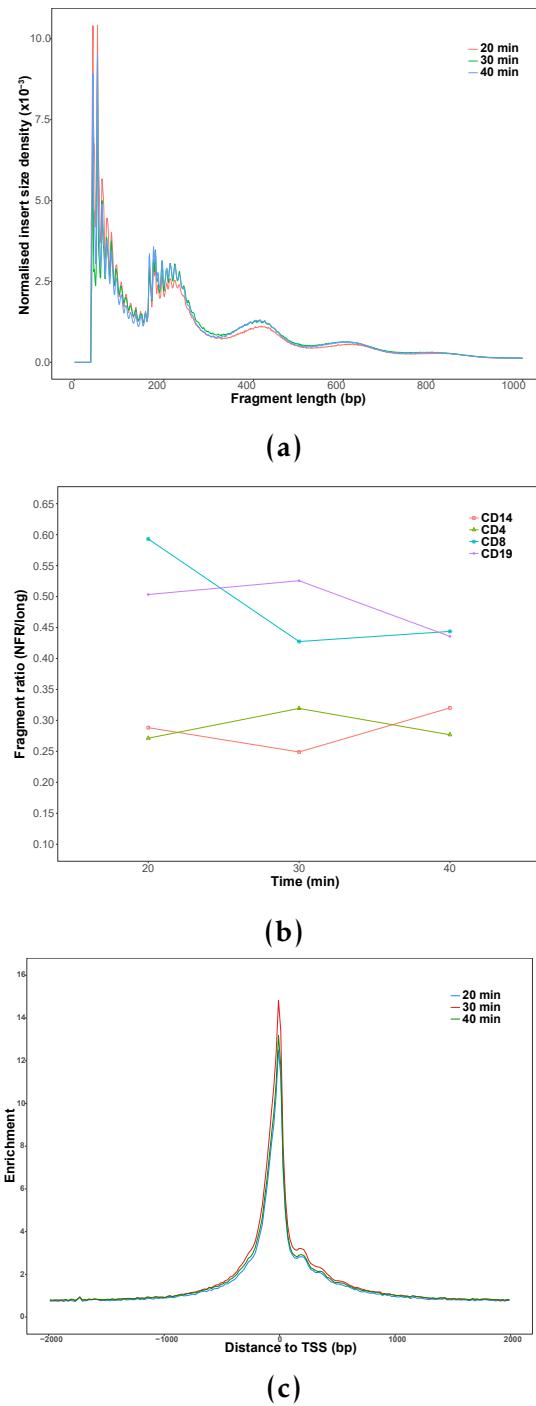
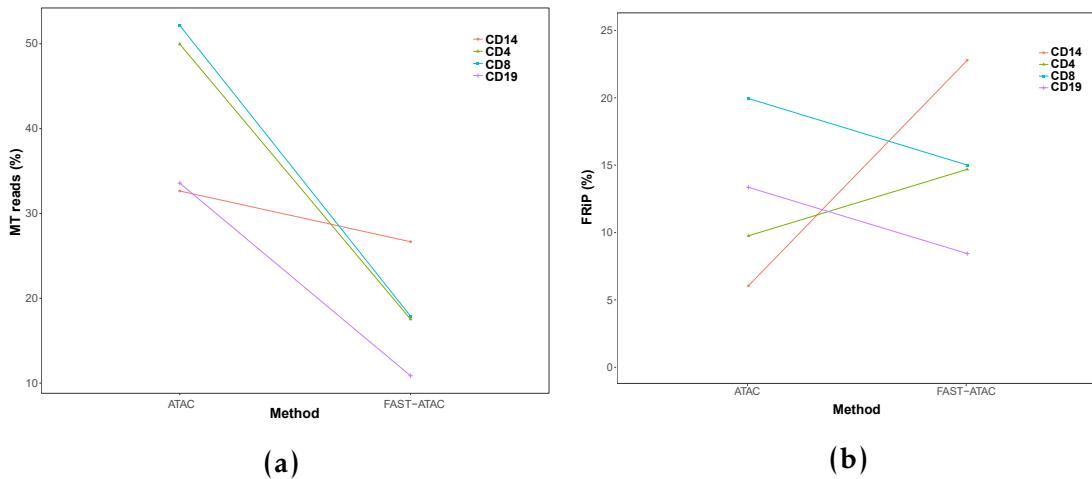


Figure 2.6: Assessment of the effect of transposition times on the ATAC-seq QC parameters

