# Chapter 1

Introduction

Our growing knowledge of genetic associations with susceptibility to psoriasis and psoriatic arthritis (PsA) has not been matched by understanding of the functional basis of these associations and translation to patient benefit. To address this challenge, it is important to understand the regulatory genomic landscape within which disease associated genetic variants may act. This thesis describes functional genomic approaches to establish genome-wide epigenetic and expression profiles for disease relevant tissues and blood-isolated immune cells in psoriasis and PsA, and explore their potential significance for disease pathogenesis and genetic variation. In this chapter, I begin by reviewing current knowledge of the pathophysiology of psoriasis and psoriatic arthritis, the role of genetic variation, and the challenge of functionally characterising genome-wide association studies (GWAS) in complex traits, including the different functional genomics approaches that can be applied.

## Psoriasis and psoriatic arthritis

Psoriasis and PsA have been described as two distinct common complex disease entities that nonetheless share certain clinical features and genetic architecture. Psoriasis is a chronic inflammatory skin disease characterised by episodes of relapse and remittance, most commonly manifesting as well-demarcated erythematous plaques with silver scale and associated with increased risk of joint, eye and systemic disorders (Nestle et al. 2009). On the other hand, PsA is a seronegative chronic inflammatory disease within the spondyloarthritis (SpA) family that usually develops after psoriasis skin manifestations (Villanova2016; Moll et al. 1973; Coates et al. 2016). Understanding similarities and differences between these conditions at the pathological level is helpful before we consider sharing and specificity at the genetic level as well as implications for the identification of new therapeutic targets.

### Epidemiology and global impact

Psoriasis represents a serious global health problem that currently affects about 100 million people worldwide, including both children and adults with no sex bias (Organization 2016). Although the mean age of onset is 33 years, a bimodal distribution with psoriasis patients being classified as early-onset/ type I (with peaks between 16 and 22 years) or late-onset / type II (between 50-60 years) has also been described (Henseler and Christophers 1985). This classification based on the age of onset also has correlates with distinctive clinical features including severity, relapse frequency and family history.

The risk of developing psoriasis shows ethnic differences (Jacobson et al. 2011) with a lower prevalence among adult African, African American and Asian populations (between 0.4 and 0.7%) compared to American and Canadian (4.6 and 4.7%, respectively). In the UK the prevalence of psoriasis ranges between 2 and 3%, affecting approximately 1.8 million people (Perera et al. 2012). On the other hand, cases of PsA in the general population varies between 0.04 and 1.2% (Perera et al. 2012) occurring in 10 to 30% of psoriasis patients, evidencing the strong association between the two diseases (Reich2008; Gelfand et al. 2005). Overall, data suggest an steady increase in both psoriasis and PsA prevalence over the last thirty years (Springate2007; Organization 2016).

Several severe comorbidities have are associated with psoriasis and PsA, with comparatively greater prevalence in PsA. For example, intraocular inflammation (uveitis) affects 8\% of PsA patients compared to only 2% of psoriasis patients (Husted et al. 2011; Oliveira et al. 2015). Other comorbidities include inflammatory bowel disease (IBD), cardiovascular disease (CVD), type 2 diabetes (T2D) and metabolic syndrome (Cohrn20017; Gelfand et al. 2006; Shapiro et al. 2007). Psoriasis

and PsA have also been associated with an increased prevalence of depression and suicidal ideation (Sampogna et al. 2012). Overall, psoriasis and PsA represent a significant burden for the economy due to treatment costs and associated morbidity. Treatment and management-associated costs for psoriasis in 2015 in the UK 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 increased in PsA.

### Psoriasis and inflammatory dermatoses

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 an important role in innate and adaptive immunity and its alterations, due to exogenous or endogenous factors, can lead to development of inflammatory skin conditions, such as psoriasis or atopic dermatitis atopic dermatistis (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 complicates its clinical classification (Perera et al. 2012). As a result, several phenotypes including chronic plaque psoriasis (psoriasis vulgaris), guttate psoriasis, pustular psoriasis, erythrodermic psoriasis and nail psoriasis have been defined (Marrakchi et al. 2011).

### PsA and spondyloarthropathies

PsA belongs to the SpA family, which includes diseases such as ankylosing spondylitis (AS), reactive arthritis (ReA), idiopathic inflammatory bowel disease (IBD) and undifferentiated SpA \parencite{Goldman2001} 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 based on the affected joints (spine/sacroilicac or peripheral) and the presence of extra-articular features \parencite{Runwaleit2009}.

Major histocompatibility (MHC) class I molecules present intracellular peptides (self or from infectious agents) to T cells, encoded by the human leukocyte antigen (HLA) A, B and C genes. HLA-B27 is the strongest genetic association for the SpA family. Studies in human families and rat models with HLA-B27 positive status have shown manifestation of different SpA, 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 in fact 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).

As a phenotype, PsA can be further subdivided in five clinical groups: 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). Overall, this phenotypic heterogeneity of PsA increases the difficulty in the design and achievement of meaningful outcomes from clinical studies.

## Pathophysiology of psoriasis and psoriatic arthritis

### Clinical presentation and diagnosis

Approximately 90% of all psoriasis cases are psoriasis vulgaris, which manifests with well demarcated plaques, erythema and scaling. Plaque formation is the result of thickening (acanthosis) and vascularisation of the epidermis and can vary in size and distribution, with the most common locations being the elbows, knees and scalp (Perera et al. 2012; Griffiths and Lancet 2007). The second most common clinical presentation is guttate psoriasis (10% of all cases) characterised by acute onset of small droplike papules usually in the trunk and proximal extremities (Vence et al. 2015). Vulgaris and guttate psoriasis represent an important burden for patient wellbeing but are in general not life-threatening forms of disease.

Early and late onset psoriasis (type I and type II) differ in clinical presentation. Type I psoriasis patients commonly present with guttate lesions followed very often by bacterial infection, particularly *Streptococcus* throat infection, and have a stronger family history with a high prevalence of HLA-C\*06:02 (85.4% of the cases) (Telfer et al. 1992). In contrast, in type II psoriasis only in 14.6% of the cases are positive for HLA-C\*06:02 and most commonly manifests as spontaneous chronic plaques (psoriasis vulgaris) (Perera et al. 2012).

For PsA, symmetric/polyarticular PsA constitutes the most common manifestation (more than 50% of the cases) followed by asymmetric/oligoarticular

PsA (around 30%), which 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). In particular, nail pitting and scalp and intergluteal skin lesions constitute a predictive biomarker for development of joint inflammation (Moll1976; McGonagle; Griffiths and Lancet 2007; medicine 2011).

The diagnosis of psoriasis and PsA is primarily based on clinical assessment of the patient‘s symptoms due to the lack of diagnostic molecular biomarkers at early stages of the disease (Villanova et al. 2013). The evaluation of skin lesion 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 in research and clinical trials (Fredriksson and Dermatology 1978; Finlay 2005). PASI quantifies the lesional burden by body part based on area of affected surface and the severity of erythema, 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 (Finlay 2005; Schmitt and Dermatology 2005; Langewouters et al. 2008).

To diagnose PsA, modified Moll and Wright criteria known as Classification Criteria for Psoriatic Arthritis (CASPAR) are most widely used (Taylor et al. 2006). A positive diagnosis based on CASPAR requires the presence of inflammatory arthritis, enthesitis, and/or spondylitis and three points from a list of associated elements. In terms of disease activity and treatment efficacy, the PsA Response Criteria (PsARC) is the preferred measure (Mease 2011; Clegg

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. The score ranges from 0 (no disease) to 72 (maximal disease).

et al. 1996). PsARC considers the number of tender joints (TJC) and swollen joints (SJC) over 68 and 66, respectively, as well as patient and physician global assessment of the individual’s general health based on a short questionnaire (Table 1.2).

PsARC Description

TJC Number of tender joints over 68

SJC Number of swollen joints over 66 Patients global health assessment Evaluation of the patient’s health

by the patient (scale 0 to 5)

Physician global health assessment Evaluation of the of the patient’s

by the physician (scale 0 to 5)

Table 1.2: Variables and scoring used in the Psoriatic Arthritis Response Criteria (PsARC). The patient’s global health assessment by the patient and the physician is scored using a 5-point Likert scale, where 0 corresponds to very good, no symptoms and 5 corresponds to very poor and severe symptoms. When used to evaluate overall improvement after 12 weeks of treatment, improvement in at least two of the four variables evaluated (one of which must be TJC or SJC score) with no worsening of any criteria is required.

### Aetiology of psoriasis and PsA

Psoriasis and PsA are chronic inflammatory diseases characterised by a dysregulated immune response initiated as the result of genetic predisposition and exposure to particular environmental cues (Figure1.1). The origin of both pathologies, as well as the connection between skin and joint inflammation, still remain controversial.

Environmental factors and disease

A variety of exposures are proposed as risk factors for the development and worsening of psoriasis and PsA. A wide range of drugs including anti-depressants, anti-hypertensives and anti-cytokine therapies have been associated with initiation, exacerbation and worsening of psoriasis (Kim et al. 2010). Bacterial and viral infections are associated with triggering and exacerbation of psoriasis, notably guttate psoriasis after group C \texit{Streptococcus} throat infection as well as HIV infection (Gudjonsson and of 2003; Valdimarsson and in 2009; Diluvio et al. 2006). In PsA, statistical association with antibody production against \textit{Streptococcus pyogenes}, \textit{Yersinia enterocolitica}, \textit{ Chlamydophila psittaci} and HIV have also been reported \parencite{Thrastardottir2018} Recent studies have also observed perturbation in the composition of the gut and skin microbiota of psoriasis and PsA patients (add reference). Physical trauma and mechanical stress can also trigger the appearance of skin lesions and digit joint inflammation (Weiss et al. 2002; Nestle et al. 2009). Smoking is associated with an odds ratio (OR) of 1.78 (95% CI 1.52-2.06) for psoriasis, in particular palmoplantar pustulosis \parencite{Armstrong2014}. Psoriasis is also associated with obesity, alcohol dependency, vitamin D deficiency and stress but evidence remains controversial. (Meglio et al. 2014).

Histopathological alterations in skin and joints

The epidermis is the most external compartment of the skin, comprising approximately 90% keratinocytes 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, keratinocytes undergo upregulation in proliferation rate (hyperplasia) that causes aberrant cell differentiation (parakeratosis), thickening of the epidermis and 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 affected joint shows a wide range of histological changes (Haddad and Chandran 2013). One of the most common structural changes is arthritis caused by the swelling and inflammation of the joints (Schett et al. 2011). As a result of this inflammation, alterations in bone remodeling lead to osteolysis with subsequent bone resorption and erosion at the affected joints (Mensah2017). Bone erosion is also the main histopathological process driving dactylatis, where bone lysis resolves in shortening of the digits (Gladman et al. 2005). Moreover, 35% of PsA patients also undergo inflammation of the connective tissue at the insertion of tendons or ligaments, a phenomenon known as enthesitis (McGonagle et al. 2011; Polachek et al. 2017). The inflammatory environment at the entheses favours bony spurs formation along the insertion sites, similar to RA, causing structural debilitation of the joints parenciteBenjamin2009,Finzel2014.

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 (Figure 1.1). Interferon (IFN)-*α* and *γ* are innate immune cytokines involved in disease initiation (Leanne2009) mainly produced by circulating plasmacytoid dendritic cells (pDCs) and myeloid DC (mDCs), respectively (Perera et al. 2012) as well as T cells in lesional skin (Hijnen et al. 2013). Increased mRNA levels for both IFNs have been detected in skin plaques and shown to contribute to lymphocyte recruitment and maintenance of DC activation (Schmid et al. 1994). TNF-*α* is another key cytokine involved in the dysregulated innate immune observed in psoriasis and PsA. TNF- *α* is produced by activated keratinocytes, mast cells, NK and also adaptive immune cell types, including T helper (Th)-1 and Th-17 lymphocytes that infiltrate skin lesions and inflamed joints (Perera et al. 2012; Lizzul et al. 2005). TNF-*α* causes activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-*κ*B), a master transcriptional regulator which induces expression of pro-inflammatory cytokines, antiapoptotic genes and genes involved in maintenance of chronic inflammation (Johansen2010; Lizzul et al. 2005). Moreover, TNF-*α* has a prominent role in bone turnover and bone remodeling, key histopathological features of PsA (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 in inflamed skin and to a lesser degree by psoriatic keratinocytes (Lee et al. 2004; Li et al. 2018). 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 also by circulating CD4+ (Tonel et al. 2010). In psoriasis, IL-23 mediates the pathogenic loop between activated keratinocytes and T cells, where activation of the IL-23 pathway importantly leads Th- 17 cell differentiation and increased IL-17 cytokine levels as a result of NF- *κ*B activation (McGeachy2009). IL-17 maintains the perpetuation of the Th-17 immune mediated response through recruitment and activation of neutrophils, induction of pro-inflammatory cytokines, including interleukine-1*β* (IL-1*β*) and interleukine-6 (IL-6), and sustains keratinocytes activation (Doyle and Arthritis 2012).

More recently, interleukin 22 (IL-22) has gained relevance as mediator of dysregulated crosstalk between the innate and adaptive immune response. IL-22 levels are increased in the skin lesions and plasma of psoriatic patients and is mainly produced by a subset of CD4+ cells known as Th-22 (Wolk et al. 2006). IL-22 contributes to some of the histological changes in skin as well as to AMP production by keratinocytes (Eyerich et al. 2009).

### 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.

*Keratinocytes*. Several studies have shown the role of keratinocytes as immune sentinels through MHC class 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 keratinocytes 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-*α* (Feldmeyer

et al. 2007; 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(Boyman et al. 2004). Together, these findings support the role of epidermal dysfunction in the initiation of the psoriatic chronic inflammatory response (Proskch2008). The relevance of keratinocytes 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)

*Dendritic cells*. 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 on activation by the keratinocytes 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-*α* secreted by pDCs, expanding up to 30-fold in lesional skin (Zaba2007). Activated mDCs mediate the Th-1 and Th-17 response as well as perpetuation of KC activation through IL-23 production (Lee et al. 2004). Studies in immunodeficient psoriasis mouse models have shown that blockage of downstream IFN-*α* signaling or IFN-*α* production by pDCs failed to induce T-cell activation and psoriasis onset (Nestle et al. 2005).