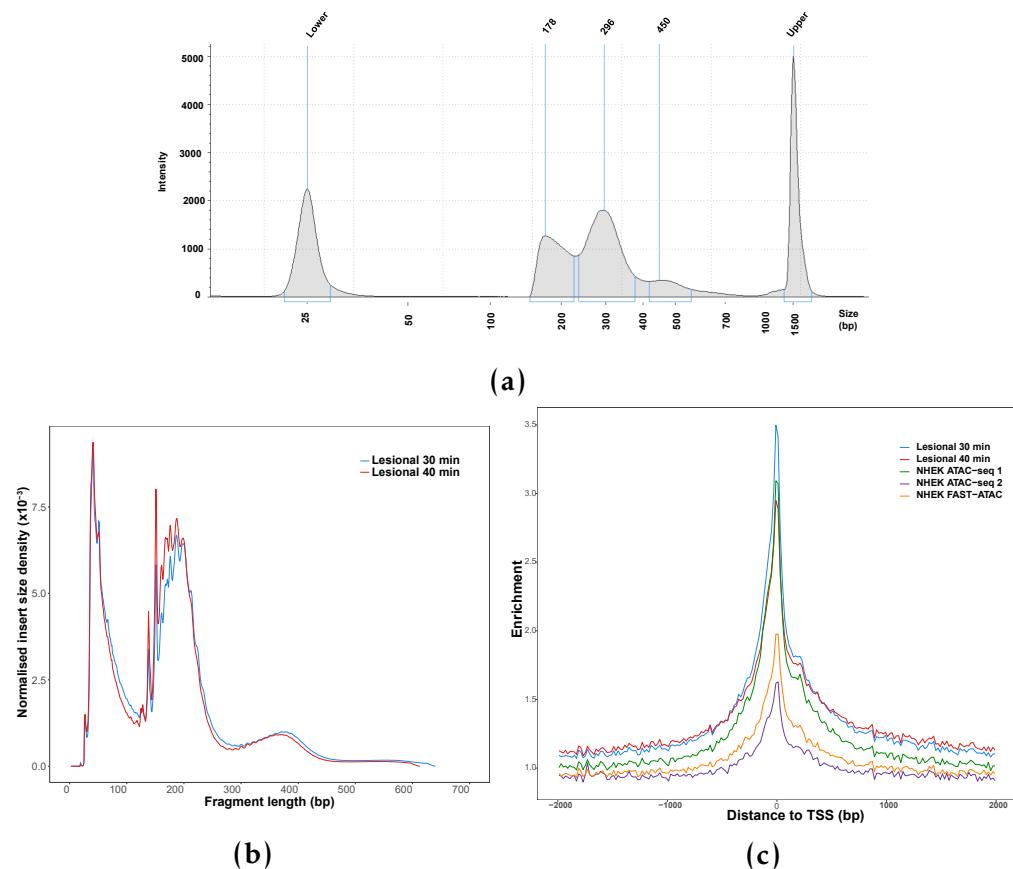


**Figure 2.7: Differences in MT DNA abundance and signal specificity between ATAC-seq and FAST-ATAC protocols**

psoriasis lesional skin biopsy. Two different transposition times (30 and 40 min) were tested. Since biopsy handling and lesional epidermal KC are particularly challenging this was considered the best system to test the performance of the standard protocol in the clinical setting of interest for the study. Two transposition times (30 and 40 min) were tested.

Although cell suspensions obtained from biopsies using trypsinisation of the epidermal sheet are 90% enriched in KC, they also contain significant amounts of dead cells and free-DNA released by apoptotic cells. In order to overcome this problem and the impact that it may have over ATAC-seq background signal, viable KC were selected by adherence assay. Biopsy cell suspensions were cultured for 3h in a 96-well plate and washed afterwards to ensure that only the viable and less differentiated KC would remain for downstream analysis. In parallel cultured NHEK were also used to assess the performance of the different ATAC-seq protocols.

Table for the conditions: done Tapestation profiles of the chosen condition. done Send the others to supplementary. QC measurements: for ATAC1, ATAC2 and NHEK, mention frag size distribution done DHS enrichment for p and q done but not convincing. The complex network of keratin filaments



**Figure 2.8: QC assessment of ATAC-seq in KC enriched cell suspension derived from a psoriatic lesional skin biopsy.** Two transposition times (30 and 40 min) were tested using the standard ATAC-seq protocol (Buenrostro *et al.*, 2013 in 50,000 cells from the same suspension.

in stratified epithelia is tightly regulated during squamous cell differentiation. Keratin 14 (K14) is expressed in mitotically active basal layer cells, along with its partner keratin 5 (K5), and their expression is down-regulated as cells differentiate.

Protocol	Lysis and transposition	Key parameters
Buenrostro et al., 2013	Two steps	0.1% NP-40 and 2.5µL Tn5
Bao et al., 2015	Two steps	0.05% NP-40 and 5µL Tn5
Corces et al., 2016	One step	C1: 0.01% digitonin, 0.5µL Tn5 C2: 0.01% digitonin, 2.5µL Tn5 C3: 0.025% digitonin, 0.5µL Tn5 C4: 0.025% digitonin, 2.5 µL Tn5