*Neutrophils.* Neutrophils are implicated 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 contribute to pDC and CD4+ T cell activation (Hu et al. 2016). Neutrophils have also been identified in recent studies as one of the main sources

**Environmental triggers**

**Genetic background**

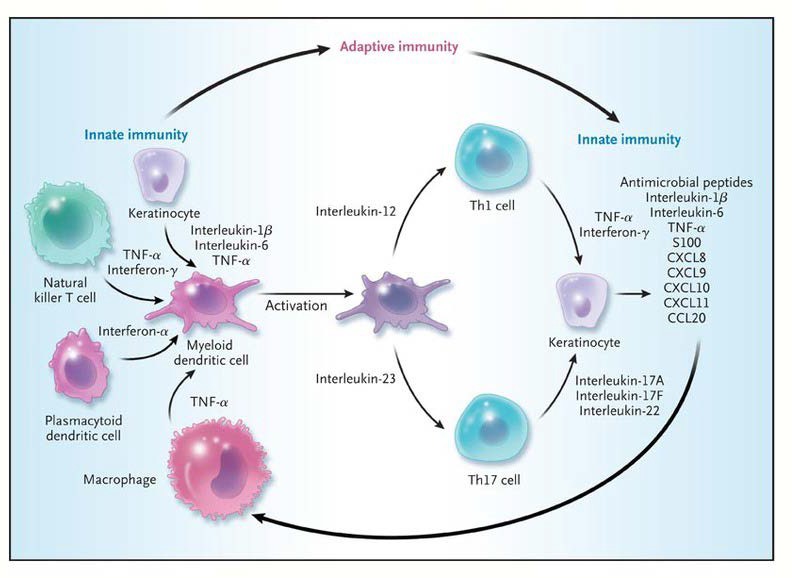


Figure 1.1: Environmental triggers and genetic predisposition leading to psoriasis and PsA (adapted from (Nestle et al. 2009)). The main cell types, cytokines and chemokines involved in the dysregulated innate and adaptive immune response found in these conditions are shown.

of IL-17 production in the skin lesions (Lin et al. 2011) and they also release a wide range of proteases some of which induce KC proliferation (Mahil2006).

*Monocytes and macrophages*. Resident macrophages in the healthy dermis undergo a 3-fold increase in psoriatic skin lesions and contribute to disease development through TNF*α* production (Perera et al. 2012; Mahil et al. 2016). Similarly, mouse models for chronic psoriasiform skin inflammation demonstrate macrophage migration into affected skin and how production of TNF-*α* contributes to maintenance of skin lesions (Stratis et al. 2006; Wang et al. 2006). Initial studies showed greater phagocytic and bactericidal activity in monocytes from psoriasis patients compared to those from healthy individuals (Bar-Eli1979). Additionally, increased circulating intermediate monocytes (CD14*high* CD16*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 RA joints (Hitchon et al. 2002).

*T cells*. T lymphocytes have been considered the most relevant cell types in the initiation and maintenance of psoriasis and PsA. There are case reports of bone marrow transplantation leading to initiation or termination of psoriasis (Gardembas1990; Eedy et al. 1990). Reduced numbers of circulating T cells but increased percentages of memory CD4+CD45RO+ and CD8+CD45RO+ have been

observed in moderate-to-severe and severe psoriasis patients vs 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+ and CD8+ , likely due to phenotype

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+ and CD8+ , likely due to phenotype

heterogeneity (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+ and CD4+ lymphocytes are found in the epidermis

and dermis, respectively (Clark et al. 2006). An increase in activated memory CD4+CD45RO+and CD8+CD45RO+ cells can be detected by the third day from the lesion appearance (Clark et al. 2006; Perera et al. 2012). *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 (Wrone-Smith and clinical 1996; Nickoloff and Wrone-Smith 1999; Perera et al. 2012). The relative importance of CD4+ versus CD8+ cells in psoriasis initiation has been explored in pre-lesional skin mouse xenografts where CD4+ but not CD8+ T cells were required in the transition from uninvolved to lesional skin (Nickoloff and Wrone-Smith 1999). Interestingly, the injection of activated CD4+ cells in mice was followed by an acute increase in activated resident CD8+ T cells. Overall, these results supported the hypothesis of skin CD4+ cells being drivers of resident T-cell activation and the population of resident activated CD8+ the main effector of the immune response. In synovial tissues of PsA patients, CD4+ are significantly more abundant than CD8+ (Diani et al. 2015). However, amongst the CD8+ populations, the memory cells are prevalent in the patients synovial fluid (SF) with a significant enrichment compared their counterparts in PsA PB and RA SF(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+ and CD8+ cells is driven

by IL-12 and IFN-*γ* (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 (Cai et al. 2012; Dolcino et al. 2015). The relevance of IL-17 has been further highlighted by the presence of CD8+ 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+ 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 *γδ* 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 (Yamamoto et al. 2015).

### Therapeutic intervention

Psoriasis and PsA are currently incurable diseases, with treatments available focused on alleviating symptoms. For instance, topical therapies are advocated in cases of mild-to-moderate psoriasis, including emollients and short-term corticosteroids (Menter et al. 2009).

Other treatments may be used in combination with corticosteroids, such as ultraviolet (UV) light therapy and vitamin D analogues, directed to inhibit T-cell and KC proliferation and stimulate KC differentiation (Rizova and Corroller 2001). In the case of PsA, for patients presenting with swelling of two or fewer joints, nonsteroidal anti-inflammatory drugs (NSAID) to control the inflammatory symptoms and intra-articular injection of glucocorticosteroids together with joint aspiration are used to reduce pain and inflammation (Coates et al. 2016). However, treatment of most forms of PsA and moderate-to-severe psoriasis require the use of systemic therapies. More severe forms of PsA require disease-modifying antirheumatic drugs (DMARDs) including the antagonist of folic acid methotrexate (MTX) and the phosphodiesterase 4 inhibitor Apremilast, which act as immunosuppressors of activated T cells and cytokine production, respectively (Keating2017; Schmitt and of 2014; Gossec et al. 2016; Polachek et al. 2017).

Remarkably, biologic systemic agents represent the most specific treatment option for severe psoriasis and PsA notably TNF-alpha inhibitors (TNFi). Three TNFi have been approved for the treatment of psoriasis: etanercept, infliximab and adalimumab (Ahil2016). In addition, certolizumab pegol and golimumab are often used in the management of PsA (Coates2016b). However, 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 patients

fail to respond to the first TNFi administrated, requiring switching to an alternative TNFi (Abramson and and 2016). New biologic therapies have been developed to target other key cytokines, such as IL-12, IL-23 (ustekinumab) or IL-17 (secukinumab and ixekizumab), which represent a substantial advance in treating patients failing to respond to TNFi (Coates2016b; Mahil et al. 2016).

## Genetics of psoriasis and psoriatic arthritis

The risk of developing psoriasis and PsA is not only influenced by environmental conditions but also by the genetic background of each individual.

### Heritability

The concordance of psoriasis is greater in monozygotic (33-55\%) compared to dizyogtic twins (13-21\%), giving a heritability estimate of 80\%, while no difference in concordance is reported for PsA, probably due to lack of statistical power and appropriate diagnosis (Pendersen2008). In the general population, approximately 40% of patients with psoriasis or PsA have a family history in first degree relatives (Gladman et al. 1986). Interestingly, the recurrence rate in first-

degree relatives has been shown to be greater in PsA (40%) compared to psoriasis (8%) in a study in the Icelandic population (Chandran et al. 2009). Altogether, this suggests differences in the heritability between the two phenotypes and a stronger genetic contribution in PsA.

### Non-GWAS and linkage studies

Linkage analysis of psoriasis and PsA in family pedigrees presenting an autosomal dominant condition yielded nine psoriasis susceptibility loci (PSORS1-9) with PSORS1 showing the strongest genetic association (International2003; Capon 2017). PSORS1 locus lies within the MHC class I region, initially associated with psoriasis susceptibility in serological studies (Rusell1972; Tiilikainen et al. 1980). Rare highly penetrant mutations have also been identified for two genes within PSORS2 (17q25): zinc finger protein 750 (*ZNF750*) and caspase domain family member 14 (*CARD14*) (Tomfohrde et al. 1994; Jordan et al. 2012a), with common variants in *CARD14* also reported implicating genetic variation in this gene in Mendelian and multi-factorial forms of disease (Jordan et al. 2012a; Tsoi et al. 2012). 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\*06:02 (uszczek et al. 2004; Williams et al. 2005; Carter et al. 2007; Yen et al. 2001). Nevertheless, the inability of independent studies to reproduce these results for regions other than PSOR1, 2 and 4, highlights the limitations of linkage studies to understand the genetics of complex diseases (Capon 2017).

### Genome-wide association studies

The first psoriasis and PsA GWAS were published in 2007, with a total of sixty-three independentgenetic associations identified at genome-

wide significance (p*<*5x10−8) (Table 1.3) which explain 28% of psoriasis and PsA

heritability (Tsoi et al. 2017). The majority of 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). Early GWAS with moderate power confirmed association with loci overlapping the PSOR1 and PSOR2 genomic regions identified by linkage studies (Cargill et al. 2007; Strange et al. 2010). 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 informativeness of GWAS was significantly enhanced with the Immunochip genotyping chip, 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* and also included meta-analysis with the largest available psoriasis cohorts at the time(Tsoi et al. 2012). This meta-analysis has since been further expanded yielding sixteen additional associations and reinforcing the importance of NF*κ*B and cytotoxicity pathways in disease pathophysiology (Tsoi et al. 2015; Tsoi et al. 2017). Meta-analysis of GWAS across Caucasian and Chinese populations has demonstrated the value of a 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 independent GWAS for psoriasis and PsA has been shown by differences in HLA-C and HLA-B alleles frequencies. Interestingly comparative higher association with HLA-B has been found in PsA individuals compared to psoriasis patients not developing joint inflammation .Specific associations for PsA with *TRAF3IP*, *IFNLR1*, *IFIH1* and *NFKBIA* (previously reported as psoriasis loci) as well as PsA-specific independent signals for *IL23R* and *TNFAIP3 have also been demonstrated* (Ellinghaus et al. 2010; Stuart et al. 2015Immunochip GWAS including only PsA phenotype revealed a specific association in chromosome 5q31 not previously reported (Bowes et al. 2015).

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 *IL12B* |
| (Nair et al. 2009) | Eur | 1,409/1,436 | Psoriasis, PsA | *IL23A*, *IL23R*, *IL12B*, *TNIP1*, *TNFIP3*, |
|  |  |  |  | *IL4* and *IL13* |
| (Stuart et al. 2010) | WA, Eur | 1,831/2,546 | Psoriasis, PsA | *NOS2*, *FBXL19*,*PSMA6-NFKBIA* |
| (Ellinghaus et al. | German | 472/1,146 | Psoriasis | *TRAF3IP2* |
| 2010) |  |  |  |  |
| (Strange et al. 2010) | Eur | 2,622/5,667 | Psoriasis, PsA | *LCE3D* (PSOR2), *IL28RA*, *REL*, *IFIH1*, |

(Zhang et al. 2008) Chinese 1,139/1,132 Psoriasis (type

I)

*ERAP1*, *TYK2* and *HLA-C/ERAP1*

epistasia

*LCE* gene family and *IL12B*

Introduction

22

Introduction

23

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (Sun et al. 2010) | Chinese |  | 8,312/12,919 | Psoriasis, PsA | *ERAP1*, *PTTG1*, *CSMD1*, *GJB2*,  *SERPINB8* , *ZNF816A* |
| (Tsoi et al. 2012)*\** | WA, Eur |  | 10,588/22,806 | Psoriasis, PsA | *CARD14* (PSOR4), *RUNX3*, *B3GNT2*, |
| (Tsoi et al. 2015)*\** | WA, Eur |  | 15,000/27,000 | Psoriasis, PsA | *ELMO1*, *STAT3*  1q31.1, 5p13.1, *PLCL2*, *NFKBIZ*, |
|  |  |  |  |  | *CAMK2G* |
| (Bowes et al. 2015) | British, | Irish, | 1,962/8,923 | PsA | 5q31 PsA-specific |
|  | Australians |  |  |  |  |
| (Stuart et al. 2015) | WA and Eur |  | 1,430/1,417 | Psoriasis, PsA | 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 | *LOC144817*, *COG6*, *RUNX1* and *TP63*; |
| (Tsoi et al. 2017)*\** | WA, Eur |  | 19,032/39,498 | Psoriasis, PsA | signals with ethnic heterogeneity  *CHUK*, *IKBKE*, *FASLG*, *KLRK1*, *PTEN* |

Overall, GWAS studies have demonstrated shared and distinct genetic architectures for psoriasis and PsA. It is important to take into account that these results are affected by imprecise phenotyping of cases, which is one of the many challenges in the systematic comparison between the two diseases.

### Relevance of non-coding variants in disease susceptibility

Approximately 88% of all GWAS associations map within non-coding regions (Welter et al. 2013). Psoriasis exome association studies in Chinese and Caucasian populations have increased the number of coding variants with putative effects on the protein structure (Tang et al. 2014; Zuo et al. 2015; Dand et al. 2017). These studies have confirmed some previously identified missense associations in *CARD14* and *ERAP1*, revealed new common coding variants at these previously associated loci and identified rare protective missense changes, for example in the *TYK2* gene(Tang et al. 2014; 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 (Tang et al. 2014).

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 in different regulatory elements, including enhancer, silencers, promoters and the 5’ and 3’ untranslated region (UTR) of genes (Ward and biotechnology 2012). 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 2014) (1.4.2).

Identification of the target gene regulated by non-coding variants represents a challenge in the field of functional genetics. This limitation 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 an such approach revealed a *cis*-eQTL involving the TF *KLF4* and a haplotype of non-coding GWAS SNPs located 14kb up-stream (Small et al. 2011). Moreover, this haplotype also showed association with genes in *trans*, highlighting downstream targets regulated by KLF4. Nonetheless, eQTL mapping alone only provides statistical suggestion of transcriptomic regulation, and additional functional assays, such as chromatin conformation and genome editing, are required to demonstrate causality (Edwards et al. 2013).

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

GWAS represent a biologically unbiased approach to shed some light on 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.

Nevertheless, it is important to bear in mind that in most of the cases non-coding variants from GWAS studies lack of functional characterisation and they tend to be associated arbitrarily to the nearest gene or the closest gene which fits into current knowledge about pathophysiology. This bias to some extent the genes that contribute to enrichment of certain pathways and the efficacy of drugs developed to target some of those genes has helped to further confirm their truly involvement in disease.

Better understanding of immune-related diseases has likewise led to identification of shared susceptibility loci and supported use of therapeutic interventions across diseases, such as anti- IL-23 and anti- IL-17 antibodies to treat psoriasis, PsA, AS and IBD (Visscher et al. 2017).

Systematic comparison of the genetic architecture across different conditions has revealed psoriasis and PsA risk loci to be shared, in the same or opposite directions, with AS, Crohn’s disease (CD), multiple sclerosis (MS),

RA and type 1 diabetes (T1D) ([https://www.immunobase.org](https://www.immunobase.org/)). Interestingly, cross-disease association studies has revealed significant overlap of the multi-trait associated loci in regulatory elements in bone marrow, NK and T cells as well as immune response pathways (Ellinghaus et al. 2016).