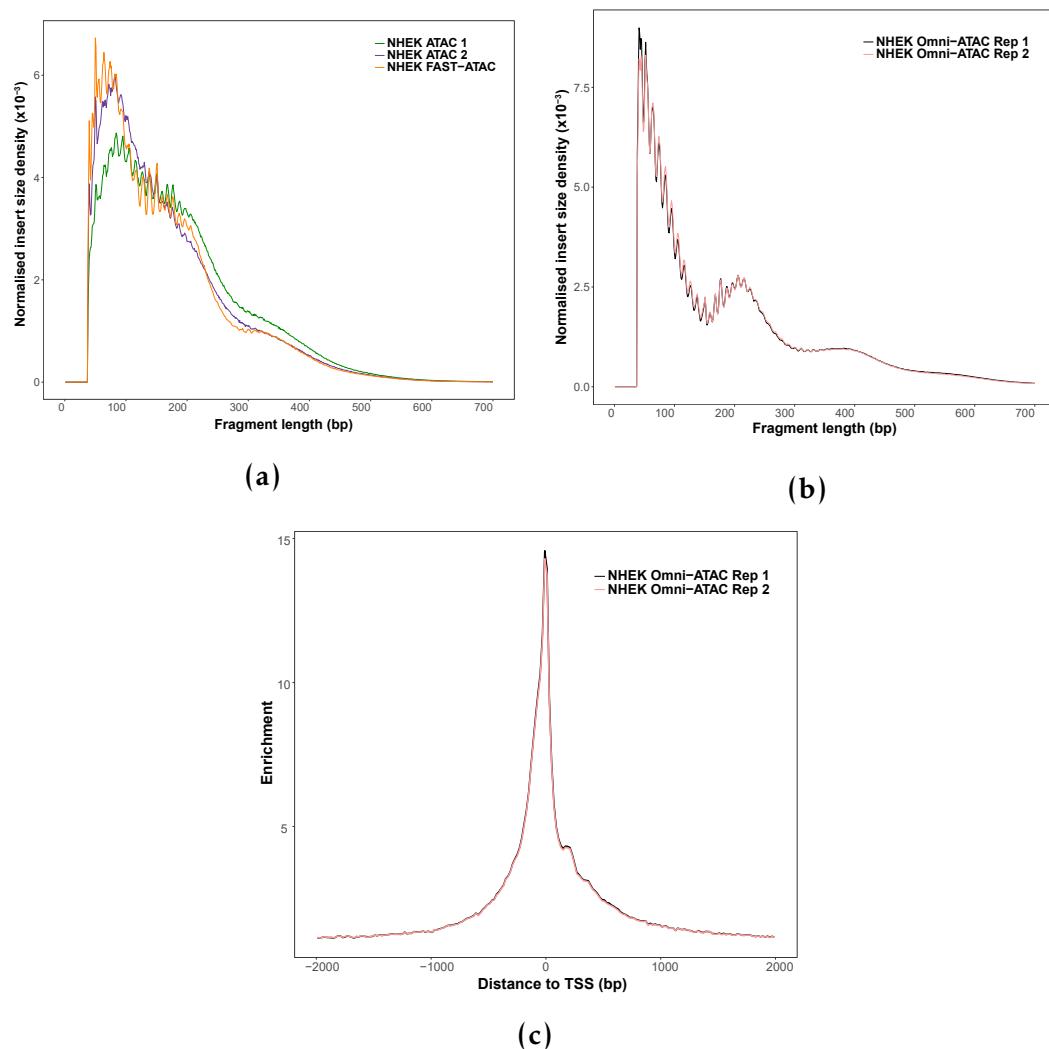
**Table 2.3: Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies.** Transposition for all the different protocols was 30 min.

Omni-ATAC Tapestation profiles of the chosen condition include it with the supplementary that includes all other tapestation profiles. done QC measurements: frag size distribution and TSS done Track including all skin samples