In the case of psoriasis and PsA, the majority of 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) as detailed below.

Antigen presentation

In psoriasis *HLA-Cw\*0602* represents the strongest GWAS association, shared with other diseases such as Hepatitis C, primary sclerosing cholangitis and Graves disease (Blais et al. 2011). No differences at the transcript level have been identified for HLA-Cw\*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 *ERAP1* gene, involved in the trimming of peptide antigens. Moreover, GWAS studies identified that *ERAP1* 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\*0602* and SNPs in the vicinity of the zeta chain of T cell receptor associated protein kinase 70 (*ZAP70*) gene (Picard et al. 2009).These epistatic phenomena, whereby association of one gene is dependent on the presence of another, have also been reported between *HLA-B\*27* and *ERAP1* in AS (Cortes2015b; Evans et al. 2011). Interestingly, the AS *ERAP1* GWAS association increases *ERAP1* expression and also alters splicing,

resulting in an ERAP1 protein isoform with increased activity (Constatino2015; Hanson et al. 2018).

Skin barrier

GWAS have highlighted KC specific genes such the *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 disease is driven by a deletion in two of the genes within this family, *LCE3B* and *LCE3C* (*LCE3C LCE3B del*)(Cid et al. 2009). 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 KC terminal differentiation (Bergboer et al. 2011). The lack of *LCE3B* and *LCE3C* expression in psoriasis patients has been hypothesised to impair the repair following skin disruption, potentially facilitating microorganism infection and triggering a dysregulated immune response (Bergboer et al. 2011). In fact, the use of UVB radiation has been shown to upregulate *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-*κ*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-kB as well as overexpression of psoriasis-associated genes including *IL6*, *TNFA* and *TNFAIP2*, among others (Jordan et al. 2012b).

NF-*κ*B and TNF pathways

The NF-*κ*B pathway is involved in the regulation of the innate and adaptive immune responseand NF-*κ*B contributes to the development of many chronic inflammatory

diseases (Liu et al. 2017). In psoriasis elevated levels of NF-*κ*B are present in lesional compared to uninvolved and normal skin (Lizzul et al. 2005).

Several psoriasis and PsA GWAS loci have been mapped to gene members of the NF-*κ*B and TNF signalling pathways including *TNIP1*, *TNFAIP3*, *NFKBIA*, *REL*, *TRAF3IP2, CHUK*, *IKBKE* and *FASLG* (H“”–u˝ffmeier2010; Nair et al. 2008; Ellinghaus et al. 2010; Wang et al. 2008; Idel et al. 2003; Bowes et al. 2012) (Tsoi et al. 2017). For example, a haplotype including missense mutations and intronic variants in *TRAF3IP2* has been reported to drive psoriasis and PsA association by reducing its affinity for TRAF interacting proteins and concomitantly altering NF-*κ*B activation and the IL- 17/IL-23 axis (Huffmeier2010; Ellinghaus et al. 2010). In addition, exome-sequencing studies have identified variants with predicted influence on protein structure and function at *TNFSF15*, a TNF ligand regulating NF-*κ*B and MAP kinases activation in endothelial cells (Dand et al. 2017; Wang et al. 2014).

Type I IFN and innate host defense

Members of the type-I IFN signalling pathway have also been associated with psoriasis and PsA highlighting the role of genes contributing to the host response to viruses and bacteria in the disease pathophysiology. Associations include *IL28RA*, *IFIH1*, *TYK2*, *RNF114*, *ELMO1* and *DDX58*, some of which have been previously reported as susceptibility loci for other immune-mediated diseases (1.3). Exome-sequencing and GWAS have identified two independent protective missense mutations predicted to impair the catalytic activity of the Janus kinases (JAK) protein member TYK2, and thus the initiation of the IFN-I downstream inflammatory cascade in psoriasis and PsA (Strange et al. 2010; Tsoi et al. 2012; Dand et al. 2017). A JAK inhibitor approved for RA is currently undergoing clinical trials in other immune- related diseases alongside the development of more specific JAK inhibitors (Baker and diseases 2017) and drugs targeting upstream type I IFN pathway members, such as *TLR7* and TLR9 (Baker and diseases 2017).

IL-17/IL-23 axis

Together with the TNF pathway, the IL-17/IL-23 axis is the most common target of biological therapeutics. In fact, some studies have reported greater efficacy of individual IL-17A or IL-23 blockade compared to TNF inhibition in

the treatment of psoriasis and PsA (Griffiths et al. 2015; Blauvelt et al. 2017). The cytokine IL-23 is formed of two subunits: IL-23A/p19 and IL12-B/p40. Transcriptional studies have shown increased levels of p40 and p19 in psoriasis lesional skin and a role for both subunits in abnormal KC differentiation (Zhu2011; Lee et al. 2004). Psoriasis and PsA GWAS associations with *IL-23R* have been reported, including a protective two SNP haplotype shared with CD (Nair et al. 2008; Strange et al. 2010; Tsoi et al. 2012). GWAS associations have also been established implicating *IL23A* and *IL12* (Cargill et al. 2007; Strange et al. 2010; Tsoi et al. 2012). Interestingly, an *IL-23* signal secondary to that reported by Tsoi *et al.,* 2012 has been specifically associated with PsA *Tsoi2012,Bowes2015*. Regarding the genetics of the Th-17 pathway, its relevance is partly explained through the cross-talk with the IL-23 response, which mediates Th-17 cell differentiation and activation. Additionally, associations implicating TFs regulating Th- 17 polarisation, such as *IRF4* and *STAT3*, have also been identified for psoriasis and PsA (Tsoi et al. 2012; Huber et al. 2008; Harris et al. 2007).

Genome-wide pathway enrichment analysis and intergenic regions

New approaches using genetic association data have disclosed relevant biological processes by conducting genome-wide pathway analysis. In psoriasis, this has revealed association of novel processes, such as retinol metabolism, transport of inorganic ions and amino acids and

post-translational protein modifications (PTMs) (Aterido2015).

As previously mentioned, the majority of the non-coding GWAS associations are located in intergenic regions and often lack functional characterisation. Therefore these variants tend to be associated to the nearest gene but may occur in intergenic regions at a distance from any gene, including chr1p36.23, chr2p15 and chr9q31.2 in psoriasis and PsA.

One of the most interesting regions is chr1p36.23, 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 immortalised KC cell line HaCaT has revealed interaction of SNPs in this locus with the promoter of the *ERRFI1* gene, an inhibitor of the epidermal growth factor receptor signaling required for normal KC proliferation (Ray-Jones et al. 2017).

### Limitations and future of GWAS

GWAS have made a great contribution to our understanding of the genetic basis of complex diseases. However, this approach has a number of limitations that need to be considered.

One of the major limitations is the challenge of fine mapping due to linkage disequilibrium (LD). An association between a genetic locus and a trait does not reveal the causal variant, which could potentially be any of the highly correlated SNPs in the same LD block as the lead SNP. This can be addressed in part by dense genotyping, statistical fine-mapping methods and incorporation of epigenetic data while ultimately application of genome editing may be needed to define the causal SNP.

Another concern is missing heritability when that explained by GWAS is compared to the estimated heritability from twin and 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 (Visscher et al. 2017). For example, in human height, most of 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 (Wray 2005). Such limitations have partly been overcome by improved genotyping arrays like Immunochip, which incorporates SNPs with MAF*<*1% (Cortes and Brown 2011). 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 structural variation such as CNV, small (*<*1Kb) insertions/deletions (indels) and inversions could all contribute to missing heritability. Incorporation of new genotyping platforms has allowed the genome-wide identification of CNV (Glessner et al. 2009; Marshall et al. 2017) while the accurate detection of translocations and inversions relies on the implementation of long read WGS technologies (Visscher et al. 2017). 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 (Zuk et al. 2012).

## Functional interpretation of GWASin complex diseases

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

GWAS report associations with disease for a particular locus but typically fail to identify the true causal variant(s) within the haplotype block (Edwards et al. 2013). 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. The integration of statistical fine-mapping with cell type and context specific epigenetic data, including chromatin accessibility, histone modifications and DNA methylation, can help to determine the chromatin state where the fine-mapped variants are located and its potential in regulating gene expression (Petronis 2010). Additionally, the incorporation of gene expression, eQTL analysis and chromatin

interaction data can establish a relationship between non-coding variants and putative gene targets. Final validation of the functional relationship between the genetic variant and the disease phenotype involves conducting appropriate cellular assays and *in vivo* experiments using animal models.

### Understanding the epigenetic landscape in complex diseases

Epigenetic modifications consist of heritable changes in the phenotype and/or gene expression that do not involve changes in the DNA sequence(Feil and Fraga 2012). These changes include a wide range of modification in the proteins which serve as scaffold for the DNA, known as histones, as well as DNA methylation and non-coding RNAs. Environmental and intrinsic factors can trigger changes in the epigenome that result in dysregulation of gene expression and, consequently, in alteration of gene function.

Genetic background can increase the predisposition to epigenetic changes caused by extrinsic factors. Studies have demonstrated differences in response to environmental factors by different mice breeds as well as greater differences in the epigenetic landscape between human dizygotic twins when compared to monozygotic (Pogribny et al. 2009; Kaminsky et al. 2009). Importantly, disease-associated GWAS variants have consistently shown enrichment for DNA regulatory elements, characterised by the combination of epigenetic marks, including accessible chromatin, histone modifications and DNA methylation (Trynka and Raychaudhuri 2013a; Trynka and Raychaudhuri 2013b; Gusev et al. 2014).

The plasticity of the epigenetic landscape is required for cell differentiation and identity and particularly important in the immune system to ensure adaptation and response to different pathogen infections (Yosef and Regev 2016). The role of cell type specificity in the epigenetic landscape has been

demonstrated in eQTL studies, where 50 to 90% of genetic variants regulating gene expression are cell type and stimulus dependent (Dimas et al. 2009; Nica et al. 2011; Fairfax et al. 2012; Fairfax et al. 2014; Raj et al. 2014; Naranbhai et al. 2015; Kasela et al. 2017). Recent methodological advances have made the personalised study and understanding of the epigenome possible by the implementation of low-cell-input high-throughput techniques coupled to next generation sequencing (NGS) (Buenrostro2013; Schmidl2015; Oudelaar et al. 2017). Understanding of cell-to-cell epigenomic heterogeneity is also being addressed with single-cell methods and may help to elucidate the impact of genetic variability in regulation of gene expression and disease mechanisms (Buenrostro et al. 2015; Cusanovich et al. 2015; Rotem et al. 2015; Nagano et al. 2013; Smallwood et al. 2014).

### The chromatin landscape

In the cell nucleus, 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 (Luger et al. 1997). In general, highly compacted DNA will remain more inaccessible for the assembly of the transcriptional machinery, consequently preventing gene expression. Chromatin accessibility can be altered by post-translational modifications (PTM) of the histone proteins that affect their affinity with the DNA within the nucleosome as well as the interaction between nucleosomes in the vicinity (Polach et al. 2000; Pepenella et al. 2014). Additionally, chromatin structure can also be influenced by adenosin triphosphate (ATP)-remodelling complexes that facilitate sliding of individual nucleosomes to neighboring DNA segments, increasing temporary chromatin accessibility at particular sites (Cosma et al. 1999). From the biochemical point of view, the signature of chromatin accessibility, histone

modifications, transcription factor occupancy and DNA methylation has been used to identifying *cis*-regulatory elements such as promoters, enhancers, silencers, insulators and locus control regions, and define the cellular chromatin landscape (Boyle et al. 2012; Kundaje et al. 2015).

Methods to ascertain chromatin accessibility

Accessible chromatin constitutes about 1% of the human genome and represents a very robust marker for histone modifications, early replication regions, transcription start sites (TSS) and TF binding sites (TFBSs) (ENCODE 2007). The informativeness of chromatin accessibility for understanding gene regulation has driven the development of several high-throughput techniques for accurately tagging these regions. Amongst those techniques, the “gold standard” is DNase I hypersensitive site mapping, which uses the non- specific double strand endonuclease DNase I to preferentially cut on nucleosome- free regions (DNase hypersensitive sites, DHSs). In this approach, isolation of the chromatin-free DNA is followed by further enzymatic digestion and DNA library preparation prior to NGS (John et al. 2013). DNase-seq also provides high quality information regarding TFBS, generating footprints that identify TF binding in relation to chromatin structure (Hesselberth et al. 2009; Boyle et al. 2010).

Another method to interrogate the accessible genome is formaldehyde- assisted isolation of regulatory elements (FAIRE-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 (Giresi et al. 2006). Both methods have enabled ENCODE to map regulatory elements in several cell lines, primary cells and tissues, revealing that 76.6% of all non-coding GWAS SNPs together with those in complete LD are located within broadly accessible chromatin tagged by DHSs (ENCODE 2007; Buck et al. 2014; Gaulton et al. 2010; Maurano et al. 2012). Indirect measurement of the chromatin

accessibility has also been performed using micrococcal nuclease-sequencing (MNase-seq). In this approach chromatin-free DNA on cross-linked nuclei is degraded and only the nucleosome-bound material is retained for downstream sequencing, providing a qualitative and quantitative comprehensive map for nucleosome positioning and also TF occupancy (Axel1975; Ponts et al. 2010). The high number of cells (5 to 10 millions or more) required by these assays for good quality data limits their application to particular biological and clinical samples.

Recently, a new technique assay for transposase-accessible chromatin using sequencing (ATAC-seq) has represented a groundbreaking step in characterisation of the genomic regulatory landscape (Buenrostro2013). ATAC-seq is based on an engineered hyperactive transposase enzyme, known as Tn5, that preferentially accesses nucleosome-free and inter-nucleosomal DNA inserting sequencing adapters at both end of those fragments (Gradman et al. 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. These two aspects make ATAC-seq a very versatile technique to interrogate the chromatin landscape in a clinical setting, where sample availability and time-efficiency are key factors (Scharer et al. 2016; Qu et al. 2015; Qu et al. 2017).

The role of histone modifications and TF occupancy in the chromatin landscape

Identifying the combination of histone modifications and binding of TF is essential to characterise regulatory regions of the genome. Histone modifications take place in the NH2terminal tail that protrudes from the nucleosome, the most common modifications being acetylation, phosphorylation and methylation. The co-localisation of different histone marks modulate the affinity for DNA-

binding proteins and the interaction with neighboring nucleosomes in varied manners, contributing to the overall chromatin accessibility landscape of the cells (Jenuwein and Science 2001; Bannister and research 2011). 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 (Hansen et al. 2008; Bannister et al. 2001).