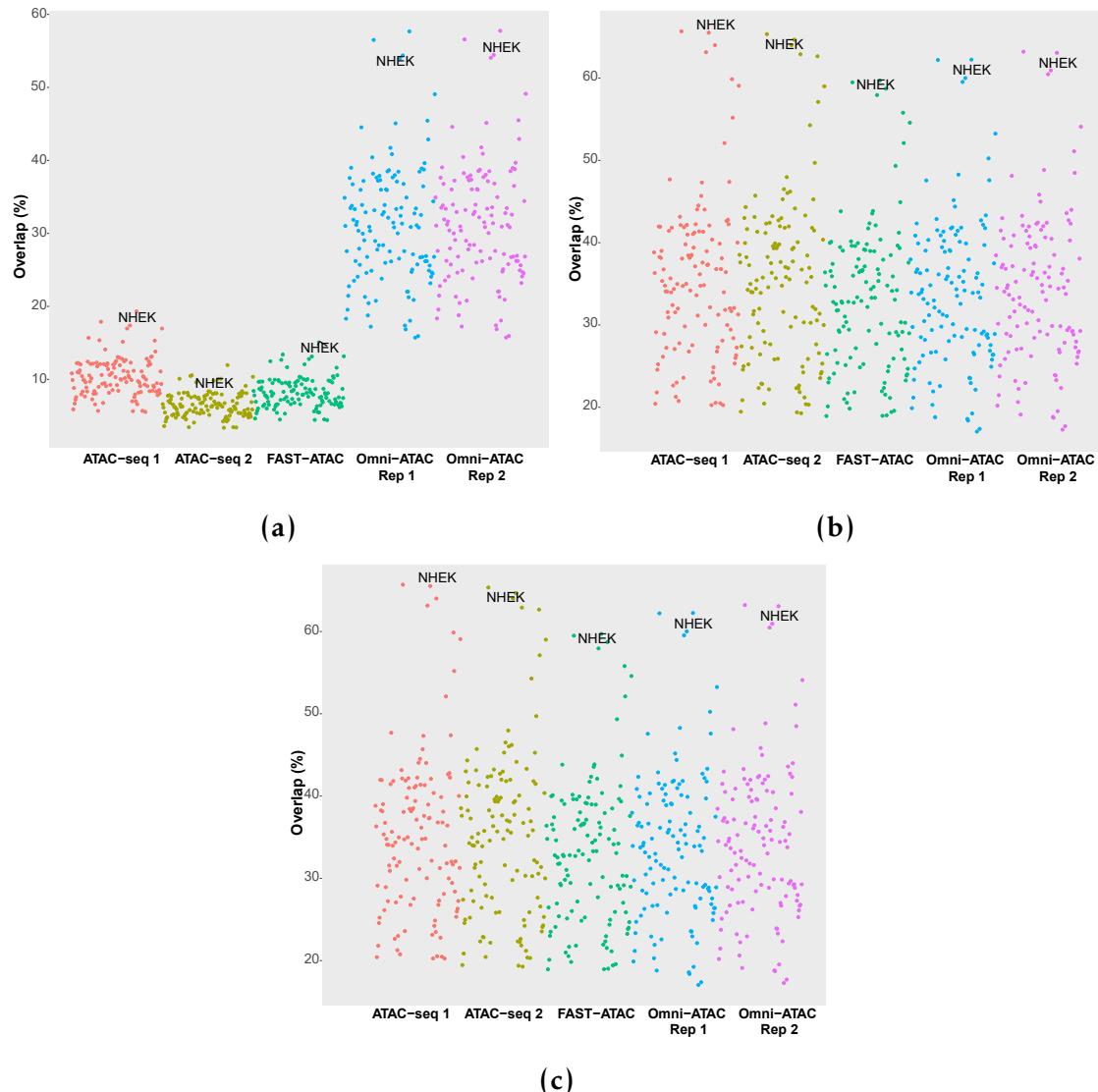
Think of what to include about the biopsies in supplementary done

#### 2.2.4 Discussion

Maybe justify in the dicussion the use of DESeq2 and limma shared based on Alasoo observation of noise effect in limma



**Figure 2.9: QC assessment of FAST-ATAC and Omni-ATAC in cultured NHEK.**



**Figure 2.10: QC assessment of Omni-ATAC in NHEK and chromatin accessibility signal for the samples generated with the different ATAC-seq protocols.**

# **Chapter 3**

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

### **3.1 Introduction**

The techniques to generate the scRNA-seq data have also evolved and SmartSeq2 and 10X Chromium are the tow main

### **3.2 Results**

#### **3.2.1 PsA patients cohort description and datasets**

For this study blood and SF samples were collected from six PsA patients, with equal number of male and female (Table 3.1). All the patients presented oligoarticular joint affection and had been first diagnosed with psoriasis. Maybe add sth about oligoarticular?

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, and joint pain of xxxx. Regarding global assessment, the mean scores for the patient and physician evaluation were X and 3, respectively, in a scale of 1 to 5. These four measurements including joints and global assessment compose the PsARC

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

---

disease activity scores, used by clinicians as the main indicator of response to treatment by recommendation of the xxxx, as previously explained in Chapter 1.

The mean age of the cohort at the time of diagnosis was 44.3 years old and the mean disease duration 8.8 years. Interestingly, PsA1728 was diagnosed at a later age compared to the other patients in the cohort (late PsA onset clinical significance??). Moreover, CRP levels, other marker of inflammation, was also measured in all the patients presenting an average of 17.45 mg/L and being particularly higher in PSA1719 and PSA1728, compared to the other patients. At the time of sample recruitment all the PsA patients were naive for treatment and only PSA1505 had been on methotrexate therapy in the past for xxx months/years (how many years ago?). Post-visit, most of the patients qualified for TNFi biologic therapy xxxx.

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

**Table 3.1: Description of PsA patients cohort recruitment and metadata.** 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	CRP (mg/l)
PSA1718	Female	17	180	Oligo	2/2	3	6
PSA1719	Male	33	24	Oligo	1/1	3	36.6
PSA1607	Male	42	108	Oligo	1/1	4	8
PSA1728	Female	72	48	Oligo	2/2	3	43.2
PsA1801	Female	53	168	Oligo	2/2	3	9.9
PsA1505	Male	35	108	Oligo	1/1	2	1
Total	-	44.3	106	-	1.5/1.5	3	17.45

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

---

For each of the patients, paired data in blood and SF was generated from bulk mononuclear cells or the isolated cell types of interest (detailed in Table 3.2 and Chapter 2.2.1). However, not all types of data including ATAC-seq, PCR gene expression array, scRNA-seq and mass cytometry were generated for all six individuals of the cohort due to project constrains.