Several types of chromatin corresponding to different regulatory elements have also been defined. Enhancers and promoters, regardless of their activation state, are tagged by high levels of H3K4me1 or H3K4me3, respectively, and both features co-localise with H3K4me2 modifications (Heintzman et al. 2007; Hon et al. 2009). H3K9ac is specifically enriched at active promoters whereas H3K27ac generally designates activation at both promoters and enhancers (Hon et al. 2009; Creyghton and the 2010). Conversely, H3K27me3 together with the heterochromatin mark H3K9me3 indicates gene repression at promoter elements (Hansen et al. 2008; Bannister et al. 2001; Pan et al. 2007). 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 2013a). Overall, functional understanding and interpretation of histone mark co-localisation remains challenging and incorporation of additional epigenetic information is usually required. Together with histone modifications, TF also play a role in nucleosome positioning as well as in acting as boundary elements to separate chromatin states (Vierstra et al. 2014; Zhang et

al. 2009; Bell and Nature 2000). TF occupancy is indirectly tagged by chromatin accessibility assays, such as DHS, through reduced cutting sensitivity of DNase I due to protein binding and steric hindrance.

Chromatin immunoprecipitation sequencing (ChIP-seq) has been widely used to precisely locate histone modifications and TF binding in the genome. This technique assays protein-DNA binding *in vivo* using Abs that specifically recognise histone modifications or TF after DNA-protein cross-linking 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 (Solomon et al. 1988; Barski et al. 2007; Johnson et al. 2007). 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; Bernstein and Nature 2010; Adams et al. 2012). Similarly to the first generation of chromatin accessibility techniques, ChIP-seq requires at least between 5 to 10 million cells per experiment, restricting 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 out as the simplest and most cost-effective method, only requiring 10,000 and 100,000 cells to assay histone modifications or TF binding, respectively (Schmidl2015). ChIPm involves 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.

DNA methylation

DNA methylation involves the transferal of a methyl group to the 5’ carbon of a cytosine that precedes a guanine nucleotide (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 (Herman and Medicine 2003). Together with histone modifications, DNA methylation has a pivotal role in the differentiation of haematopoietic stem cells and the maturation and activation of immune cells (Sellars et al. 2015; Lai et al. 2013). The pathogenicity of changes in the methylome has been studied in a range of diseases including RA, SLE, psoriasis and PsA (Lei et al. 2009; Liu et al. 2013; Zhang et al. 2010). For example, regulation of TNF-*α* production upon inflammatory stimuli involves a complex network of DNMTs that alter the methylation signature at the locus (Sullivan et al. 2007).

Chromatin interactions and gene expression

The functional understanding of non-coding variants has benefited from eQTL studies. Nevertheless, eQTLs only provide indirect evidence of the effect of a SNP on regulating expression of a particular gene. 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). Chromatin is organised into topologically associating domains (TADs) of several hundred kb insulated from other TADs by the binding of CTCF protein, amongst others (Nora et al. 2017). Chromatin loops between promoters and the corresponding regulatory elements mostly take place within the same TAD and are highly cell- and context-specific (Smith et al. 2016). Hence, 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. As an example, obesity risk

non-coding variants located within the *FTO* gene appeared to regulate expression through chromatin looping of the *IRX3* gene, located 1Mb downstream (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 (Davies et al. 2017). 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; Oudelaar et al. 2017). 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 to enhance the specificity of the assay while reducing sequencing depth and input material (Mumbach et al. 2016).

### 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 conditional on studying the relevant disease tissue, which sometimes remains challenging due to a lack of pathophysiological understanding of disease mechanisms or difficulties in accessing it. In immune-mediated diseases, PBMC differential gene expression (DGE) analysis between patients and controls has enabled identification on relevant pathways and biochemical functions including psoriasis and PsA (Miao et al. 2013; Junta and 2009; Baechler et al. 2003; Assassi et al. 2010; Batliwalla et al. 2005). Similarly, the growing evidence supporting cell type and context specificity 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 (Katschke et al. 2001; Dozmorov et al. 2015; Jabbari et al. 2012).

Likewise, the extensive overlap of GWAS variants with non-coding regions potentially dysregulating gene expression has highlighted the importance of performing context-specific eQTL studies. In this respect, consortia such as the Genotype Tissue Expression (GTEx) have generated publicly accessible comprehensive tissue-specific eQTL studies that have greatly contributed to the functional understanding of GWAS risk alleles in many complex diseases (Londsdale2013; Fagny et al. 2017).

Long non-coding RNAs and enhancer RNAs

In addition to protein coding mRNAs, non-coding RNAs have been demonstrated to have a role in regulation of gene expression. One category of non-coding RNAs are the long non-coding RNAs (lncRNAs), transcripts between 200 and 100Kb long that undergo splicing, 5’ capping and 3’ poly-adenylation (Derrien et al. 2012). 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 acting as a decoy for other non-coding RNAs and regulatory proteins

(Pandei2008; Faghihi et al. 2008; Gong and Maquat 2011; Carrieri et al. 2012; Kino et al. 2010).

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 cell activation and host- pathogen interactions (Rossetto2009; Pang et al. 2009). Moreover, differential case-control gene expression analyses have underscored the contribution of lncRNAs in several chronic inflammatory conditions, including RA, SLE and psoriasis (Muller2014; Li2014; Shi et al. 2014; Ahn et al. 2016).

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 (Fantom et al. 2014). Although traditionally chromatin segmentation maps have defined enhancers as DNA regions with particular epigenetic 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 enhancers has been demonstrated to be an excellent proxy to identify functionally active regulatory region, which have also been successfully validated by reporter assays (Anderssen2014; Fantom et al. 2014).

Another class of non-coding RNAs are micro-RNAs (miRs), 21 to 24 nucleotides long (Lee et al. 2002). 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 (Ameres et al. 2010; Braun et al. 2011; Petersen et al. 2006) with 30 and 80% of human genes predicted to be under transcriptional control of miRNAs (Lewis et al. 2005; Friedman et al. 2008).

Methods to assay gene expression

RNA-seq involves reverse- transcription of the extracted RNA into cDNA and PCR amplification preserving relative abundance of each transcript, followed by library preparation and NGS (Mortazavi et al. 2008). Systematic comparison has shown superior dynamic range of detection for RNA-seq compared to micro-arrays, particularly for low abundance transcripts (Zhao et al. 2014). Furthermore, RNA-seq allows the capture of additional information to the expression profile, including the identification of new exons, alternative splicing events and allele-specific expression (ASE). Quantification of ASE for transcripts in individuals heterozygous for exonic SNP haplotypes in a particular gene through RNA-seq has provided direct evidence for local/*cis*-eQTLs driven by allele-specific mechanism, showing significant differences in haplotype

transcript abundance for up to 88% of the genes with an associated *cis*-eQTL (Pickrell et al. 2010). 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 (Tang et al. 2009; Tang et al. 2010).

Variations of the RNA-seq methodology such as cap analysis of gene expression (CAGE) and other 5’end RNA-sequencing methods have enabled the precise identification of TSS and the associated promoters for each transcript (Yamashita et al. 2011; Fantom et al. 2014). The CAGE data generated by the functional annotation of the mammalian genome 5’ (FANTOM5) Consortium includes thousands of eRNAs and has contributed to better definition of enhancers and their spatial and temporal specificity in hundreds of human primary cells and tissues (Andersson et al. 2014).