<b>Sample ID</b>	<b>% FAST-ATAC</b>	<b>RNA PCR array</b>	<b>scRNA-seq</b>	<b>mass cytometry</b>
PSA1718	Yes	Yes	No	Yes
PSA1719	Yes	Yes	No	Yes
PSA1607	Yes	No	Yes	Yes
PSA1728	No	Yes	No	Yes
PSA1801	No	No	Yes	Yes
PSA1605	No	No	Yes	Yes

**Table 3.2: Datasets generated for the PsA cohort samples.** Four types of data were generated in a paired way between blood and SF from the same individual. The types of data available varies between individuals due to project constrains. FAST-ATAC data was generated for CD14<sup>+</sup>, mCD4<sup>+</sup>, mCD8<sup>+</sup> and NK cells. RNA PCR array was performed in CD14<sup>+</sup>, mCD4<sup>+</sup> and mCD8<sup>+</sup>. scRNA-seq data was generated using 10X technology in bulk PBMCs, bulk SFMCs and sorted mCD4<sup>+</sup> and mCD8<sup>+</sup> from both tissues.

### **3.2.2 Differences in the chromatin accessibility landscape between circulating and SF immune cells**

#### **Quality control of open chromatin regions**

The twenty four PsA samples form four cell types and two different tissues (PB and SF) were sequenced to a median of 158M reads (79M paired-end) per sample. After filtering for low quality mapping, duplicates and MT reads, the median total number of reads were 70.2M, 50.6M, 46.6 and 66.7M for CD14<sup>+</sup>, mCD4<sup>+</sup>, mCD8<sup>+</sup> and NK cells, respectively (Figure 3.1 a). The differences between cell types and samples in the median of total reads remaining after filtering was inversely related to the percentage MT and duplicated reads identified (Figure 3.1 b). For example, mCD1<sup>+</sup> and mCD8<sup>+</sup>, presented the lower

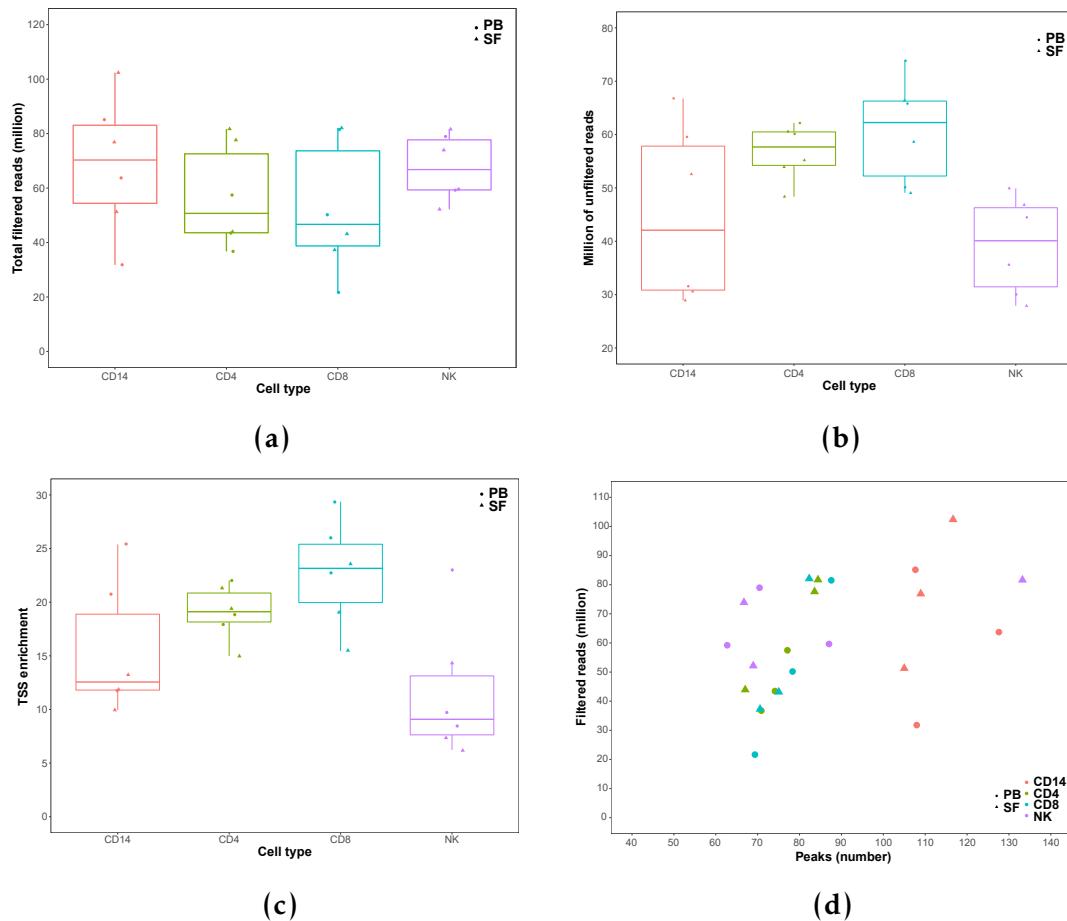
## **Cross-tissue comparison of chromatin accessibility, gene expression signature and immunophenotypes in PsA**

---

median of total number of reads after filtering concomitantly with the greater percentage of MT and duplicated reads. In combination, MT and duplicated reads accounted for a median of 42, 57.6, 62.2 and 40% in CD14<sup>+</sup>, mCD4<sup>+</sup>, mCD8<sup>+</sup> and NK cells, respectively, importantly contributing to the loss of reads in this experiment. As previously mentioned, the MT DNA in ATAC-seq is one of the main sources of read loss, which is more accessible to the Tn5transposase due to the absence of nucleosomes. Although the FAST-ATAC protocol represented an improvement, the percentage of MT reads across amongst all the samples ranged between 2.1 and 25.4%. Similarly, despite initial optimisation of the number of PCR cycles used in the library amplification, the duplicated reads still represented between 22.9 to 55% of the total number of the pre-filtered reads.

Regarding sample quality determination, TSS enrichment analysis showed variation in the levels of background noise across cell types and highlighted the variability in performance of FAST-ATAC (Figure 3.1 c). A trend towards greater TSS enrichment in PB samples compared to SF can be observed in all four cell types. In terms of cell types,mCD4<sup>+</sup> and mCD8<sup>+</sup> presented the best signal-to-noise ratios, with median of 19.1 and 23.1 fold enrichment, respectively. In contrast, NK was the cell type with the lower TSS enrichment values. Particularly, the fold enrichment for PSA1719 and PSA1607 were 7.3 and 6.2, respectively, both just above the 6 to 10 acceptable range from ENCODE. If a bigger sample size was available it would be appropriate to drop this samples from the differential analysis since greater background levels will reduce the power of this approach.

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



**Figure 3.1: QC of FAST-ATAC PsA samples in four cell types.**

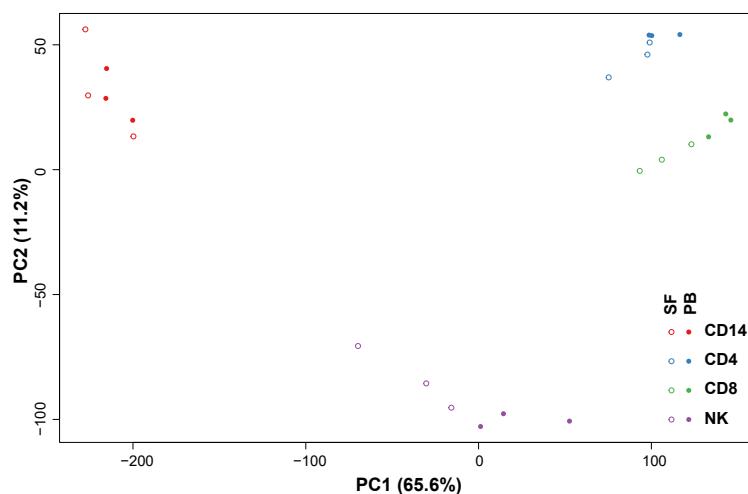
When identifying open chromatin regions through peak calling and standard filtering for FDR<0.01 (not the IDR sample-specific filtering), the number of peaks ranged from  $\sim 62 \times 10^3$  to  $\sim 133 \times 10^3$  peaks per sample (Figure 3.1 d). A clear positive correlation between the number of called peaks and number of reads after filtering could be observed in the data. For example, CD14<sup>+</sup> was the cell type with greatest number of called peaks ( $108.4 \times 10^3$ ) as well as the greater median of reads remaining after filtering when compared to the other three cell types (Figure 3.1 a). For the NK, the two samples with the greatest TSS enrichment (PSA1718 SF and PB) showed greater number of called peaks when compared to the other NK samples with similar number of reads. This observation was consistent with the correlation between sample quality and the number of identified accessible chromatin regions previously demonstrated in

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

Chapter ??ch:Results1). Overall, appropriate number of peaks were called in all the samples and no concerning outliers were identified.

### Open chromatin reflects cell type specificity and functional relevance

In order to determine the ability of the open chromatin identified by the in house pipeline in the PsA sample cohort, a combined master list including all four cell types and the two tissues was built. Following Chapter 2.2.1 and Chapter 2, the combined master list contained open chromatin regions identified in at least 30% of the samples (in this case 7 samples) regardless cell type and tissue to avoid any bias.