### Transcriptional regulation in complex diseases

Non-coding GWAS variants can exert pathogenic effects by affecting one or many of the previously described mechanisms responsible for the fine regulation of gene expression in homeostatic conditions. Intronic SNPs can influence mRNA splicing through exon skipping, resulting in truncated but functional proteins. For instance, exon skipping caused by an intronic risk allele at the TNF Receptor Superfamily Member 1A (*TNFRSF1A*) associated with MS results in a soluble isoform of the TNFRS1A protein with TNF antagonistic function (Gregory et al. 2012). On the other hand, non-coding variants at enhancers, silencers and promoters can dysregulate gene expression by altering affinity at TFBS, histone modifications and chromatin accessibility. For instance, in thyroid autoimmunity, the risk allele of an intronic SNP in the thyroid stimulating hormone receptor *(*TSHR) gene reduces *TSHR* protein expression in IFN-*α* stimulated thyroid cells (Stefan et al. 2014). The risk SNP increases the affinity of the repressor promyelocytic leukemia zinc finger protein (*PLZF*) that

recruits histone acetylases (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 (Zhang et al. 2012).

Alternatively, non-coding SNPs can regulate gene expression by creating a new promoter-like element, as in the *α*- thalassemia disease, where this phenomenon leads to dysregulated downstream activation of all *α*-like globin genes in erythroid cells (Gobbi et al. 2006). Genetic variants at eRNAs can also affect regulation of gene expression as it has been demonstrated in the nuclear receptor for anti-diabetic drugs PPAR*γ* in mice (Soccio et al. 2015). Lastly, non- coding variants placed in UTRs and intergenic regions can affect binding of miRNAs and lncRNA to the target genes. This is the case of a CD associated variant at the 3’UTR of the gene immunity related GTPase M *IRGM* which reduces binding of the miR-196, increasing its mRNA stability and translation, ultimately resulting in disrrupted 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).

### Integration and interpretation of genomic data

The evolution of different omics methods towards generation of paired datasets at a high-throughput scale presents 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 effect that involve several genes and signaling pathways

through alteration of epigenetic features and subsequent dysregulation of gene expression.

Tools such as RegulomeDB allow the querying of a large number of publicly available epigenetic and functional datasets, including DHSs, TFBS, histone modification and DNA-protein interactions, at the SNP level (Boyle et al. 2012). Other powerful tools include the University of California Santa Cruz (UCSC) genome browser, a resource to display in-house and publicly accessible annotation data (Kent et al. 2002). In addition to this, international consortia generating large- scale epigenetic and expression data such as ENCODE, Blueprint, Roadmap Epigenome, GTEx or FANTOM have created comprehensive website resources for browsing and downloading data (Lonsdale2013; Adams2012 ; ENCODE 2007; Fantom et al. 2014). These collaborations have also led to the integration of epigenetic datasets and assembling of cell type specific chromatin states maps. This consists of the segmentation and labelling of the genome with a chromatin state based on concurrence of several epigenetic marks using Hidden Markov Model algorithms such as ChromHMM, amongst others (Kundaje2015 ; Ernst and Kellis 2010; Ernst et al. 2011; Hoffman et al. 2013).

In addition to data integration, the other main bottleneck encountered by functional genomics is determining the clinical relevance of 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 (e.g ontologies, signalling pathways or functional elements) within the entities of interest. For instance, 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 (Kanehisa and Goto 2000; Fabregat

et al. 2018). Such annotation sources may be used to interpret, for example, a set of differentially expressed genes or a list of genes obtained through annotation of non-coding regions using proximity, chromatin interaction data or eQTL studies. Similarly, this type of analysis can be used to find enrichment of genomic regions of interest for a varied collection of epigenomic features tagging regulatory elements in relevant cell types.

From the number of tools designed to perform this type of analysis, eXploring Genomic Relations (XGR) is particularly powerful (Fang2016 ). XGR is an open source R package and web-app that allows handling of different types of input data (SNPs, genes and regions). XGR integrates a wide range of ontologies and up to date publicly available functional data to 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.

### The use of fine-mapping to prioritise functional causal variants

The aim of fine-mapping is to reduce the size of GWAS genomic intervals and yield a minimal set of SNPs containing the causal variant that will explain most of the association for that particular locus (Spain and genetics 2015). 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 density at previously associated immune-relevant loci in a cost-effective manner (Trynka et al. 2011). 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 et al. 2012). More recently, the UK10K project has improved the quality of imputation specifically for rare variants with MAF between 0.01% and 0.5% (Chou et al. 2016).

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). BFs are 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 report a credible set of SNPs that account for 95 or 99% of the overall PP in each loci(Bunt et al. 2015).

The inclusion of functional data from publicly available sources as priors of the approximate Bayesian model has demonstrated a reduction of the number of SNPs in the credible set and also increased the proportion of successfully fine- mapped loci (Bunt et al. 2015; Kichaev and Genetics 2015). The integration of fine-mapping data generated with the Bayesian probabilistic identification of causal SNPs (PICS) method and a map of genomic regulatory elements, revealed that approximately 60% of the top fine-mapped SNPs overlapped enhancer elements (importantly stimulus-specific) and were very close but not

within TF binding sites (TFBS) (Farh2015).

### 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 unequivocally addresses the functional mechanisms conferring pathogenic effect in a cell type and context specific manner. To overcome this, a wide range of experimental approaches can be performed to functionally validate 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, involve for example transfection of constructs containing the promoter or enhancer element of interest followed by luciferase expression (Niimi et al. 2002). 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 (Vernes et al. 2007). 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 (Kheradpour et al. 2013). In addition to this, mass spectrometry (MS) techniques have been used to perform allele-specific quantitative proteomics and have revealed allele-dependent binding of TF and co-regulators at the T2D *PPARG*

GWAS locus (Lee et al. 2017). *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 been shown to have limitations that need to be taken into account when interpreting the results (Ermann and Glimcher 2012).

Both, *in vitro* and *in vivo*, models for functional studies have benefited from the development the genome-editing technology clustered regularly interspaced short palindromic repeats (CRISPR/Cas) (Cong et al. 2013). 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 edit certain primary cells is being overcome by the use of human induced pluripotent stem cells (hiPSCs), which can undergo terminal differentiation into the cell type of interest after CRISPR modification (Ding et al. 2013).

**Aims and objectives**

The aim of this thesis is to investigate the epigenetic regulatory landscape in psoriasis and PsA to identify disease and cell type specific changes in putative regulatory regions and differences in gene expression, with the longer-term goal of advancing understanding of the pathophysiology of these diseases and informing interpretation of genetic associations arising from GWAS. Specific objectives are:

1. To establish ATAC-seq protocol(s) appropriate for cells and tissues of interest and samples taken in the clinic, with optimization of required methodologies and analytical tools including quality control measurements; peak filtering; differential analysis; transposition time; approaches to reduce mitochondrial DNA contamination; application to skin biopsies; and utility of cryopreservation and fixation.
2. To determine chromatin accessibility, histone modification and gene expression differences between psoriasis patients and controls in peripheral blood for four major circulating immune cell types in peripheral blood (CD14+ monocytes, CD4+ and CD8+ T cells, and CD19+ B cells) and to complement this with analysis of differential gene expression in lesional and uninvolved epidermis isolated from psoriatic skin biopsies.
3. To identify cell subsets contributing to pathophysiological relevant pathways in PsA by assaying chromatin accessibility and gene expression in CD14+ monocytes, mCD4+, mCD8+ and NK cells for peripheral blood and synovial fluid, determining transcriptional differences at the single-cell level in cell types of interest and relating findings to PsA GWAS.