**Figure 3.2: Combined PCA analysis of all four cell types isolated from blood and SF.**

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

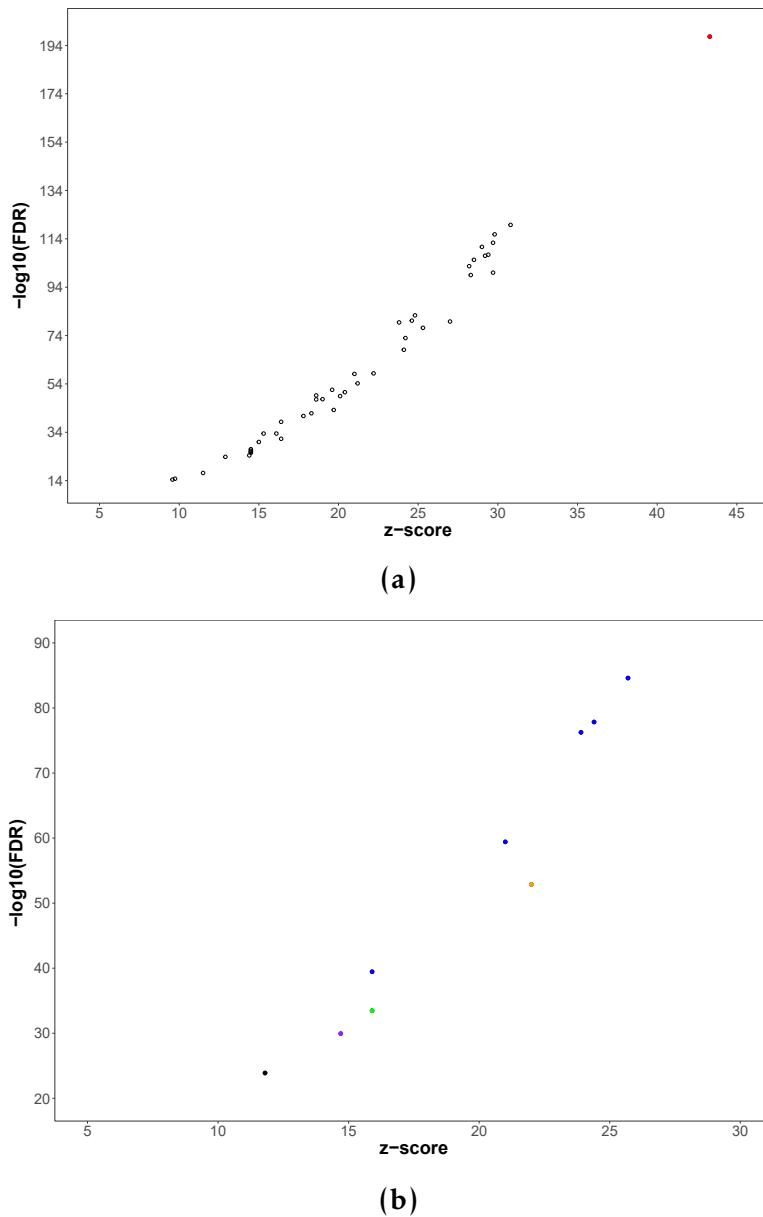


Figure 3.3: Enrichment of eQTLs in the combined cell types PsA accessible chromatin master list xxxx

### Differential open chromatin analysis between blood and SF

Differential chromatin accessibility analysis was performed using a paired design between SF and PB for each of the four cell types (Table 3.4). In all for analysis an 80% cut-off for background noise was applied in the count matrix as previously explained in Chapter 3. Only DOCs identified with DESeq2 and also shared by quantile normalisation and limma voom were taken forward.

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

---

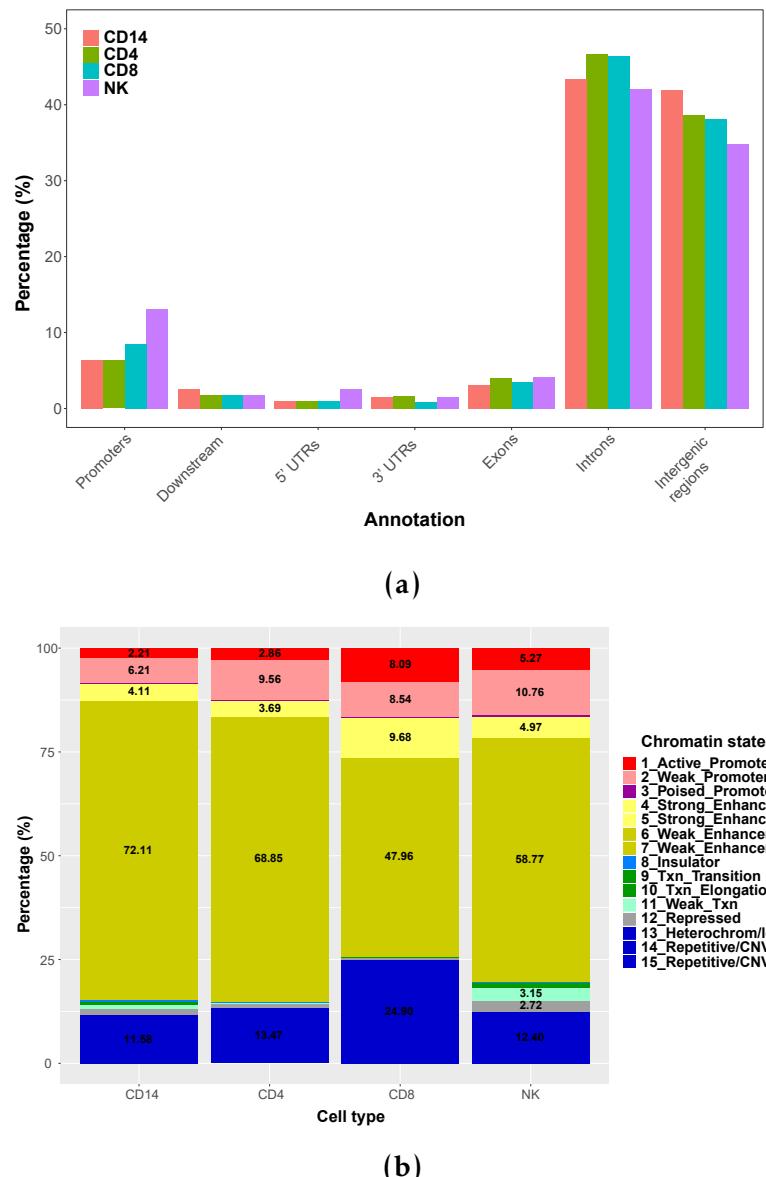
The CD14<sup>+</sup> monocytes and NK showed a greater proportion of differentially accessible regions (23.3 and 8.9%, respectively) compared to CD4<sup>+</sup> and CD8<sup>+</sup> T cells. In CD14<sup>+</sup> monocytes 3,779 out of 5,285 DOCs were more accessible in SF versus the 1,506 that were more open in the cells isolated from PB. Conversely, the number of DOCS more open in each of the tissues were evenly distributed between SF and PB for CD4<sup>+</sup>, CD8<sup>+</sup> and NK.

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

**Table 3.3: xxxxxxxxxxx**

When performing genomic annotation of the DOCs, intronic and intergenic regions represented together 80% or more of all the DOCs in the four cell types (Figure 3.4 a). DOCs annotated in universal promoter regions represented approximately between 5 to 15%, constituting the third most represented genomic feature. DOCs were also annotated with the fifteen states cell type-specific chromatin segmentation maps from the Epigenome Roadmap (Figure 3.4 b). For all four cell types between 44.96 and 72.11% of the DOCs were annotated as weak enhancers, which represented the most prominent category. This was consistent with the predominance of introns and intergenic regions which is the preferred location for enhancers, and also highlighted the cell type specificity of the differences in open chromatin, since enhancers are more cell type specific than promoters.

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



**Figure 3.4: Annotation with genomic regions and chromatin states of the PsA DOCs from the four cell types differential analysis.** xxxx

Interestingly, from the DOCs located within a gene entity the majority were located within introns instead of untranslated regions (UTRs) of the genes and have also been annotated as weak or strong enhancers according to the cell type specific chromatin segmentation map (Table 3.4). For all four cell types, a number of gene entities contained more than one DOC showing the same direction of chromatin accessibility between SF and PB. For example, in CD14<sup>+</sup> two DOCS located at the 5' and 3'UTRs for IL7R gene where found to be

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

---

more accessible in SF compared to PB. Similarly, more accessible chromatin in SF compared to PB was identified in five regions of the *IL15* gene annotated as promoter and enhancers in CD14<sup>+</sup> monocytes. Differences at both gene locations were sCD14<sup>+</sup> cell type specific.

Cell type	DOCs in gene body	Gene with > one DOC	Enhancers	Introns
CD14 <sup>+</sup>	2,357	744	1,775	1,920
CD4 <sup>+</sup>	700	99	504	577
CD8 <sup>+</sup>	831	118	503	666
NK	1,246	235	782	937

**Table 3.4: Summary results of the chromatin accessibility analysis between SF and PB in PsA samples.**

Enrichment of psoriasis and PsA GWAS hits for the differentially open regions between SF and PB in the four cell types was performed using XGR co-localisation and permutation analysis. Although no significant results were found at the SNP level (lead SNPs and SNPs in LD  $r^2 \geq 8$ ), significant enrichment (2-fold enrichment and empirical p-val 0.043) was observed for psoriasis GWAS LD blocks with the CD14<sup>+</sup> DOCs only.

Here I am planning an overview of the differential analysis including -total number of doc split between open in SF and open in PB, -annotation according to chromatin accessibility in terms of all the hits -mention how many of them are annotated in gene bodies and if more than one region within the same gene and choose an example of an interesting gene, include UCSC track showing the differences

-enrichment for TFBS and others, either global or per open in SF and PB, maybe combined plots for all cell types kind of fig6 of XGR paper

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

---

**DOCs highlight relevant functional pathways in a cell type and tissue specific manner**

-pathway enrichment analysis if possible per open chromatin in each cell type  
-maybe include an A2 pathway which is different and unique between open in SF and PB in one cell type

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

Array data

## **3.3 Discussion**

fGWAS analysis as Matthias did would be of interest but needs appropriate GWAS data I am going to try using XGR to do some of this